Refining the chemical and kinetic decoupling description of thermally produced dark matter

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm ProPhys

der Georg-August University School of Science (GAUSS)

vorgelegt von

Tobias Binder

aus Heilbronn-Neckargartach

Göttingen, 2019

Betreuungsausschuss:

Prof. Dr. Laura Covi,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Prof. Dr. Karl-Henning Rehren,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Mitglieder der Prüfungskommission:

Referent:
Prof. Dr. Laura Covi,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Korreferent:
Prof. Dr. Karl-Henning Rehren,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

2. Korreferent:
Prof. Dr. Björn Garbrecht,
Fakultät für Physik,
Technische Universität München

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Stan Lai
II. Physikalisches Institut,
Georg-August-Universität Göttingen

Prof. Dr. David J. E. Marsh
Institut für Astrophysik,
Georg-August-Universität Göttingen

Prof. Dr. Marcus Müller,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Prof. Dr. Steffen Schumann,
Institut für Theoretische Physik,
Georg-August-Universität Göttingen

Tag der mündlichen Prüfung: 13. März 2019

Contents

Abstract **3**

1 Introduction **4**

2 Boltzmann formalism **7**
- 2.1 Phase-space density . . . 7
- 2.2 Number density and relic abundance . . . 9
- 2.3 Kinetic decoupling temperature and mass of first protohalos . . . 12

3 Thermal decoupling **15**
- 3.1 Standard chemical and kinetic decoupling . . . 15
- 3.2 *Publication:* Early kinetic decoupling of dark matter: when the standard way of calculating the thermal relic density fails . . . 17

4 Light mediators **33**
- 4.1 Challenges on small scales and possible solutions . . . 33
- 4.2 Light mediators as an alternative solution to the small scale crisis . . . 35
- 4.3 *Publication:* Matter power spectrum in hidden neutrino interacting dark matter models . . . 38
- 4.4 *Publication:* Reannihilation of self-interacting dark matter . . . 68

5 Long-range interactions at finite temperature **93**
- 5.1 Simplified model . . . 93
- 5.2 Two-particle effective in-medium potential . . . 94
- 5.3 Chemical equilibration rate . . . 96
- 5.4 *Publication:* Dark matter Sommerfeld-enhanced annihilation and bound-state decay at finite temperature . . . 99
- 5.5 Boltzmann equation for ionization equilibrium . . . 142
- 5.6 Comparison to coannihilation . . . 144

6 Summary and Conclusion **146**

Bibliography **149**

Acknowledgments **160**

Curriculum vitae **161**

Abstract

The first part of this thesis deals with refined theoretical predictions of the thermal relic abundance for weakly interacting massive particles (WIMPs). Methods are developed in a model independent way to describe the discovered exception where kinetic equilibrium, the main assumption entering the standard prediction, is not maintained during the chemical decoupling process. The impact of early kinetic decoupling is quantified for the Scalar Singlet model and the results show that the refined prediction can differ from the standard computation of the thermal relic abundance by up to an order of magnitude. Furthermore, many previous works show within the classical Boltzmann approach that attractive long-range interactions can lead to an enhancement of the annihilation cross-section at the time of chemical decoupling. In this thesis, it is investigated how to describe long-range interactions in the presence of a hot and dense plasma background in the framework of non-equilibrium quantum field theory. The main result is a novel number density equation expressed in terms of thermal correlation functions which allows to study chemical decoupling including various long-range effects and plasma background corrections in a self-consistent way. Consistency is proven with previous literature results under certain limits. The second part of the thesis deals with long-range forces in the context of Self-Interacting Dark Matter (SIDM), considered in literature as a possible solution of structure formation problems at small scales. In this thesis, new types of long-range interactions are identified which could alleviate in a similar parameter region several structure formation issues simultaneously. It is shown that a particular type of long-range force could additionally relax tensions between the cosmic microwave background and low-redshift astronomical observations.

Chapter 1

Introduction

Dark Matter (DM) is nowadays an indispensable component of the concordance model of cosmology. There is overwhelming gravitational evidence for this non-luminous form of matter on a variety of length scales, ranging from the smallest visible objects such as ultra-faint dwarf galaxies up to the largest scales which are probed by observations of the cosmic microwave background (CMB) radiation [1, 2]. One leading hypothesis is that DM consists of weakly interacting massive particles (WIMPs), expected to be thermally produced in the early Universe. However, during the last decade, null results in various experimental searches for WIMPs have shadowed the expectation that the observed cosmic abundance of DM is made up by thermally produced WIMPs with a mass and coupling strength to Standard Model (SM) particles at the electroweak order. This thesis reconsiders the standard description of thermal production and subsequent decoupling under different aspects.

In the traditional picture, WIMPs are initially in thermal equilibrium with SM particles and chemically decouple once their production rate falls below the cosmic expansion rate. After chemical decoupling, the number density is conserved in a comoving volume. The number density has a direct relation to the fractional energy density. To make a viable theoretical prediction, the latter quantity should match the observed energy fraction of DM which is nowadays measured to an unprecedented precision, reaching sub-percentage accuracy [1, 2]. Sophisticated numerical dark matter codes, e.g. DarkSUSY [3, 4] and micrOMEGAs [5, 6], theoretically predict the relic abundance with an expected accuracy compatible with the experimental error. Often one has to deal with models where the chemical reaction network is large, as it occurs in fundamental extensions of the SM like in the supersymmetric scenarios. A common approach is to use the concept of detailed balance, which simplifies the complexity of the network and allows for an acceptable runtime for scanning over a large amount of model parameters. While very often the assumptions of detailed balance are well justified, in a part of this thesis the main assumption of local thermal equilibrium, entering in many computations, is scrutinized.

Chemical reactions are often not the only processes. For annihilating WIMPs, there must be from symmetry considerations also scattering reactions with SM particles. The latter keep DM in kinetic equilibrium until the scattering rate falls below the Hubble expansion rate. In the standard picture, kinetic decoupling takes place at much later times, i.e., when number changing processes can be ignored. This allows to assume that kinetic equilibrium is maintained during chemical decoupling, which is typically the case. The temperature at which DM kinetically decouples is directly related to the minimum mass of the first protohalos that have formed in the early Universe. Since the concordance model has no such characteristic mass scale, the abundance of small scale halos allows for a complementary test of the underlying micro physics between dark matter and SM particles. In the case of a neutralino WIMP, a mixture of superpartners of the gauge and Higgs fields, the kinetic decoupling temperature takes values between several MeV up to a few GeV, which corresponds to a minimum halo mass between $10^{-3} M_\odot$ and $10^{-11} M_\odot$ [7], respectively. The abundance of DM halos today on such small mass scales is hard to probe. This thesis deals with models featuring late kinetic decoupling, where the minimum mass of the first protohalos reaches sub-galactic scales of about $10^8 M_\odot$. The suppression of the DM halo abundance on such scales can be predicted through the computation of the matter power spectrum, a measurable quantity, which allows to constrain [8] or potentially detect [9] the effect through the Lyman-α forest. The latter is a series of absorption lines caused by neutral hydrogen energy level transitions. Further ideas to measure the abundance of only gravitationally visible dark matter halos are through modifications of star motions in stellar streams (see, e.g., [10, 11]).

Several works have reported that structure formation on sub-galactic scales could be incompatible with the predictions made by the concordance model. Halos in cold and collisionless DM simulations seem to fail to produce the observed abundance and density structure of smaller galaxies (see, e.g., [12]). Other works point out that this "small scale crisis" is not a problem at all and can be explained within the concordance model. A literature review on this debate including alternative solutions is given in section 4.1. Next to the challenges on small scales, there are discrepancies between CMB and low-redshift astrophysical observations. The collaboration of the European space satellite Planck reports a smaller value of the expansion rate today ($H_0 = 67.27 \pm 0.60\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, [2], 2018) compared to the value inferred from local measurements by, e.g., the Hubble space telescope (HST) ($H_0 = 73.52 \pm 1.62\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$, [13], 2018).[1] Another tension concerns the matter fluctuation amplitude σ_8 at $8h^{-1}\,\mathrm{Mpc}$ or the total fractional amount of matter Ω_m (both correlate). Some galaxy weak lensing results [15, 16] prefer lower values compared to the results of the Planck 2015 analysis [1] (modest tension). In the recent Planck 2018 report [2] it is shown that latter tensions relaxed, however, see [17].

While simplified particle physics models of DM have been proposed in literature as a possible solution to the small scale crisis [18–20], this thesis investigates if those models could simultaneously explain the discrepancies between CMB and astrophysical observations. The key ingredient is a light dark mediator at the MeV scale which acts as a long-range force between heavy dark matter particles. The long-range force induces, e.g., strong DM self-interactions. These could alter the inner halo density structure, leading to characteristic deviations from cold and collisionless DM predictions.

In general, long-range interactions introduce a variety of quantum mechanical effects. These are a major topic of this thesis and not only in the context of strongly self-interacting DM models. In the original work by J. Hisano et al. [21], it was shown that if electroweakly charged DM particles are heavy enough, the electroweak gauge bosons of the SM effectively act as attractive long-range forces, leading to a strongly enhanced annihilation cross-section of slowly moving DM particles in the galactic center. The enhancement of the annihilation cross-section through attractive long-range interactions is commonly referred as the Sommerfeld enhancement [22]. In ref. [23], it has been pointed out that the Sommerfeld effect can also become important to include in the relic abundance computation. An enhanced annihilation cross-section at chemical decoupling lowers the predicted relic abundance, which allows for larger masses of the annihilating DM particles to compensate for the effect. In the original work, it was shown that the Sommerfeld enhancement reduces the pure Wino neutralino density by up to 50 %, pushing the allowed Wino mass to a few TeV [23]. Nowadays, the effect is widely studied in different WIMP scenarios. The interesting physical aspect is that the effect can lift the allowed WIMP mass above the production threshold at the Large Hadron Collider and therefore the model could evade strong exclusion limits.

---

[1] The tension has increased (in favour of our proposed solution) since the publication of the article in section 4.4, where older values from Planck 2015 [1] and HST 2016 [14] were used.

Related quantum mechanical effects, caused by attractive long-range interactions, are bound-state solutions. If existing, they strongly increase the size of the chemical network in the relic abundance computation by opening reaction channels for conversion processes between scattering and bound states. These processes are for example bound-state formation via radiative emission [24] and energy level transitions [25, 26]. In addition to the Sommerfeld enhancement, they further contribute to a depletion of the relic abundance and therefore allow for even heavier WIMP masses (see, e.g., [27, 28]).

While the classical Boltzmann equation is often used to compute the relic abundance of DM, it is unclear if that approach is a sufficient description for long-range interactions. A coupling of the force-carriers to the dense primordial plasma introduces possible thermal effects like charge screening, which influences the effective interaction among DM particles. The status of the literature is the following. The computational methods for describing effective interactions inside a plasma medium are well developed far beyond the naive screening picture. However, the description of out-of-chemical equilibrium transitions including thermal effects is based on a method, which is strictly speaking only applicable to the linear regime close to chemical equilibrium. This thesis uses non-equilibrium quantum field theory techniques, to derive a number density equation which allows to describe out-of-chemical reactions beyond the linear regime and includes thermal corrections to long-range interactions.

The thesis is structured as follows. Section 2 is a review of the standard Boltzmann formalism for describing chemical and kinetic decoupling. In section 3, the main assumption of kinetic equilibrium during chemical decoupling is scrutinized. In section 4, we turn to light mediator models, map them into the linear matter power spectrum, and investigate if they can potentially solve the small scale crisis and tensions between CMB and astrophysical observations simultaneously. The whole section 5 is devoted to thermal effects for long-range interactions. The model is introduced, the main literature on this subject is reviewed, and a more general number density equation is derived from non-equilibrium quantum field theory. The latter method is introduced in the dedicated part "Real-time formalism prerequisites" (chapter II in the publication reprinted in section 5.4). The thesis is summarized and concluded in section 6. Natural units $c = \hbar = k_B = 1$ are used.

Chapter 2

Boltzmann formalism

This chapter is devoted as a basic introduction into the Boltzmann formalism. The latter is the common approach to describe the thermal history of DM. Section 2.1 introduces the Boltzmann equation on phase-space density level to describe annihilation and elastic scattering processes. The classical number density equation for annihilating dark matter is shown in section 2.2. Section 2.3 introduces the Fokker-Planck method to describe kinetic decoupling of DM, as well as the relation between kinetic decoupling temperature and the minimum mass of the first protohalos.

2.1 Phase-space density

The evolution of the phase-space density f_1 for a certain DM species labelled by 1 is governed by the Boltzmann equation, which can be written as (see, e.g., [29]):

$$E(\partial_t - Hp\partial_p)f_1(t,p) = \hat{C}[f_1]. \tag{2.1}$$

Here, the energy is $E = \sqrt{m^2 + p^2}$ and the Hubble expansion rate H reads

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\sum_i \rho_i, \tag{2.2}$$

where a is the scale factor, ρ_i are energy densities, and G is the gravitational constant. The phase-space density is normalized to the particle number density as

$$n(t) = g_1 \frac{4\pi}{(2\pi)^3}\int_0^\infty \mathrm{d}p\, p^2 f_1(t,p), \tag{2.3}$$

where g_1 are the number of internal degrees of freedom.

The form of the collision operator $\hat{C}$ depends on the type of interactions. In the following, *binary interactions* are assumed to dominate at the leading order. In particular, annihilation and elastic scattering processes are considered. The collision term can be written as a sum over two terms $\hat{C}[f_1] = C_{\rm an} + C_{\rm sc}$, given by:

$$C_{\rm an} = -\frac{1}{2g_1}\int {\rm d}\Pi_2{\rm d}\Pi_3{\rm d}\Pi_4(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)|\mathcal{M}|^2_{\rm an}J_{\rm an}, \tag{2.4}$$

$$J_{\rm an} = f_1f_2\left(1\pm f_3^{\rm eq}\right)\left(1\pm f_4^{\rm eq}\right) - f_3^{\rm eq}f_4^{\rm eq}\left(1\pm f_1\right)\left(1\pm f_2\right), \tag{2.5}$$

$$C_{\rm sc} = -\frac{1}{2g_1}\int {\rm d}\Pi_2{\rm d}\Pi_3{\rm d}\Pi_4(2\pi)^4\delta^4(p_1+p_3-p_2-p_4)|\mathcal{M}|^2_{\rm sc}J_{\rm sc}, \tag{2.6}$$

$$J_{\rm sc} = f_1f_3^{\rm eq}\left(1\pm f_2\right)\left(1\pm f_4^{\rm eq}\right) - f_2f_4^{\rm eq}\left(1\pm f_1\right)\left(1\pm f_3^{\rm eq}\right), \tag{2.7}$$

where the on-shell integration is ${\rm d}\Pi_i = \frac{{\rm d}^3\mathbf{p}_i}{(2\pi)^3 2E_i}$. DM particles are labelled as 1 and 2, while bath particles (BP) are labelled as 3 and 4. $C_{\rm an}$ describes annihilation processes of the form ${\rm DM}(1) + {\rm DM}(2) \leftrightarrow {\rm BP}(3) + {\rm BP}(4)$, and $C_{\rm sc}$ consists of *elastic scattering processes*, i.e., reactions of the form DM(1) + BP(3) ↔ DM(2) + BP(4). The matrix elements $|\mathcal{M}|^2_i$ are summed over initial and final particle internal degrees of freedom. BPs are assumed to be in thermodynamic equilibrium, which implies for their phase-space density to be of the form

$$f^{\rm eq}_{3/4} = \frac{1}{e^{\beta E_{3/4}} \mp 1}, \tag{2.8}$$

where the upper (lower) sign applies to the Bose-Einstein (Fermi-Dirac) statistics and $\beta \equiv 1/T$ is the inverse temperature T. For the purpose of this section, the nature of the BPs is left open. In concrete applications later, the BPs will be either SM particles or belong to the dark sector as well, such as dark radiation.

For typical thermal production scenarios, quantum statistical effects of the DM distribution can be neglected to a good approximation. This is because out-of-chemical equilibrium transitions take place when DM enters the non-relativistic regime and therefore the dilute limit $T \ll E_{1/2}$ applies. Consequently, $f_{1/2} \ll 1$ and the Bose-enhancement or Pauli-blocking factors of DM particles can be neglected in both collision terms. This classical limit does also apply to the BPs in the annihilation term, which allows to adopt a Maxwell-Boltzmann distribution. Using energy conservation one can rewrite the classical distribution of BPs in the annihilation term as $f_3^{\rm eq}f_4^{\rm eq} = e^{-\beta(E_3+E_4)} = e^{-\beta(E_1+E_2)} = f_1^{\rm eq}f_2^{\rm eq}$. In these classical limits, the collision terms read:

$$C_{\rm an} = -\frac{1}{2g_1}\int {\rm d}\Pi_2{\rm d}\Pi_3{\rm d}\Pi_4(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)|\mathcal{M}|^2_{\rm an}\Big[f_1f_2 - f_1^{\rm eq}f_2^{\rm eq}\Big], \tag{2.9}$$

$$C_{\rm sc} = -\frac{1}{2g_1}\int {\rm d}\Pi_2{\rm d}\Pi_3{\rm d}\Pi_4(2\pi)^4\delta^4(p_1+p_3-p_2-p_4)|\mathcal{M}|^2_{\rm sc}\Big[f_1f_3^{\rm eq}\left(1\pm f_4^{\rm eq}\right) - f_2f_4^{\rm eq}\left(1\pm f_3^{\rm eq}\right)\Big]. \tag{2.10}$$

Integration over final state momenta in the annihilation term are now factorized from the phase-space distributions. A similar expression can be achieved in the scattering term by using the identity

$$f_4^{\rm eq}(1\pm f_3^{\rm eq}) = e^{-(E_1-E_2)/T}f_3^{\rm eq}(1\pm f_4^{\rm eq}), \tag{2.11}$$

leading to a more convenient form of the Boltzmann equation for the DM phase-space density:

$$
\begin{aligned}
\left(\partial_t - Hp\partial_p\right) f_1 = g_2 &\int \frac{\mathrm{d}^3\mathbf{p}_2}{(2\pi)^3}\frac{(p_1p_2)}{E_1E_2}(\sigma v_{\mathrm{rel}})\left[e^{-\beta(E_1+E_2)} - f_1f_2\right] \\
&+ g_2\int \frac{\mathrm{d}^3\mathbf{p}_2}{(2\pi)^3} W\left[e^{-\beta(E_1-E_2)}f_2 - f_1\right]. \qquad (2.12)
\end{aligned}
$$

This form allows to read off stationary solutions, discussed in the next sections. The annihilation cross-section σ, relative velocity v_{rel} and W were defined as:

$$
\sigma \equiv \frac{1}{\sqrt{(p_1p_2)^2 - m_1^4}}\frac{1}{4g_1g_2}\int \mathrm{d}\Pi_3\mathrm{d}\Pi_4\,(2\pi)^4\,\delta^4(p_1+p_2-p_3-p_4)|\mathcal{M}|^2_{\mathrm{an}}, \qquad (2.13)
$$

$$
v_{\mathrm{rel}} \equiv \frac{\sqrt{(p_1p_2)^2 - m_1^4}}{(p_1p_2)}, \qquad (2.14)
$$

$$
W \equiv \frac{1}{4g_1g_2E_1E_2}\int \mathrm{d}\Pi_3\mathrm{d}\Pi_4(2\pi)^4\delta^4(p_1+p_3-p_2-p_4)|\mathcal{M}|^2_{\mathrm{sc}}f_3^{\mathrm{eq}}(1\pm f_4^{\mathrm{eq}}). \qquad (2.15)
$$

The Boltzmann Eq. (2.12) is a partial integro-differential equation, describing the time evolution of the DM phase-space density under the influence of an expanding background, momentum dependent annihilation and elastic scattering processes. In general, a fully numerical treatment is required to obtain a solution. Instead of solving for the phase-space density directly, one can alternatively transform the partial differential equation into a set of coupled ordinary differential equations. This transformation can be done by integrating Eq. (2.12) over certain *momentum moments*, resulting in differential equations for primitive variables like mean number density or mean kinetic energy. Albeit being numerically less challenging, the difficulty within this approach is that each individual momentum moment equation is in general coupled to other moments, reflecting the in principle infinite *Boltzmann hierarchy*. Within reasonable assumptions about the properties of the system, which will be discussed in the subsequent sections, it can be possible to obtain a closer of the hierarchy within the lowest momentum moments.

2.2 Number density and relic abundance

In the standard case of thermally produced dark matter particles, it is assumed that scattering rates at early times are much larger compared to the annihilation and Hubble expansion rate, leading to *kinetic equilibration*. In kinetic equilibrium, the DM phase-space distribution has a thermal shape: $f_{1/2} \propto e^{-\beta E_{1/2}}$. This is indeed a stationary point of the scattering term, i.e., the second line in Eq. (2.12) vanishes for thermal distributions. Furthermore, the stationary point of the scattering term is assumed to be an attractor. The normalization of the phase-space density, as given in Eq. (2.3), fixes the proportionality factor, leading to

$$f_{1/2} = \frac{n}{n_{\rm eq}(\beta)} e^{-\beta E_{1/2}}, \text{ where} \tag{2.16}$$

$$n_{\rm eq}(\beta) = g_1 \int \frac{{\rm d}^3\mathbf{p}_1}{(2\pi)^3} e^{-\beta E_1} = \frac{g_1}{2\pi^2} m_1^3 K_2(m_1\beta)/(m_1\beta). \tag{2.17}$$

$K_i(x)$ are the modified Bessel functions of order i and symmetric DM is considered. This form of the phase-space density is consistent with the picture that DM in kinetic equilibrium is thermodynamically described as a dilute ideal gas in a grand canonical state with time dependent chemical potential $\beta\mu = \ln[n/n_{\rm eq}]$ and statistical occupation $f_{1/2} = e^{-\beta(E_{1/2}-\mu)}$.

The assumption of kinetic equilibrium closes the Boltzmann hierarchy at the number density level. It can be seen by inserting Eq. (2.16) into Eq. (2.12) and integrating over $g_1 \int \frac{{\rm d}^3\mathbf{p}_1}{(2\pi)^3}$ ("zeroth momentum moment"). This results in the standard Boltzmann equation for the DM number density, given by [30, 31]:

$$\dot{n} + 3Hn = -\langle \sigma v_{\rm rel}\rangle \left[n^2 - n_{\rm eq}^2\right]. \tag{2.18}$$

The thermally averaged cross-section reads [31]:

$$\begin{aligned} \langle \sigma v_{\rm rel}\rangle &\equiv \frac{g_1 g_2}{n_{\rm eq}^2(\beta)} \int \frac{{\rm d}^3\mathbf{p}_1}{(2\pi)^3} \frac{{\rm d}^3\mathbf{p}_2}{(2\pi)^3} \frac{(p_1p_2)}{E_1E_2} (\sigma v_{\rm rel}) e^{-\beta(E_1+E_2)} \\ &= \frac{2m_1\beta}{{K_2}^2(m_1\beta)} \int_1^\infty d\tilde{s}\, (\sigma v_{\rm rel}) \sqrt{\tilde{s}-1}(2\tilde{s}-1) K_1\Big(2\sqrt{\tilde{s}}m_1\beta\Big). \end{aligned} \tag{2.19}$$

Here, $\tilde{s} \equiv s/(4m_1^2)$, where $s = (p_1 + p_2)^2$ is the s-Mandelstam variable and (p_1p_2) are contracted four momenta. From the first to the second line, five of six integrals were performed analytically which allows for efficient numerical integration. It is a simple expression for the relativistic thermal average of the annihilation cross-section, where the latter quantity only depends on the s-Mandelstam variable.

The Boltzmann Eq. (2.18) applies to a broad class of DM models and is often used in literature to investigate the number density evolution. The quadratic form of the right hand side is characteristic for 2 ↔ 2 annihilation processes and can also occur in association with other number changing reactions featuring a different form. Decaying dark matter would have a linear dependence on the number density while Strongly-Interacting-Massive-Particles (SIMPs) [32] introduce a cubic dependence.

If annihilation processes are efficient, the right hand side of Eq. (2.18) forces DM to be in chemical equilibrium, i.e., $n = n_{\rm eq}$ is an attractor solution. Once annihilation processes become ineffective, the number density is conserved in a comoving volume and scales as $n \propto a^{-3}$. The details of the solution in between these asymptotic regimes is part of section 3.1.

In some cases and for later reference, it is simpler to consider the non-relativistic limit of Eq. (2.18), where finally one has to replace the equilibrium number density and the thermal average by

$$n_{\mathrm{eq}} \simeq g_1 \left(\frac{m_1 T}{2\pi}\right)^{3/2} e^{-\beta m_1}, \tag{2.20}$$

$$\langle \sigma v_{\mathrm{rel}} \rangle \simeq \frac{(m_1 \beta)^{3/2}}{2\sqrt{\pi}} \int_0^\infty (\sigma v_{\mathrm{rel}})\, e^{-\beta m_1 v_{\mathrm{rel}}^2/4}\, v_{\mathrm{rel}}^2 \, \mathrm{d}v_{\mathrm{rel}}. \tag{2.21}$$

The non-relativistic number density equation is often referred in the literature as the *Lee-Weinberg equation* [30]. The non-relativistic thermal average has the advantage that it can be computed analytically for annihilation cross-sections were it is appropriate to expand in the terms of relative velocity. For example, if $(\sigma v_{\mathrm{rel}})$ is velocity independent, the thermal average is simply given by $\langle \sigma v_{\mathrm{rel}} \rangle = (\sigma v_{\mathrm{rel}})$.

The number density of DM particles can not be measured directly. Instead it is the total DM energy density ρ_{DM}, which gravitationally influences, e.g., the evolution of the CMB temperature anisotropies. The Planck satellite measurements [1, 2] of the CMB temperature anisotropies have nowadays determined the fractional amount of the dark matter energy density $\Omega_{\mathrm{DM}} \equiv \rho_{\mathrm{DM}}/\rho_{\mathrm{cr}}$ to an unprecedented precision.[2] The observed value of the energy fraction at present day is $\Omega_{\mathrm{DM}} h^2 = 0.1188 \pm 0.0010$ (TT,TE,EE+lowP+lensing+ext. analysis in ref. [1]), reaching sub-percentage accuracy when combining CMB and external data, e.g., measurements of the baryon acoustic oscillation.

In general, the present fractional energy density of a non-relativistic single species χ can be written as:

$$\Omega_\chi = \left.\frac{m_\chi n_\chi}{\rho_{\mathrm{cr}}}\right|_{T_{\mathrm{CMB}}}. \tag{2.22}$$

This relates the observable abundance to the theoretically predicted value of n_χ today (at the present day CMB temperature $T_{\mathrm{CMB}} = 2.7255 \pm 0.0006$ K [33]). For annihilating dark matter, n_χ would be the solution of the number density Eq. (2.18). The total abundance is the sum over all species contributions and should match the observed value.

The number density equation for annihilating DM can not be solved analytically due to the non-linear structure. Following standard estimates in ref. [34], we obtain for a constant WIMP annihilation cross-section:

$$\Omega_\chi h^2 \simeq 2.1 \times 10^{-6} \left(\frac{x_{\mathrm{cd}}}{25}\right) \left(\frac{\sqrt{g_{\mathrm{eff}}}}{g_s/10}\right) \left[(100\,\mathrm{GeV})^2\, (\sigma v_{\mathrm{rel}})_0\right]^{-1}. \tag{2.23}$$

For an annihilation cross-section of the form $(\sigma v_{\mathrm{rel}})_0 \sim (\alpha/m_\chi)^2$ with a fine structure constant $\alpha \sim 10^{-2}$ and a DM mass $m_\chi \sim 200\,\mathrm{GeV}$, the predicted value of the relic abundance would be of the order $\Omega_\chi h^2 \sim \mathcal{O}(0.1)$, which is close to the observed value. The observation that the correct relic abundance can be obtained for a fine structure constant value and a DM mass around the electroweak scale, is called the *WIMP miracle*. The chemical decoupling temperature $x_{\mathrm{cd}} \equiv m_\chi/T \sim 25$ has only a mild logarithmic dependence on the WIMP cross-section and other quantities like effective degrees of freedom (see [34] for further details).

---

[2] The critical density today is given by $\rho_{\mathrm{cr}} = 3H_0^2/8\pi G$, where the Hubble expansion rate today H_0 is commonly expressed in terms of the dimensionless quantity h as $H_0 = 100 \times h$ km/s/Mpc.

Here, g_s and g_{eff} are the SM entropy and effective degrees of freedom, respectively. In the estimate they are assumed to be constant and to have a value of the order $\mathcal{O}(100)$ at typical WIMP chemical decoupling temperatures.

Not all of the DM abundance has to be made up by a single species χ. Therefore, the predicted abundance should always fulfil $\Omega_\chi h^2 \leq \Omega_{\text{DM}} h^2$ to be compatible with the observed CDM fractional energy density. This *overclosure bound* is important for identifying the viable parameter region of a model. In the WIMP miracle example, where the annihilation cross-section is given by $(\sigma v_{\text{rel}})_0 \sim (\alpha/m_\chi)^2$, the border of overclosure leads to an upper limit on the WIMP mass m_χ for fixed coupling value. For larger mass values above this limit, the cross-section would be too small, the WIMP would chemically decouple too early, and the Universe would contain too much DM as can be seen from Eq. (2.23).

2.3 Kinetic decoupling temperature and mass of first protohalos

The case is considered where scattering processes, keeping DM in kinetic equilibrium, start to become inefficient compared to the Hubble expansion rate. An assumption in the case of thermally produced DM particles is that this out-of-kinetic equilibrium transition takes place when annihilation processes can be neglected, i.e., the first term on the right hand side of Eq. (2.12) is dropped in the following. Even without the annihilation term, it is in general a difficult task to find a closed set of momentum moments describing kinetic decoupling for an arbitrary phase-space distribution. However, within a few assumptions on the particle properties it is possible to simplify the scattering collision term in such a way that one single primitive variable is already sufficient to describe the evolution of the system.

If DM is non-relativistic during the decoupling from the relativistic heat bath, momentum transfer per typical collision is much smaller compared to the DM momentum. In this case, the scattering collision term can be expanded up to the second order in small momentum transfer compared to the DM momentum, resulting in the DM non-relativistic limit in a simple Fokker-Planck type form [35, 36]:

$$C_{\text{sc}} \simeq \gamma(T)\frac{m_1}{2}\left[T m_1 \partial_p^2 + \left(p + 2T\frac{m_1}{p}\right)\partial_p + 3\right] f_1. \tag{2.24}$$

A stationary solution of this Fokker-Planck operator is the non-relativistic Maxwell-Boltzmann distribution $f_1 \propto e^{-\beta p^2/(2m_1)}$. The *momentum transfer rate* γ is given by [37, 38][3]

---

[3] In the work by Bertschinger [35], the Fokker-Planck operator Eq. (2.24) was used but for specific $\gamma(T)$ (for bino-type models). Bringmann et al. [36] derived Eq. (2.24) in the correct formal expansion up to the second order in small momentum transfer, resulting in an broader applicable $\gamma(T)$ but without Mandelstam t-average and the scattering amplitude at $t = 0$. In an unpublished thesis by Kasahara [37], higher order terms were taken into account by t-averaging the amplitude as in Eq. (2.25), which was published by Gondolo

$$\gamma(T) \equiv \frac{1}{3 m_1 T g_1} \int \frac{\mathrm{d}^3\mathbf{p}_3}{(2\pi)^3} f_3^{\mathrm{eq}} \left(1 \pm f_3^{\mathrm{eq}}\right) \frac{|\mathbf{p}_3|}{E_3} \int_{-4\mathbf{p}_3^2}^{0} \mathrm{d}t(-t) \frac{\mathrm{d}\sigma}{\mathrm{d}t}, \tag{2.25}$$

where

$$\frac{\mathrm{d}\sigma}{\mathrm{d}t} \equiv \frac{1}{64\pi} \frac{1}{|\mathbf{p}_3|^2 m_1^2} |\mathcal{M}|^2_{s=m_1^2+2m_1|\mathbf{p}_3|}. \tag{2.26}$$

Defining the DM temperature as

$$T_\chi \equiv \frac{g_1}{n} \frac{2}{3} \int \frac{\mathrm{d}^3\mathbf{p}_1}{(2\pi)^3} \frac{\mathbf{p}_1^2}{2m_1} f_1 \tag{2.27}$$

and integrating the Boltzmann equation in the Fokker-Planck approximation over $\int \frac{\mathrm{d}^3\mathbf{p}_1}{(2\pi)^3} \frac{\mathbf{p}_1^2}{2m_1}$, called the "second momentum moment", results in an ordinary differential equation for the DM temperature [36]:

$$\dot{T}_\chi + 2HT_\chi = -\gamma(T)\left[T_\chi - T\right]. \tag{2.28}$$

The solution of this equation describes the temperature evolution of non-relativistic DM coupled to heat bath particles with temperature T. In the tight coupling regime, where $\gamma \gg H$, the right hand side dominates and $T_\chi = T$ is an attractor solution. After kinetic decoupling ($\gamma \ll H$) the left hand side determines the asymptotic behaviour, leading to $T_\chi \propto a^{-2}$ as expected for non-relativistic particles. In this thesis, we define the *kinetic decoupling temperature* T_{kd} as the temperature which solves $\gamma/H = 1$ (where both rates are comparable).[4] The kinetic decoupling temperature has a direct relation to the small scale structure of DM, introduced in the following.

Primordial density perturbations in the dark matter fluid are initial seeds, leading to the formation of cosmological structure such as galaxies. In the cold and collisionless DM (CDM) paradigm, structure formation takes place hierarchically. Density modes on smaller scales enter the horizon first. After horizon entry, the regions containing more DM start to grow and form the first gravitationally bound objects in the Universe, called protohalos. The

---

et al. in ref. [38]. The t-average allows to describe kinetic decoupling for more general scattering amplitudes. In the article reprinted in section 4.3, collaborators derived the Fokker-Planck collision term including t-average in an alternative formal expansion as in [36, 37]. In the deep non-relativistic regime and assuming heat bath particles to be massless, the results of the article reprinted in section 4.3 coincide with Eq. (2.24) and Eq. (2.25).

[4] T_{kd} can also be defined as the temperature which solves $y_{\mathrm{eq}}(x) = y(x)|_{x\to\infty}$, see [7, 36]. Both conventions lead on a level of 10% to the same result for the models considered here in this thesis.

CDM paradigm predicts a certain amount of these clumps down to essentially arbitrarily small scales.

Since thermally produced dark matter differs from pure CDM at early times, one might expect that the microphysics of interacting DM modifies the formation and evolution history of the smallest protohalos. In fact, there are mainly two mechanisms, relevant for this thesis, generating a significant damping of the amplitude of density perturbations below a certain length scale (see [39] for a rather general classification).

The first mechanism describes the effect of collisionless damping. After DM has kinetically decoupled from the relativistic heat bath, it can travel freely from regions containing a higher density to lower density regions. Thereby it washes out the density contrast and reduces the amplitude of the perturbations, where the efficiency depends on the thermal velocity at decoupling. It has been shown that this free-streaming effect leads to an exponential suppression of the density contrast above a characteristic comoving wave number denoted by $k_{\rm fs}$ [40–42]. This wave number can be estimated analytically (see pp. 13 in ref. [42]) as a function of the DM mass and the kinetic decoupling temperature $T_{\rm kd}$. The total dark matter mass inside a sphere with characteristic radius $\pi/k_{\rm fs}$ is given by (see, e.g., [7]):

$$M_{\rm fs} = \frac{4\pi}{3}\left(\frac{\pi}{k_{\rm fs}}\right)^3 \rho_{\rm DM}\Big|_{\rm kd} = 2.9\times 10^{-6}\left(\frac{1+\ln\left(g_{\rm eff}^{1/4}T_{\rm kd}/50\text{ MeV}\right)/19.1}{\left(\frac{m}{100\text{ GeV}}\right)^{1/2} g_{\rm eff}^{1/4}\left(\frac{T_{\rm kd}}{50\text{ MeV}}\right)^{1/2}}\right)^3 M_\odot, \tag{2.29}$$

where $g_{\rm eff}$ is evaluated at the kinetic decoupling (kd) temperature. The quantity $M_{\rm fs}$ is an estimate of the critical mass of the first protohalos, below which the halo abundance starts to become exponentially suppressed compared to pure CDM due to the efficient damping of density perturbations by the free-streaming effect.

The second damping mechanism describes the evolution of density modes that enter the horizon while DM is in kinetic equilibrium with the radiation fluid. During kinetic equilibrium, the density perturbations of DM inside the horizon can not grow but oscillate due to the competition between radiative pressure and gravitational contraction [35, 43]. In the latter references, it has been pointed out that this dark acoustic oscillation effect can even dominate over the damping induced by free-streaming, especially if DM decouples when being highly non-relativistic. The critical protohalo mass, below which the halo abundance is suppressed due to dark acoustic oscillation, can be estimated by the mass inside a sphere with the radius of the Hubble horizon at the time of DM kinetic decoupling (see, e.g., [7]):

$$M_{\rm ao} = \frac{4\pi}{3}\left(\frac{1}{H}\right)^3 \rho_{\rm DM}\Big|_{\rm kd} = 3.4\times 10^{-6}\left(\frac{g_{\rm eff}^{1/4}T_{\rm kd}}{50\text{MeV}}\right)^{-3} M_\odot. \tag{2.30}$$

Since the damping mechanisms of free streaming and dark acoustic oscillations are within these estimates physically independent, the minimum cut-off mass can be defined as [7]

$$M_{\rm cut} \equiv \max[M_{\rm fs}, M_{\rm ao}]. \tag{2.31}$$

In the article reprinted in section 4.3, we consider DM models that feature cut-off masses at the dwarf galaxy mass scale. Analytic scaling patterns of $M_{\rm cut}$ in terms of model parameters are derived.

Chapter 3

Thermal decoupling

In section 3.1, the standard chemical and kinetic decoupling description of thermally produced DM is shown. The article reprinted in section 3.2 points out and investigates a new exception, where the standard description fails.

3.1 Standard chemical and kinetic decoupling

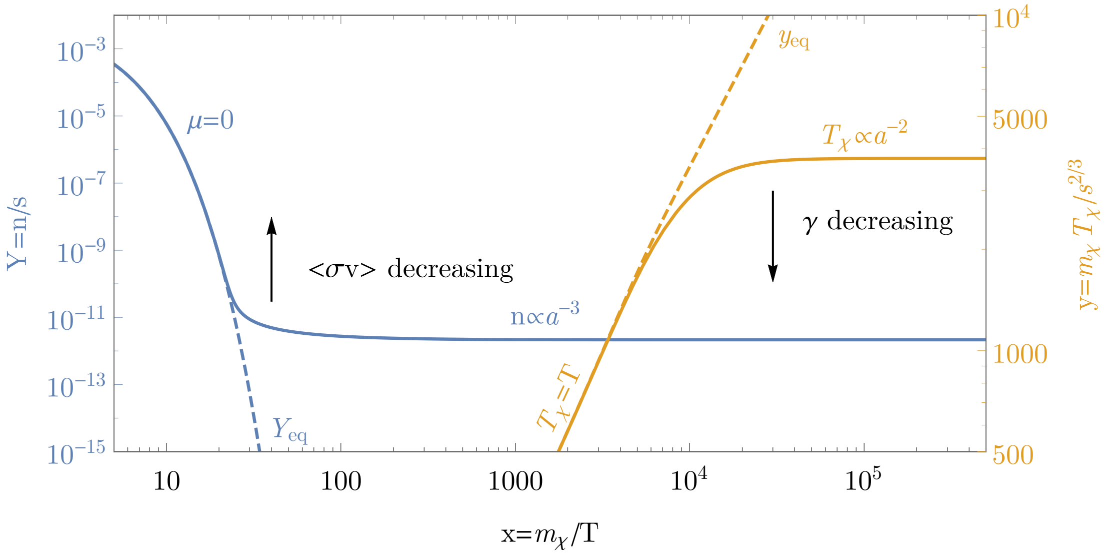


Fig. 1: Typical decoupling history of thermally produced dark matter shown. Solid lines are the numerical solution of Eq. (3.1) and (3.2), while dashed lines denote the equilibrium values.

The standard chemical and kinetic decoupling process for thermally produced dark matter is introduced in this section. Dimensionless variables are introduced as $Y \equiv n/s$, $y =$

$m_\chi T_\chi / s^{2/3}$ and $x \equiv m_\chi / T$ in order to bring Eq. (2.18) and (2.28) into standard form (see, e.g., [31, 36]):

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = -\frac{\langle \sigma v_{\mathrm{rel}} \rangle s}{\tilde{H} x} \left[Y^2 - Y_{\mathrm{eq}}^2(x) \right] , \tag{3.1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}x} = -\frac{\gamma(x)}{\tilde{H} x} \left[y - y_{\mathrm{eq}}(x) \right] . \tag{3.2}$$

To arrive at those equations, entropy conservation $a^{-3}\partial_t(a^3 s) = 0$ was used, where $s = (2\pi^2/45) g_s T^3$. The effective entropy degrees of freedom g_s are temperature dependent and $\tilde{H} \equiv H / \left[1 + \tilde{g}(x)\right]$, where

$$\tilde{g} \equiv \frac{1}{3} \frac{T}{g_s} \frac{\mathrm{d}g_s}{\mathrm{d}T} . \tag{3.3}$$

In the radiation dominated epoch, the Hubble expansion rate can be written as $H = \sqrt{g_{\mathrm{eff}} 4\pi^3/45 T^2/m_{\mathrm{pl}}}$, where $m_{\mathrm{pl}} \approx 1.22 \times 10^{19}$ GeV is the Planck mass and g_{eff} are the effective degrees of freedom. Throughout this thesis, the values of g_s and g_{eff} are adopted from [44].

In figure 1, the numerical solution of Eq. (3.1) and (3.2) is shown. For a quantitative discussion, the DM mass $m_\chi = 100$ GeV, the annihilation cross-section $\langle \sigma v_{\mathrm{rel}} \rangle = \text{const}$ and momentum transfer rate $\gamma \propto T^6$ were chosen to represent a typical situation. At the initial time $x \lesssim 10$ the coupling of DM to the BPs is tight. Scattering and annihilation processes are very efficient leading to kinetic and chemical equilibrium. In equilibrium, the DM temperature is equal to the heat bath temperature ($y = y_{\mathrm{eq}}$) and the chemical potential μ equals zero ($Y = Y_{\mathrm{eq}}$). These are the attractor solutions of Eq. (3.1) and (3.2), shown by the dashed lines in figure 1. Any deviation from the equilibrium values will be efficiently restored in the high temperature regime.

The chemical equilibrium is maintained as long as the annihilation rate $\Gamma_{\mathrm{an}} \equiv \langle \sigma v_{\mathrm{rel}} \rangle Y s$ is much larger compared to H. When DM enters the non-relativistic regime during chemical equilibrium, the value of Γ_{an} decreases exponentially fast in x due to the Boltzmann suppression in Y_{eq} [see, e.g., Eq. (2.20)]. At some point, the annihilation rate falls behind H and the number density starts to freeze out, i.e., it approaches $Y = \text{const}$ or equivalently $n \propto a^{-3}$. The transition point, where the number density starts to deviate from its equilibrium value, is called chemical decoupling and happens in figure 1 at about $x \sim 25$. The final number density increases for decreasing annihilation cross-section because DM decouples earlier.

Kinetic decoupling takes place at a time when the momentum transfer rate becomes comparable to the Hubble expansion rate. In the example shown, this happens around $x \sim 10^4$. After kinetic decoupling, the temperature of non-relativistic DM scales as $T_\chi \propto a^{-2}$ or $y = \text{const}$. The decoupling process itself runs on rather longer time scales compared to the chemical one and takes place earlier for decreasing momentum transfer rate.

In the article reprinted in the next section, the main assumption of kinetic equilibrium during chemical decoupling is scrutinized.

3.2 *Publication:* Early kinetic decoupling of dark matter: when the standard way of calculating the thermal relic density fails

Reprinted article with permission from the authors:
Tobias Binder, Torsten Bringmann, Michael Gustafsson, and Andrzej Hryczuk.
Editor's suggestion, Physical Review D **96**, 115010 (2017)
https://doi.org/10.1103/PhysRevD.96.115010
Copyright 2017 by the American Physical Society.

Author contributions T. Bringmann suggested to explore early kinetic decoupling and conducted the project. M. G. developed the numerical code solving the Boltzmann equation on phase-space density level. A. H. computed the details of the Scalar Singlet model and implemented the coupled system of momentum moments. T. Binder and T. Bringmann supported the numerical code developments during an early phase, contributed with analytic computations, and constructed the relativistic generalization of the coupled system of momentum moments. T. Binder suggested to use a semi-relativistic version of the Fokker-Planck operator and worked out the details. M. G. and A. H. produced all numerical results. All authors contributed in writing the article, while T. Bringmann wrote most of the manuscript.

PHYSICAL REVIEW D **96,** 115010 (2017)

Early kinetic decoupling of dark matter: When the standard way of calculating the thermal relic density fails

Tobias Binder,[1,*] Torsten Bringmann,[2,†] Michael Gustafsson,[1,‡] and Andrzej Hryczuk[2,§]

[1]*Institute for Theoretical Physics, Georg-August University Göttingen, Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany*

[2]*Department of Physics, University of Oslo, Box 1048, NO-0316 Oslo, Norway*

(Received 31 August 2017; published 15 December 2017)

Calculating the abundance of thermally produced dark matter particles has become a standard procedure, with sophisticated methods guaranteeing a precision that matches the percent-level accuracy in the observational determination of the dark matter density. Here, we point out that one of the main assumptions in the commonly adopted formalism, namely local thermal equilibrium during the freeze-out of annihilating dark matter particles, does not have to be satisfied in general. We present two methods for how to deal with such situations, in which the kinetic decoupling of dark matter happens so early that it interferes with the chemical decoupling process: (i) an approximate treatment in terms of a coupled system of differential equations for the leading momentum moments of the dark matter distribution, and (ii) a full numerical solution of the Boltzmann equation in phase space. For illustration, we apply these methods to the case of scalar singlet dark matter. We explicitly show that even in this simple model the prediction for the dark matter abundance can be affected by up to 1 order of magnitude compared to the traditional treatment.

DOI: 10.1103/PhysRevD.96.115010

I. INTRODUCTION

The leading hypothesis for the cosmological dark matter (DM) [1,2] is a new type of elementary particle [3]. One of the most attractive options to explain the present abundance of these particles consists in the possibility that they have been thermally produced in the early universe. This is particularly interesting for the scenario originally studied by Lee and Weinberg [4], as well as others [5–8], in which nonrelativistic DM particles initially are kept in thermal equilibrium with the heat bath through frequent annihilation and creation processes with standard model (SM) particles. Once the interaction rate starts to fall behind the expansion rate of the universe, the DM number density begins to "freeze-out" and remains covariantly conserved. For weakly interacting massive particles (WIMPs), elementary particles with masses and interaction strengths at the electroweak scale, this scenario automatically leads to a relic abundance in rough agreement with the observed DM density—a fact sometimes referred to as the *WIMP miracle*.

The by now standard treatment [9,10] of calculating the resulting DM abundance in these scenarios implements an efficient and highly accurate method of solving the Boltzmann equation for a given (effective) invariant DM annihilation rate. This approach fully captures, in particular, the three famous exceptions to the original relic density calculations pointed out in a seminal paper by Griest and Seckel [11], namely coannihilations, threshold effects and resonances. The main assumption entering this formalism is that, during the freeze-out process, DM is still kept in *local* thermal equilibrium with the heat bath by frequent scattering processes with relativistic SM particles. For many WIMP candidates, this is indeed satisfied to a high accuracy and kinetic decoupling typically only happens much later than the chemical decoupling [12].

Here we point out that exceptions to this standard lore do exist, even in very simple scenarios, where kinetic decoupling happens so early that it cannot be neglected during the freeze-out process. We develop both semianalytical and fully numerical methods to solve the Boltzmann equation and to compute the DM relic abundance in these cases. Technically, one of the challenges that had to be overcome for obtaining sufficiently accurate results was to extend the highly nonrelativistic Boltzmann equation, as discussed previously in the literature, to the semirelativistic regime. Numerically, we also succeeded to resolve the evolution of the full phase-space distribution accurately enough to test, for the first time, the underlying assumptions for the standard way of calculating the relic density of WIMPs or other self-annihilating DM candidates (for a recent example where the relic density is instead set by inelastic scattering, rather than self-annihilation, see Ref. [13]). We illustrate our general results by a detailed discussion of the scalar singlet model [14–16], for which we find a DM relic

---

[*] tobias.binder@theorie.physik.uni-goettingen.de
[†] torsten.bringmann@fys.uio.no
[‡] michael.gustafsson@theorie.physik.uni-goettingen.de
[§] a.j.hryczuk@fys.uio.no

© 2017 American Physical Society

density that differs by up to an order of magnitude from the standard treatment.

This article is organized as follows. In Sec. II, we start with a general description of the underlying Boltzmann equation that governs the DM phase-space evolution. We then briefly review the standard treatment of solving for the DM number density (Sec. II A), extend this by deriving a *coupled* system of evolution equations for the number density and the velocity dispersion (Sec. II B), and finally introduce our framework for a fully numerical solution (Sec. II C). Section III is devoted to a thorough application of these methods to the scalar singlet model. We comment on our results in Sec. IV, and discuss potential other areas of application, before we conclude in Sec. V. In two Appendices we discuss in detail the evolution of the singlet DM phase-space density for selected parameter points (Appendix A) and comment on the semirelativistic form of the scattering operator in the Boltzmann equation (Appendix B).

II. THERMAL PRODUCTION OF DARK MATTER

Let us denote the DM particle by χ, and its phase-space density by $f_\chi(t,\mathbf{p})$. The evolution of f_χ is governed by the Boltzmann equation which, in an expanding Friedmann-Robertson-Walker universe, is given by [17,18]

$$E(\partial_t - H\mathbf{p}\cdot\nabla_{\mathbf{p}})f_\chi = C[f_\chi]. \tag{1}$$

Here, $H=\dot{a}/a$ is the Hubble parameter, a the scale factor, and the collision term $C[f_\chi]$ contains all interactions between DM and SM particles f. For WIMPs, we are to leading order interested in two-body processes for DM annihilation and elastic scattering, $C = C_{\rm ann} + C_{\rm el}$, where

$$\begin{aligned}C_{\rm ann} = {}&\frac{1}{2g_\chi}\int\frac{d^3\tilde{p}}{(2\pi)^3 2\tilde{E}}\int\frac{d^3k}{(2\pi)^3 2\omega}\int\frac{d^3\tilde{k}}{(2\pi)^3 2\tilde{\omega}}\\ &\times(2\pi)^4\delta^{(4)}(\tilde{p}+p-\tilde{k}-k)\\ &\times[|\mathcal{M}|^2_{\bar{\chi}\chi\leftarrow\bar{f}f}\,g(\omega)g(\tilde{\omega}) - |\mathcal{M}|^2_{\bar{\chi}\chi\rightarrow\bar{f}f}\,f_\chi(E)f_\chi(\tilde{E})],\end{aligned} \tag{2}$$

and

$$\begin{aligned}C_{\rm el} = {}&\frac{1}{2g_\chi}\int\frac{d^3k}{(2\pi)^3 2\omega}\int\frac{d^3\tilde{k}}{(2\pi)^3 2\tilde{\omega}}\int\frac{d^3\tilde{p}}{(2\pi)^3 2\tilde{E}}\\ &\times(2\pi)^4\delta^{(4)}(\tilde{p}+\tilde{k}-p-k)|\mathcal{M}|^2_{\chi f\leftrightarrow\chi f}\\ &\times[(1\mp g^\pm)(\omega)g^\pm(\tilde{\omega})f_\chi(\tilde{\mathbf{p}}) - (\omega\leftrightarrow\tilde{\omega},\mathbf{p}\leftrightarrow\tilde{\mathbf{p}})].\end{aligned} \tag{3}$$

In the above expressions, $|\mathcal{M}|^2$ refers to the respective squared amplitude, summed over *all* spin and other internal degrees of freedom, as well as all SM particles f. We assume the SM particles to be in thermal equilibrium, such that their phase-space distribution is given by $g^\pm(\omega) = 1/[\exp(\omega/T)\pm 1]$. Note that we have neglected Bose enhancement and Pauli blocking factors for f_χ here, as we assume DM to be nonrelativistic; momentum conservation then implies that, in $C_{\rm ann}$, we can also neglect these factors for the SM particles.

Assuming *CP* invariance, and using the fact that in thermal equilibrium annihilation and creation processes should happen with the same frequency, the annihilation term given by Eq. (2) can be further simplified to [9]

$$\begin{aligned}C_{\rm ann} = {}&g_\chi E\int\frac{d^3\tilde{p}}{(2\pi)^3}v\sigma_{\bar{\chi}\chi\rightarrow\bar{f}f}\\ &\times[f_{\chi,\rm eq}(E)f_{\chi,\rm eq}(\tilde{E}) - f_\chi(E)f_\chi(\tilde{E})],\end{aligned} \tag{4}$$

where $v = v_{\rm M\o l}\equiv(E\tilde{E})^{-1}[(p\cdot\tilde{p})^2 - m_\chi^4]^{1/2}$ is the Møller velocity, which in the rest frame of one of the DM particles coincides with the lab velocity $v_{\rm lab} = [s(s-4m_\chi^2)]^{1/2}/(s-2m_\chi^2)$.

The scattering term, on the other hand, is in general considerably more difficult to manage. Analytic expressions have, however, been obtained in the highly nonrelativistic limit of the DM particles, and assuming that the momentum transfer in the scattering process is much smaller than the DM mass [12,18–22]:

$$C_{\rm el}\simeq\frac{m_\chi}{2}\gamma(T)\left[Tm_\chi\partial_p^2 + \left(p + 2T\frac{m_\chi}{p}\right)\partial_p + 3\right]f_\chi, \tag{5}$$

where the momentum exchange rate is given by

$$\gamma(T) = \frac{1}{48\pi^3 g_\chi m_\chi^3}\int d\omega\, g^\pm\partial_\omega(k^4\langle|\mathcal{M}|^2\rangle_t), \tag{6}$$

with

$$\langle|\mathcal{M}|^2\rangle_t\equiv\frac{1}{8k^4}\int_{-4k_{\rm cm}^2}^{0}dt(-t)|\mathcal{M}|^2 = 16\pi m_\chi^2\sigma_T, \tag{7}$$

and $k_{\rm cm}^2 = (s-(m_\chi-m_f)^2)(s-(m_\chi+m_f)^2)/(4s)$ evaluated at $s = m_\chi^2+2\omega m_\chi+m_f^2$. Here, $\sigma_T = \int d\Omega(1-\cos\theta)d\sigma/d\Omega$ is the standard *transfer cross section* for elastic scattering. In Appendix B, we discuss how the scattering term is expected to change in the semirelativistic case, i.e. when the assumption of highly nonrelativistic DM is slightly relaxed. For reference, we will in the following use

$$C_{\rm el}\simeq\frac{E}{2}\gamma(T)\left[TE\partial_p^2 + \left(p + 2T\frac{E}{p} + T\frac{p}{E}\right)\partial_p + 3\right]f_\chi \tag{8}$$

when explicitly addressing this regime.

A. The standard treatment

In order to calculate the DM relic abundance, we can integrate the Boltzmann Eq. (1) over $\mathbf{p}$. This results in

$$\frac{dn_\chi}{dt} + 3Hn_\chi = g_\chi \int \frac{d^3p}{(2\pi)^3E} C_{\rm ann}[f_\chi], \tag{9}$$

which has to be solved for the DM number density

$$n_\chi = g_\chi \int d^3p/(2\pi)^3 f_\chi(\mathbf{p}) \tag{10}$$

(note that $C_{\rm el}$ vanishes once it is integrated over). In order to evaluate the rhs of this equation, the usual assumption [9] is that during chemical freeze-out one can make the following ansatz for the DM distribution:

$$f_\chi = A(T) f_{\chi,\rm eq} = \frac{n_\chi}{n_{\chi,\rm eq}} f_{\chi,\rm eq}, \tag{11}$$

where $A(T) = 1$ in full equilibrium, i.e. before chemical freeze-out. This is motivated by the fact DM-SM scattering typically proceeds at a much faster rate than DM-DM annihilation, because the number density of relativistic SM particles is not Boltzmann suppressed like that of the nonrelativistic DM particles. In that case, DM particles are kept in *local* thermal equilibrium even when the annihilation rate starts to fall behind the Hubble expansion and chemical equilibrium can no longer be maintained.

Approximating furthermore $f_{\chi,\rm eq}(E) \simeq \exp(-E/T)$, i.e. neglecting the impact of quantum statistics for nonrelativistic particles, five of the six integrals in Eq. (9) can be performed analytically. This by now standard treatment, as established by Gondolo and Gelmini [9], results in the often-quoted expression

$$\frac{dn_\chi}{dt} + 3Hn_\chi = \langle\sigma v\rangle (n_{\chi,\rm eq}^2 - n_\chi^2), \tag{12}$$

where $n_{\chi,\rm eq} = g_\chi m_\chi^2 T K_2(m_\chi/T)/(2\pi^2)$ and

$$\langle\sigma v\rangle \equiv \frac{g_\chi^2}{n_{\chi,\rm eq}^2} \int \frac{d^3p}{(2\pi)^3} \frac{d^3\tilde p}{(2\pi)^3} \sigma v_{\bar\chi\chi\to\bar f f} f_{\chi,\rm eq}(\mathbf{p}) f_{\chi,\rm eq}(\tilde{\mathbf{p}}) \tag{13}$$

$$= \int_1^\infty d\tilde s \sigma_{\bar\chi\chi\to\bar f f} v_{\rm lab} \frac{2m_\chi\sqrt{\tilde s - 1}(2\tilde s - 1)K_1(\frac{2\sqrt{\tilde s}m_\chi}{T})}{T K_2{}^2(m_\chi/T)}. \tag{14}$$

Here, K_i are the modified Bessel functions of order i, and we have introduced $\tilde s \equiv s/(4m_\chi^2)$. While there are various ways to state the final result for $\langle\sigma v\rangle$, the form given above stresses that physically one should indeed think of this quantity as a thermal average of $\sigma v_{\rm lab}$ rather than any other combination of cross section and velocity (in the sense that we strictly have $\langle\sigma v\rangle = \sigma v_{\rm lab}$ for $\sigma v_{\rm lab} =$ const; for e.g., $\sigma v_{\rm CMS} =$ const, on the other hand, with $v_{\rm CMS} = 2(1 - 4m_\chi^2/s)$, we instead have $\langle\sigma v\rangle \to \sigma v_{\rm CMS}$ only in the limit $T \to 0$).

By introducing dimensionless variables

$$x \equiv m_\chi/T, \tag{15}$$

$$Y \equiv n_\chi/s, \tag{16}$$

and assuming entropy conservation, finally, the above Boltzmann equation for the number density, Eq. (12), can be brought into an alternative form that is particularly suitable for numerical integration:

$$\frac{Y'}{Y} = \frac{sY}{x\tilde H}\langle\sigma v\rangle \left[\frac{Y_{\rm eq}^2}{Y^2} - 1\right]. \tag{17}$$

Here, $s = (2\pi^2/45) g_{\rm eff}^s T^3$ denotes the entropy density, $' \equiv d/dx$ and $\tilde H \equiv H/[1 + \tilde g(x)]$, where

$$\tilde g \equiv \frac{1}{3}\frac{T}{g_{\rm eff}^s}\frac{dg_{\rm eff}^s}{dT}. \tag{18}$$

The value of Y today, $Y_0 \equiv Y(x \to \infty)$, can then be related to the observed DM abundance by [9]

$$\Omega_\chi h^2 = 2.755 \times 10^{10} \left(\frac{m_\chi}{100\ \rm GeV}\right)\left(\frac{T_{\rm CMB}}{2.726\ \rm K}\right)^3 Y_0. \tag{19}$$

We note that Eq. (17) is the basis for the implementation of relic density calculations in all major numerical codes [23–29].

B. Coupled Boltzmann equations

The main assumption that enters the standard treatment reviewed above is contained in Eq. (11), i.e. the requirement that during chemical freeze-out, or in fact during any period when the comoving DM density changes, local thermal equilibrium with the heat bath is maintained. If that assumption is not justified, one has in principle to solve the full Boltzmann equation in phase space, Eq. (1), numerically (see next subsection). As first pointed out in Ref. [30], however, it sometimes suffices to take into account the second moment of Eq. (1), instead of only the zeroth moment as in the previous subsection. This leads to a relatively simple *coupled* system of differential equations that generalizes Eq. (17).

The starting point is to define, in analogy to Y for the zeroth moment of f_χ, a dimensionless version of the *second* moment of f_χ:

$$y \equiv \frac{m_\chi}{3s^{2/3}}\left\langle\frac{\mathbf{p}^2}{E}\right\rangle = \frac{m_\chi}{3s^{2/3}}\frac{g_\chi}{n_\chi}\int\frac{d^3p}{(2\pi)^3}\frac{\mathbf{p}^2}{E} f_\chi(\mathbf{p}). \tag{20}$$

For a thermal distribution, the DM particles thus have a *temperature*

$$T_\chi = y s^{2/3}/m_\chi. \tag{21}$$

We note that for nonthermal distributions we could still view this last equation as an alternative *definition* of the DM "temperature," or velocity dispersion, in terms of the second moment of f_χ as introduced above. This allows, e.g., a convenient characterization of kinetic decoupling as the time when T_χ no longer equals T but instead starts to approach the asymptotic scaling of $T_\chi = T_{\rm kd}(a/a_{\rm eq})^{-2}$ for highly nonrelativistic DM [12,18].

Integrating Eq. (1) over $g_\chi \int d^3p/(2\pi)^3/E$ and $g_\chi \int d^3p/(2\pi)^3 \mathbf{p}^2/E^2$, respectively, we find

$$\frac{Y'}{Y} = \frac{m_\chi}{x\tilde{H}} C_0, \tag{22}$$

$$\frac{y'}{y} = \frac{m_\chi}{x\tilde{H}} C_2 - \frac{Y'}{Y} + \frac{H}{x\tilde{H}} \frac{\langle p^4/E^3\rangle}{3T_\chi}, \tag{23}$$

where

$$\langle p^4/E^3\rangle \equiv n_\chi^{-1} g_\chi \int \frac{d^3p}{(2\pi)^3} \frac{\mathbf{p}^4}{E^3} f_\chi(\mathbf{p}) \tag{24}$$

and we introduced the moments of the collision term as

$$m_\chi n_\chi C_0 \equiv g_\chi \int \frac{d^3p}{(2\pi)^3 E} C[f_\chi], \tag{25}$$

$$m_\chi n_\chi \left\langle \frac{\mathbf{p}^2}{E} \right\rangle C_2 \equiv g_\chi \int \frac{d^3p}{(2\pi)^3 E} \frac{\mathbf{p}^2}{E} C[f_\chi]. \tag{26}$$

Plugging in $C = C_{\rm ann} + C_{\rm el}$ as provided in Eqs. (4) and (5), finally, we arrive at a coupled set of equations that constitutes one of our main results[1]:

$$\frac{Y'}{Y} = \frac{sY}{x\tilde{H}} \left[\frac{Y_{\rm eq}^2}{Y^2} \langle \sigma v \rangle - \langle \sigma v \rangle_{\rm neq} \right], \tag{27}$$

$$\frac{y'}{y} = \frac{\gamma(T)}{x\tilde{H}} \left[\frac{y_{\rm eq}}{y} - 1 \right] + \frac{sY}{x\tilde{H}} [\langle \sigma v \rangle_{\rm neq} - \langle \sigma v \rangle_{2,\rm neq}] + \frac{sY}{x\tilde{H}} \frac{Y_{\rm eq}^2}{Y^2} \left[\frac{y_{\rm eq}}{y} \langle \sigma v \rangle_2 - \langle \sigma v \rangle \right] + \frac{H}{x\tilde{H}} \frac{\langle p^4/E^3\rangle_{\rm neq}}{3T_\chi}. \tag{28}$$

Here, in addition to $\langle \sigma v \rangle$ in Eq. (13), we also introduced another, temperature-weighted thermal average:

$$\langle \sigma v \rangle_2 \equiv \frac{g_\chi^2}{T n_{\chi,\rm eq}^2} \int \frac{d^3p\, d^3\tilde{p}}{(2\pi)^6} \frac{p^2}{3E} \sigma v_{\bar{\chi}\chi \to \bar{f}f} f_{\chi,\rm eq}(\mathbf{p}) f_{\chi,\rm eq}(\tilde{\mathbf{p}}) \tag{29}$$

$$\begin{aligned} &= \int_1^\infty d\tilde{s}\, \sigma_{\bar{\chi}\chi \to \bar{f}f} v_{\rm lab} \frac{4\tilde{s}(2\tilde{s}-1)x^3}{3{K_2}^2(x)} \\ &\quad \times \int_1^\infty d\epsilon_+ e^{-2\sqrt{\tilde{s}}x\epsilon_+} \Bigg[\epsilon_+ \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)} \\ &\quad + \frac{1}{2\sqrt{\tilde{s}}} \log\left(\frac{\sqrt{\tilde{s}}\epsilon_+ - \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)}}{\sqrt{\tilde{s}}\epsilon_+ + \sqrt{(\tilde{s}-1)(\epsilon_+^2-1)}} \right) \Bigg], \end{aligned} \tag{30}$$

where we have used $\epsilon_+ \equiv (E+\tilde{E})/\sqrt{s}$. The "out-of-equilibrium average" $\langle \sigma v \rangle_{2,\rm neq}$ is defined as in Eq. (29), but for arbitrary n_χ, $f_\chi(\mathbf{p})$—and hence also $1/T \to 1/T_\chi$ in the normalization; the last equality, Eq. (30), thus does not hold in this case. Correspondingly, $\langle \sigma v \rangle_{\rm neq}$ is defined in analogy to Eq. (13), but equals in general not the expression given in Eq. (14).

Two comments about this central result are in order. The first comment, more important from a practical point of view, is that the set of Eqs. (22) and (23) includes higher moments of f_χ, and hence does not close with respect to the variables Y and y. Concretely, we need additional input to determine the quantities $\langle \sigma v \rangle_{\rm neq}$, $\langle \sigma v \rangle_{2,\rm neq}$ and $\langle p^4/E^3\rangle_{\rm neq}$ in Eqs. (27) and (28) in terms of only y and Y. We will make the following ansatz for these quantities:

$$\langle \sigma v \rangle_{\rm neq} = \langle \sigma v \rangle|_{T = ys^{2/3}/m_\chi}, \tag{31}$$

$$\langle \sigma v \rangle_{2,\rm neq} = \langle \sigma v \rangle_2|_{T = ys^{2/3}/m_\chi}, \tag{32}$$

$$\langle p^4/E^3\rangle_{\rm neq} = \left[\frac{g_\chi}{2\pi^2 n_{\chi,\rm eq}(T)} \int dp \frac{p^6}{E^3} e^{-\frac{E}{T}} \right]_{T = ys^{2/3}/m_\chi}. \tag{33}$$

These expressions would, in particular, result from a DM phase-space distribution of the form

$$f_\chi = \frac{n_\chi(T)}{n_{\chi,\rm eq}(T_\chi)} \exp\left(-\frac{E}{T_\chi} \right) \bigg|_{T_\chi = ys^{2/3}/m_\chi}, \tag{34}$$

which describes a situation in which the DM particles follow a Maxwellian velocity distribution with a temperature different from that of the heat bath (as expected, e.g., if the DM particles exhibit significant self-scattering

---

[1] This extends the results presented in [30]. Compared to that reference, we have kept terms proportional to $Y_{\rm eq}$ (see also [31]) and adopted a fully relativistic temperature definition in Eqs. (20) and (21). The latter indeed turns out to be important outside the highly nonrelativistic regime and is the origin of the last term in Eq. (28), as well as the corrected form of $\langle \sigma v \rangle_2$—which now (unlike in its original form) can be seen as a proper thermal average in the sense that a constant $\sigma v_{\rm lab}$ leads to $\langle \sigma v \rangle_2 = \sigma v_{\rm lab}$ for *all* values of T (i.e. not only for $T \to 0$).

We note that both $\langle p^4/E^3\rangle$ and the integral over ϵ_+ can be expressed in terms of a series of Bessel functions when expanding E in the denominator around $E = m$. Since this series does not converge very fast for the relatively small values of x that we will be interested in here, however, we do not display these series.

[30,32–34]). We emphasize, however, that from the point of view of solving the coupled set of Eqs. (27) and (28), there is no need to make such a relatively strong assumption about $f_\chi(\mathbf{p})$: *any* form of f_χ that leads to (very) similar results for the quantities given in Eqs. (31)–(33) will also lead to (very) similar results for $Y(x)$ and $y(x)$. In other words, we expect our coupled system of Boltzmann equations to agree with the full numerical solution discussed in the next section–concerning the evolution of Y and y–if and only if the ansatz in Eqs. (31)–(33) coincides with the corresponding averages numerically determined from the "true" phase-space distribution. As we will see later, this is indeed very often the case.

The second comment concerns the first term on the rhs of Eq. (28), which is proportional to the second moment of the elastic scattering term given in Eq. (5). As that latter expression is valid only to lowest order in $p^2/E^2 \sim p^2/m_\chi^2 \sim 1/x$, we had for consistency also to neglect any higher-order corrections in these quantities to the elastic scattering part of C_2 when deriving our final result. As discussed in Appendix B, in fact, there is no simple way of determining the next-to-leading order corrections to C_{el}. If we use our default semirelativistic scattering term given in Eq. (8), however, including the resulting corrections from subleading orders corresponds to replacing in Eq. (28),

$$T_\chi\left[\frac{y_{\mathrm{eq}}}{y}-1\right] = T - T_\chi \rightarrow T - T_\chi + \frac{1}{6}\left\langle\frac{p^4}{E^3}\right\rangle - \frac{5}{6}T\left\langle\frac{p^2}{E^2}\right\rangle + \frac{1}{3}T\left\langle\frac{p^4}{E^4}\right\rangle. \tag{35}$$

By construction, see Appendix B, this operator must still be an attractor to the equilibrium solution, and hence be proportional to (some power of) $T - T_\chi$; for the ansatz of Eq. (34), e.g., this can easily be verified directly. In practice, this replacement has very little impact on the evolution of Y and y, even at times as early as $x \sim 10$. We can think of the resulting small differences as a measure of the intrinsic uncertainty associated to our treatment of the scattering term.

C. The full phase-space density evolution

We now turn to solve the Boltzmann Eq. (1) at the full phase-space density level. This is numerically more challenging, but allows to assess the validity of the assumptions in previous sections and to track deviations (as we will see can occur) from the standard Maxwell Boltzmann velocity distribution. To achieve this, we start by reexpressing Eq. (1) in the two dimensionless coordinates,

$$x(t,p) \equiv m_\chi/T \quad \text{and} \quad q(t,p) \equiv p/T,$$

where the monotonic temperature $T(t)$ replaces as before the time parameter t via our $x(T)$, and q is now the "momentum" coordinate that depends on both t and p. In these variables, we can rewrite the Liouville operator on the lhs of Eq. (1) as

$$(\partial_t - H\mathbf{p}\cdot\nabla_{\mathbf{p}}) = \partial_t - Hp\partial_p = \tilde{H}(x\partial_x - \tilde{g}q\partial_q). \tag{36}$$

Here, we used the fact that the system is isotropic and assumed, as in the previous sections, that entropy is conserved. With the collision terms given in Eqs. (4) and (8), the Boltzmann equation for f_χ now becomes

$$\begin{aligned}\partial_x f_\chi(x,q) = {} & \frac{m_\chi^3}{\tilde{H}x^4}\frac{g_{\bar\chi}}{2\pi^2}\int d\tilde{q}\tilde{q}^2\frac{1}{2}\int d\cos\theta\, v_{\text{Møl}}\sigma_{\bar\chi\chi\rightarrow\bar f f} \\ & \times[f_{\chi,\mathrm{eq}}(q)f_{\chi,\mathrm{eq}}(\tilde{q}) - f_\chi(q)f_\chi(\tilde{q})] \\ & + \frac{\gamma(x)}{2\tilde{H}x}\left[x_q\partial_q^2 + \left(q + \frac{2x_q}{q} + \frac{q}{x_q}\right)\partial_q + 3\right]f_\chi \\ & + \tilde{g}\frac{q}{x}\partial_q f_\chi,\end{aligned} \tag{37}$$

where $x_q \equiv \sqrt{x^2+q^2}$ and θ is the angle between $\mathbf{q}$ and $\tilde{\mathbf{q}}$.

The benefits of this rewriting are twofold. First, the interpretation of the Boltzmann equation becomes very transparent, in the sense that this "comoving" phase-space density $f_\chi(x,q)$ clearly stays unaltered for $\tilde{g}(x) = 0$ and vanishing annihilation and scattering rates (being proportional to $\sigma_{\bar\chi\chi\rightarrow\bar f f}$ and γ, respectively). The new coordinates thus absorb how momentum and DM density change exclusively due to the Hubble expansion. (For nonvanishing $\tilde{g}$, these quantities continue to scale in the same way with the scale factor a, but taking into account that $a \propto {g_{\mathrm{eff}}^s}^{-1/3}T^{-1}$.) Second, the use of a comoving momentum $q \equiv p/T$ significantly helps numerical calculations that extend over a large range in $x = m_\chi/T$. In fact, $f_\chi(x,q)$ is expected to stay unchanged in shape both in the early semirelativistic and kinetically coupled regime, where $f_\chi \sim e^{-p/T_\chi} = e^{-q}$ given that $T_\chi = T$, as well as in the late nonrelativistic kinetically decoupled regime, where $f_\chi \sim e^{-p^2/(2mT_\chi)} \propto e^{-q^2/(2m)}$ given that $T_\chi \propto T^2$ in this case—at least as long as $\tilde{g} = 0$ and the DM phase-space distribution remains close to Maxwellian as in Eq. (34).

Let us stress that here, unlike for our discussion in the previous subsection, it is indeed mandatory to use the semirelativistic form of Eq. (8) for the scattering operator when discussing the evolution of the phase-space density, in the sense that it must drive the distribution function $f_\chi(q)$ towards the fully relativistic form $\propto e^{-E/T}$ [and not as Eq. (5) to the nonrelativistic approximation $\propto e^{-\frac{q^2}{2m}/T}$]. The importance of this can be seen by comparing the second and the third line of Eq. (37). The factor in the second line will always drive DM annihilation to occur unless an equilibrium distribution f_{eq} is reached. The term in the third line determines towards which equilibrium shape the scattering operator will drive the DM distribution $f_\chi(q)$. If the scattering attractor distribution would not match the $f_{\rm eq}(q)$ of the second line, then scattering could artificially drive annihilation to occur. For more discussions of the semirelativistic aspects of the scattering term, see Appendix B.

We then use a technique that discretizes the unbounded momentum variable q into a finite number of q_i with $i \in \{1, 2, \ldots, N\}$. This allows to rewrite our *integro partial differential equation* into a set of N coupled ordinary differential equations (ODEs):

$$\begin{aligned}\frac{d}{dx} f_i = {}& \frac{m_\chi^3}{\tilde{H} x^4} \frac{g_{\bar{\chi}}}{2\pi^2} \sum_{j=1}^{N-1} \frac{\Delta \tilde{q}_j}{2} [\tilde{q}_j^2 \langle v_{\text{Møl}} \sigma_{\bar{\chi}\chi\to\bar{f}f} \rangle_{i,j}^\theta (f_i^{\rm eq} f_j^{\rm eq} - f_i f_j) \\ & + \tilde{q}_{j+1}^2 \langle v_{\text{Møl}} \sigma_{\bar{\chi}\chi\to\bar{f}f} \rangle_{i,j+1}^\theta (f_i^{\rm eq} f_{j+1}^{\rm eq} - f_i f_{j+1})] \\ & + \frac{\gamma(x)}{2\tilde{H} x} \left[x_{q,i} \partial_q^2 f_i + \left(q_i + \frac{2 x_{q,i}}{q_i} + \frac{q_i}{x_{q,i}} \right) \partial_q f_i + 3 f_i \right] \\ & + \tilde{g} \frac{q_i}{x} \partial_q f_i, \end{aligned} \tag{38}$$

where $f_i \equiv f_\chi(x, q_i)$, and the derivatives $\partial_q f_i$ and $\partial_q^2 f_i$ are determined numerically by finite differentials using several neighboring points to f_i. $\langle v_{\text{Møl}} \sigma_{\bar{\chi}\chi\to\bar{f}f} \rangle_{i,j}^\theta$ is the velocity-weighted cross section averaged over θ (which is evaluated analytically or numerically) as a function of q_i and $\tilde{q}_j$, and $\Delta \tilde{q}_j \equiv \tilde{q}_{j+1} - \tilde{q}_j$. Finally, the DM number density in Eq. (10) is determined by trapezoidal integration.

Numerous numerical tests have been performed to ensure stability of our solutions to the ODEs of Eq. (38) and that imposed conditions on the now emerged boundary points (at q_1 and q_N) are physically sound. It turns out that very small step sizes over a large range in q are required for solving these stiff ODEs. We typically used the range $q_1 = 10^{-6}$ to $q_N = 50$ with about thousand steps in between, and set the two last terms of Eq. (38) to zero at q_N while using forward derivatives to evaluate them at q_1. By the use of the ODE15s code in MATLAB, and by analytically deriving internally required Jacobians, we are able to efficiently calculate the full phase-space evolution for the freeze-out after optimizing numerical settings. On the time scale of a few minutes we can derive the relic abundance for a given DM model. The code is general enough to be adapted to any standard single WIMP setup.

III. SCALAR SINGLET DARK MATTER

The simplest example of a renormalizable model providing a WIMP DM candidate is the scalar singlet model [14–16], originally proposed as DM made of "scalar phantoms" by Silveira and Zee [14]. In this model, the only addition to the standard model is a real gauge-singlet scalar field S which is stabilized by a $\mathbb{Z}_2$ symmetry and never obtains a nonvanishing vacuum expectation value. The simplicity of the model has in itself triggered considerable interest [35–44], with a further boost of attention after the discovery of the Higgs boson [45–58]. Recently, the GAMBIT [59] collaboration presented the so far most comprehensive study of this model by performing a global fit taking into account experimental constraints from both direct, indirect and accelerator searches for DM [60].

Interestingly, the resulting parameter region with the highest profile likelihood in this global fit is the one where the scalar singlet mass m_S is about half that of the SM Higgs mass, m_h, and where the DM abundance today is set by the resonant annihilation of two DM particles through an almost on-shell Higgs boson. As we will see, it is exactly in this parameter region that the standard way of calculating the relic density, as implemented in all previous studies of this model, fails because kinetic decoupling happens so early that it essentially coincides with chemical decoupling. Instead, the formalism introduced in the previous section provides a reliable calculation of the relic abundance of scalar singlet DM.

A. Model setup

The model symmetries, along with the requirement of renormalizability, uniquely determine the form of the Lagrangian to be

$$\mathcal{L}_{\rm SZ} = \mathcal{L}_{\rm SM} + \frac{1}{2} \partial_\mu S \partial^\mu S - \frac{1}{2} \mu_S^2 S^2 - \frac{1}{2} \lambda_S S^2 H^\dagger H - \frac{1}{4!} \lambda_{SS} S^4, \tag{39}$$

where H is the Standard Model Higgs doublet. The S boson mass receives contributions from both the bare mass term, μ_S, and from electroweak symmetry breaking, leading to $m_S = \sqrt{\mu_S^2 + \frac{1}{2} \lambda_S v_0^2}$, where $v_0 = 246.2$ GeV is the Higgs vacuum expectation value. We adopt the Higgs mass and the total width from decay to SM particles to be $m_h = 125.09\,\text{GeV}$ [61] and $\Gamma_{h,SM} = 4.042\,\text{MeV}$ [62]. For the moment, we neglect the quartic self-coupling λ_{SS}, but will later comment on its potential (minor) impact on relic density calculations.

The annihilation cross section of DM pairs to SM particles, apart from hh final states, is given by [49]

$$\sigma v_{\rm CMS} = \frac{2 \lambda_S^2 v_0^2}{\sqrt{s}} |D_h(s)|^2 \Gamma_{h\to{\rm SM}}(\sqrt{s}), \tag{40}$$

where $\Gamma_{h\to{\rm SM}}(\sqrt{s})$ is the partial decay width of a Standard Model Higgs boson of mass $\sqrt{s}$, and

$$|D_h(s)|^2 = \frac{1}{(s - m_h^2)^2 + m_h^2 \Gamma_h^2}. \tag{41}$$

The total Higgs width Γ_h in the above propagator, but not elsewhere, includes not only all SM channels but also the

$h \to SS$ channel if it is open. For $\Gamma_{h\to \mathrm{SM}}(\sqrt{s})$, as in [49], we use tabulated values for $\sqrt{s} < 300$ GeV from [62] and analytic expressions at higher $\sqrt{s}$. Note however that the latter high $\sqrt{s}$ region has no impact on the relic density in the studied scalar singlet mass range. Likewise, the channel $SS \to hh$ lies outside our kinematic region of interest.

For the elastic scattering processes, we take into account DM scattering with all SM fermions. Being mediated only by a Higgs in the t-channel, the corresponding squared amplitude takes a particularly simple form,

$$|\mathcal{M}_{Sf\to Sf}|^2 = \frac{N_f \lambda_S^2 m_f^2}{2} \frac{4m_f^2 - t}{(t - m_h^2)^2}, \tag{42}$$

where m_f is the mass of the SM fermion and the color factor is $N_f = 3$ for quarks and $N_f = 1$ for leptons. Averaging over the transferred momentum, as in Eq. (7), we thus find

$$\langle|\mathcal{M}|^2\rangle_t = \sum_f \frac{N_f \lambda_S^2 m_f^2}{8k^4} \left[\frac{2k_{\mathrm{cm}}^2 - 2m_f^2 + m_h^2}{1 + m_h^2/(4k_{\mathrm{cm}}^2)} - (m_h^2 - 2m_f^2)\log\left(1 + 4k_{\mathrm{cm}}^2/m_h^2\right)\right]. \tag{43}$$

Note that the sum here runs over all relevant fermions and antifermions *separately*.

The hierarchical Yukawa structure of the Higgs couplings leads to the scattering rate being dominated by the interactions with the heaviest fermions that for a given temperature are still sufficiently abundant in the plasma. In the range of DM masses m_S that we are interested in, freeze-out happens around $T \sim \mathcal{O}(1\ \mathrm{GeV})$, which is not far from the temperature of the QCD phase transition. Consequently, the details of this transition and the SM plasma can have a significant impact on the scattering rate; a study which goes beyond the scope of this work. Therefore, we follow the literature and adopt two extreme scenarios that can be thought of as bracketing the actual size of the scattering term:

(A) All quarks are free and present in the plasma down to temperatures of $T_c = 154$ MeV (largest scattering scenario, as adopted in [20]).

(B) Only light quarks (u, d, s) contribute to the scattering, and only for temperatures above $4T_c \sim 600$ MeV, below which hadronization effects start to become sizable [63] (smallest scattering scenario, as adopted in [12]).

Finally, we adopt the recent results from Drees *et al.* [64] for the effective number of relativistic degrees of freedom $g_{\mathrm{eff}}(T)$ that enter the calculation of the Hubble rate during radiation domination, $H = \sqrt{4\pi^3 g_{\mathrm{eff}}/45T^2/m_{\mathrm{Pl}}}$, as well as the entropy degrees of freedom entering for example in the calculation of $\tilde{g}(T)$ as defined in Eq. (18).

B. Relic density of scalar singlet dark matter

Let us first compute the relic density following the standard treatment adopted in the literature. To this end, we numerically solve Eq. (17) for a given set of parameters (m_S, λ_S) and determine the resulting asymptotic value of Y_0. The blue dashed line in Fig. 1 shows the contour in this plane that results in Y_0 corresponding to a relic density of $\Omega h^2 = 0.1188$, cf. Eq. (19). We restrict our discussion to values of m_S in the kinematic range where $\langle\sigma v\rangle$ is enhanced due to the Higgs propagator given in Eq. (41), and the coupling λ_S that results in the correct relic density is hence correspondingly decreased. This curve agrees with the corresponding result obtained in Ref. [49].

For comparison, we show in the same figure the required value of λ_S that results when instead solving the coupled system of Boltzmann equations (27) and (28), or when numerically solving the full Boltzmann equation as described in Sec. II C. Here, the solid (dashed) line shows the situation for the B (A) scenario for scatterings on quarks. Outside the resonance region, the coupled Boltzmann equations lead to identical results compared to the standard approach, indicating that kinetic decoupling indeed happens much later than chemical decoupling and that the assumption of local thermal equilibrium during chemical freeze-out thus is satisfied. For DM masses inside the resonance region, on the other hand, we can see that the

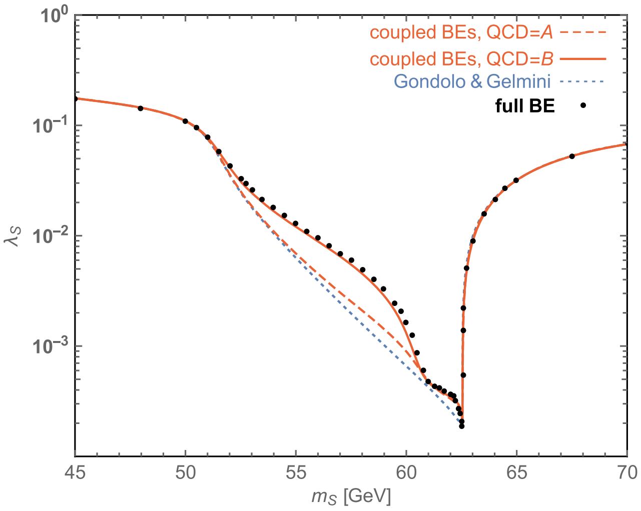


FIG. 1. The required value of the singlet-Higgs coupling λ_S, as a function of the scalar singlet mass m_S, in order to obtain a relic density of $\Omega h^2 = 0.1188$. The blue dashed line shows the standard result as established by Gondolo and Gelmini [9], based on the assumption of local thermal equilibrium during freeze-out. For comparison, we also plot the result of solving instead the coupled system of Boltzmann equations (27) and (28) for the maximal ("B") and minimal ("A") quark scattering scenarios defined in the main text (red solid and dashed lines, respectively). Finally, we show the result of fully solving the Boltzmann equation numerically, for the maximal quark scattering scenario and with no DM self-interactions included ("full BE").

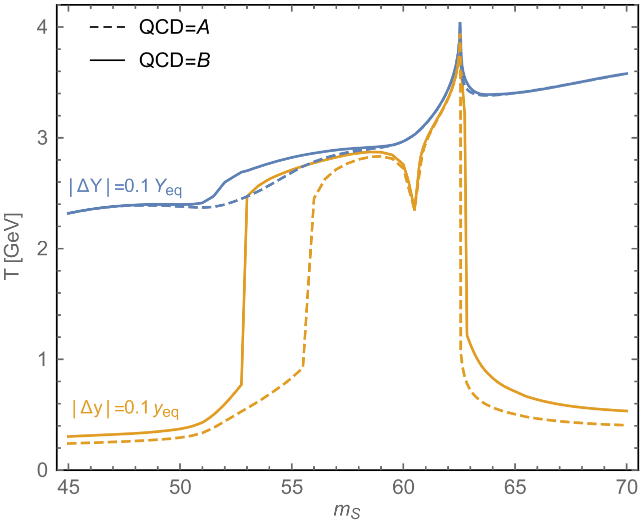


FIG. 2. Temperatures at which DM number density and velocity dispersion (“temperature”) start to deviate from their equilibrium values, defined for the purpose of this figure as $|Y - Y_{\rm eq}| = 0.1 Y_{\rm eq}$ and $|y - y_{\rm eq}| = 0.1 y_{\rm eq}$, respectively. These curves are based on solving the coupled system of Boltzmann equations (27) and (28), for the same parameter combinations as in Fig. 1 (resulting thus in the correct relic density).

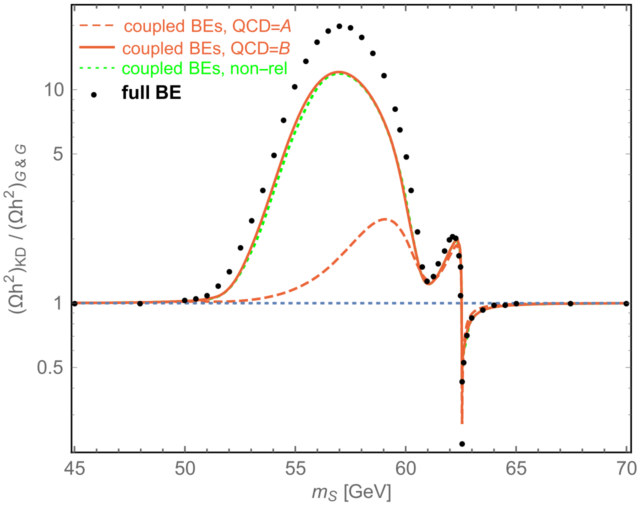


FIG. 3. The impact of the improved treatment of the kinetic decoupling on the relic density for parameter points that would satisfy the relic density constraint in the standard approach (dotted line in Fig. 1), both for the minimal (solid) and maximal (dashed) scenario for scattering with quarks. The numerical result (full BE) implements minimal quark scattering; note that this does not take into account the effect of DM self-interactions (while the other curves are consistent with assuming a maximal self-scattering rate). The green dashed curve shows the impact of implementing the elastic scattering term in the highly nonrelativistic limit, cf. Eq. (5).

two methods can give significantly different results, implying that this assumption must be violated. For the same reason, a smaller scattering rate (as in scenario B) leads to an even larger deviation from the standard scenario than the maximal scattering rate adopted in scenario A.

This interpretation is explicitly confirmed in Fig. 2, where we plot the temperatures at which the DM number density and temperature start to deviate from the equilibrium values: in the parameter range that we focus on here, kinetic decoupling happens indeed very close to chemical decoupling. The reason for this very early kinetic decoupling is straightforward to understand as the result of a strongly suppressed momentum transfer rate $\gamma(T)$, compared to the annihilation rate, due to two independent effects: (i) the small coupling λ_S needed to satisfy the relic density requirement, without a corresponding resonant enhancement of $\gamma(T)$, and (ii) the scattering rate being proportional to the Yukawa coupling squared, which favors scattering with Boltzmann-suppressed heavy fermions. We note that the latter point also explains the relatively large difference between the two extreme quark scattering scenarios used here for illustration (in scenario B, the largest Yukawa couplings do not contribute to the scattering).

In order to emphasize the importance of our improved treatment of the decoupling history, we plot in Fig. 3 also the *ratio* of the resulting relic density to that of the standard approach (for parameter values satisfying the relic density constraint for the latter, i.e. corresponding to the blue dashed curve in Fig. 1). Let us stress that, compared to the observational uncertainty in this quantity of about 1%, these corrections are by no means small even in the minimal scattering scenario A. In the same figure, we also compare our result for the coupled system of Boltzmann equations (27) and (28) to the full numerical solution of the Boltzmann equation in phase space, as described in Sec. II C (black dots). Before getting back to these results, let us briefly comment on the green dashed line in Fig. 3, which implements the highly nonrelativistic scattering term $C_{\rm el}$ of Eq. (5), and hence *not* the replacement (35) in Eq. (28) which we otherwise adopt as our default. Clearly, the impact of this choice is very limited for this approach. We note that the quantitative importance of the relativistic correction term proportional to $\langle p^4/E^3\rangle$ in Eq. (28) lies in the same ballpark, affecting the relic density by at most ~10% in the region very close to the resonance (and below the percent level elsewhere).

In Appendix A we discuss in depth the time evolution of both the coupled Boltzmann equations and the full phase-space density in the resonance region. Let us here just mention that the characteristic features of the curves displayed in Figs. 2 and 3 can indeed all more or less directly be understood in terms of the highly enhanced annihilation rate in a relatively narrow kinematic region around the resonance, $\sqrt{s} \sim m_h \pm \Gamma_h$. As the full numerical solution reveals, furthermore, the shape of $f_\chi(p)$ can in some cases be quite different from the Maxwell-Boltzmann form (34) that is consistent with the coupled system of Boltzmann equations (27) and (28). Whether this has a

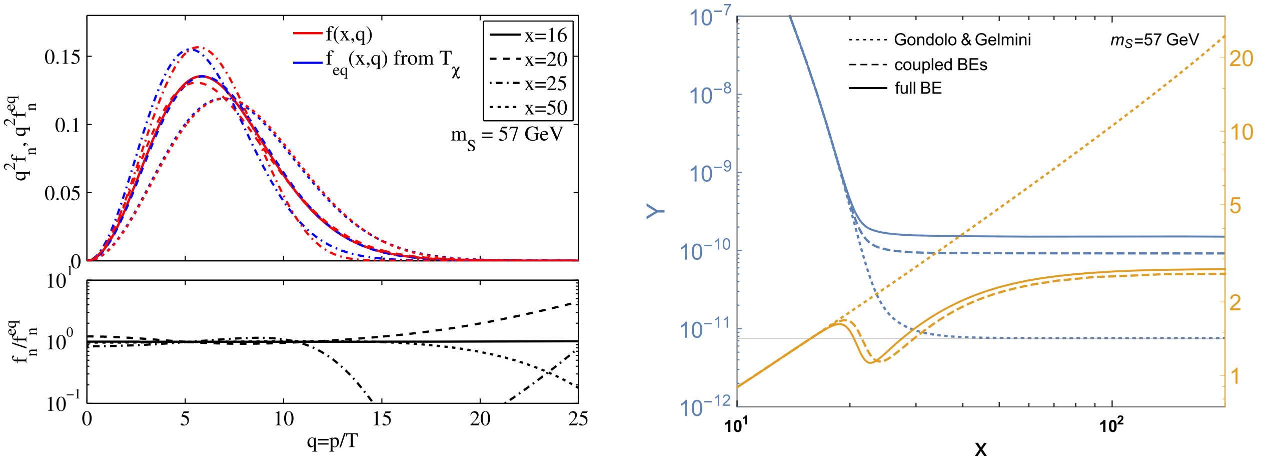


FIG. 4. Phase-space distributions and their evolution for a scalar singlet DM particle with $m_S = 57$ GeV. Left panel: Unit normalized phase-space distributions $f_n(q)$ from our full numerical solution of the Boltzmann equation (red lines) and thermal equilibrium distributions $f_n^{\rm eq}(q)$ (blue lines) at four different temperatures $x = m_S/T = 16$ (solid), 20 (dashed), 25 (dot dashed) and 50 (dotted). The equilibrium distributions $f_n^{\rm eq}$ are Maxwell-Boltzmann distributions evaluated at the temperatures T_χ, as defined in Eqs. (20) and (21). The bottom part shows the fractional deviation from the respective thermal distribution $f_n(q)/f_n^{\rm eq}(q)$. Right panel: The evolution of Y (blue) and y (yellow), assuming a Higgs-scalar coupling that leads to the correct relic density in the standard approach (dotted line in Fig. 1). We show these curves for the standard case (dotted lines), the approach using coupled Boltzmann equations (dashed) and the full numerical result (solid). The thin gray line indicates the asymptotic value of Y corresponding to the observed relic density.

noticeable impact on the resulting relic density (like for $m_S \sim 57$ GeV) or not (like for $m_S \sim m_h/2$) again mostly depends on whether or not the shape is affected for those momenta that can combine to $\sqrt{s} \sim m_h$ during chemical freeze-out.

For illustration, we pick a DM mass of $m_S = 57$ GeV and show in Fig. 4 the full phase-space distribution for a few selected values of x (left panel) as well as the relevant evolution of Y and y (right panel). For models with DM masses in this range, the relatively large difference between full solution and coupled equations (as visible in Fig. 3) can mostly be understood in terms of the *dip* in the *ratio* of DM phase-space distributions at intermediate values of $q = p/T$ that starts to develop for $x \gtrsim 20$. Concretely, the fact that the actual distribution for those momenta is slightly suppressed compared to a distribution fully characterized only by its second moment, as in Eq. (34), causes the DM particles to annihilate less efficiently, $\langle \sigma v \rangle_{\rm neq} < \langle \sigma v \rangle$, because this is the momentum range probed by the resonance for these x values. This in turn leads to the DM particles falling out of chemical equilibrium earlier, and hence a larger asymptotic value of Y. The reason for this momentum suppression to develop in the first place is also to be found in the particularly efficient annihilation close to the resonance, which leads to a depletion of DM particles with corresponding momenta because the scattering rate is no longer sufficiently large to redistribute the phase-space distribution to a thermal shape. We note that the bulk part of this effect is actually well captured by the coupled Boltzmann system, cf. the dashed vs solid lines in the right panel of Fig. 4. For further details, we refer again to Appendix A.

IV. DISCUSSION

From the above discussion, we have learned that very early kinetic decoupling is not just a theoretical possibility. It can appear in simple WIMP models, like the scalar singlet case, and affect the DM relic density in a significant way. We note that the size of the latter effect is, as expected, directly related to the size of the momentum exchange rate and hence to just *how* early kinetic decoupling happens compared to chemical decoupling. Let us stress that, from a general point of view, this is a much more important message connected to our choice of considering two scattering scenarios than the question of which of those scenarios is more realistic for the specific model we have studied here.

We have also seen that the coupled system of Boltzmann equations (27) and (28) provides a qualitatively very good description for the resulting DM abundance, see in particular Fig. 1, even though for high-precision results it seems mandatory to actually solve the full Boltzmann equation in phase space. As discussed in Appendix A, differences can arise when the true phase-space distribution is not of the Maxwellian form assumed in Eq. (34)—though the two methods *can* actually still give almost identical results for the relic abundance even when the two distribution differ vastly. The question of under which conditions the coupled system of equations provides an accurate description of the relic density is thus a somewhat subtle one, and requires a careful discussion of the velocity dependence of the annihilation term in the Boltzmann equation.

An exception to this general complication is a DM *self-interaction* rate large enough to force the DM distribution into the form given by Eq. (34) [30,32–34] and hence render the coupled system of Boltzmann equations (27) and (28) *exactly* correct (up to, as discussed, corrections due to quantum statistics). Sizable self-scattering rates can for example arise due to corresponding contact interactions, like the quartic coupling λ_{SS} in the scalar singlet case, or by adding light mediators that couple to the DM particle (which was indeed the first time such a coupled system of Boltzmann equations was considered [30], albeit in a different context). For the case of resonant annihilation, furthermore, the same resonance also mediates an enhanced self-interaction. For future work, it would hence be worthwhile to extend our numerical framework to even include those DM self-interaction processes. For the scalar singlet case, in particular, we expect that adding the process $SS \to h^* \to SS$ would bring all numerical results for the full Boltzmann equation—e.g., those shown in Fig. 3—even closer to those resulting from the coupled system of Boltzmann equations.

Let us finally stress that both the coupled Boltzmann equations and the numerical setup that we have described here are very general, and can be used to consistently study early kinetic decoupling for a much larger range of models than the scalar singlet case. Obvious applications are other scenarios where resonant annihilation and/or annihilation to heavy final states is important in setting the relic abundance, see also Ref. [31]. Further examples where the ratio of the scattering rate to the annihilation rate can be smaller than usual, hence potentially leading to early kinetic decoupling, include Sommerfeld-enhanced annihilation [30,34,65,66] (if the light mediators are not abundant enough to take part in the scattering process) and annihilation to DM bound states [67,68]. Quite in general, our methods provide a powerful means to check whether the DM particles are indeed in local thermal equilibrium with the heat bath around the time when their abundance freezes out—which is the usual assumption, though rarely explicitly tested, not only in WIMP-like scenarios but also when so-called semiannihilations [69] are important in setting the relic density, when computing the relic abundance for modified expansion histories [70,71], or in scenarios that go beyond simple 2 → 2 annihilation processes [72–74].

V. CONCLUSIONS

The standard way of calculating the thermal relic density of self-annihilating DM particles rests on the assumption of local thermal equilibrium during freeze-out, and that hence kinetic decoupling occurs much later than chemical decoupling. Here, we demonstrated for the first time that departure from kinetic equilibrium can instead happen much earlier, even *simultaneously* with the departure from chemical equilibrium.

By introducing a coupled system of equations for the DM number density and its “temperature”, or rather velocity dispersion, we improved the standard way of calculating the relic density in such cases. For an even higher accuracy in predicting the DM abundance, we also found a way of solving the full Boltzmann equation numerically. The latter approach has the additional advantage of obtaining the full phase-space distribution, rather than only the number density, which in particular allows to test in detail the assumption of a Maxwellian velocity distribution adopted in the standard approach. A numerical solver for the coupled system of Boltzmann equations, s. (27) and (28), will be available in an upcoming version of DarkSUSY [28] and our implemented solver for the full Boltzmann uation at the phase-space level, . (38), will be released separately.[2]

Applied to the simplest renormalizable WIMP model—the scalar singlet, extensively discussed in the literature—we somewhat surprisingly found that the relic abundance predicted in the standard approach can differ by up to an order of magnitude from the correct treatment presented in this paper. This is rather remarkable not only in view of the simplicity of this model, but also because the affected region in parameter space happens to coincide with the best-fit region resulting from most recent global scans. We thus expect our results to have a noticeable phenomenological impact, and that our treatment will prove useful also when applied to other examples of relic density calculations in cases where the standard assumption of local thermal equilibrium during freeze-out is not exactly satisfied.

ACKNOWLEDGMENTS

We thank Mateusz Duch, Joakim Edsjö, Bohdan Grządkowski, Andreas Hohenegger and Ayuki Kamada for very useful conversations during the preparation of this work. A. H. is supported by the University of Oslo through the Strategic Dark Matter Initiative (SDI). M. G. and T. Binder have received funding from the European Union’s Horizon 2020 research and innovation program under Grants Agreements No. 690575 and No. 674896. T. Binder gratefully acknowledges financial support from the German Science Foundation (DFG RTG 1493).

Note added.—Recently, we became aware of a dedicated study on resonant DM annihilation [31], which also found that DM can kinetically decouple much earlier than usual in this case.

APPENDIX A: PHASE-SPACE DENSITY EVOLUTION OF THE SCALAR SINGLET

In Sec. III B, we investigated the impact of our improved treatment of the Boltzmann equation on the expected DM relic abundance in the scalar singlet model. Here, we supplement this by discussing in some more detail the evolution of the DM phase-space density. The main focus

---

[2] Please contact any of the authors if you need these numerical routines prior to their public release.

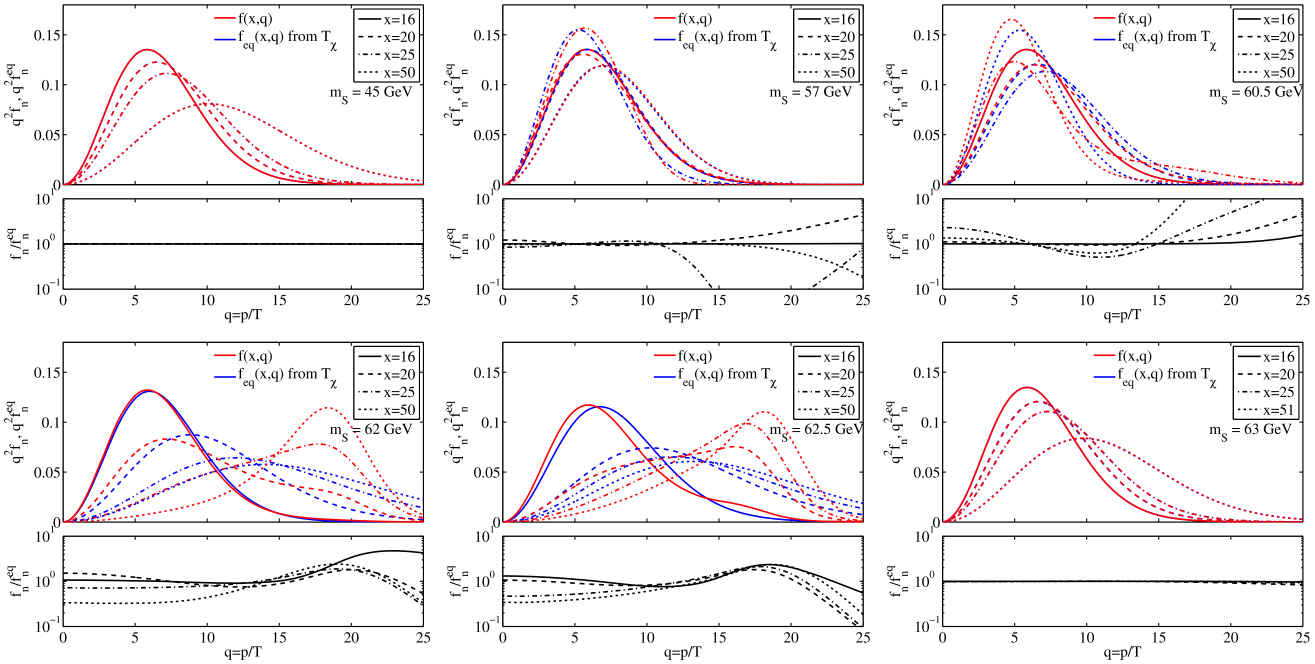


FIG. 5. Same as the left panel of Fig. 4 in the main text, but now for comparison for various DM masses $m_S = 45, 57, 60.5, 62, 62.5,$ 63 GeV. Note that for cases where the equilibrium distributions appear to be missing in the top panels, it is just because it agrees very well with the actual phase-space distribution (as also visible in the fractional deviation plotted in the bottom panels).

of this discussion, however, will be a more thorough qualitative understanding of the specific features seen in Figs. 2 and 3, and the underlying interplay of chemical and early kinetic decoupling. Specifically, we can distinguish three mass regimes:

(1) A regime with $53\,\mathrm{GeV} \lesssim m_S \lesssim 60.5\,\mathrm{GeV}$, which we will refer to as *subresonant* because f_χ starts to deviate from its equilibrium value, $f_{\chi,\mathrm{eq}} = \exp(-E/T)$, at a temperature where the typical DM momenta are too small to hit the resonance, i.e. $\sqrt{s} \lesssim m_h - \Gamma_h$. As a result, we have $\langle\sigma v\rangle_{(\mathrm{neq})} < \langle\sigma v\rangle_{2,(\mathrm{neq})}$[3] during the whole freeze-out process in this regime—this is because $p^2 f_\chi(p)$ peaks at a higher value of p than $f_\chi(p)$, which brings its bulk distribution closer to (or even on) the cross-section resonance.

(2) A regime with $60.5\ \mathrm{GeV} \lesssim m_S \lesssim 62.5$ GeV that we will refer to as *resonant*. Here, we have $\langle\sigma v\rangle > \langle\sigma v\rangle_2$ around the time when the DM particles start to leave thermal equilibrium, because the larger mass combines with the relevant momenta to $s \sim m_h^2$. At slightly later times, on the other hand, still relevant in changing the DM abundance, the DM momenta have redshifted so much that we are back to a situation where typically $\sqrt{s} \lesssim m_h - \Gamma_h$ and hence $\langle\sigma v\rangle < \langle\sigma v\rangle_2$.

(3) Finally, there is a *superresonant regime* with $62.5\ \mathrm{GeV} \lesssim m_S \lesssim 65$ GeV, where decoupling occurs at such high temperatures that we have $\langle\sigma v\rangle > \langle\sigma v\rangle_2$ during the whole time it takes for $Y(x)$ to reach its asymptotic value (determining the relic density).

To help our discussion, let us look at a selection of benchmark points with scalar snglet masses $m_S = 45, 57,$ 60.5, 62, 62.5, 63 GeV and coupling constants $\lambda_S(m_S)$ that result in the correct relic density in the standard approach (dotted line in Fig. 1). In Fig. 5, we show the DM distribution function for these benchmark points that we find with our full numerical approach, for selected values of x, and in Fig. 6 the full evolution of $Y(x)$ and $y(x)$ for the different approaches. These figures thus extend the information in Fig. 4 by covering a range of DM masses.

The first thing to note, as exemplified by the benchmark points with $m_S = 45$ GeV and $m_S = 63$ GeV, is that for masses sufficiently far away from the resonance we find a phase-space distribution which remains almost exactly Maxwellian in shape. For these points, we therefore find as expected a very good agreement for the evolution of $Y(x)$ and $y(x)$ when comparing the numerical solution and the coupled Boltzmann approach, as well as with Y in the standard Gondolo and Gelmini setup (which assumes $T = T_\chi$). We note that this provides an important consistency check for both methods.

An example for a model in the *subresonant* region is the case with $m_S = 57$ GeV, which we discussed in the main

---

[3] For the sake of better readability, we will suppress the subscript "neq" for the remainder of this section. We note that, once chemical decoupling has started, the contribution of thermal averages without this subscript is suppressed by a factor of Y_{eq}/Y in Eqs. (27) and (28).

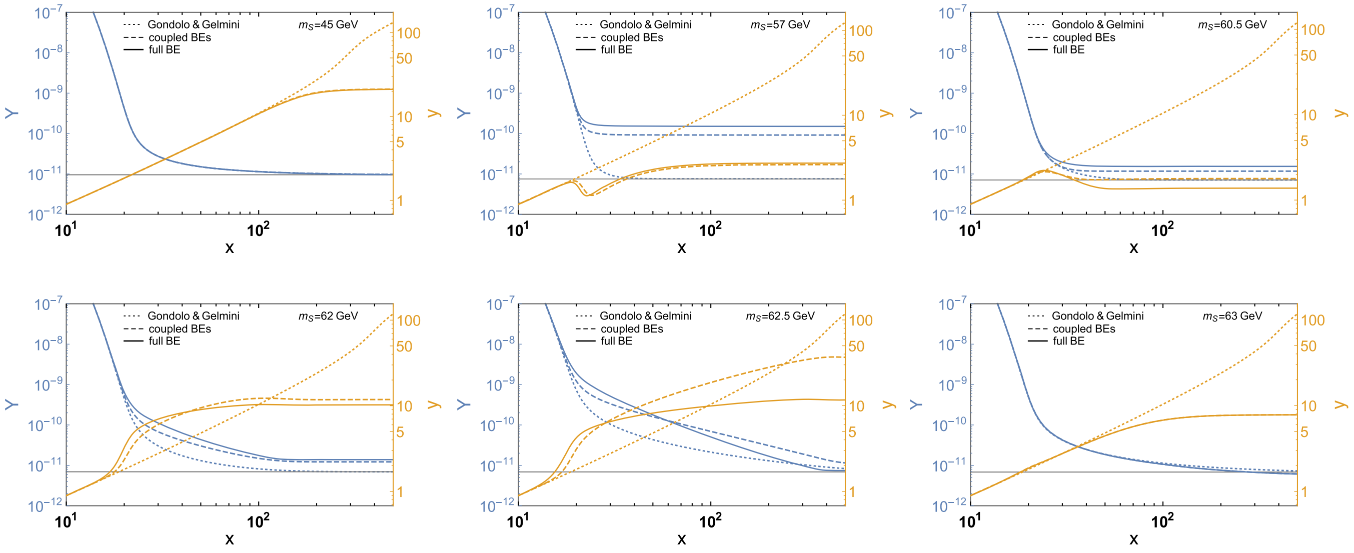


FIG. 6. Evolution of $Y(x)$ and $y(x)$, for the same DM masses as shown in Fig. 5.

text. Here, the resonant annihilation depletes $f_\chi(q)$ for momenta *just above the peak* of the distribution, leading to a relative decrease with respect to a thermal distribution at these momenta, and hence a decrease in the DM velocity dispersion (also known as "temperature"). This effect is visible in Fig. 5 starting with a slight suppression at $q \sim 8$ for the curve with $x = 20$ (note that the relative enhancement at larger values of q is not relevant for our discussion given that f_χ is already highly suppressed here), and results in the decrease in the evolution of y seen in Fig. 6. The latter can also directly be understood from inspection of Eq. (28): in the subresonant regime we have $\langle\sigma v\rangle < \langle\sigma v\rangle_2$, which drives y to smaller values after decoupling (with a strength proportional to Y—which explains why the scattering term $\propto \gamma(T)$ can increase y again, slightly, once the DM abundance has decreased sufficiently). A second effect of this depletion in $f_\chi(q)$ is that $\langle\sigma v\rangle_{\rm neq}$ decreases, which in turn leads to an earlier chemical decoupling and hence an increased relic density. The difference between the numerical and the coupled Boltzmann approach can in this case thus exclusively be understood as resulting from the slight offset in the $y(x)$ curves during the freeze-out [which in turn results from the fact that the scattering term is not strong enough to maintain an exact Maxwellian shape of $f_\chi(q)$ when the velocity dispersion decreases as explained above].

As we increase the DM mass, we leave the subresonant regime and enter the *resonant regime*, with the transition point marked by the benchmark model with m_S=60.5GeV. We note that this transition is also clearly visible in Fig. 2, as a sharp decrease in the temperature at which the DM velocity dispersion deviates from its equilibrium value. The origin of this feature is *not* an actual delay of kinetic decoupling, but that DM annihilation now starts to deplete $f_\chi(q)$ *below* the peak of the would-be Maxwellian distribution.[4] This leads to an increase of the velocity dispersion, once equilibrium is left, rather than a decrease as in the subresonant regime. This effect is very clearly seen in Figs. 5 and 6, up to DM masses at the higher end of this regime, where the influence of the resonance starts to become less important because we have $\sqrt{s} \lesssim m_h + \Gamma_h$ only for DM momenta well below the peak of the phase-space distribution.

In the *superresonant regime* with $m_S \gtrsim m_h/2$, finally, we have necessarily $\sqrt{s} \gtrsim m_h$. A resonantly enhanced annihilation rate is thus only possible for a very small portion of phase space, with almost vanishing relative DM momenta. This implies not only that we always have $\langle\sigma v\rangle > \langle\sigma v\rangle_2$ in this regime, but also that the effect of the resonance rapidly becomes negligible.

Lastly, it is interesting to note that for $\sqrt{s} \gtrsim m_h$ the annihilation rate effectively features a $1/v^2$ velocity dependence. This is similar to resonant Sommerfeld-enhanced annihilation, which leads to a suppressed relic density after a prolonged freeze-out phase [30]. This can clearly be seen in the evolution of $Y(x)$ in Fig. 6, for $m_S \sim m_h/2$, where the differences between the numerical and the coupled Boltzmann approach are mostly due to the late-time differences in $y(x)$—which in turn come about because of the rather significant differences in $f_\chi(q)$ at large values of x (cf. Fig. 5).

---

[4] In a similar way, the sharp rise around $m_S \sim 54$ GeV in Fig. 2 should not be interpreted as a feature in the momentum exchange rate $\gamma(T)$. Rather, it can be understood as the point where the shape of the $y(x)$ evolution starts to develop from something close to the one in the top left panel in Fig. 6 into something that is much closer to the one in the top center panel (which in turn is driven by the annihilation terms, as explained in the text). As a result, the temperature at which y departs from $y_{\rm eq}$ increases very quickly as the mass increases beyond this transition point.

APPENDIX B: SEMIRELATIVISTIC KINETIC THEORY

In this Appendix, we discuss how to generalize the highly nonrelativistic elastic scattering term in Eq. (5) to incorporate the most important relativistic corrections needed for the numerical implementation of the full Boltzmann equation. Throughout, we refer to this result as "semirelativistic" scattering.

The starting point is to expand the full collision term $C_{\rm el}$ in small momentum transfer compared to the typical DM momentum—similar to what is done in order to arrive at Eq. (5), but not only keeping lowest-order terms in $\mathbf{p}^2/m_\chi^2 \sim T/m_\chi$. From this, we can derive a Fokker-Planck scattering operator in a relativistic form (for details, see [21]):

$$C_{\rm el} \simeq \frac{E}{2}\nabla_{\mathbf{p}} \cdot [\gamma(T,\mathbf{p})(ET\nabla_{\mathbf{p}} + \mathbf{p})f_\chi]. \tag{B1}$$

Being a total divergence, this scattering operator manifestly respects number conservation, as it should. Another important property, which one can directly read off from the part inside the brackets, is that it features a stationary point given by the relativistic Maxwell-Boltzmann distribution,

$$f_\chi^{\rm eq} \propto e^{-E/T}. \tag{B2}$$

The nonrelativistic limit of Eq. (B1) gives the scattering operator (5), but in this limit the stationary point would instead be the nonrelativistic version $f_\chi^{\rm eq} \propto \exp[-p^2/(2m_\chi T)]$—which would cause a problem in the full BE as this does not correspond to the *actual* equilibrium distribution fed into the annihilation term of Eq. (37).

In general, the momentum transfer rate $\gamma(T,\mathbf{p})$ in Eq. (B1) depends on the DM momentum $\mathbf{p}$. However, the stationary point is independent of γ, which motivates us to restrict ourselves to the leading order term $\gamma(T) \equiv \gamma(T,\mathbf{0})$, neglecting any momentum dependence, and use the nonrelativistic limit in Eq. (B1) *only* to evaluate the momentum transfer rate $\gamma(T)$ as it appears in Eq. (6). To this order, we could thus also replace the leading E in Eq. (B1) by m_χ; here, we choose to still keep it as it leads to a much more compact analytical form of the equation governing the DM temperature (see below). Explicitly performing the first partial derivative in $C_{\rm el}$ then leads to the final form of our semirelativistic Fokker-Planck operator as given by Eq. (8). This operator is our default choice for the numerical implementation of the full Boltzmann equation.

As already pointed out in Sec. II C, it is mandatory for the full phase-space calculation to have a scattering operator with a fixpoint that matches the equilibrium distribution of Eq. (B2) assumed in the annihilation term. For the coupled integrated Boltzmann system, on the other hand, this issue is fully addressed by using the relativistic temperature definition of Eq. (21)—rather than its nonrelativistic version typically adopted in the literature in the context of kinetic decoupling—because this automatically leads to the correct fixpoint $T_\chi = T$ for both the semirelativistic Eq. (8) and, to the lowest order, for the nonrelativistic version Eq. (5); see the discussion in Sec. II B.

Another advantage of our semirelativistic Fokker-Planck operator is that the differential equation for T_χ, often quoted when discussing kinetic decoupling, takes a very simple form even beyond the highly nonrelativistic limit. To see this, let us for the moment ignore the impact of annihilations, and take the second moment of the Boltzmann equation with this operator (using the relativistic definition of T_χ). This leads to

$$\dot{T}_\chi + 2\left(1 - \frac{\langle p^4/E^3\rangle}{6T_\chi}\right)HT_\chi = \gamma\left[T\left(1 - \frac{5}{6}\langle p^2/E^2\rangle + \frac{2}{6}\langle p^4/E^4\rangle\right) - T_\chi\left(1 - \frac{\langle p^4/E^3\rangle}{6T_\chi}\right)\right], \tag{B3}$$

which of course is equivalent to Eq. (28) in the main text, when neglecting the annihilation terms and implementing the replacement given in Eq. (35). Let us repeat that the rhs of the above equation only takes this particular form with our default choice of the semirelativistic Fokker-Planck term, whereas the moment appearing on the left-hand side is an exact result. This equation is in general not closed in terms of T_χ. However, if we make the ansatz of a Maxwellian DM phase-space distribution, cf. Eq. (34), we get a relation between the different momentum moments,

$$5\langle p^2/E^2\rangle - 2\langle p^4/E^4\rangle = \langle p^4/E^3\rangle/T_\chi, \tag{B4}$$

such that the differential equation closes in terms of T_χ. Indeed, introducing

$$2(1-w) \equiv \frac{g_\chi}{3T_\chi n_\chi}\int \frac{d^3p}{(2\pi)^3}\frac{\mathbf{p}^4}{E^3}f_\chi(\mathbf{p}) = \frac{\langle p^4/E^3\rangle}{3T_\chi}, \tag{B5}$$

it takes a very simple form:

$$\dot{T}_\chi + 2w(T_\chi)HT_\chi = w(T_\chi)\gamma(T)(T - T_\chi). \tag{B6}$$

This generalizes the highly nonrelativistic result [18], for which $w \to 1$ and we hence find the familiar scaling $T_\chi \propto T^2$ after kinetic decoupling (i.e. when $\gamma \ll H$). In the ultrarelativistic limit, on the other hand, we have $w(T_\chi) \to 1/2$ and the likewise familiar scaling of $T_\chi \propto T$ for relativistic particles. We note that in the region $x \gtrsim 10$ relevant for early kinetic decoupling, the correction to the nonrelativistic limit is already sizable; e.g., $w(x = 10) \approx 0.8$.

[1] G. Hinshaw *et al.* (WMAP Collaboration), Astrophys. J. Suppl. Ser. **208**, 19 (2013); arXiv:1212.5226.
[2] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016); arXiv:1502.01589.
[3] G. Bertone, D. Hooper, and J. Silk, Phys. Rep. **405**, 279 (2005); arXiv:hep-ph/0404175.
[4] B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).
[5] P. Hut, Phys. Lett. B **69**, 85 (1977).
[6] K. Sato and M. Kobayashi, Prog. Theor. Phys. **58**, 1775 (1977).
[7] D. A. Dicus, E. W. Kolb, and V. L. Teplitz, Phys. Rev. Lett. **39**, 168 (1977); **39**, 973(E) (1977).
[8] S. Wolfram, Phys. Lett. B **82**, 65 (1979).
[9] P. Gondolo and G. Gelmini, Nucl. Phys. **B360**, 145 (1991).
[10] J. Edsjö and P. Gondolo, Phys. Rev. D **56**, 1879 (1997); arXiv:hep-ph/9704361.
[11] K. Griest and D. Seckel, Phys. Rev. D **43**, 3191 (1991).
[12] T. Bringmann, New J. Phys. **11**, 105027 (2009); arXiv:0903.0189.
[13] R. T. D'Agnolo, D. Pappadopulo, and J. T. Ruderman, Phys. Rev. Lett. **119**, 061102 (2017); arXiv:1705.08450.
[14] V. Silveira and A. Zee, Phys. Lett. B **161**, 136 (1985).
[15] J. McDonald, Phys. Rev. D **50**, 3637 (1994); arXiv:hep-ph/0702143.
[16] C. P. Burgess, M. Pospelov, and T. ter Veldhuis, Nucl. Phys. **B619**, 709 (2001); arXiv:hep-ph/0011335.
[17] E. W. Kolb and M. S. Turner, Front. Phys. **69**, 1 (1990).
[18] T. Bringmann and S. Hofmann, J. Cosmol. Astropart. Phys. 04 (2007) 016; 03 (2016) E02; arXiv:hep-ph/0612238.
[19] J. Kasahara, Neutralino dark matter: The mass of the smallest halo and the golden region, Ph.D. thesis, The University of Utah, 2009.
[20] P. Gondolo, J. Hisano, and K. Kadota, Phys. Rev. D **86**, 083523 (2012); arXiv:1205.1914.
[21] T. Binder, L. Covi, A. Kamada, H. Murayama, T. Takahashi, and N. Yoshida, J. Cosmol. Astropart. Phys. 11 (2016) 043; arXiv:1602.07624.
[22] T. Bringmann, H. T. Ihle, J. Kersten, and P. Walia, Phys. Rev. D **94**, 103529 (2016); arXiv:1603.04884.
[23] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, Comput. Phys. Commun. **149**, 103 (2002); arXiv:hep-ph/0112278.
[24] P. Gondolo, J. Edsjö, P. Ullio, L. Bergström, M. Schelke, and E A. Baltz, J. Cosmol. Astropart. Phys. 07 (2004) 008; arXiv:astro-ph/0406204.
[25] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, Comput. Phys. Commun. **176**, 367 (2007); arXiv:hep-ph/0607059.
[26] A. Arbey and F. Mahmoudi, Comput. Phys. Commun. **181**, 1277 (2010); arXiv:0906.0369.
[27] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, Comput. Phys. Commun. **185**, 960 (2014); arXiv:1305.0237.
[28] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio, and L. Bergstrom, DarkSUSY 6: An Advanced Tool to Compute Dark Matter Properties Numerically (to be published).
[29] T. Bringmann *et al.* (GAMBIT Dark Matter Workgroup), arXiv:1705.07920.
[30] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, Phys. Rev. D **85**, 123512 (2012); arXiv:1202.5456.
[31] M. Duch and B. Grzadkowski, J. High Energy Phys. 09 (2017) 159; arXiv:1705.10777.
[32] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. Lett. **104**, 151301 (2010); arXiv:0911.0422.
[33] M. R. Buckley and P. J. Fox, Phys. Rev. D **81**, 083522 (2010); arXiv:0911.0422.
[34] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. D **82**, 083525 (2010); arXiv:1005.4678.
[35] C. E. Yaguna, J. Cosmol. Astropart. Phys. 03 (2009) 003; arXiv:0810.4267.
[36] S. Profumo, L. Ubaldi, and C. Wainwright, Phys. Rev. D **82**, 123514 (2010); arXiv:1009.5377.
[37] C. Arina and M. H. G. Tytgat, J. Cosmol. Astropart. Phys. 01 (2011) 011; arXiv:1007.2765.
[38] Y. Mambrini, Phys. Rev. D **84**, 115017 (2011); arXiv:1108.0671.
[39] R. N. Lerner and J. McDonald, Phys. Rev. D **80**, 123507 (2009); arXiv:0909.0520.
[40] M. Herranen, T. Markkanen, S. Nurmi, and A. Rajantie, Phys. Rev. Lett. **115**, 241301 (2015); arXiv:1506.04065.
[41] F. Kahlhoefer and J. McDonald, J. Cosmol. Astropart. Phys. 11 (2015) 015; arXiv:1507.03600.
[42] S. Profumo, M. J. Ramsey-Musolf, and G. Shaughnessy, J. High Energy Phys. 08 (2007) 010; arXiv:0705.2425.
[43] V. Barger, P. Langacker, M. McCaskey, M. Ramsey-Musolf, and G. Shaughnessy, Phys. Rev. D **79**, 015018 (2009); arXiv:0811.0393.
[44] J. M. Cline and K. Kainulainen, J. Cosmol. Astropart. Phys. 01 (2013) 012; arXiv:1210.4196.
[45] A. Djouadi, O. Lebedev, Y. Mambrini, and J. Quevillon, Phys. Lett. B **709**, 65 (2012); arXiv:1112.3299.
[46] K. Cheung, Y.-L. S. Tsai, P.-Y. Tseng, T.-C. Yuan, and A. Zee, J. Cosmol. Astropart. Phys. 10 (2012) 042; arXiv:1207.4930.
[47] M. Endo and Y. Takaesu, Phys. Lett. B **743**, 228 (2015); arXiv:1407.6882.
[48] A. Djouadi, A. Falkowski, Y. Mambrini, and J. Quevillon, Eur. Phys. J. C **73**, 2455 (2013); arXiv:1205.3169.
[49] J. M. Cline, K. Kainulainen, P. Scott, and C. Weniger, Phys. Rev. D **88**, 055025 (2013); **92**, 039906(E) (2015); arXiv:1306.4710.
[50] A. Urbano and W. Xue, J. High Energy Phys. 03 (2015) 133; arXiv:1412.3798.
[51] X.-G. He and J. Tandean, J. High Energy Phys. 12 (2016) 074; arXiv:1609.03551.
[52] M. Escudero, A. Berlin, D. Hooper, and M.-X. Lin, J. Cosmol. Astropart. Phys. 12 (2016) 029; arXiv:1609.09079.
[53] A. Goudelis, Y. Mambrini, and C. Yaguna, J. Cosmol. Astropart. Phys. 12 (2009) 008; arXiv:0909.2799.
[54] N. Craig, H. K. Lou, M. McCullough, and A. Thalapillil, J. High Energy Phys. 02 (2016) 127; arXiv:1412.0258.
[55] H. Han, J. M. Yang, Y. Zhang, and S. Zheng, Phys. Lett. B **756**, 109 (2016); arXiv:1601.06232.
[56] P. Ko and H. Yokoya, J. High Energy Phys. 08 (2016) 109; arXiv:1603.04737.
[57] A. Beniwal, F. Rajec, C. Savage, P. Scott, C. Weniger, M. White, and A. G. Williams, Phys. Rev. D **93**, 115016 (2016); arXiv:1512.06458.
[58] A. Cuoco, B. Eiteneuer, J. Heisig, and M. Krämer, J. Cosmol. Astropart. Phys. 06 (2016) 050; arXiv:1603.08228.

[59] P. Athron *et al.* (GAMBIT Collaboration), arXiv:1705.07908.
[60] P. Athron *et al.* (GAMBIT Collaboration), Eur. Phys. J. C **77**, 568 (2017); arXiv:1705.07931.
[61] C. Patrignani *et al.* (Particle Data Group), Chin. Phys. C **40**, 100001 (2016).
[62] S. Dittmaier *et al.* (LHC Higgs Cross Section Working Group), arXiv:1101.0593.
[63] D. Boyanovsky, H. J. de Vega, and D. J. Schwarz, Annu. Rev. Nucl. Part. Sci. **56**, 441 (2006); arXiv:hep-ph/0602002.
[64] M. Drees, F. Hajkarim, and E. R. Schmitz, J. Cosmol. Astropart. Phys. 06 (2015) 025; arXiv:1503.03513.
[65] J. B. Dent, S. Dutta, and R. J. Scherrer, Phys. Lett. B **687**, 275 (2010); arXiv:0909.4128.
[66] J. Zavala, M. Vogelsberger, and S. D. M. White, Phys. Rev. D **81**, 083502 (2010); arXiv:0910.5221.
[67] J. L. Feng, M. Kaplinghat, H. Tu, and H.-B. Yu, J. Cosmol. Astropart. Phys. 07 (2009) 004; arXiv:0905.3039.
[68] B. von Harling and K. Petraki, J. Cosmol. Astropart. Phys. 12 (2014) 033; arXiv:1407.7874.
[69] F. D'Eramo and J. Thaler, J. High Energy Phys. 06 (2010) 109; arXiv:1003.5912.
[70] F. D'Eramo, N. Fernandez, and S. Profumo, J. Cosmol. Astropart. Phys. 05 (2017) 012; arXiv:1703.04793.
[71] K. Redmond and A. L. Erickcek, Phys. Rev. D **96**, 043511 (2017); arXiv:1704.01056.
[72] Y. Hochberg, E. Kuflik, T. Volansky, and J. G. Wacker, Phys. Rev. Lett. **113**, 171301 (2014); arXiv:1402.5143.
[73] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, Phys. Rev. Lett. **115**, 021301 (2015); arXiv:1411.3727.
[74] E. Kuflik, M. Perelstein, N. R.-L. Lorier, and Y.-D. Tsai, Phys. Rev. Lett. **116**, 221302 (2016); arXiv:1512.04545.

Chapter 4

Light mediators

Light mediator models were proposed in the literature as a solution to structure formation problems on small scales. Section 4.1 reviews the problems and possible solutions from a neutral point of view. In Section 4.2, phenomenological aspects of light mediator models are introduced. These are the Sommerfeld enhancement, strong velocity-dependent self-interactions and late kinetic decoupling. Section 4.3 and 4.4 contain the reprinted research articles, including investigations about late kinetic decoupling and a second epoch of annihilation caused by Sommerfeld resonances, respectively.

4.1 Challenges on small scales and possible solutions

The cosmological concordance model is a highly successful theory. Nevertheless, several studies pointed out certain discrepancies between experimental observations and predictions of the *abundance and structure* of CDM halos in simulations. Collectively, those mismatches are called the small scale crisis [12], described in the following.

- CDM-only computer simulations show that the radial dependence of DM halo densities is typically characterized by a Navarro-Frank-White (NFW) profile [45], which rises steeply towards the central region. The predicted cusp in the inner parts is however in conflict with some of the density profiles inferred from fits to measured galactic rotation curves. Various galaxies are better fitted by a more cored density profile in the central region, recognized early on [46, 47]. In more recent literature, this discrepancy is called the Core-Cusp (CC) problem [48–50].

- Measurements of galactic rotation curves also show a spread in inner circular velocities among dwarf galaxies of equal maximum rotation speed. It has been pointed out that this diversity is not predicted by ΛCDM simulations, and the problem remains even when including baryonic feedbacks [51].

- Furthermore, N-body simulation predict a rich abundance of CDM subhalos within Milky-Way-size host halos. However, the number of known satellites and dwarf galaxies within the virial radius of the Milky Way is significantly lower compared to the predicted number of dark subhalos, called the missing satellite (MS) problem [52–54].

- The too-big-to-fail (TBTF) problem further sharpens the situation: there is a mismatch between the observed mass enclosed within the half-light radius of bright dwarf spheroidal galaxies and the prediction of CDM N-body only simulation [55, 56].

In contrast, other works point out that all problems described above can be explained within the ΛCDM paradigm (cf., e.g., [57–61]) by baryonic physics, environmental effects, or adopting a certain relation between stellar and halo masses (abundance matching) [62]. Sophisticated hydrodynamical simulations, based on the EAGLE [63], IllustrisTNG [64], and FIRE [65] collaborations, differ in some details in the modelling of baryonic processes, however, there probably is overall consensus that the population and properties of DM halos in such simulations can be brought in agreement with observation (see, e.g., [57, 66]).

In parallel, a variety of alternative solutions to the small scale challenges were offered. A prominent example is warm dark matter (WDM, [67, 68]). Due to the light mass of the order ~ 1 keV, WDM has a large thermal velocity and erases substructure via the free-streaming effect. This leads to a cut-off mass at the order of dwarf galaxy scales, which was conceived to solve the MS and TBTF problem [69, 70]. The thermal velocity would also lead to the formation of density cores. Another example is fuzzy dark matter (fDM) which could potentially solve the MS, TBTF and CC problem [71, 72]. fDM are ultra-light bosonic particles forming a condensate, where the interesting mass region regarding the solution of small scale challenges would be of the order $m_{\text{fDM}} = \mathcal{O}(10^{-21})$ eV. They are produced in a special way (e.g., via the misalignment mechanism) such that they are non-relativistic. Substructure is erased below their enormously large de Broglie wavelength and their halos feature solitonic cores (see, e.g., [72]). However, WDM runs into a Catch 22 problem [73] which might be also the case for fDM [74]. Measurements of the Lyman-α forest [75–77] set a lower bound on the WDM and fDM mass. Within current limits, the minimum allowed mass for both cases would be too large in order to simultaneously address the CC problem [73, 74, 78].

Self-interacting dark matter [79, 80] (SIDM) is cold DM but equipped with a self-scattering cross-section of the order $\sigma/m_{\text{DM}} = \mathcal{O}(0.1 - 10)$ cm^2/g. This self-interaction strength is sufficient enough to thermalize the inner parts of DM halos, leading to an appealing solution of the CC [81–87] and the TBTF problem [88–90] without relying on strong baryonic feedbacks in simulations. Moreover, it has been pointed out that baryons gravitationally interact with SIDM in the right way to potentially explain the diversity problem of galactic rotation curves [91, 92], supported also from combined baryon and SIDM simulations [93, 94].

Observations of the merging galaxy cluster 1E0657-56 set an upper limit on the self-scattering cross-section as sever as $\sigma/m_{\text{DM}} \lesssim 1.25$ cm^2/g [95]. In ref. [96], a statement is given that this upper bound is strongly overstated and needs careful revision. SIDM candidates with a more constant self-scattering cross-section are atomic DM (see, e.g. ref. [97]) or Strongly-Interacting-Massive-Particles (SIMPs) [32]. A mild velocity-dependent self-scattering cross-section (decreasing towards larger relative velocities) might be needed for SIDM particles in order to allow for the observed ellipticity of halos in massive clusters of galaxies [98, 99].

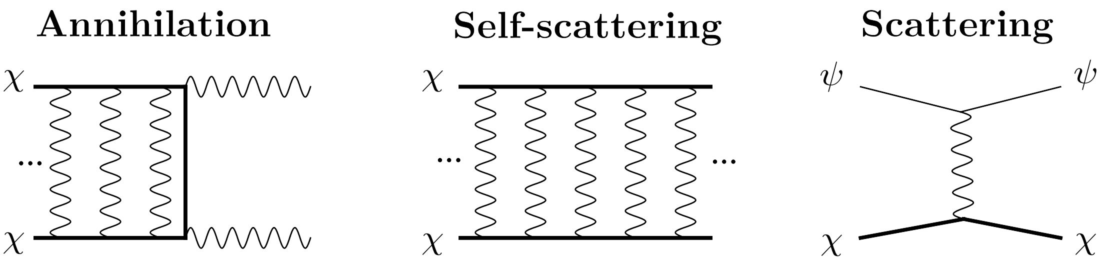


Fig. 2: Graphics illustrate long-range effects induced by light mediators. Left plot shows ladder diagrams which modify the tree-level annihilation process of DM into two light vector mediators. Middle plot shows ladder diagrams which alter the velocity dependence of the self-scattering cross-section. On the right, Coulomb-like scattering keeps DM long time in kinetic equilibrium.

4.2 Light mediators as an alternative solution to the small scale crisis

It is intriguing to see that SIDM, featuring a velocity-dependent self-scattering cross-section, could be already realized in simple particle models where DM is a Dirac fermion χ and interacts either with a light scalar or vector mediator [20] as:

$$\mathcal{L}_V \supset g_\chi \bar{\chi}\gamma^\mu \chi \phi_\mu, \tag{4.1}$$

$$\mathcal{L}_S \supset g_\chi \bar{\chi}\chi\phi. \tag{4.2}$$

Effectively, both mediator types induce a screened Yukawa force between two slowly moving DM particles, described by the potential:

$$V_Y = \pm\frac{\alpha_\chi}{r}e^{-m_\phi r}. \tag{4.3}$$

Here, m_ϕ is the mass of the mediator and $\alpha_\chi = g_\chi^2/(4\pi)$ is the fine-structure constant. This potential enters the two-body Schrödinger equation and the self-scattering cross-section can be computed from the phase-shift method of standard quantum mechanical scattering theory. In the case of Yukawa interaction (scalar mediator) the Yukawa potential is always attractive, whereas in the vector mediator case the potential is attractive for particle-antiparticle scattering and repulsive otherwise. It has been shown in ref. [19] that in all cases, a mediator mass of the order $m_\phi = \mathcal{O}(10-1)$ MeV leads to a self-scattering cross-section of the required order $\sigma/m_\chi = \mathcal{O}(0.1-10)$ cm^2/g at velocity scales typical for dwarf galaxies, possible for a broad range of DM masses $m_\chi = \mathcal{O}(1-10^4)$ GeV and for realistic coupling values g_χ set by the relic density requirement. DM can be thermally produced from the light mediators which are singlets under the SM gauge groups in these simplified models.

The long-range property of the mediators can also lead to a Sommerfeld enhanced annihilation cross-section, which increases towards lower relative velocities, and can become relevant in the relic abundance computation already at the time of chemical decoupling. A proper treatment of the long-range force requires resummation of t-channel exchanges of the mediator, as illustrated in fig. 2 together with the related effect for the self-scattering. DM annihilation into two light vector mediators is to leading order in relative velocity expansion

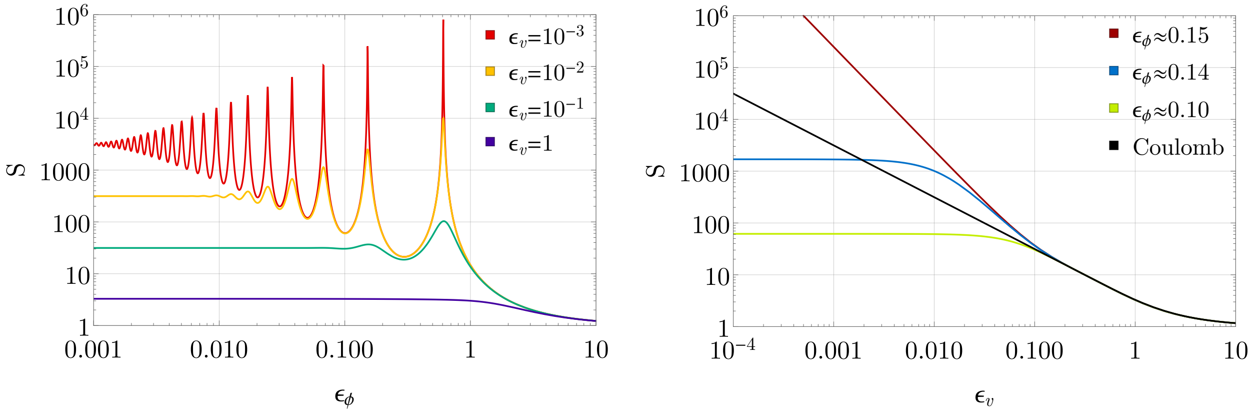


Fig. 3: Graphics show the Sommerfeld factor for an attractive Yukawa and Coulomb potential.

of the tree-level cross-section an s-wave (zero angular momentum) process, for which the total cross-section can be written as (see, e.g., [20, 100]):

$$(\sigma v_{\rm rel}) = (\sigma v_{\rm rel})_0 |\psi(r=0)|^2. \tag{4.4}$$

The s-wave tree-level annihilation cross-section is the short-range part and given by $(\sigma v_{\rm rel})_0 = \pi\alpha_\chi^2/m_\chi^2$. Here, ψ is the solution of the two-body Schrödinger equation for positive energies which accounts for the resummation of ladder exchanges. The absolute value squared of the wave-function at the origin is called the Sommerfeld factor $S \equiv |\psi(0)|^2$.

In fig. 3 the Sommerfeld factor S is shown for an attractive Yukawa potential. The Schrödinger equation is solved numerically, with the same boundary conditions as in scattering theory, to find the wave-function value at the origin. Standard procedures to obtain the solution are provided, e.g., in ref. [101]. In the left plot, the enhancement factor is presented as a function of $\epsilon_\phi \equiv m_\phi/(\alpha_\chi m_\chi)$ for fixed $\epsilon_v \equiv v_{\rm rel}/(2\alpha_\chi)$. For a Yukawa screening length larger than the Bohr radius, i.e., $1/m_\phi \gtrsim 1/(\alpha_\chi m_\chi)$ or $\epsilon_\phi \lesssim 1$, the potential is long-ranged and the Sommerfeld factor leads to a large boost of the annihilation cross-section. For decreasing ϵ_v, there appears a certain peaked structure. At the resonance location, the Sommerfeld enhancement for a Yukawa potential can even exceed the value for a Coulomb potential, as shown in the right plot of fig. 3. For a Coulomb potential, analytic solutions are known and for zero angular momentum given by [102]:

$$|\psi(0)|^2 = \frac{2\pi\alpha}{v_{\rm rel}} \frac{1}{1-e^{-\frac{2\pi\alpha}{v_{\rm rel}}}}. \tag{4.5}$$

For ϵ_ϕ values between resonances, the enhancement factor starts to saturate at some point and approaches a constant value.

The resonant structure can be explained by bound-state contributions. In the Yukawa potential case, there exists in general only a finite number of bound states. For a screening radius smaller than the Bohr radius $\epsilon_\phi \gtrsim 1$, no bound states exist. This corresponds to the regime where the enhancement factor has negligible contribution and the potential is short-range. The first s-wave bound state with principle quantum number $n = 1$ (ground state) starts to exist for ϵ_ϕ values smaller than $\epsilon_\phi \sim 0.6$, where the latter value corresponds to the first resonant peak. Precisely at the location of the first pole, the ground state has zero binding energy [21], leading to a resonant enhancement of the annihilation cross-section for the scattering states. At the second pole, the bound state with principle quantum number $n = 2$ has zero binding energy and exists for smaller ϵ_ϕ values. The limit $\epsilon_\phi \to 0$ results in the Coulomb potential, where infinitely many bound states exist. Thus, for small enough ϵ_v, the left plot of fig. 3 allows to count the number of existing bound-state solutions of the Schrödinger equation. In the article reprinted in section 5.4, the Sommerfeld enhancement is represented in terms of a two-particle spectral function, where the resonant structure of fig. 3 can be better understood.

Turning to small scale structure formation, the MS problem can not be solved by SIDM alone. Moreover, in the simplified models introduced above, the mediators would be stable and shift matter-radiation equality to the time when they become non-relativistic, which is clearly ruled out for MeV-scale masses. In ref. [18], the light vector mediator is additionally coupled to other effectively massless Dirac fermions ψ, called dark radiation, which allows the mediator to decay into a pair of ψ particles. Furthermore, the additional particle introduces the possibility for Coulomb-like scattering processes of the form $\chi\psi \to \chi\psi$ with a light vector mediator in the t-channel, as illustrated in fig. 2. The interesting observation is that for a MeV-scale vector mediator, DM could kinetically decouple extremely late such that the effect of dark acoustic oscillations would erase the substructure at the dwarf galaxy scale which might solve the MS problem. The parameter region, where the MS problem could be alleviated, has overlap with the region where χ would be a SIDM candidate, which means that all structure formation problems could be ameliorated simultaneously [18]. This simplified model has no Catch 22 problem since the self-scattering cross-sections can be made independently of the kinetic decoupling temperature. Lyman-α forest measurements set a lower limit on the latter quantity, derived in ref. [8], which will be discussed in the article reprinted in section 4.4.

An embedding of the vector mediator case into the SM is challenging. If the vector mediator couples to standard model particles via kinetic mixing with the hypercharge field strength or mass mixing with the Z-boson, (almost) the entire parameter space in the SIDM region would be ruled out by CMB observations [103, 104]. The main reason is that at the time of recombination, the DM annihilation into light vector mediators, which subsequently decay into charged SM particles, would be too efficient due to the Sommerfeld enhancement. The energy injection heats the photons which distorts the CMB, in conflict with current limits [1]. Asymmetric dark matter might be a way out of this conflict [105, 106].

In the next two sections, it is assumed that DM, the light mediators, and ψ particles are all decoupled from the SM. Mediator types are varied in the paper reprinted in section 4.3 to investigate if there are further scenarios featuring late kinetic decoupling. In the article reprinted in section 4.4, it is investigated if Sommerfeld enhanced annihilation cross-sections in a totally closed dark sector could lead to an identifiable gravitational signature.

4.3 *Publication:* Matter power spectrum in hidden neutrino interacting dark matter models

Reprinted article with permission from the authors:

Tobias Binder, Laura Covi, Ayuki Kamada, Hitoshi Murayama, Tomo Takahashi, and Naoki Yoshida.

Journal of Cosmology and Astroparticle Physics **1611**, 043 (2016)

https://doi.org/10.1088/1475-7516/2016/11/043

Copyright 2016 IOP Publishing.

Author contributions The Fokker-Planck formalism for DM interacting with light degrees of freedom, the cosmological perturbation theory, and the implementation of the modified evolution equations into CAMB [107] was developed by A. K. during his Ph.D. before the project started, under the supervision of H. M. , T. T. and N. Y. . H. M. and N. Y. founded and supervised the project. A. K. and T. B. recognized together the importance of t-averaging the scattering amplitude. T. B. suggested to study the specific type of dark radiation interacting DM models. A. K. adjusted his modified CAMB code for the concrete particle setup and produced the matter power spectra. T. B. and L. C. worked on the particle physics input. T. B. did the analytic computations of the cut-off masses, the relic density and BBN and CMB constraints. A. K. and T. T. worked together on estimating the validity region of the perfect fluid approximation. T. B. and A. K. wrote the manuscript. All authors discussed the results and contributed to the revision of the article.

Journal of Cosmology and
Astroparticle Physics

OPEN ACCESS

Matter power spectrum in hidden neutrino interacting dark matter models: a closer look at the collision term

To cite this article: Tobias Binder *et al* JCAP11(2016)043

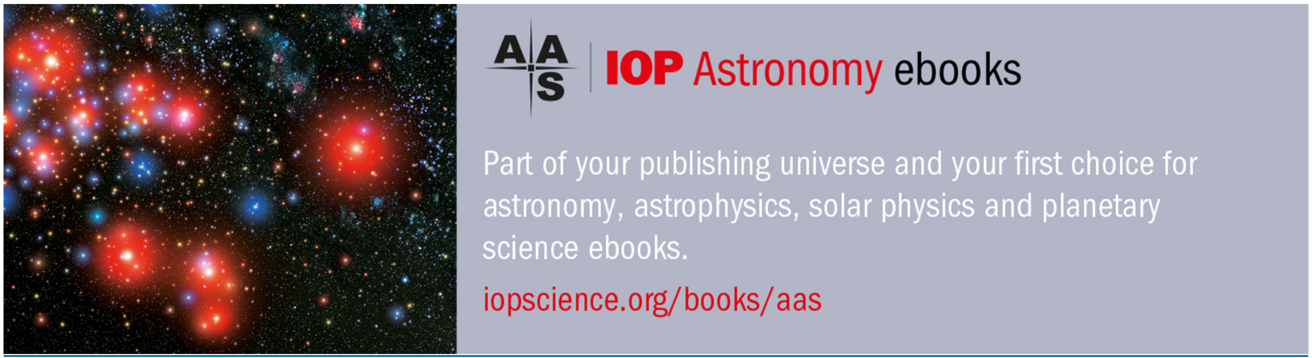
View the article online for updates and enhancements.

Related content

- Complementary test of the dark matter self-interaction in dark U(1) model by direct and indirect dark matter detection
  Chian-Shu Chen, Guey-Lin Lin and Yen-Hsun Lin
- Fermi Bubbles under Dark Matter Scrutiny Part II: Particle Physics Analysis
  Wei-Chih Huang, Alfredo Urbano and Wei Xue
- Coulomb's law corrections and fermion field localization in a tachyonic de Sitter thick braneworld
  Roberto Cartas-Fuentevilla, Alberto Escalante, Gabriel Germán et al.

Recent citations

- Self-heating of strongly interacting massive particles
  Ayuki Kamada *et al*
- Reannihilation of self-interacting dark matter
  Tobias Binder *et al*
- On dark matter-dark radiation interaction and cosmic reionization
  Subinoy Das *et al*


AAS | IOP Astronomy ebooks

Part of your publishing universe and your first choice for astronomy, astrophysics, solar physics and planetary science ebooks.

iopscience.org/books/aas

This content was downloaded from IP address 134.76.38.153 on 21/09/2018 at 13:14

Journal of Cosmology and Astroparticle Physics
An IOP and SISSA journal

Matter power spectrum in hidden neutrino interacting dark matter models: a closer look at the collision term

Tobias Binder,[a] **Laura Covi,**[a] **Ayuki Kamada,**[b] **Hitoshi Murayama,**[c,d,e] **Tomo Takahashi**[f] **and Naoki Yoshida**[c,g,h]

[a]Institute for Theoretical Physics, Georg-August University Göttingen,
Friedrich-Hund-Platz 1, Göttingen, D-37077 Germany
[b]Department of Physics and Astronomy, University of California,
Riverside, California 92521, U.S.A.
[c]Kavli Institute for the Physics and Mathematics of the Universe (WPI),
University of Tokyo Institutes for Advanced Study, University of Tokyo,
Kashiwa 277-8583, Japan
[d]Department of Physics, University of California, Berkeley,
Berkeley, California 94720, U.S.A.
[e]Theoretical Physics Group, Lawrence Berkeley National Laboratory,
Berkeley, California 94720, U.S.A.
[f]Department of Physics, Saga University,
Saga 840-8502, Japan
[g]Department of Physics, University of Tokyo,
Tokyo 113-0033, Japan
[h]CREST, Japan Science and Technology Agency,
4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

E-mail: tobias.binder@theorie.physik.uni-goettingen.de,
laura.covi@theorie.physik.uni-goettingen.de, ayuki.kamada@ucr.edu,
hitoshi.murayama@ipmu.jp, tomot@cc.saga-u.ac.jp,
naoki.yoshida@phys.s.u-tokyo.ac.jp

Received March 7, 2016
Revised July 6, 2016
Accepted November 9, 2016
Published November 21, 2016

Article funded by SCOAP^3. Content from this work may be used under the terms of the Creative Commons Attribution 3.0 License. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

doi:10.1088/1475-7516/2016/11/043

Abstract. Dark Matter (DM) models providing possible alternative solutions to the small-scale crisis of the standard cosmology are nowadays of growing interest. We consider DM interacting with light hidden fermions via well-motivated fundamental operators showing the resultant matter power spectrum is suppressed on subgalactic scales within a plausible parameter region. Our basic description of the evolution of cosmological perturbations relies on a fully consistent first principles derivation of a perturbed Fokker-Planck type equation, generalizing existing literature. The cosmological perturbation of the Fokker-Planck equation is presented for the first time in two different gauges, where the results transform into each other according to the rules of gauge transformation. Furthermore, our focus lies on a derivation of a broadly applicable and easily computable collision term showing important phenomenological differences to other existing approximations. As one of the main results and concerning the small-scale crisis, we show the equal importance of vector and scalar boson mediated interactions between the DM and the light fermions.

Keywords: cosmological perturbation theory, dark matter theory, particle physics - cosmology connection, power spectrum

ArXiv ePrint: 1602.07624

Contents

1 Introduction **1**

2 Fokker-Planck equation **3**
2.1 Collision term 3
2.2 Perturbation theory in the synchronous gauge 6

3 Neutrino interacting dark matter **9**
3.1 Simplified neutrino model 10
3.2 Minimal halo mass 10
3.3 Matter power spectrum 14

4 Summary and outlook **15**

A Perturbation theory in the conformal Newtonian gauge **17**

B Impact of the higher order terms in the Boltzmann hierarchy **18**

C Thermal history calculation **20**
C.1 Relic abundance 20
C.2 Minimal halo mass of the pseudo scalar and pseudo vector operators 21

1 Introduction

Precise measurements of cosmic microwave background (CMB) anisotropies have been building strong evidence for the existence of a new form of matter, called dark matter (DM) [1, 2]. However, its nature has not yet been uncovered and one of the most important subjects both in astrophysics and in particle physics. Recently vigorous efforts have been devoted to cosmological probes of interaction strengths between the DM and other long-lived particles [3–16]. Interestingly, such probes are not limited to couplings of the DM to standard model (SM) particles (e.g., baryons, photons, and neutrinos). Couplings to hidden particles are equally subject to searches. In this paper, we restrict our discussion to hidden light particles, which we call neutrinos for simplicity. However, the formulation developed and given in this paper is applicable to other models with DM couplings to SM particles.

Interacting DM models are not only within the scope of precise measurements of the large-scale structure of the Universe. They also have their motivation in apparent discrepancies between predictions from DM-only N-body simulations and observations on subgalactic scales. Such discrepancies are called the *small-scale crisis* collectively: the *missing satellite problem* [17, 18]; the *cusp vs core problem* [19–21]; the *too big to fail* problem [22, 23]. The simulations assume that the DM consists of particles with negligible thermal velocities and faint interactions, called cold dark matter (CDM). The small-scale crisis may imply alternatives to the CDM paradigm, while it has to be clarified by state-of-the-art hydrodynamic simulations what role baryonic processes play in the formation and evolution of subgalactic objects [24, 25]. One famous alternative is called the warm dark matter (WDM) model, in which sizable thermal velocities of DM particles suppress the formation of subgalactic objects [26]. Interacting DM effectively reduces the abundance of substructures in a galactic halo to a similar degree to that some WDM models do [27–36].

Although there is a growing interest in interacting DM models, it is still unclear what the evolution equations of cosmological perturbations are in such models. This is because it is difficult to handle and simplify collision terms of the Boltzmann equations. Some works start with the relativistic Navier-Stokes equation for the DM imperfect fluid in particle flow manifest (Eckart's) formulation [37, 38]. They determine fluid variables with the help of the Chapman-Enskog method to estimate damping scales in matter power spectra in interacting DM models. Others just add a collision term in the evolution equations of cosmological perturbations by analogy to the well-known Thomson scattering term for baryons and photons [39]. One plausible way is to reduce the collision term to the Fokker-Planck equation by assuming the momentum transfer in each collision is smaller than the typical DM momentum. Such formulation is developed for the traditional bino-like DM in [40]. However, the overall factor, i.e., the reaction rate, in the Fokker-Planck equation is controversial so far. A systematic expansion of the collision term in terms of the momentum transfer leads to a reaction rate proportional to the invariant amplitude at a zero momentum transfer $t \to 0$ [41–43]. On the other hand, in [44, 45], the reaction rate is given by t-averaging like $\int dt(-t)d\sigma/dt$.

The two formulations introduced above result in different phenomenological consequences. We consider a simple model, where the SM sector is extended by a Dirac DM, a Dirac (hidden) neutrino, and a mediator. A similar scenario is investigated in [28]. When the mediator is a scalar, the reaction rate with a zero momentum transfer is negligible and does not change the matter power spectra on and above subgalactic scales within a plausible range of model parameters. A subgalactic damping scale can be achieved by a vector mediator within this formulation [29]. On the other hand, both vector and scalar mediators can suppress the resultant matter power spectra with the t-averaged reaction rate. We address this point by calculating the resultant matter power spectra in the latter formulation numerically. To this end, we derive the evolution equations of cosmological perturbations in two gauges: the conformal Newtonian gauge and the synchronous gauge [46]. We provide an explicit form of gauge transformations between them. We also show a derivation of the t-averaged reaction rate. It may be useful because a similar derivation is given only in an unpublished thesis [44]. In the recent ETHOS (effective theory of structure formation) papers [34, 35], they study the structure formation in interacting DM models based on the former treatment of the collision term, while they also mention the importance of the t-averaging in some models. The ETHOS paper and this paper are complementary to each other in a treatment of the collision term.

The paper is organized as follows. In section 2 we start from first principles and give a detailed derivation of the Fokker-Planck equation with the t-averaged reaction rate. Furthermore, the evolution equations of cosmological perturbations in the synchronous gauge are derived for the most general case of an imperfect fluid. We show explicitly in appendix A that our results transform into the form of the conformal Newtonian gauge according to the rules of gauge transformation. In section 3, we give an introduction of the neutrino interacting DM model first. Then, we summarize our analytic results for scalar and vector mediators: the relic density of the DM; the t-averaged reaction rate; the resultant smallest mass of halos. In appendix C, we present details of our calculations of the chemical decoupling and also summarize the results for models with pseudo scalar and pseudo vector mediators. Finally, we show the resultant matter power spectra to stress that not only a vector, but also a

JCAP11(2016)043

scalar mediator can lead to a sizable suppression of matter power spectra. In appendix B, we provide an estimation of a critical wavenumber below which the perfect fluid approximation appears valid.

Throughout this paper, we use the *Planck* 2013 cosmological parameters [2]: $\Omega_{\mathrm{m}} = 0.3175$, $H_0 = 67.11$, $\ln(10^{10} A_{\mathrm{s}}) = 3.098$, and $n_{\mathrm{s}} = 0.9624$. Updating these input parameters to the *Planck* 2015 ones would not change our results significantly.

2 Fokker-Planck equation

In this section, the perturbed Fokker-Planck equation is derived. Our starting point is the classical Boltzmann equation with the collision term. We expand it assuming the momentum transfer per collision is smaller than the typical DM momentum. Within this approximation the collision term satisfies *detailed balance* and respects number conservation. As a further result of this expansion method, the momentum transfer rate can easily be computed by first taking a t-average and secondly a thermal average of the differential scattering cross section. As an important result of the formulation used, the t-average is a direct consequence of the expansion method. Other methods like in [42] expand the scattering amplitude in terms of a small momentum transfer and keep only the zero order. But this approximation shows a completely different phenomenology for certain DM theories as will be shown as an explicit example in section 3. As part of this section, we develop the evolution equations of linear cosmological perturbations in the synchronous gauge. A comparison to previous works is given. The results are equivalent to those in the conformal Newtonian gauge under the gauge transformation law as we show for the first time in appendix A.

2.1 Collision term

In this subsection, we derive a Fokker-Planck equation for the DM phase space distribution function f, partially inspired by the unpublished thesis [44]. Our starting point is the classical Boltzmann equation for the DM,

$$\left[P^{\mu}\partial_{x^{\mu}} - \Gamma^{\mu}_{\kappa\lambda}P^{\kappa}P^{\lambda}\partial_{P^{\mu}}\right] f = C[f]\,, \tag{2.1}$$

where P_{μ} is the conjugate momentum of the spatial coordinate x^{μ}. When we handle the collision term $C[f]$, it is convenient to take a local inertial frame X^{μ}, where the metric is $\mathrm{diag}(-1,+1,+1,+1)$ and the proper momentum is denoted by $p^{\mu} = (E, \mathbf{p})$. We normalize the distribution function such that $\sum_s \int d^3\mathbf{p}/(2\pi)^3 (p^{\mu}/E) f = n^{\mu}$, where s are spin degrees of freedom and n^{μ} is the DM number current. If we assume the DM particles to interact elastically with particles in a thermal bath, i.e., DM(1)+TP(2) ↔ DM(3)+TP(4), the collision term takes the form of

$$C[f_1] = \frac{1}{2}\sum_{s_3}\int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\left[-S^{\mathrm{eq}}(p_1,p_3) f_1(1\mp f_3) + S^{\mathrm{eq}}(p_3,p_1) f_3(1\mp f_1)\right], \tag{2.2}$$

where

$$\begin{aligned} S^{\mathrm{eq}}(p_1,p_3) = {} & \sum_{s_2}\int \frac{d^3\mathbf{p}_2}{(2\pi)^3 2E_2}\sum_{s_4}\int \frac{d^3\mathbf{p}_4}{(2\pi)^3 2E_4}(2\pi)^4\delta^4(p_1+p_2-p_3-p_4) \\ & \times\overline{|\mathcal{M}(1+2\to 3+4)|^2} f_2^{\mathrm{eq}}(1\mp f_4^{\mathrm{eq}})\,, \end{aligned} \tag{2.3}$$

$$\begin{aligned} S^{\mathrm{eq}}(p_3,p_1) = {} & \sum_{s_2}\int \frac{d^3\mathbf{p}_2}{(2\pi)^3 2E_2}\sum_{s_4}\int \frac{d^3\mathbf{p}_4}{(2\pi)^3 2E_4}(2\pi)^4\delta^4(p_3+p_4-p_1-p_2) \\ & \times\overline{|\mathcal{M}(3+4\to 1+2)|^2} f_4^{\mathrm{eq}}(1\mp f_2^{\mathrm{eq}})\,. \end{aligned} \tag{2.4}$$

Here, $\overline{|\mathcal{M}|^2}$ is the spin-averaged invariant amplitude squared, and $f^{\rm eq}$ is a thermal distribution,

$$f^{\rm eq} = (\exp\{(-p \cdot u - \mu)/T\} \pm 1)^{-1} \tag{2.5}$$

with a temperature $T \simeq T_0(\tau) + T_1(x)$, a reference four velocity $u^\mu \simeq (1, \mathbf{u}(x))$, and a chemical potential μ.

If the elastic scattering is T*-inversion invariant*, $\overline{|\mathcal{M}|^2}$'s are identical between the forward and backward scatterings,

$$\overline{|\mathcal{M}(1+2 \to 3+4)|^2} = \overline{|\mathcal{M}(3+4 \to 1+2)|^2} = \overline{|\mathcal{M}|^2}\,. \tag{2.6}$$

In the presence of four-momentum conservation $\delta^4(p_1 + p_2 - p_3 - p_4)$, thermal distributions satisfy

$$f_2^{\rm eq}(1 \mp f_4^{\rm eq}) = \exp\{-(p_1 - p_3) \cdot u/T\} f_4^{\rm eq}(1 \mp f_2^{\rm eq})\,. \tag{2.7}$$

From (2.6) and (2.7), we obtain the following relation:

$$S^{\rm eq}(p_1, p_3) = \exp\{-(p_1 - p_3) \cdot u/T\} S^{\rm eq}(p_3, p_1). \tag{2.8}$$

Thus, the collision term is

$$C[f_1] = \frac{1}{2} \sum_{s_3} \int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3} S^{\rm eq}(p_3, p_1) \big[f_3(1 \mp f_1) - \exp\{-(p_1 - p_3) \cdot u/T\} f_1(1 \mp f_3) \big]\,. \tag{2.9}$$

We can easily check that the above expression satisfies the so-called detailed balance, i.e., $C[f_1] = 0$ if $f_1 = f_1^{\rm eq}$ and $f_3 = f_3^{\rm eq}$, which follows from the T-inversion invariance.

We assume that the momentum transfer $\tilde{\mathbf{q}} = \mathbf{p}_3 - \mathbf{p}_1$ is smaller than the typical DM momentum p_{1i} and expand the collision term up to the second order,

$$f_3 \simeq f_1 + \tilde{\mathbf{q}}_i \frac{\partial f_1}{\partial \mathbf{p}_{1i}} + \frac{1}{2} \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_j \frac{\partial^2 f_1}{\partial \mathbf{p}_{1i} \partial \mathbf{p}_{1j}}\,, \quad \exp\{-(p_1 - p_3) \cdot u/T\} = 1 + A_i \tilde{\mathbf{q}}_i + B_{ij} \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_j\,, \tag{2.10}$$

where

$$A_i = -\frac{\mathbf{v}_{1i} - \mathbf{u}_i}{T}\,, \quad B_{ij} = \frac{1}{2} \left(\frac{\partial A_i}{\partial \mathbf{p}_{1j}} + A_i A_j \right)\,, \tag{2.11}$$

with the velocity of the particle $\mathbf{v} = \mathbf{p}/E$. After collecting terms, we obtain

$$\big[f_3(1 \mp f_1) - \exp\{-(p_1 - p_3)/T\} f_1(1 \mp f_3) \big] \simeq \alpha_i \tilde{\mathbf{q}}_i + \frac{1}{2} \left(\frac{\partial \alpha_i}{\partial \mathbf{p}_{1j}} + \alpha_i A_j \right) \tilde{\mathbf{q}}_i \tilde{\mathbf{q}}_j\,, \tag{2.12}$$

where

$$\alpha_i = \frac{\partial f_1}{\partial \mathbf{p}_{1i}} - A_i f_1 (1 \mp f_1)\,. \tag{2.13}$$

The collision term is

$$C[f_1] \simeq \frac{1}{2} \left\{ \alpha_i \beta_i + \frac{1}{2} \left(\frac{\partial \alpha_i}{\partial \mathbf{p}_{1j}} + \alpha_i A_j \right) \gamma_{ij} \right\}, \tag{2.14}$$

where

$$\beta_i = \sum_{s_3}\int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3} S^{\rm eq}(p_3, p_1)\tilde{\mathbf{q}}_i\,, \quad \gamma_{ij} = \sum_{s_3}\int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3} S^{\rm eq}(p_3, p_1)\tilde{\mathbf{q}}_i\tilde{\mathbf{q}}_j\,. \tag{2.15}$$

It should be noted that $\alpha_i = 0$ and thus $C[f_1] = 0$ if $f_1 = f_1^{\rm eq}$, which implies that the detailed balance is maintained in the approximation.

We expand $S^{\rm eq}(p_3, p_1)$ in terms of $\tilde{\mathbf{q}}$, noting that $p_3^\mu = (\sqrt{E_1^2 + 2\mathbf{p}_{1i}\tilde{\mathbf{q}}_i + \tilde{\mathbf{q}}^2}, \mathbf{p}_1 + \tilde{\mathbf{q}})$. The scalar function $S^{\rm eq}$ depends on $\tilde{\mathbf{q}}$ only through $\mathbf{p}_{1i}\tilde{\mathbf{q}}_i$ and $\tilde{\mathbf{q}}^2$. Since we keep the terms only up to the second order in terms of $\tilde{\mathbf{q}}$, the expansion in terms of $\tilde{\mathbf{q}}^2$ leads to higher order terms in $C[f_1]$, which are to be neglected in our treatment. Therefore we expand $S^{\rm eq}$ only in terms of $\mathbf{p}_{1i}\tilde{\mathbf{q}}_i$ as follows:[1]

$$S^{\rm eq}(p_3, p_1) \simeq S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2) + S_1^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\mathbf{p}_{1i}\tilde{\mathbf{q}}_i\,, \tag{2.16}$$

$$S^{\rm eq}(p_1, p_3) \simeq S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2) + \frac{\partial S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)}{\partial \mathbf{p}_{1i}}\tilde{\mathbf{q}}_i - S_1^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\mathbf{p}_{1i}\tilde{\mathbf{q}}_i\,, \tag{2.17}$$

where $S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)$ and $S_1^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)$ are the expansion coefficients defined by (2.16).

Substituting (2.16) and (2.17) into (2.15) and using (2.8) and (2.10), we obtain

$$\begin{aligned}
\beta_i &= \sum_{s_3}\int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3}\frac{1}{2}\left[S^{\rm eq}(p_3, p_1) + \exp\{(p_1 - p_3)\cdot u/T\}S^{\rm eq}(p_1, p_3)\right]\tilde{\mathbf{q}}_i \\
&\simeq \sum_{s_3}\int \frac{d^3\mathbf{p}_3}{2(2\pi)^3}\left[\left(-\frac{\mathbf{p}_{1j}\tilde{\mathbf{q}}_j}{E_1^3}\right)S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\tilde{\mathbf{q}}_i + \frac{1}{2E_1}(-A_j\tilde{\mathbf{q}}_j)S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\tilde{\mathbf{q}}_i \right.\\
&\quad \left. + \frac{1}{2E_1}\frac{\partial S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)}{\partial \mathbf{p}_{1j}}\tilde{\mathbf{q}}_i\tilde{\mathbf{q}}_j\right] \\
&\simeq \frac{E_1}{2}\frac{\partial}{\partial \mathbf{p}_{1j}}\left(\frac{1}{E_1}\tilde{\gamma}_{ij}\right) - \frac{1}{2}A_j\tilde{\gamma}_{ij}\,,
\end{aligned} \tag{2.18}$$

where

$$\tilde{\gamma}_{ij} = \frac{1}{3}\delta_{ij}\sum_{s_3}\int \frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_1} S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\tilde{\mathbf{q}}^2\,. \tag{2.19}$$

In the second equality of (2.18), we have dropped the term proportional to $S_0^{\rm eq}(p_1, \tilde{\mathbf{q}}^2)\tilde{\mathbf{q}}_i$, since it vanishes after integrated in terms of $d^3\mathbf{p}_3 = d^3\tilde{\mathbf{q}}$. In addition, in the last equality, we have replaced $\tilde{\mathbf{q}}_i\tilde{\mathbf{q}}_j$ with $(1/3)\delta_{ij}\tilde{\mathbf{q}}^2$ for the same reason. It should be noted that, $\gamma_{ij} = \tilde{\gamma}_{ij}$ holds up to the second order in $\tilde{\mathbf{q}}$.

In practice it is more convenient when evaluating (2.19) to replace the perturbative quantities $S_0^{\rm eq}(\mathbf{p}_1, \tilde{\mathbf{q}}^2)$ and $-\tilde{\mathbf{q}}^2$ with their non-perturbative counterparts $S^{\rm eq}(p_3, p_1)$ and $t/(1 - \mathbf{v}_{1i}\mathbf{v}_{1i}/3) = -(p_3 - p_1)^2/(1 - \mathbf{v}_{1i}\mathbf{v}_{1i}/3)$, respectively. These resulting coefficients only differ through higher order terms and amount to an alternate perturbative expansion. Substituting (2.18) and (2.19) into (2.14), we obtain a Fokker-Planck-type equation for f_1 since the collision term becomes

$$C[f_1] \simeq E_1\frac{\partial}{\partial \mathbf{p}_{1i}}\left[\gamma\left(E_1 T\frac{\partial f_1}{\partial \mathbf{p}_{1i}} + (\mathbf{p}_{1i} - E_1\mathbf{u}_i)f_1(1 \mp f_1)\right)\right], \tag{2.20}$$

---

[1] In fact, if we take $\tilde{\mathbf{q}} \to 0$, $S^{\rm eq}$ diverges owing to a delta function of zero $\delta(0)\delta^3(\mathbf{p}_4 - \mathbf{p}_2)$ in the integrand. The expansion just in terms of $\mathbf{p}_{1i}\tilde{\mathbf{q}}_i$ also allows us to avoid such a divergence.

where the momentum transfer rate is

$$\gamma = \frac{1}{6E_1 T(1-\mathbf{v}_{1i}\mathbf{v}_{1i}/3)} \sum_{s_2} \int \frac{d^3\mathbf{p}_2}{(2\pi)^3} f_2^{\rm eq}(1 \mp f_2^{\rm eq}) \int_{-4\mathbf{p}_{\rm cm}^2}^{0} dt(-t)\frac{d\sigma}{dt} v\,. \tag{2.21}$$

Here, $d\sigma/dt$ is the differential cross section and v is the relative velocity of the initial particles. The center of mass momentum is evaluated by $4s\mathbf{p}_{\rm cm}^2 = \{s-(m_1-m_2)^2\}\{s-(m_1+m_2)^2\}$, where $s = -(p_1+p_2)^2$ and m is the mass of the particle. This equation satisfies two important requirements. First, it maintains the detailed balance: if $f_1 = f_1^{\rm eq}$, then $C[f_1] = 0$. Second, it conserves the DM number,

$$\partial_{X^\mu} n_1^\mu = \sum_{s_1} \int \frac{d^3\mathbf{p}_1}{(2\pi)^3} \frac{C[f_1]}{E_1} = 0\,. \tag{2.22}$$

If the DM particles decouple from the thermal bath when they are relativistic, momentum transfer in each collision is as large as the typical momentum of the DM, which may spoil our approximation approach, i.e., the Fokker-Planck equation. It may be useful to give the non-relativistic limit. Then, the Fokker-Planck equation is

$$C[f_1] = m_1 \frac{\partial}{\partial \mathbf{p}_{1i}} \left[\gamma \left(m_1 T \frac{\partial f_1}{\partial \mathbf{p}_{1i}} + (\mathbf{p}_{1i} - m_1 \mathbf{u}_i) f_1 \right) \right], \tag{2.23}$$

where the momentum transfer rate is

$$\gamma = \frac{1}{6m_1 T} \sum_{s_2} \int \frac{d^3\mathbf{p}_2}{(2\pi)^3} f_2^{\rm eq}(1 \mp f_2^{\rm eq}) \int_{-4\mathbf{p}_2^2}^{0} dt(-t)\frac{d\sigma}{dt} v\,. \tag{2.24}$$

This expression is the same as given in [44, 45]. The cross section is essentially independent of $\mathbf{p}_1$ since we consider the case that DM is non-relativistic, or in other words, $|\mathbf{p}_1|$ is much smaller than m_1. We focus on such a case in the following.

Before closing this subsection, let us discuss the relation between [41–43] and the present paper. The main difference is the presence of t-averaging in the momentum transfer rate of (2.24). Once we set $t \to 0$ in $d\sigma/dt$, we can evaluate the t-integral analytically to reproduce the result in [41–43]. The t-averaging originates from the approximation in (2.16) and (2.17) and the replacement of the perturbative quantities with the non-perturbative counterparts after (2.19). In this respect, our formulation is not a systematic expansion in terms of the momentum transfer like that in [42]. However, in some cases, the expansion of invariant amplitudes is not a good approximation since the leading order does not give the dominant contribution. One such example is the scalar operator of DM-neutrino interaction investigated in the present paper. There, the leading order is suppressed by a factor of $m_\nu^2/(-t)$ when compared to the next-to-leading order.

2.2 Perturbation theory in the synchronous gauge

Now we develop a linear theory in the synchronous gauge:

$$ds^2 = a^2\left[-d\tau^2 + (\delta_{ij} + h_{ij})d\mathbf{x}^i d\mathbf{x}^j\right]. \tag{2.25}$$

Up to the first order of cosmological perturbations, the Fokker-Planck equation is given by

$$\dot f + \frac{\mathbf{q}_i}{m_\chi}\frac{\partial}{\partial \mathbf{x}_i} f - \frac{1}{2}\dot h_{ij}\mathbf{q}_i \frac{\partial}{\partial \mathbf{q}_j} f = (\gamma_0 + \gamma_1) a \frac{\partial}{\partial \mathbf{q}_i}\left[(\mathbf{q}_i - a m_\chi \mathbf{u}_i) f + a^2 m_\chi (T_0 + T_1)\frac{\partial}{\partial \mathbf{q}_i} f\right], \tag{2.26}$$

JCAP11(2016)043

with $\gamma = \gamma_0(\tau) + \gamma_1(x)$ and the comoving momentum $q = ap$.[2] The homogeneous and isotropic part, i.e., the leading order is

$$\dot{f}_0 = \gamma_0 a \frac{\partial}{\partial \mathbf{q}_i}\left[\mathbf{q}_i f_0 + a^2 m_\chi T_0 \frac{\partial}{\partial \mathbf{q}_i} f_0\right] . \tag{2.27}$$

A solution,

$$f_0 = \frac{\bar{n}}{g_\chi}\left(\frac{2\pi}{m_\chi T_{\chi 0}}\right)^{3/2} \exp\left(-\frac{\mathbf{q}^2}{2a^2 m_\chi T_{\chi 0}}\right) , \tag{2.28}$$

is parametrized by the DM temperature $T_{\chi 0}(\tau)$ and the DM number density per spin degree of freedom $\bar{n}/g_\chi$ with $g_\chi = 2s_\chi + 1$. Its evolution is described by

$$\frac{d\ln(a^2 T_{\chi 0})}{d\tau} = 2\gamma_0 a\left(\frac{T_0}{T_{\chi 0}} - 1\right) . \tag{2.29}$$

The DM temperature is tightly coupled to the temperature of the thermal bath $T_{\chi 0} = T_0 \propto 1/a$ before the kinetic decoupling $\gamma/H > 1$. After they decouple, the DM particles start to stream freely and the temperature decreases adiabatically $T_{\chi 0} \propto 1/a^2$.

The first order perturbation follows:

$$\begin{aligned}\dot{f}_1 + \frac{\mathbf{q}_i}{m_\chi}\frac{\partial}{\partial \mathbf{x}_i} f_1 - \frac{1}{2}\dot{h}_{ij}\mathbf{q}_j \frac{\partial}{\partial \mathbf{q}_i} f_0 = \gamma_1 a L_{\rm FP}[f_0] - \gamma_0 a^2 m_\chi \mathbf{u}_i \frac{\partial}{\partial \mathbf{q}_i} f_0 + \gamma_0 a^3 m_\chi T_1 \frac{\partial^2}{\partial \mathbf{q}^2} f_0 \\ + \gamma_0 a L_{\rm FP}[f_1] .\end{aligned} \tag{2.30}$$

Here, we define the Fokker-Planck operator by

$$L_{\rm FP}[f] = \frac{\partial}{\partial \mathbf{q}_i}\left[\mathbf{q}_i f + a^2 m_\chi T_0 \frac{\partial}{\partial \mathbf{q}_i} f\right] . \tag{2.31}$$

In the Fourier space $\mathbf{k}_i = k\hat{\mathbf{k}}_i$, these equations are rewritten as

$$\begin{aligned}\dot{f}_1 + \frac{i\mathbf{k}_i\mathbf{q}_i}{am_\chi} f_1 - \gamma_0 a L_{\rm FP}[f_1] = \dot{\eta}\frac{\mathbf{q}^2}{2a^2 m_\chi T_{\chi 0}} f_0 - \frac{\dot{h} + 6\dot{\eta}}{2k^2}(\mathbf{k}_i\mathbf{q}_i)^2 \frac{1}{2a^2 m_\chi T_{\chi 0}} f_0 - \frac{i\mathbf{k}_i\mathbf{q}_i}{aT_{\chi 0}}\gamma_0 a \frac{\theta_{\rm TP}}{k^2} f_0 \\ + \left[\gamma_1 a\left(\frac{T_0}{T_{\chi 0}} - 1\right) + \gamma_0 a \frac{T_1}{T_{\chi 0}}\right]\left(\frac{\mathbf{q}^2}{2a^2 m_\chi T_{\chi 0}} - 3\right) f_0 .\end{aligned} \tag{2.32}$$

Hereafter we consider only the scalar perturbations, defining $\theta_{\rm TP}$, η, and h such that $\theta_{\rm TP} = i\mathbf{k}_i\mathbf{u}_i$ and $h_{ij} = \hat{\mathbf{k}}_i\hat{\mathbf{k}}_j h + \left(\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j - \frac{1}{3}\delta_{ij}\right) 6\eta$ (the same notation as in [46]).

In order to handle the Fokker-Planck operator, we expand f_1 in terms of eigenfunctions of the Fokker-Planck operator,

$$L_{\rm FP}\phi_{n\,\ell\,m} = -(2n + \ell)\phi_{n\,\ell\,m} , \quad \phi_{n\,\ell\,m} = e^{-y} S_{n\,\ell}(y) Y_{\ell\,m}(\hat{\mathbf{q}}) , \tag{2.33}$$

with $y = \mathbf{q}^2/(2a^2 m_\chi T_0)$, $\mathbf{q}_i = |\mathbf{q}|\hat{\mathbf{q}}_i$, and a dimensionless function $S_{n\,\ell}(y) = y^{\ell/2} L_n^{\ell+1/2}(y)$. Here $Y_{\ell\,m}$ and L_n^α denote the spherical harmonics and the Laguerre polynomials, respectively. Noting the rotational symmetry, we can write

$$f_1(\mathbf{k}, \mathbf{q}, \tau) = \frac{1}{(2\pi a^2 m_\chi T_0)^{3/2}} e^{-y} \sum_{n,\ell=0}^{\infty} (-i)^\ell (2\ell + 1) S_{n\,\ell}(y) P_\ell(\hat{\mathbf{k}}_i\hat{\mathbf{q}}_i) f_{n\ell}(k, \tau) , \tag{2.34}$$

---

[2] Hereafter, for notational simplicity, we respectively use m_χ and p for the DM mass and the proper momentum instead of m_1 and p_1 that are used in the previous subsection.

with the Legendre polynomial P_ℓ, and vice versa,

$$f_{n\ell}(k,\tau) = i^\ell \frac{\sqrt{\pi}}{2} \frac{n!}{\Gamma(n+\ell+3/2)} \int d^3\mathbf{q}\, S_{n\,\ell}\left(\frac{\mathbf{q}^2}{2a^2 m_\chi T_0}\right) P_\ell(\hat{\mathbf{k}}_i \hat{\mathbf{q}}_i) f_1(\mathbf{k},\mathbf{q},\tau)\,. \tag{2.35}$$

After a lengthy but straightforward calculation, we obtain the Boltzmann hierarchy:

$$\begin{aligned}
&\dot{f}_{n\ell} + (2n+\ell)(\gamma_0 a + R) f_{n\ell} - 2nR f_{n-1\ell} \\
&+k\sqrt{\frac{2T_0}{m_\chi}}\left\{\frac{\ell+1}{2\ell+1}\left[\left(n+\ell+\frac{3}{2}\right) f_{n\ell+1} - n f_{n-1\ell+1}\right] + \frac{\ell}{2\ell+1}(f_{n+1\ell-1} - f_{n\ell-1})\right\} \\
&= \delta_{\ell 0}\left\{-\frac{1}{2}A_n \dot{h} + \frac{1}{3}B_n \dot{h} - 2B_n\left[\gamma_1 a\left(\frac{T_0}{T_{\chi 0}} - 1\right) + \gamma_0 a \frac{T_1}{T_{\chi 0}}\right]\right\} \\
&+\delta_{\ell 1}\frac{1}{3}A_n k\sqrt{\frac{2m_\chi}{T_0}}\gamma_0 a \frac{\theta_{\mathrm{TP}}}{k^2} + \delta_{\ell 2}\frac{2}{15}\frac{T_{\chi 0}}{T_0}A_n(\dot{h} + 6\dot{\eta})\,.
\end{aligned} \tag{2.36}$$

Here we introduce three new quantities:

$$R = \frac{d\ln(aT_0^{1/2})}{d\tau}\,, \quad A_n = \left(1 - \frac{T_{\chi 0}}{T_0}\right)^n\,, \quad B_n = n\frac{T_{\chi 0}}{T_0}\left(1 - \frac{T_{\chi 0}}{T_0}\right)^{n-1}\,. \tag{2.37}$$

The first quantity is essentially proportional to the Hubble expansion rate: $R = aH/2$. Only a few of the second and third quantities are non-zero before the kinetic decoupling ($T_{\chi 0} = T_0$): $A_0 = 1$ and $B_1 = 1$, while the others vanish. Higher orders of the second quantity become non-zero after the kinetic decoupling ($T_{\chi 0} \ll T_0$): $A_n = 1$, while $B_n (= nT_{\chi 0}/T_0)$ is tiny.

Although we need to solve the full Boltzmann hierarchy to obtain a rigorous result, just taking some small moments of n and ℓ can give the *fluid* approximation (see discussion in subsection 3.3). The perturbations $f_{n\ell}$ with small n and ℓ can be interpreted as primitive variables of the DM imperfect fluid (i.e., mass density ρ, bulk velocity potential θ, pressure P, and anisotropic inertia σ):

$$\bar{\rho}(1+\delta) = -T^0{}_0 = a^{-4}\sum_{s_\chi}\int \frac{d^3\mathbf{q}}{(2\pi)^3} m_\chi\, f\,, \tag{2.38}$$

$$(\bar{\rho}+\bar{P})\theta = i\mathbf{k}_i T^i{}_0 = a^{-4}\sum_{s_\chi}\int \frac{d^3\mathbf{q}}{(2\pi)^3} i\mathbf{k}_i\mathbf{q}_i\, f\,, \tag{2.39}$$

$$\bar{P} + \delta P = \frac{1}{3}T^i{}_i = a^{-4}\sum_{s_\chi}\int \frac{d^3\mathbf{q}}{(2\pi)^3}\frac{\mathbf{q}^2}{3m_\chi}\, f\,, \tag{2.40}$$

$$(\bar{\rho}+\bar{P})\sigma = -\left(\hat{\mathbf{k}}_i\hat{\mathbf{k}}_j - \frac{1}{3}\delta_{ij}\right)T^i{}_j = -a^{-4}\sum_{s_\chi}\int \frac{d^3\mathbf{q}}{(2\pi)^3}\frac{\mathbf{q}^2}{m_\chi}\left[(\hat{\mathbf{k}}_i\hat{\mathbf{q}}_i)^2 - \frac{1}{3}\right] f\,. \tag{2.41}$$

Substituting the exact form of $f = f_0(\tau) + f_1$, we obtain

$$\bar{\rho} = m_\chi \bar{n}\,, \quad \bar{P} = \frac{T_{\chi 0}}{m_\chi}\bar{\rho}\,, \tag{2.42}$$

$$\delta = f_{00}\,, \qquad \theta = 3k\sqrt{\frac{T_0}{2m_\chi}} f_{01}\,, \quad \delta P = \frac{T_0}{T_{\chi 0}}\bar{P}(f_{00} - f_{10})\,, \quad \sigma = 5\frac{T_0}{m_\chi} f_{02}\,. \tag{2.43}$$

The dynamics of the DM imperfect fluid is described by the following equations:

$$\begin{aligned}
\dot{\delta} &= -\theta - \frac{1}{2}\dot{h}\,, && (2.44)\\
\dot{\theta} &= -\frac{\dot{a}}{a}\theta - k^2\sigma + k^2\frac{T_{\chi 0}}{m_\chi}\frac{\delta P}{\bar{P}} + \gamma_0 a(\theta_{\mathrm{TP}} - \theta)\,, && (2.45)\\
\dot{\sigma} &= -2\frac{\dot{a}}{a}\sigma - k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\left(\frac{21}{4}f_{03} + f_{11}\right) + \frac{4}{3}\frac{T_0}{m_\chi}\theta + \frac{2}{3}\frac{T_0}{m_\chi}(\dot{h} + 6\dot{\eta}) - 2\gamma_0 a\sigma\,, && (2.46)\\
\dot{\delta P} &= -5\frac{\dot{a}}{a}\delta P - \frac{5}{6}\bar{P}\dot{h} + \frac{5}{4}k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\bar{\rho}f_{11} - \frac{5}{3}\frac{T_0}{T_{\chi 0}}\bar{P}\theta\\
&\quad -2\gamma_0 a\delta P + 2\gamma_0 a\frac{T_0}{T_{\chi 0}}\bar{P}\delta + 2\bar{P}\left[\gamma_1 a\left(\frac{T_0}{T_{\chi 0}} - 1\right) + \gamma_0 a\frac{T_1}{T_{\chi 0}}\right]. && (2.47)
\end{aligned}$$

The pressure perturbation δP can be decomposed into isentropic $c_\chi^2\delta$ and entropy π perturbations:

$$\frac{\delta P}{\bar{\rho}} = c_\chi^2\delta + \pi\,. \tag{2.48}$$

The sound speed squared of the DM fluid is

$$c_\chi^2 = \frac{T_{\chi 0}}{m_\chi}\left(1 - \frac{1}{3}\frac{d\ln T_{\chi 0}}{d\ln a}\right). \tag{2.49}$$

The evolution of the DM imperfect fluid can be rewritten as

$$\begin{aligned}
\dot{\delta} &= -\theta - \frac{1}{2}\dot{h}\,, && (2.50)\\
\dot{\theta} &= -\frac{\dot{a}}{a}\theta - k^2\sigma + k^2(c_\chi^2\delta + \pi) + \gamma_0 a(\theta_{\mathrm{TP}} - \theta)\,, && (2.51)\\
\dot{\sigma} &= -2\frac{\dot{a}}{a}\sigma - k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\left(\frac{21}{4}f_{03} + f_{11}\right) + \frac{4}{3}\frac{T_0}{m_\chi}\theta + \frac{2}{3}\frac{T_0}{m_\chi}(\dot{h} + 6\dot{\eta}) - 2\gamma_0 a\sigma\,, && (2.52)\\
\dot{\pi} &= -2\frac{\dot{a}}{a}\pi + \frac{5}{4}k\left(\frac{2T_0}{m_\chi}\right)^{3/2} f_{11} - \frac{1}{a^2}\frac{d(a^2c_\chi^2)}{d\tau}\delta - \left(\frac{5}{3}\frac{T_0}{m_\chi} - c_\chi^2\right)\theta - \frac{1}{2}\left(\frac{5}{3}\frac{T_{\chi 0}}{m_\chi} - c_\chi^2\right)\dot{h}\\
&\quad -2\gamma_0 a\left[\pi - \frac{T_1}{m_\chi} - \left(\frac{T_0}{m_\chi} - c_\chi^2\right)\delta\right] + 2\gamma_0 a\left(\frac{T_0}{T_{\chi 0}} - 1\right)\frac{T_{\chi 0}}{m_\chi}\frac{\gamma_1}{\gamma_0}\,. && (2.53)
\end{aligned}$$

3 Neutrino interacting dark matter

The section starts with the introduction of the neutrino interacting DM model via a MeV-scale boson. This particle combination leads in a valid parameter region to a possible solution to all three small-scale crisis problems if the mediator is of vector type [29]. We reproduce and confirm these results by using the method that is derived in the previous section to describe the DM kinetic decoupling. The used method has a different expansion of the collision term when compared to the aforementioned reference and to others like [42].

Furthermore, by using this alternative description we explicitly show a suppression of the power spectrum for other types of mediators as well. The suppression is sizable enough to reduce the abundance of dwarf galaxies but unexpected from the point of view of the above literature. In particular, scalar and vector mediators share an analogue phenomenology within our model set-up and the parameter region is relatively similar concerning the minimal size of the first protohalos. Approximation methods to follow the evolution of cosmological perturbations are also given. Finally, the matter linear power spectrum for scalar and vector interactions are presented, showing a suppression of powers on subgalactic scales.

3.1 Simplified neutrino model

A simplified model extends SM by a DM fermion and additional light fermions (denoted by ν). The DM fermion and the additional light fermions are assumed to be of Dirac type, coupled by a MeV-scale boson denoted by ϕ. In particular, this choice allows us to write down the following set of renormalizable dimension four operators without derivatives:

$$\mathcal{L}_{\rm S} \supset g_\chi \bar{\chi}\phi\chi + g_\nu \bar{\nu}\phi\nu \,, \tag{3.1}$$

$$\mathcal{L}_{\rm V} \supset g_\chi \bar{\chi}\gamma^\mu\chi\phi_\mu + g_\nu \bar{\nu}\gamma^\mu\nu\phi_\mu \,, \tag{3.2}$$

$$\mathcal{L}_{\rm PS} \supset g_\chi \bar{\chi}\phi\gamma^5\chi + g_\nu \bar{\nu}\phi\gamma^5\nu \,, \tag{3.3}$$

$$\mathcal{L}_{\rm PV} \supset g_\chi \bar{\chi}\gamma^\mu\gamma^5\chi\phi_\mu + g_\nu \bar{\nu}\gamma^\mu\gamma^5\nu\phi_\mu \,. \tag{3.4}$$

Here, we assume parity conservation in the interaction Lagrangian and consider each operator type separately. There are four parameters: the DM mass m_χ, the light mediator mass m_ϕ, the DM-mediator coupling g_χ, and the light fermion-mediator coupling g_ν. Specifically, extensions of the simplified model (3.2) into ultraviolet complete models and their constraints have already been investigated by many authors in connection with the small-scale crisis (for an exemplary list of references, see [47–50]).

For simplicity and by analogy to previous works we call the light fermions hidden neutrinos. In the early Universe, the DM and the hidden neutrinos are assumed to be in thermal equilibrium, where a temperature difference when compared to the SM sector hides the additional light fermions. Further, the light boson is in thermal equilibrium with the neutrino sector during the DM chemical freeze-out. For all operators the parameters are chosen such that the relic density of the DM is dominantly determined through $\chi\bar{\chi} \to \phi\phi$ annihilation and not via direct s-channel neutrino production. This is because for the vector, scalar, and pseudo scalar interactions, we assume $g_\nu \ll g_\chi$ (see [51] for a list of possible natural explanations). In this scenario, the DM relic abundance for all operators is independent of the neutrino coupling g_ν. In appendix C.1 we provide for all operators the full calculus of the annihilation cross section and the relic abundance. Due to a more complicated but less illuminating phenomenology, we discuss the results for the pseudo scalar and pseudo vector operators in appendix C.2.

3.2 Minimal halo mass

Elastic scattering via a MeV-scale boson keeps the DM for a long time in kinetic equilibrium with the hidden neutrino sector. During kinetic equilibrium, the DM density perturbations do not grow but oscillate. This phenomena is known as acoustic oscillations and has been shown in [3–15, 27–31, 33–36] to be the dominant damping mechanism of the density perturbations in the case of a late kinetic decoupling.

In the cosmological perturbation theory, the mode that enters the horizon at the kinetic decoupling defines a cutoff in the linear matter power spectrum of density fluctuations. Only the DM density modes that enter the horizon thereafter can significantly grow and collapse later into halos. Thus, fluctuations on shorter scales are damped. The minimal mass of first protohalos can be estimated by the mass inside a sphere with radius of Hubble horizon at the time of the kinetic decoupling:

$$M_{\rm cut} = \rho_{\rm m} \frac{4\pi}{3} \left(\frac{1}{H}\right)^3 = 2.2 \times 10^8 r^3 \left(\frac{1\,{\rm keV}}{T_\nu^{\rm kd}}\right)^3 M_\odot \,, \tag{3.5}$$

where the matter density $\rho_{\rm m}$ and the Hubble expansion rate H are evaluated at the kinetic decoupling. Here, we allow for a different light fermion temperature from the photon temperature to hide the additional neutrinos. The ratio between the two temperatures is defined as $r = T_\nu^{\rm kd}/T_\gamma^{\rm kd}$, where the superscript kd means the corresponding value at the DM kinetic decoupling that occurs when the momentum transfer rate γ equals to the Hubble expansion rate H.

In the following, we derive an approximation method to estimate the kinetic decoupling temperature $T_\nu^{\rm kd}$ in order to calculate the corresponding cutoff mass according to (3.5). The general expression for γ (2.24) is adjusted to describe the scattering of the DM with the light fermions. Dividing it by the Hubble expansion rate and by introducing the following dimensionless variables $x = |\mathbf{p}_\nu|/T_\nu$, $y = T_\nu/m_\chi$, and $z = m_\phi/m_\chi$, one ends up with the following form:

$$\frac{\gamma}{H} = \left(\frac{T_\nu}{T_\gamma}\right)^2 \frac{m_{\rm pl}}{m_\chi} \sqrt{\frac{45}{4\pi^3}} \frac{N_\nu}{48\pi} \frac{1}{\sqrt{g_{\rm eff}}} y^{-2} \int_0^\infty {\rm d}x\, f_\nu^{\rm eq}(x) \left(1 - f_\nu^{\rm eq}(x)\right) g(xy, z) \,, \tag{3.6}$$

where we multiply by the number of light fermion species N_ν. The phase space distribution function $f_\nu^{\rm eq}(x)$ is the usual equilibrium Fermi-Dirac distribution, where we neglected the mass of the light fermions:

$$f_\nu^{\rm eq}(x) = \frac{1}{\exp\left(|\mathbf{p}_\nu|/T_\nu\right) + 1} \,. \tag{3.7}$$

Furthermore, the dimensionless quantity $g(xy, z)$ is defined as the t-averaged scattering amplitude squared:

$$g(xy, z) = \frac{1}{m_\chi^4 (4\pi)^2} \int_{-4\mathbf{p}_\nu^2}^0 {\rm d}t(-t) \sum_{s_2, s_3, s_4} \overline{|\mathcal{M}|^2} \,, \tag{3.8}$$

where in this convention,

$$\overline{|\mathcal{M}|^2} = \frac{1}{16} \sum_{s_1, s_2, s_3, s_4} |\mathcal{M}|^2_{s \to m_\chi^2 + 2 m_\chi E_\nu} \tag{3.9}$$

is the invariant amplitude squared that are averaged over the initial and final spin states.

Equation (3.6) is the basic formula for the kinetic decoupling description of the neutrino interacting DM. In the following, we derive analytic estimates for the scalar and vector operators, which are valid in a broad range of parameters and derive their corresponding $M_{\rm cut}$ scaling patterns. In the case of the pseudo scalar and pseudo vector operators, this approximation that we call the *effective propagator* description is only valid in a small parameter space, and thus (3.6) has to be solved numerically at some point. The results are given in appendix C.2. The DM-neutrino scattering amplitudes for the scalar and vector operators are given by:

$$\text{Vector operator:} \quad \sum_{s_1,s_2,s_3,s_4} |\mathcal{M}|^2 = g_\chi^2 g_\nu^2 \frac{8\left(8E_\nu^2 m_\chi^2 + 4E_\nu m_\chi t + t(2m_\chi^2 + t)\right)}{(t - m_\phi^2)^2}\,, \tag{3.10}$$

$$\text{Scalar operator:} \quad \sum_{s_1,s_2,s_3,s_4} |\mathcal{M}|^2 = g_\chi^2 g_\nu^2 \frac{4t(t - 4m_\chi^2)}{(t - m_\phi^2)^2}\,. \tag{3.11}$$

In the parameter region we are interested in, it turns out that the mass of the mediator is much larger than the kinetic decoupling temperature. In this case, the Mandelstam t in the boson propagator denominator of the scattering amplitudes can be neglected. We call this approximation the effective propagator description. The propagator denominator can be simplified in such a way because $t \in [0, -4\mathbf{p}_\nu^2]$ and the neutrino momentum is further limited by the phase space distribution: $|\mathbf{p}_\nu| \simeq T_\nu$. So t can be neglected in the denominator of the propagator as long as $T_\nu^{\rm kd} \ll m_\phi$, which is the case in the parameter region of the scalar and vector operators.

Within the effective propagator framework, $g(xy, z)$ is only a polynomial function in its variables and the integral in (3.6) has even an analytic expression. To the leading order in T_ν, we find for the vector operator

$$\frac{\gamma}{H} = 17.2 \times \left(\frac{r}{r_0}\right)^2 \left(\frac{N_\nu}{6}\frac{\alpha_\chi}{0.035}\frac{\alpha_\nu}{10^{-4}}\right)\left(\frac{m_\chi}{1\,\mathrm{TeV}}\right)^{-1}\left(\frac{m_\phi}{1\,\mathrm{MeV}}\right)^{-4}\left(\frac{T_\nu}{1\,\mathrm{keV}}\right)^4\,, \tag{3.12}$$

and for the scalar operator

$$\frac{\gamma}{H} = 16.7 \times \left(\frac{r}{r_0}\right)^2 \left(\frac{N_\nu}{6}\frac{\alpha_\chi}{0.17}\frac{\alpha_\nu}{10^{-5}}\right)\left(\frac{m_\chi}{1\,\mathrm{TeV}}\right)^{-1}\left(\frac{m_\phi}{1\,\mathrm{MeV}}\right)^{-4}\left(\frac{T_\nu}{1\,\mathrm{keV}}\right)^4\,, \tag{3.13}$$

with $\alpha_{\chi/\nu} = g_{\chi/\nu}^2/(4\pi)$.

To estimate the kinetic decoupling temperature, we set $\gamma/H = 1$ in the last two equations, which are solved for $T_\nu^{\rm kd}$.[3] The corresponding minimal halo masses that are derived from the kinetic decoupling temperature according to (3.5) is given by

$$(M_{\rm cut})_{\rm V} = 6.8 \times 10^8\, M_\odot \left(\frac{r}{r_0}\right)^{9/2}\left(\frac{N_\nu}{6}\frac{\alpha_\nu}{10^{-4}}\frac{\alpha_\chi}{0.035}\right)^{3/4}\left(\frac{m_\chi}{1\,\mathrm{TeV}}\right)^{-3/4}\left(\frac{m_\phi}{1\,\mathrm{MeV}}\right)^{-3}\,, \tag{3.14}$$

$$(M_{\rm cut})_{\rm S} = 6.6 \times 10^8\, M_\odot \left(\frac{r}{r_0}\right)^{9/2}\left(\frac{N_\nu}{6}\frac{\alpha_\nu}{10^{-5}}\frac{\alpha_\chi}{0.17}\right)^{3/4}\left(\frac{m_\chi}{1\,\mathrm{TeV}}\right)^{-3/4}\left(\frac{m_\phi}{1\,\mathrm{MeV}}\right)^{-3}\,, \tag{3.15}$$

where we normalize r to the SM neutrino temperature ratio: $r_0 = (4/11)^{1/3}$.

To be consistent with constraints on additional radiation components, we use the combined results of Big Bang nucleosynthesis (BBN) and CMB constraints given in [52] to derive an upper bound for our model within the 1σ error bar:

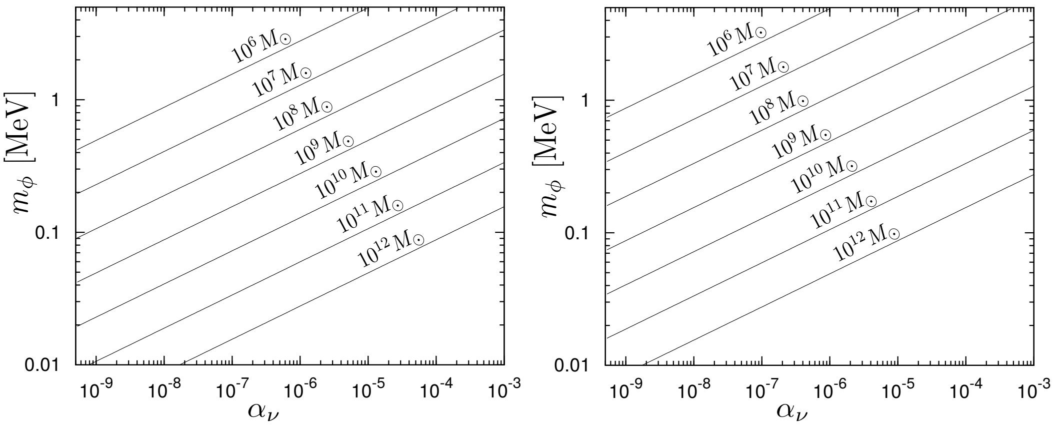
$$\frac{r}{r_0} < \left(\frac{0.51}{N_\nu + \frac{4}{7}g_{\rm pol}}\right)^{1/4}\,. \tag{3.16}$$

---

[3] This defines our kinetic decoupling temperature. Another definition of $T_{\rm kd}$ is used in the literature [42], which also has a direct map into the non-linear $M_{\rm cut}$ given recently in [34]. With our definition, $M_{\rm cut}$ is smaller by less than a factor of three when compared to the aforementioned literature.

$(N_\nu, g_{\rm pol})$	$(r/r_0)^{9/2} \times (N_\nu/6)^{3/4}$
$(2, \{0, 1, 3\})$	$< (0.09, 0.07, 0.05)$
$(6, \{0, 1, 3\})$	$< (0.06, 0.05, 0.05)$

Table 1. Upper bounds on $(r/r_0)^{9/2} \times (N_\nu/6)^{3/4}$ derived from [52] are shown. We separate two extreme cases: the mediator is still relativistic at BBN $g_{\rm pol} = \{1, 3\}$; its contribution to the radiation components can be neglected ($g_{\rm pol} = 0$). The factors on the right column reduce the cutoff masses (3.14) and (3.15) by at least one order of magnitude.



(a) Cutoff mass for the vector mediator

(b) Cutoff mass for the scalar mediator

Figure 1. Contour line of a constant $M_{\rm cut}$ is shown for the vector (left) and scalar (right) mediators within the effective propagator framework. The other parameters are chosen according to the normalization values in (3.14) and (3.15). In the parameter region shown, the results obtained from the effective propagator description and the exact numerical results obtained by integrating (3.6) coincide.

Here, we consider the possibility of having a sub-MeV scale mediator contribution to the radiation components at BBN. In table 1 we summarize the upper bounds for two extreme scenario: the mediator does not contribute ($g_{\rm pol} = 0$); the mediator is still relativistic at BBN and contributes via its internal degrees of freedom ($g_{\rm pol} = \{1, 3\}$ for the scalar and massive vector mediators, respectively).

First of all, these cutoff masses (3.14) and (3.15) have the same scaling dependence, and thus differ only by a numerical constant and depend mostly on the boson mass. Using the relic density constraint on α_χ given by (C.5) and (C.6), we see that $M_{\rm cut}$ is essentially independent of the DM mass. In figure 1, contour lines of a constant $M_{\rm cut}$ are shown for the scalar and vector interactions in the (m_ϕ, α_ν)-plane. In order to account for the missing satellite problem and to be consistent with Ly-α forest bounds, the cutoff mass has to be roughly in between $10^7\,M_\odot \lesssim M_{\rm cut} \lesssim 5 \times 10^{10}\,M_\odot$ [29]. We provide the corresponding $M_{\rm cut}$ contour plots for the pseudo scalar and pseudo vector operators and their discussion in appendix C.2.

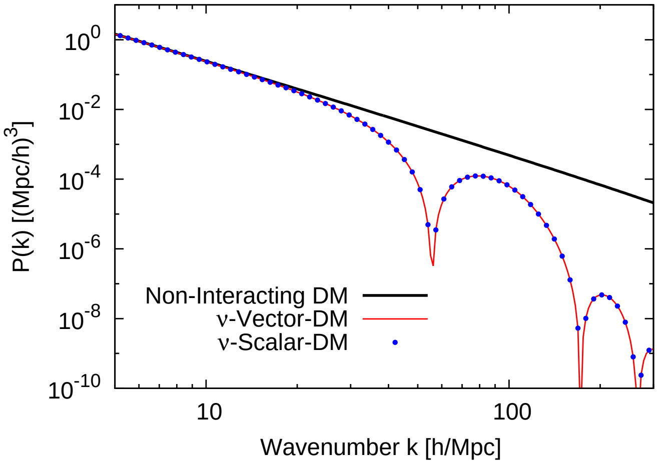


Figure 2. This figure shows the linear matter power spectra at present for the standard CDM (black line) and the neutrino interacting DM via the vector (red line) and scalar (blue dots) mediators. The wavenumber of $k = 50h$ /Mpc corresponds to a halo mass of $M = \rho_{\rm m} 4\pi/3(\pi/k)^3 \simeq 10^8\, M_\odot$. In both the interacting DM cases, we adjust the free neutrino coupling parameters α_ν in (3.14) and (3.15) to give the same cutoff mass ($M_{\rm cut} = 6.4 \times 10^8\, M_\odot$) and ignore the small effects of the DM-neutrino interactions on the neutrino perturbations (*back-reaction*).

3.3 Matter power spectrum

The minimal halo masses derived in the previous subsection imply that the scalar operator leaves a similar suppression in the resultant matter power spectra to the case of the vector operator. In order to see this explicitly, let us consider our model where the DM scatters light fermions via the scalar operator. The scattering amplitude has a pure t-dependence given by (3.11). In other collision term expansion methods like that in [42], the scattering rate would be declared to be zero at the leading order. But as already shown in the previous subsection, we find that DM models with a scalar interaction can also account for the missing satellite problem.

To emphasize that scalar interactions are as important as vector interactions regarding the small-scale crisis problems, we adjust the free neutrino coupling parameters α_ν in (3.14) and (3.15) to give the same cutoff mass and show that their linear matter power spectra are close to each other in figure 2. Here, we modify the public code `CAMB` [53] suitably to follow the coevolutions of cosmological perturbations of the DM (subsection 2.2) and the other components (e.g., baryons, photons, neutrinos, and gravitational potential). The small effects of the DM-neutrino interactions on the neutrino perturbations are neglected and the *perfect fluid* approximation (explained below) is used. Clearly, the shape of the power spectrum shows the characteristic features of the dark acoustic oscillations and the power on small scales is suppressed when compared to the CDM prediction.

Additionally, we check the validity of the perfect fluid assumption by comparing the results to the case of an imperfect fluid. To obtain a closed set of equations, we need to develop an approximation for f_{03} and f_{11} (see (2.50)–(2.53)). One way is setting them to be zero, defining the *imperfect fluid* approximation. This is valid when $T_\chi/m_\chi \ll 1$, i.e., the free streaming of the DM particles is negligible after they decouple kinetically $\gamma/H < 1$ (see appendix B). Actually, we can also take $\sigma = 0$ and $\pi = 0$ for the adiabatic perturbations in

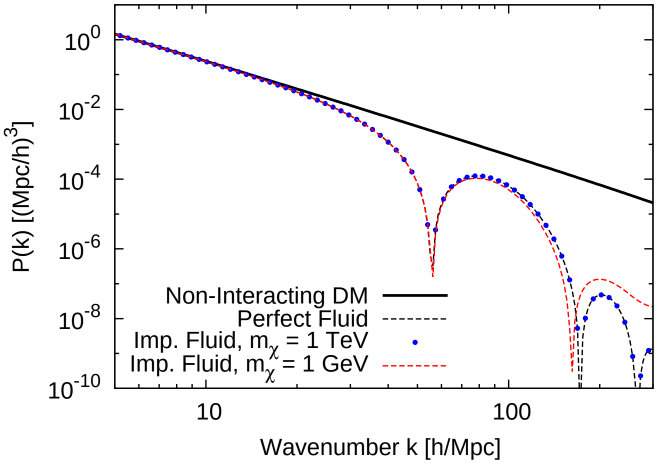


Figure 3. This figure compares the linear matter power spectrum in the perfect and imperfect fluid approximations. We take the same model with the scalar mediator as in figure 2. We take $m_\chi = 1\,\text{TeV}$ in both the perfect and imperfect fluid approximations. When the DM mass is lowered to $m_\chi = 1\,\text{GeV}$ and γ is kept fixed, the resultant matter power spectrum in the imperfect fluid approximation starts to differ from that in the perfect fluid approximation at wavenumbers larger than $k \gtrsim 100\,h/\text{Mpc}$.

the same limit, defining the perfect fluid approximation. Before the kinetic decoupling, all the variables $f_{n\ell}$ but f_{00} and f_{01} remain zero due to the damping term $\sim \gamma_0 f_{n\ell}$ in (2.36). The former, corresponding to δ, does not have the damping term. The latter, corresponding to θ, has the source term $\sim \gamma_0(\theta_{\text{TP}} - \theta)$. One non-trivial check is to compare the resultant power spectra in the perfect and imperfect fluid approximations.

When the results from the perfect and imperfect fluid approximations deviate from each other, it does not necessarily mean that the imperfect fluid approximation gives a better description, but it just indicates that the perfect fluid approximation is not valid. To check if the imperfect fluid approximation gives a valid description or not, we need to compare the result from the treatment incorporating the full Boltzmann hierarchy, which is beyond the scope of this paper. Let us stress that the above deviation does not correspond to a limitation of the Fokker-Planck equation, which is valid as long as the momentum transfer in each collision is smaller than the typical DM momentum.

For a smaller DM mass with γ being fixed, we find differences in their power spectrum above a certain critical wavenumber as shown in figure 3. This is because the free streaming is sizable after the kinetic decoupling for the lighter DM. The results from the perfect fluid approximation are reliable below the critical wavenumber. On smaller scales, however, we may need to solve the full Boltzmann hierarchy (2.36). In appendix B, we give a more detailed discussion on the impact of higher order terms in the Boltzmann hierarchy and give a rough estimate of the critical wavenumber, where the results from the perfect and imperfect fluid approximations start to deviate.

4 Summary and outlook

In summary, we presented a consistent formulation that allows one to start from an underlying DM model and calculate its linear matter power spectrum. Regarding the small-scale crisis, the method is broadly applicable to essentially generic radiation interacting DM models that lead to a power spectrum suppression when compared to the standard cosmology on subgalactic scales.

In this paper, we focused on the case where the DM is in kinetic equilibrium with light and hidden fermions for a long time and the decoupling process was investigated for mediators of fundamentally different type. The new message is that not only a vector mediator at the MeV scale may solve all three small-scale problems at the same time [29], but we find that new classes of interactions may also solve at least the missing satellite problem. This result was unexpected from the point of view of previous literature [41–43], where the leading contribution to the momentum transfer rate is assumed to come from the scattering amplitude evaluated at Mandelstam $t = 0$. We explicitly derived an expansion method of the collision term where the scattering amplitude is t-averaged in the final form of the momentum transfer rate. This results in a different phenomenology from that in the previous literature for scattering amplitudes proportional to Mandelstam t, e.g., in the scalar, pseudo scalar, and pseudo vector interactions between the DM and the hidden fermions.

With this new insight, the classification of possible DM-radiation interactions, which are suppressing the abundance of dwarf galaxies, has to be revisited. During the preparation of this work, we have been informed that Bringmann et al. [54] have independently derived similar results concerning the possibility of kinetic decoupling at late times with in new classes of interactions. As a consequence, our work and studies by the latter authors may extend the list of realistic WIMP-like DM theories accounting for small-scale discrepancies.

As an important subtlety, we also discussed the validity of the perfect fluid approximation for the calculation of the power spectrum. We derive the consistent equations needed to be solved for an imperfect fluid treatment and compare the power spectra obtained from the perfect and imperfect fluid approximations. As indicated from figure 3, the perfect fluid approximation is limited by free streaming effects on the smallest scales. This may infer that we need to solve the full Boltzmann hierarchy to have reliable results for some models where the DM mass is small.

Our formulation, as a fundamental building block, in combination with N-body simulations would allow one to map DM models into the observational non-linear small-scale structure. We plan to combine baryonic feedback and DM-induced small-scale suppression to investigate the observational outcome. At present or in close future, this kind of sophisticated simulations are expected to shed more light on whether the small-scale crisis will be related to fundamental properties of DM or not. Even if the DM-radiation interaction does not resolve the small-scale crisis, our work and others can help to constraint DM models from a new perspective.

Acknowledgments

A.K. would like to thank Paolo Gondolo for kindly providing Junya Kasahara's thesis. T.B. sincerely thanks Torsten Bringmann for reading our early draft and sharing valuable comments. L.C. would like to thank Giorgio Arcadi for useful discussions. T.B. and L.C. acknowledge partial support from the European Union FP7 ITN INVISIBLES (Marie Curie Actions, PITN-GA-2011-289442). The work of T.T. is partially supported by JSPS KAKENHI Grant Number 15K05084 and MEXT KAKENHI Grant Number 15H05888. N.Y. acknowledges the financial supports from JST CREST and from JSPS Grant-in-Aid for Scientific Research (25287050, 25610050).

A Perturbation theory in the conformal Newtonian gauge

In this appendix, we develop a linear theory in the conformal Newtonian gauge and show its equivalence to the synchronous gauge. The explicit form of the gauge transformation is presented. The conformal Newtonian gauge is given by:

$$ds^2 = a^2 \left[-(1+2\Phi) d\tau^2 + (1-2\Psi) d\mathbf{x}^2 \right] \,. \tag{A.1}$$

Up to the first order of cosmological perturbations, the Fokker-Planck equation in the conformal Newtonian gauge is given by:

$$\begin{aligned} &\dot f + \frac{\mathbf{q}_i}{m_\chi}\frac{\partial}{\partial \mathbf{x}_i} f + \left(\dot\Psi \mathbf{q}_i - m_\chi \frac{\partial}{\partial \mathbf{x}_i}\Phi \right) \frac{\partial}{\partial \mathbf{q}_i} f \\ &= (\gamma_0+\gamma_1) a (1+\Phi) \frac{\partial}{\partial \mathbf{q}_i}\left[(\mathbf{q}_i - a m_\chi \mathbf{u}_i) f + a^2 m_\chi (T_0+T_1) \frac{\partial}{\partial \mathbf{q}_i} f \right] \,. \end{aligned} \tag{A.2}$$

Between the conformal Newtonian and synchronous gauges, the collision term differs by a factor of $(1+\Phi)$.[4] This is because in the conformal newtonian gauge, the gravitational potential Φ put the conformal time forward/back in relative to the local inertial time. The first order perturbation follows:

$$\begin{aligned} &\dot f_1 + \frac{\mathbf{q}_i}{m_\chi}\frac{\partial}{\partial \mathbf{x}_i} f_1 + \left(\dot\Psi \mathbf{q}_i - m_\chi \frac{\partial}{\partial \mathbf{x}_i}\Phi \right) \frac{\partial}{\partial \mathbf{q}_i} f_0 \\ &= (\gamma_1 + \gamma_0\Phi) a L_{\rm FP}[f_0] - \gamma_0 a^2 m_\chi \mathbf{u}_i \frac{\partial}{\partial \mathbf{q}_i} f_0 + \gamma_0 a^3 m_\chi T_1 \frac{\partial^2}{\partial \mathbf{q}^2} f_0 + \gamma_0 a L_{\rm FP}[f_1] \,. \end{aligned} \tag{A.3}$$

In the Fourier space, these equations are rewritten as

$$\begin{aligned} \dot f_1 + \frac{i\mathbf{k}_i \mathbf{q}_i}{a m_\chi} f_1 - \gamma_0 a L_{\rm FP}[f_1] = {}& \dot\Psi \frac{\mathbf{q}^2}{2a^2 m_\chi T_{\chi 0}} f_0 - \frac{i\mathbf{k}_i\mathbf{q}_i}{a T_{\chi 0}} \left(\Phi + \gamma_0 a \frac{\theta_{\rm TP}}{k^2} \right) f_0 \\ & + \left[(\gamma_1 + \gamma_0 \Phi) a \left(\frac{T_0}{T_{\chi 0}} - 1 \right) + \gamma_0 a \frac{T_1}{T_{\chi 0}} \right] \left(\frac{\mathbf{q}^2}{2a^2 m_\chi T_{\chi 0}} - 3 \right) f_0 \,. \end{aligned} \tag{A.4}$$

We obtain the Boltzmann hierarchy,

$$\begin{aligned} &\dot f_{n\ell} + (2n+\ell)(\gamma_0 a + R) f_{n\ell} - 2nR f_{n-1\ell} \\ &\quad + k\sqrt{\frac{2T_0}{m_\chi}} \left\{ \frac{\ell+1}{2\ell+1} \left[\left(n + \ell + \frac{3}{2} \right) f_{n\ell+1} - n f_{n-1\ell+1} \right] + \frac{\ell}{2\ell+1} (f_{n+1\ell-1} - f_{n\ell-1}) \right\} \\ &= \delta_{\ell 0} \left\{ 3A_n \dot\Psi - 2B_n \left[\dot\Psi + (\gamma_1 + \gamma_0\Phi) a \left(\frac{T_0}{T_{\chi 0}} - 1 \right) + \gamma_0 a \frac{T_1}{T_{\chi 0}} \right] \right\} \\ &\quad + \delta_{\ell 1} \frac{1}{3} A_n k \sqrt{\frac{2m_\chi}{T_0}} \left(\Phi + \gamma_0 a \frac{\theta_{\rm TP}}{k^2} \right) \,. \end{aligned} \tag{A.5}$$

---

[4] The factor $(1+\Phi)$ is missing in the corresponding equation of [40].

This description is equivalent to that in the synchronous gauge through the gauge transformation of

$$\begin{aligned} f_{n0}(\mathrm{Syn}) &= f_{n0}(\mathrm{Con}) + (3A_n - 2B_n)\frac{\dot{a}}{a}\alpha + B_n \frac{d\ln(a^2 T_{\chi 0})}{d\tau}\alpha\,, & \text{(A.6)}\\ f_{n1}(\mathrm{Syn}) &= f_{n1}(\mathrm{Con}) - \frac{1}{3}A_n k\sqrt{\frac{2m_\chi}{T_0}}\alpha\,, & \text{(A.7)}\\ T_1(\mathrm{Syn}) &= T_1(\mathrm{Con}) - \dot{T}_0\alpha\,, & \text{(A.8)}\\ \gamma_1(\mathrm{Syn}) &= \gamma_1(\mathrm{Con}) - \dot{\gamma}_0\alpha\,, & \text{(A.9)} \end{aligned}$$

with the parameter $\alpha = (\dot{h} + 6\dot{\eta})/(2k^2)$. Here, it should be noted again that the above gauge transformation works only with the factor of the collision term $(1 + \Phi)$ in the conformal Newtonian gauge. The dynamics of the DM imperfect fluid is described by the following equations:

$$\begin{aligned} \dot{\delta} &= -\theta + 3\dot{\Psi}\,, & \text{(A.10)}\\ \dot{\theta} &= -\frac{\dot{a}}{a}\theta - k^2\sigma + k^2\frac{T_{\chi 0}}{m_\chi}\frac{\delta P}{\bar{P}} + k^2\Phi + \gamma_0 a(\theta_{\mathrm{TP}} - \theta)\,, & \text{(A.11)}\\ \dot{\sigma} &= -2\frac{\dot{a}}{a}\sigma - k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\left(\frac{21}{4}f_{03} + f_{11}\right) + \frac{4}{3}\frac{T_0}{m_\chi}\theta - 2\gamma_0 a\sigma\,, & \text{(A.12)}\\ \dot{\delta P} &= -5\frac{\dot{a}}{a}\delta P + 5\bar{P}\dot{\Psi} + \frac{5}{4}k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\bar{\rho} f_{11} - \frac{5}{3}\frac{T_0}{T_{\chi 0}}\bar{P}\theta & \\ &\quad -2\gamma_0 a\delta P + 2\gamma_0 a\frac{T_0}{T_{\chi 0}}\bar{P}\delta + 2\bar{P}\left[(\gamma_1 + \gamma_0\Phi)a\left(\frac{T_0}{T_{\chi 0}} - 1\right) + \gamma_0 a\frac{T_1}{T_{\chi 0}}\right]. & \text{(A.13)} \end{aligned}$$

The evolution of the DM imperfect fluid can be rewritten with the isentropic and entropy perturbations:

$$\begin{aligned} \dot{\delta} &= -\theta + 3\dot{\Psi}\,, & \text{(A.14)}\\ \dot{\theta} &= -\frac{\dot{a}}{a}\theta - k^2\sigma + k^2(c_\chi^2\delta + \pi) + k^2\Phi + \gamma_0 a(\theta_{\mathrm{TP}} - \theta)\,, & \text{(A.15)}\\ \dot{\sigma} &= -2\frac{\dot{a}}{a}\sigma - k\left(\frac{2T_0}{m_\chi}\right)^{3/2}\left(\frac{21}{4}f_{03} + f_{11}\right) + \frac{4}{3}\frac{T_0}{m_\chi}\theta - 2\gamma_0 a\sigma\,, & \text{(A.16)}\\ \dot{\pi} &= -2\frac{\dot{a}}{a}\pi + \frac{5}{4}k\left(\frac{2T_0}{m_\chi}\right)^{3/2} f_{11} - \frac{1}{a^2}\frac{d(a^2c_\chi^2)}{d\tau}\delta - \left(\frac{5}{3}\frac{T_0}{m_\chi} - c_\chi^2\right)\theta + 3\left(\frac{5}{3}\frac{T_{\chi 0}}{m_\chi} - c_\chi^2\right)\dot{\Psi} & \\ &\quad -2\gamma_0 a\left[\pi - \frac{T_1}{m_\chi} - \left(\frac{T_0}{m_\chi} - c_\chi^2\right)\delta\right] + 2\gamma_0 a\left(\frac{T_0}{T_{\chi 0}} - 1\right)\frac{T_{\chi 0}}{m_\chi}\left(\frac{\gamma_1}{\gamma_0} + \Phi\right). & \text{(A.17)} \end{aligned}$$

B Impact of the higher order terms in the Boltzmann hierarchy

In this appendix we take a closer look at the higher order terms in the Boltzmann hierarchy. As discussed in subsection 3.3, they represent the free streaming of DM particles and are important for the case of a smaller DM mass. Once we solve the full Boltzmann hierarchy directly, we can see their effects on resultant matter power spectra quantitatively. It is, however, challenging and beyond the scope of this paper. Instead we give an estimate of the critical wavenumber, below which the perfect fluid approximation appears trustworthy.

Before the kinetic decoupling ($\gamma/H \gg 1$), the higher order terms are negligible. This is because the friction term ($\propto \gamma_0 a$) in the Boltzmann hierarchy (2.36) leads $f_{n\ell}$ to a rapid damping:

$$\dot{f}_{n\ell} + (2n+\ell)\gamma_0 a f_{n\ell} = -2\delta_{n1}\delta_{\ell 0}\gamma_0 a \frac{T_1}{T_{\chi 0}} + \delta_{\ell 1}\frac{1}{3}A_n k\sqrt{\frac{2m_\chi}{T_0}}\gamma_0 a \frac{\theta_{\rm TP}}{k^2}\,. \tag{B.1}$$

Here we have used $T_{\chi 0} = T_0$, which results in $B_1 = 1$ and $B_n = 0$ ($n \neq 1$) as discussed below (2.37). Exceptions are f_{00}, f_{01}, and f_{10} since the first does not have the friction term in its evolution equation and the last two have the source terms (right-handed side) induced by the collision ($\propto \gamma_0 a$) in their evolution equations. Through (2.43), f_{00}, f_{01}, and f_{10} are respectively related with the density perturbation δ, the bulk velocity θ, and the entropy perturbation π. From (2.53) with a rapid momentum transfer,

$$\dot{\pi} = -2\gamma_0 a\left(\pi - \frac{T_1}{m_\chi} + \frac{T_0}{3m_\chi}\delta\right)\,, \tag{B.2}$$

we can see that the entropy perturbation is proportional to the isocurvature perturbation $S_{\rm TP,DM} = \delta(s/n)/(\bar{s}/\bar{n}) = 3T_1/T_0 - \delta$: $\pi = 1/3(T_0/m_\chi)S_{\rm TP,DM}$. As long as DM particles and those in a thermal bath are tightly coupled to each other, thereby forming one fluid, $S_{\rm TP,DM}$ vanishes for the adiabatic perturbations. Thus only f_{00} and f_{01}, or in other words, δ and θ are non-zero. The perfect fluid approximation is valid before the kinetic decoupling.

After the kinetic decoupling ($\gamma/H \ll 1$), higher order terms become sizable. They, however, do not change the resultant matter power of long wavelength modes as follows. In this limit, we can neglect the term proportional to $k\sqrt{T_0/m_\chi}$:

$$\dot{f}_{n\ell} + (2n+\ell)Rf_{n\ell} - 2nRf_{n-1\ell} = -\frac{1}{2}\delta_{\ell 0}\dot{h} + \delta_{\ell 1}\frac{1}{3}k\sqrt{\frac{2m_\chi}{T_0}}\gamma_0 a\frac{\theta_{\rm TP}}{k^2} + \delta_{\ell 2}\frac{2}{15}\frac{T_{\chi 0}}{T_0}(\dot{h}+6\dot{\eta})\,. \tag{B.3}$$

Here we have used $A_n = 1$ and $B_n \ll 1$ after the kinetic decoupling (see the discussion below (2.37)). Noting that $f_{n-1\ell}$ affects the evolution of $f_{n\ell}$ through the term of $-2nRf_{n-1\ell}$, we can see that the higher order terms f_{n0} become of the order of $f_{00} = \delta$ within a few Hubble time after the kinetic decoupling. This, however, does not affect the evolution of δ and thus does not change the resultant matter power. This is because $f_{n-1\ell}$ affects the evolution of $f_{n\ell}$ but not vice versa.

From the above observations, we infer that the impact of higher order terms is suppressed by a factor of $k\sqrt{T_0/m_\chi}/(aH)$. Thus we can estimate the critical wavenumber by equating the factor with unity. This ratio scales in proportion to $a^{1/2}$ (a^0, or in other words, constant) in the radiation (matter) dominated era, and hence it takes a maximum value of $k/k_{\rm eq}\sqrt{T_0(a_{\rm eq})/m_\chi}$ with the wavenumber $k_{\rm eq}$ and the scale factor $a_{\rm eq}$ at the matter radiation equality. As a result, we infer that for

$$k \ll 430\,/{\rm Mpc} \times \left(\frac{r}{r_0}\right)^{-1/2}\left(\frac{m_\chi}{\rm GeV}\right)^{1/2}\,, \tag{B.4}$$

the fluid approximation is trustworthy (see discussion below (3.5) for the definition of r). In figure 3, the deviation between the results from the perfect and imperfect fluid approximations can be seen above $k \simeq 100\,h/{\rm Mpc}$ for $m_\chi = 1\,{\rm GeV}$. This appears compatible with the above estimation.

C Thermal history calculation

In this appendix, the annihilation cross section and the annihilation cross section for all operators. The minimal halo masses and the momentum transfer rates of the pseudo scalar and pseudo vector operators are presented as well, showing a different kind of phenomenology when compared to the scalar and vector ones.

C.1 Relic abundance

In our simplified model, the DM abundance is dominantly determined via annihilation process into two mediators ϕ. The invariant amplitude for this process is a sum over t- and u-channel diagrams. In order to calculate the DM relic abundance, the cross section times relative velocity $(\sigma v_{\rm rel})$ is expanded up to the second order in terms of the relative velocity $v_{\rm rel}$ and the mass ratio z. For each operator, the expanded annihilation cross section is given by:

$$(\sigma v_{\rm rel})_{\rm V} = \frac{g_\chi^4}{16\pi m_\chi^2}\left(1-\frac{1}{2}z^2+\mathcal{O}(z^4)\right)+\frac{g_\chi^4}{16\pi m_\chi^2}\left(\frac{19}{24}z^2+\mathcal{O}(z^4)\right)v_{\rm rel}^2+\mathcal{O}(v_{\rm rel}^4)\,, \tag{C.1}$$

$$(\sigma v_{\rm rel})_{\rm S} = \frac{3g_\chi^4}{128\pi m_\chi^2}\left(1+\frac{11}{18}z^2+\mathcal{O}(z^4)\right)v_{\rm rel}^2+\mathcal{O}(v_{\rm rel}^4)\,, \tag{C.2}$$

$$(\sigma v_{\rm rel})_{\rm PV} = \frac{g_\chi^4}{16\pi m_\chi^2}\left(1-\frac{1}{2}z^2+\mathcal{O}(z^4)\right)+\frac{g_\chi^4}{12\pi m_\chi^2}\left(z^{-4}+\mathcal{O}(z^{-2})\right)v_{\rm rel}^2+\mathcal{O}(v_{\rm rel}^4)\,, \tag{C.3}$$

$$(\sigma v_{\rm rel})_{\rm PS} = \frac{g_\chi^4}{384\pi m_\chi^2}\left(1-\frac{1}{2}z^2+\mathcal{O}(z^4)\right)v_{\rm rel}^2+\mathcal{O}(v_{\rm rel}^4)\,. \tag{C.4}$$

The scalar, vector, and pseudo scalar cross sections are consistent with the ones obtained in [55]. In the case of the pseudo vector interaction, we find the leading term to be proportional to z^{-4}. We discuss this subtlety in subsection C.2.

We estimate the DM freeze-out temperature x_f, following basically the method used in [57], and determine the relic abundance for each operator, which is given approximately by:

$$\text{V: } \Omega_\chi h^2 = \frac{0.12}{2}\left(\frac{\alpha_\chi}{0.035}\right)^{-2}\left(\frac{m_\chi}{1\,\text{TeV}}\right)^2\left(\frac{x_f}{26.1}\right)\,, \tag{C.5}$$

$$\text{S: } \Omega_\chi h^2 = \frac{0.12}{2}\left(\frac{\alpha_\chi}{0.17}\right)^{-2}\left(\frac{m_\chi}{1\,\text{TeV}}\right)^2\left(\frac{x_f}{26.8}\right)^2\,, \tag{C.6}$$

$$\text{PV: } \Omega_\chi h^2 = \frac{0.12}{2}\left(\frac{r}{r_0}\right)\left(\frac{\alpha_\chi}{8.4\times10^{-12}}\right)^{-2}\left(\frac{m_\chi}{100\,\text{MeV}}\right)^2\left(\frac{z}{10^{-3}}\right)^4\left(\frac{x_f}{13.4}\frac{r_0}{r}\right)^2\,, \tag{C.7}$$

$$\text{PS: } \Omega_\chi h^2 = \frac{0.12}{2}\left(\frac{r}{r_0}\right)\left(\frac{\alpha_\chi}{4.7\times10^{-5}}\right)^{-2}\left(\frac{m_\chi}{100\,\text{MeV}}\right)^2\left(\frac{x_f}{13.4}\frac{r_0}{r}\right)^2\,. \tag{C.8}$$

In the case of the scalar and vector operators, we assume the DM, ϕ, and the light fermions to have the same temperature as the SM particles at the DM freeze-out. In the case of the pseudo scalar and pseudo vector operators, we had to lower the DM mass in order to get a cutoff mass around $\sim 10^8\,M_\odot$ as shown in subsection C.2. The DM freeze-out in this case occurs at a time close to BBN. Thus, the temperature of ϕ and the light fermions has to differ from the SM radiation temperature in order to be hidden and not to be in conflict with observation. This subtlety is taken into account in the relic abundance of (C.7) and (C.8). Throughout this paper, we ignore the logarithmic dependence of the freeze-out temperature

x_f/r on the model parameters and drop the last factor in (C.5)–(C.8) if the relic density constraint is used to reduce one of the parameters.

Furthermore, we remark that due to the presence of a light mediator and its long range property one has to include the Sommerfeld effect for DM annihilation in principle. This may lead to an $\mathcal{O}(1)$ correction of the DM coupling in order to produce the correct relic abundance, but including the effect is beyond the scope of this paper.

C.2 Minimal halo mass of the pseudo scalar and pseudo vector operators

In the case of the pseudo scalar and pseudo vector operators, the parameter space of interest spoils partially the effective propagator description, and thus γ/H does not have a simple power law dependence on the neutrino temperature like in the scalar and vector cases. Nevertheless, we derive analytically the scaling pattern of the cutoff mass from the effective propagator description, and compare it to the cutoff mass derived from the exact numerical evaluation of (3.6). The DM-neutrino scattering amplitudes for the pseudo scalar and pseudo vector operators are given by:

$$\text{Pseudo scalar operator:} \quad \sum_{s_1,s_2,s_3,s_4} |\mathcal{M}|^2 = g_\chi^2 g_\nu^2 \frac{4t^2}{(t-m_\phi^2)^2}\,, \tag{C.9}$$

$$\text{Pseudo vector operator:} \quad \sum_{s_1,s_2,s_3,s_4} |\mathcal{M}|^2 = g_\chi^2 g_\nu^2 \frac{8\left(8E_\nu^2 m_\chi^2 + 4E_\nu m_\chi t - t(2m_\chi^2 - t)\right)}{(t-m_\phi^2)^2}\,. \tag{C.10}$$

Pseudo scalar operator. The DM-neutrino scattering amplitude (C.9) via a pseudo scalar mediator has a pure t^2-dependence. Within the effective propagator framework, γ/H depends therefore on a different power of T_ν when compared to the scalar and vector operators:

$$\frac{\gamma}{H} = 2.0 \times \left(\frac{r}{r_0}\right)^2 \left(\frac{N_\nu}{6}\frac{\alpha_\chi}{4.7\times10^{-5}}\frac{\alpha_\nu}{10^{-6}}\right)\left(\frac{m_\chi}{100\,\mathrm{MeV}}\right)^{-3}\left(\frac{m_\phi}{10\,\mathrm{keV}}\right)^{-4}\left(\frac{T_\nu}{1\mathrm{keV}}\right)^{6}\,. \tag{C.11}$$

Inserting the relic density constraint for α_χ given by (C.8), we find the scaling pattern of the cutoff mass:

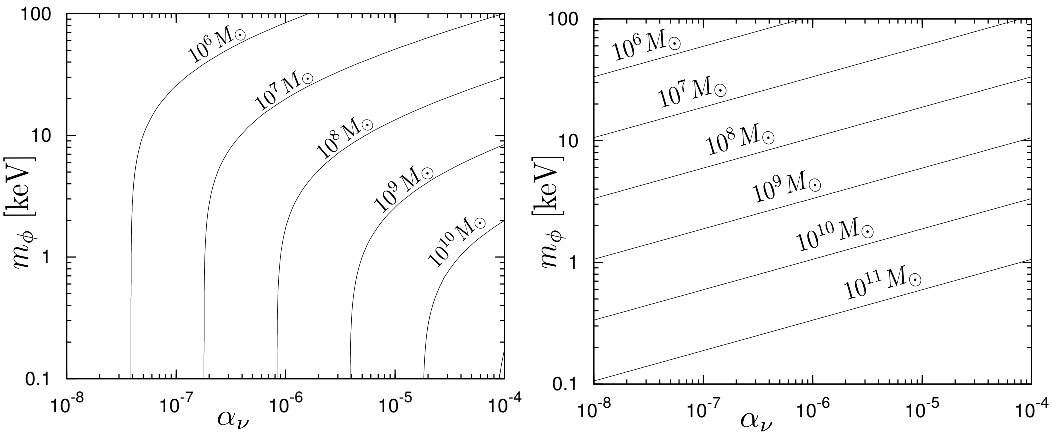
$$(M_\mathrm{cut})_\mathrm{PS} = 1.1\times10^8\,M_\odot \left(\frac{r}{r_0}\right)^{15/4}\left(\frac{N_\nu}{6}\frac{\alpha_\nu}{10^{-6}}\right)^{1/2}\left(\frac{m_\chi}{100\,\mathrm{MeV}}\right)^{-1}\left(\frac{m_\phi}{10\,\mathrm{keV}}\right)^{-2}\,. \tag{C.12}$$

Note that the mass of the mediator is close to the temperature $\sim 1\,\mathrm{keV}$ for subgalacitc cutoff masses. This spoils our effective propagator description as can be seen by comparing the exact numerical result with the effective description in figure 4.

Pseudo vector operator. The DM annihilation cross section via a pseudo vector mediator shows a z^{-4} enhancement in (C.3). At a first look, the limit $z \to 0$ in the cross section seems to diverge and give rise to unitarity violation [56]. By embedding the model into a local U(1) gauge theory where both the mass of the DM and the gauge boson mass arise due to the spontaneous symmetry breaking via an additional scalar field, we show explicitly that this is not the case and the parameter region that we use to produce subgalactic cutoffs is in the perturbative regime.

We denote the additional scalar by Φ and the local U(1) gauge invariant action reads

$$\mathcal{L} = i\bar{\chi}\not{D}_{+}\chi + |D_{\mu,-2}\Phi|^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \lambda_Y\left(\bar{\chi}_L\Phi\chi_R + \bar{\chi}_R\Phi^\star\chi_L\right) - V(\Phi)\,, \tag{C.13}$$



(a) Exact numerical result

(b) Effective propagator description

Figure 4. Contour line of a constant $M_{\rm cut}$ is shown for the pseudo scalar operator within the exact (left) and effective propagator frameworks (right). The DM parameters chosen are $m_\chi = 100\,\mathrm{MeV}$ and α_χ satisfying the relic density constraint. The effective propagator description is only valid in the upper right quarter of figure 4(b).

where $\not{D}_+ = \not{\partial} + ig_\chi\not{\phi}\gamma^5$, $D_{\mu,-2} = \partial_\mu - i2g_\chi\phi_\mu$, and $V(\Phi) = -\mu^2\Phi^\star\Phi + \frac{\lambda}{2}\left(\Phi^\star\Phi\right)^2$ and the fields transform such that

$$\chi \to e^{i\gamma^5\alpha(x)}\chi\,,\;\; \phi_\mu \to \phi_\mu - \frac{1}{g_\chi}\partial_\mu\alpha(x)\,,\;\; \Phi \to e^{-2i\alpha(x)}\Phi\,. \tag{C.14}$$

The vacuum expectation value of the field Φ in this potential is given by $v = \sqrt{\mu^2/\lambda}$. We expand the scalar field around its minimum $\Phi(x) = v + \left(h(x) + i\Phi_2(x)\right)/\sqrt{2}$ and get the following relevant quantities after symmetry breaking: $m_\chi = \lambda_Y v$, $m_\phi^2/2 = 4g_\chi^2 v^2$, scalar mass $m_h = \sqrt{2\lambda v^2} = \sqrt{2}\mu$, Yukawa interaction $-\lambda_Y/\sqrt{2}\bar{\chi}h\chi = -m_\chi/(v\sqrt{2})\bar{\chi}h\chi$ and scalar-gauge boson interaction $+4\sqrt{2}g_\chi^2 v\, h\phi_\mu\phi^\mu$.

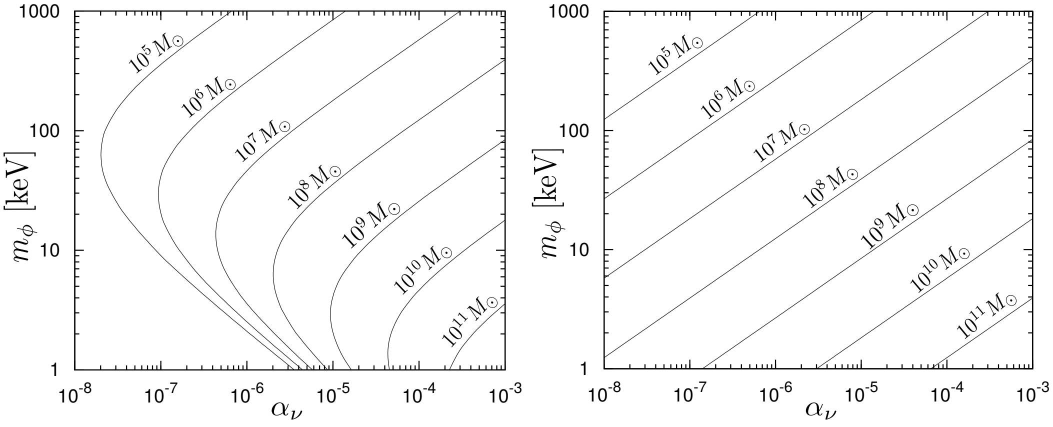
The invariant amplitude of DM annihilation into two gauge bosons contains three terms:

$$\begin{aligned}\mathcal{M} = {} & \epsilon^\star_\mu(k_2)\epsilon^\star_\nu(k_1)\bar{v}(p_2)\left(-ig_\chi\gamma^\mu\gamma^5\right)\frac{i\left(\not{p}_1 - \not{k}_1 + m_\chi\right)}{\left(p_1 - k_1\right)^2 - m_\chi^2}\left(-ig_\chi\gamma^\nu\gamma^5\right)u(p_1)\\ & + \epsilon^\star_\mu(k_1)\epsilon^\star_\nu(k_2)\bar{v}(p_2)\left(-ig_\chi\gamma^\mu\gamma^5\right)\frac{i\left(\not{p}_1 - \not{k}_2 + m_\chi\right)}{\left(p_1 - k_2\right)^2 - m_\chi^2}\left(-ig_\chi\gamma^\nu\gamma^5\right)u(p_1)\\ & + 2\times\bar{v}(p_2)\left(-i\frac{m_\chi}{\sqrt{2}v}\right)u(p_1)\frac{i}{\left(p_1 + p_2\right)^2 - m_h^2}\left(+i4\sqrt{2}g_\chi^2 v\right)\epsilon^\star_\mu(k_2)\epsilon^{\star,\mu}(k_1)\,, \end{aligned}\tag{C.15}$$

and the total cross section results in

$$\begin{aligned}(\sigma v_{\rm rel}) = {} & \frac{g_\chi^4}{16\pi m_\chi^2}\left(1 - \frac{1}{2}z^2 + \mathcal{O}(z^4)\right)\\ & + \frac{g_\chi^4}{16\pi(y^2-4)^2 m_\chi^2}\left(\frac{4}{3}(y^4+8)z^{-4} - \frac{16}{3}(2y^2+1)z^{-2} + \mathcal{O}(z^0)\right)v_{\rm rel}^2 + \mathcal{O}(v_{\rm rel}^4)\,, \end{aligned}\tag{C.16}$$

where $y = m_h/m_\chi$. Now, the limit of $m_\phi \to 0$ ($z \to 0$), effectively meaning $g_\chi \to 0$, results in a finite value of the annihilation cross section that is proportional to λ_Y^4. In the following, we



(a) Exact numerical result (b) Effective propagator description

Figure 5. Contour line of a constant $M_{\rm cut}$ is shown for the pseudo vector operator within the exact (left) and effective propagator frameworks (right). The DM parameters chosen are $m_\chi = 100\,{\rm MeV}$ and α_χ satisfying the relic density constraint. The effective description is valid only in the upper right quarter of figure 5(b).

show that all parameters are in the perturbative regime and the scalar contribution (C.15) can be ignored in the low energy expansion, so that (C.16) reduces to (C.3).

When we choose $z \sim 10^{-3}$, due to the z^{-4} enhancement in the annihilation cross section (C.4), the DM coupling is forced to be tiny in order to satisfy the relic abundance constraint. A choice of $m_\chi = 100\,{\rm MeV}$ leads to $g_\chi = 1.0 \times 10^{-5}$. With these choices, we derive $\lambda_Y = 0.03$, $v = \frac{m_\chi}{\lambda_Y} = 3.4\,{\rm GeV}$, and $y = \frac{\sqrt{2\lambda}}{\lambda_Y} \lesssim 49$. If we take y of $\mathcal{O}(10)$, we see directly that the leading term in (C.16) is indeed given by

$$(\sigma v_{\rm rel}) \simeq \frac{g_\chi^4}{12\pi m_\chi^2} z^{-4} v_{\rm rel}^2 \,. \tag{C.17}$$

This result is the same as the leading one in (C.3) and the relic abundance given by (C.7), where the scalar contribution has been ignored, holds.

Using the effective propagator description, we derive γ/H:

$$\frac{\gamma}{H} = 2.1 \times \left(\frac{r}{r_0}\right)^2 \left(\frac{N_\nu}{6} \frac{\alpha_\chi}{8.4 \times 10^{-12}} \frac{\alpha_\nu}{10^{-4}}\right) \left(\frac{m_\chi}{100\,{\rm MeV}}\right)^{-1} \left(\frac{m_\phi}{100\,{\rm keV}}\right)^{-4} \left(\frac{T_\nu}{1\,{\rm keV}}\right)^4 , \tag{C.18}$$

and the cutoff mass scaling pattern for the pseudo vector operator:

$$(M_{\rm cut})_{\rm PV} = 1.4 \times 10^8\, M_\odot \left(\frac{r}{r_0}\right)^{15/8} \left(\frac{N_\nu}{6} \frac{\alpha_\nu}{10^{-4}}\right)^{3/4} \left(\frac{m_\chi}{100\,{\rm MeV}}\right)^{-3/2} \left(\frac{m_\phi}{100\,{\rm keV}}\right)^{-3/2} , \tag{C.19}$$

where the relic density constraint on α_χ (C.7) is inserted into (3.6). Note that the cutoff mass depends now on the DM mass unlike in the scalar and vector operator cases. In figure 5, the exact numerical solution of γ/H is compared to the cutoff derived from (C.19), showing the valid range of the parameter space for the effective propagator description.

References

[1] WMAP collaboration, G. Hinshaw et al., *Nine-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results*, *Astrophys. J. Suppl.* **208** (2013) 19 [`arXiv:1212.5226`] [INSPIRE].

[2] Planck collaboration, P.A.R. Ade et al., *Planck 2013 results. XVI. Cosmological parameters*, *Astron. Astrophys.* **571** (2014) A16 [`arXiv:1303.5076`] [INSPIRE].

[3] C. Boehm, P. Fayet and R. Schaeffer, *Constraining dark matter candidates from structure formation*, *Phys. Lett.* **B 518** (2001) 8 [`astro-ph/0012504`] [INSPIRE].

[4] X. Chen, S. Hannestad and R.J. Scherrer, *Cosmic microwave background and large scale structure limits on the interaction between dark matter and baryons*, *Phys. Rev.* **D 65** (2002) 123515 [`astro-ph/0202496`] [INSPIRE].

[5] K. Sigurdson, M. Doran, A. Kurylov, R.R. Caldwell and M. Kamionkowski, *Dark-matter electric and magnetic dipole moments*, *Phys. Rev.* **D 70** (2004) 083501 [*Erratum ibid.* **D 73** (2006) 089903] [`astro-ph/0406355`] [INSPIRE].

[6] C. Boehm and R. Schaeffer, *Constraints on Dark Matter interactions from structure formation: Damping lengths*, *Astron. Astrophys.* **438** (2005) 419 [`astro-ph/0410591`] [INSPIRE].

[7] G. Mangano, A. Melchiorri, P. Serra, A. Cooray and M. Kamionkowski, *Cosmological bounds on dark-matter–neutrino interactions*, *Phys. Rev.* **D 74** (2006) 043517 [`astro-ph/0606190`] [INSPIRE].

[8] P. Serra, F. Zalamea, A. Cooray, G. Mangano and A. Melchiorri, *Constraints on neutrino-dark matter interactions from cosmic microwave background and large scale structure data*, *Phys. Rev.* **D 81** (2010) 043507 [`arXiv:0911.4411`] [INSPIRE].

[9] R.J. Wilkinson, J. Lesgourgues and C. Boehm, *Using the CMB angular power spectrum to study Dark Matter-photon interactions*, *JCAP* **04** (2014) 026 [`arXiv:1309.7588`] [INSPIRE].

[10] F.-Y. Cyr-Racine and K. Sigurdson, *Cosmology of atomic dark matter*, *Phys. Rev.* **D 87** (2013) 103515 [`arXiv:1209.5752`] [INSPIRE].

[11] F.-Y. Cyr-Racine, R. de Putter, A. Raccanelli and K. Sigurdson, *Constraints on large-scale dark acoustic oscillations from cosmology*, *Phys. Rev.* **D 89** (2014) 063517 [`arXiv:1310.3278`] [INSPIRE].

[12] C. Dvorkin, K. Blum and M. Kamionkowski, *Constraining dark matter-baryon scattering with linear cosmology*, *Phys. Rev.* **D 89** (2014) 023519 [`arXiv:1311.2937`] [INSPIRE].

[13] R.J. Wilkinson, C. Boehm and J. Lesgourgues, *Constraining dark matter-neutrino interactions using the CMB and large-scale structure*, *JCAP* **05** (2014) 011 [`arXiv:1401.7597`] [INSPIRE].

[14] M. Escudero, O. Mena, A.C. Vincent, R.J. Wilkinson and C. Boehm, *Exploring dark matter microphysics with galaxy surveys*, *JCAP* **09** (2015) 034 [`arXiv:1505.06735`] [INSPIRE].

[15] Y. Ali-Haïmoud, J. Chluba and M. Kamionkowski, *Constraints on Dark Matter Interactions with Standard Model Particles from Cosmic Microwave Background Spectral Distortions*, *Phys. Rev. Lett.* **115** (2015) 071304 [`arXiv:1506.04745`] [INSPIRE].

[16] J. Lesgourgues, G. Marques-Tavares and M. Schmaltz, *Evidence for dark matter interactions in cosmological precision data?*, *JCAP* **02** (2016) 037 [`arXiv:1507.04351`] [INSPIRE].

[17] B. Moore et al., *Dark Matter Substructure within Galactic Halos*, *Astrophys. J.* **524** (1999) L19 [`astro-ph/9907411`] [INSPIRE].

[18] A.V. Kravtsov, *Dark Matter Substructure and Dwarf Galactic Satellites*, *Adv. Astron.* **2010** (2010) 281913 [`arXiv:0906.3295`] [INSPIRE].

[19] B. Moore, T. Quinn, F. Governato, J. Stadel and G. Lake, *Cold collapse and the core catastrophe*, *Mon. Not. Roy. Astron. Soc.* **310** (1999) 1147 [`astro-ph/9903164`] [INSPIRE].

[20] F. Donato et al., *A constant dark matter halo surface density in galaxies*, *Mon. Not. Roy. Astron. Soc.* **397** (2009) 1169 [`arXiv:0904.4054`] [INSPIRE].

[21] W.J.G. de Blok, *The Core-Cusp Problem*, *Adv. Astron.* **2010** (2010) 789293 [`arXiv:0910.3538`] [INSPIRE].

[22] M. Boylan-Kolchin, J.S. Bullock and M. Kaplinghat, *Too big to fail? The puzzling darkness of massive Milky Way subhaloes*, *Mon. Not. Roy. Astron. Soc.* **415** (2011) L40 [`arXiv:1103.0007`] [INSPIRE].

[23] M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *The Milky Way's bright satellites as an apparent failure of* Λ*CDM*, *Mon. Not. Roy. Astron. Soc.* **422** (2012) 1203 [`arXiv:1111.2048`] [INSPIRE].

[24] A.J. Benson, C.S. Frenk, C.G. Lacey, C.M. Baugh and S. Cole, *The effects of photoionization on galaxy formation – II. Satellite galaxies in the Local Group*, *Mon. Not. Roy. Astron. Soc.* **333** (2002) 177 [`astro-ph/0108218`] [INSPIRE].

[25] T. Okamoto and C.S. Frenk, *The origin of failed subhaloes and the common mass scale of the Milky Way satellite galaxies*, *Mon. Not. Roy. Astron. Soc.* **399** (2009) L174 [`arXiv:0909.0262`] [INSPIRE].

[26] P. Bode, J.P. Ostriker and N. Turok, *Halo formation in Warm Dark Matter Models*, *Astrophys. J.* **556** (2001) 93 [`astro-ph/0010389`] [INSPIRE].

[27] C. Boehm, A. Riazuelo, S.H. Hansen and R. Schaeffer, *Interacting dark matter disguised as warm dark matter*, *Phys. Rev.* **D 66** (2002) 083505 [`astro-ph/0112522`] [INSPIRE].

[28] L.G. van den Aarssen, T. Bringmann and Y.C. Goedecke, *Thermal decoupling and the smallest subhalo mass in dark matter models with Sommerfeld-enhanced annihilation rates*, *Phys. Rev.* **D 85** (2012) 123512 [`arXiv:1202.5456`] [INSPIRE].

[29] L.G. van den Aarssen, T. Bringmann and C. Pfrommer, *Is Dark Matter with Long-Range Interactions a Solution to All Small-Scale Problems of* Λ *Cold Dark Matter Cosmology?*, *Phys. Rev. Lett.* **109** (2012) 231301 [`arXiv:1205.5809`] [INSPIRE].

[30] A. Kamada, N. Yoshida, K. Kohri and T. Takahashi, *Structure of dark matter halos in warm dark matter models and in models with long-lived charged massive particles*, *JCAP* **03** (2013) 008 [`arXiv:1301.2744`] [INSPIRE].

[31] C. Boehm, J.A. Schewtschenko, R.J. Wilkinson, C.M. Baugh and S. Pascoli, *Using the Milky Way satellites to study interactions between cold dark matter and radiation*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) L31 [`arXiv:1404.7012`] [INSPIRE].

[32] M.R. Buckley, J. Zavala, F.-Y. Cyr-Racine, K. Sigurdson and M. Vogelsberger, *Scattering, damping and acoustic oscillations: Simulating the structure of dark matter halos with relativistic force carriers*, *Phys. Rev.* **D 90** (2014) 043524 [`arXiv:1405.2075`] [INSPIRE].

[33] J.A. Schewtschenko, R.J. Wilkinson, C.M. Baugh, C. Boehm and S. Pascoli, *Dark matter-radiation interactions: the impact on dark matter haloes*, *Mon. Not. Roy. Astron. Soc.* **449** (2015) 3587 [`arXiv:1412.4905`] [INSPIRE].

[34] F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger and C. Pfrommer, *ETHOS — an effective theory of structure formation: From dark particle physics to the matter distribution of the Universe*, *Phys. Rev.* **D 93** (2016) 123527 [`arXiv:1512.05344`] [INSPIRE].

[35] M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann and K. Sigurdson, *ETHOS — An Effective Theory of Structure Formation: Dark matter physics as a possible explanation of the small-scale CDM problems*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1399 [`arXiv:1512.05349`] [INSPIRE].

[36] J.A. Schewtschenko, C.M. Baugh, R.J. Wilkinson, C. Boehm, S. Pascoli and T. Sawala, *Dark matter-radiation interactions: the structure of Milky Way satellite galaxies*, *Mon. Not. Roy. Astron. Soc.* **461** (2016) 2282 [`arXiv:1512.06774`] [INSPIRE].

[37] S. Hofmann, D.J. Schwarz and H. Stöcker, *Damping scales of neutralino cold dark matter*, *Phys. Rev.* **D 64** (2001) 083507 [`astro-ph/0104173`] [INSPIRE].

[38] A.M. Green, S. Hofmann and D.J. Schwarz, *The first WIMPy halos*, *JCAP* **08** (2005) 003 [astro-ph/0503387] [INSPIRE].

[39] A. Loeb and M. Zaldarriaga, *Small-scale power spectrum of cold dark matter*, *Phys. Rev.* **D 71** (2005) 103520 [astro-ph/0504112] [INSPIRE].

[40] E. Bertschinger, *Effects of cold dark matter decoupling and pair annihilation on cosmological perturbations*, *Phys. Rev.* **D 74** (2006) 063509 [astro-ph/0607319] [INSPIRE].

[41] P. Gondolo, J. Edsjo, P. Ullio, L. Bergström, M. Schelke and E.A. Baltz, *DarkSUSY: Computing supersymmetric dark matter properties numerically*, *JCAP* **07** (2004) 008 [astro-ph/0406204] [INSPIRE].

[42] T. Bringmann and S. Hofmann, *Thermal decoupling of WIMPs from first principles*, *JCAP* **04** (2007) 016 [hep-ph/0612238] [INSPIRE].

[43] T. Bringmann, *Particle models and the small-scale structure of dark matter*, *New J. Phys.* **11** (2009) 105027 [arXiv:0903.0189] [INSPIRE].

[44] J. Kasahara, *Neutralino dark matter: the mass of the smallest halo and the golden region*, Ph.D. Thesis, The University of Utah, Salt Lake City U.S.A. (2009).

[45] P. Gondolo, J. Hisano and K. Kadota, *The effect of quark interactions on dark matter kinetic decoupling and the mass of the smallest dark halos*, *Phys. Rev.* **D 86** (2012) 083523 [arXiv:1205.1914] [INSPIRE].

[46] C.-P. Ma and E. Bertschinger, *Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges*, *Astrophys. J.* **455** (1995) 7 [astro-ph/9506072] [INSPIRE].

[47] B. Dasgupta and J. Kopp, *Cosmologically Safe eV-Scale Sterile Neutrinos and Improved Dark Matter Structure*, *Phys. Rev. Lett.* **112** (2014) 031803 [arXiv:1310.6337] [INSPIRE].

[48] T. Bringmann, J. Hasenkamp and J. Kersten, *Tight bonds between sterile neutrinos and dark matter*, *JCAP* **07** (2014) 042 [arXiv:1312.4947] [INSPIRE].

[49] P. Ko and Y. Tang, *$\nu\Lambda$ MDM: A model for sterile neutrino and dark matter reconciles cosmological and neutrino oscillation data after BICEP2*, *Phys. Lett.* **B 739** (2014) 62 [arXiv:1404.0236] [INSPIRE].

[50] J.F. Cherry, A. Friedland and I.M. Shoemaker, *Neutrino Portal Dark Matter: From Dwarf Galaxies to IceCube*, arXiv:1411.1071 [INSPIRE].

[51] P.J. Fox and E. Poppitz, *Leptophilic dark matter*, *Phys. Rev.* **D 79** (2009) 083528 [arXiv:0811.0399] [INSPIRE].

[52] K.M. Nollett and G. Steigman, *BBN and the CMB constrain neutrino coupled light WIMPs*, *Phys. Rev.* **D 91** (2015) 083505 [arXiv:1411.6005] [INSPIRE].

[53] A. Lewis, A. Challinor and A. Lasenby, *Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models*, *Astrophys. J.* **538** (2000) 473 [astro-ph/9911177] [INSPIRE].

[54] T. Bringmann, H.T. Ihle, J. Kersten and P. Walia, *Suppressing structure formation at dwarf galaxy scales and below: late kinetic decoupling as a compelling alternative to warm dark matter*, arXiv:1603.04884 [INSPIRE].

[55] Z.-P. Liu, Y.-L. Wu and Y.-F. Zhou, *Sommerfeld enhancements with vector, scalar, and pseudoscalar force carriers*, *Phys. Rev.* **D 88** (2013) 096008 [arXiv:1305.5438] [INSPIRE].

[56] F. Kahlhoefer, K. Schmidt-Hoberg, T. Schwetz and S. Vogl, *Implications of unitarity and gauge invariance for simplified dark matter models*, *JHEP* **02** (2016) 016 [arXiv:1510.02110] [INSPIRE].

[57] M. Cannoni, *Relativistic and nonrelativistic annihilation of dark matter: a sanity check using an effective field theory approach*, *Eur. Phys. J.* **C 76** (2016) 137 [arXiv:1506.07475] [INSPIRE].

4.4 *Publication:* Reannihilation of self-interacting dark matter

Reprinted article with permission from the authors:

Tobias Binder, Michael Gustafsson, Ayuki Kamada, Stefan Marinus Rodrigues Sandner, and Max Wiesner.

Physical Review D **97**, 123004 (2018)

https://doi.org/10.1103/PhysRevD.97.123004

Copyright 2018 by the American Physical Society.

Author contributions T. B. and A. K. designed the early project idea and shared it with M. G. and two Bachelor students, S. M. R. S. and M. W. . Under the advise of M. G. and T. B. , S. M. R. S. and M. W. made new observations during their Bachelor studies on which parts of the article are based and developed a numerical code for studying reannihilation. T. B. optimized and further extended their code, produced the main results and did all analytic computations. A. K. conceived the idea to use the angular size of the sound horizon as well as other CMB quantities to constrain reannihilation and collaborated with T. B. who worked out the numerical analysis. M. G. made an important contribution regarding the shift of Sommerfeld resonances and did the numerical phase-space density computation of the Boltzmann equation. M. G. and A. K. contributed in writing and editing the article, while T. B. wrote most of the manuscript.

PHYSICAL REVIEW D **97,** 123004 (2018)

Reannihilation of self-interacting dark matter

Tobias Binder,[1,*] Michael Gustafsson,[1,†] Ayuki Kamada,[2,‡] Stefan Marinus Rodrigues Sandner,[3,§] and Max Wiesner[4,∥]

[1]*Institute for Theoretical Physics, Georg-August University Göttingen,*
Friedrich-Hund-Platz 1, Göttingen D-37077, Germany
[2]*Center for Theoretical Physics of the Universe, Institute for Basic Science (IBS), Daejeon 34126, Korea*
[3]*DESY, Notkestraße 85, 22607 Hamburg, Germany*
[4]*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 σ_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 ΛCDM 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 $\sigma/m_{\mathrm{DM}} \sim 1\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ on subgalactic scales (with a typical rotation velocity $v_0 \sim 30\ \mathrm{km/s}$) [28]. This cross section must presumably decrease with velocity to allow for a smaller impact on galaxy cluster scales (with $v_0 \sim 10^3\ \mathrm{km/s}$) 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

© 2018 American Physical Society

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 (ϕ) induces both the desired strong DM (χ) 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.[1] 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 σ_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. 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,

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

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

where g_χ and g_l are the coupling constants. The DM particle χ is a spin $1/2$ Dirac fermion with mass m_χ, and the vector ϕ_μ or scalar ϕ 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, ϕ, and χ form a dark sector, which effectively decouples from the SM plasma, leads to an additional free parameter, namely, the temperature ratio,

$$r \equiv \frac{T_l}{T_\gamma}, \tag{3}$$

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

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

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, $r_{\mathrm{BBN}} = 0.5$, 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.

---

[1] 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.

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 $\sigma/m_{\mathrm{DM}} \sim 1\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ 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 $\sigma/m_{\mathrm{DM}} \lesssim 0.1\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ [51]. A merging cluster system like a bullet cluster also provides a good test for SIDM. The reported tight constraint is $\sigma/m_{\mathrm{DM}} \lesssim 0.7\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ [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

$$\langle\sigma_T\rangle_{v_0} = \frac{4\pi}{(\sqrt{2\pi}v_0)^3}\int \sigma_T e^{-v_{\mathrm{rel}}^2/(2v_0^2)}\, v_{\mathrm{rel}}^2 \mathrm{d}v_{\mathrm{rel}}, \qquad (5)$$

where σ_T is the transfer cross section:

$$\sigma_T \equiv \int \mathrm{d}\Omega (1-\cos\theta)\frac{\mathrm{d}\sigma_{\chi\chi\to\chi\chi}}{\mathrm{d}\Omega}. \qquad (6)$$

For the Yukawa-potential scattering, induced by our light mediators, we will use the ETHOS fitting functions for σ_T in the classical regime ($m_\chi v_{\mathrm{rel}} \gtrsim m_\phi$)—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 $\langle\sigma_T\rangle_{30\ \mathrm{km/s}}/m_\chi \in [0.1, 10]\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ (sizable 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 σ_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 σ_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:

$$\begin{aligned} M^{\mathrm{cut}} &\equiv \rho_m \frac{4\pi}{3}\left(\frac{1}{H}\right)^3\bigg|_{\mathrm{DM\ \mathbf{k}inetic\ \mathbf{d}ecoupling}} \\ &= 2.2\times 10^8 r_{\mathrm{kd}}^3\left(\frac{1\ \mathrm{keV}}{T_l^{\mathrm{kd}}}\right)^3 M_\odot. \end{aligned} \qquad (7)$$

Here, ρ_m is the total matter density and H is the Hubble expansion rate. We see that a kinetic decoupling temperature $T_l^{\rm kd}$ 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 ϕ 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 χ and l dominates over that between χ and ϕ. The kinetic decoupling temperature $T_l^{\rm kd}$ can be defined as the temperature when the Hubble expansion rate H equals the momentum transfer rate γ.[2] In Refs. [24,27] the momentum transfer rate is derived to be

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

and explicit expressions of the elastic $l-\chi$ scattering cross section $d\sigma/dtv_{\rm rel}$ 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 $M^{\rm cut}$ for similar coupling constants and particle masses [27]. The suppression of the halo mass function mainly depends on the mediator mass m_ϕ and for both models a cutoff mass around the dwarf galactic scale can be achieved for m_ϕ 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-α 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 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 $T_l^{\rm kd}/r\gtrsim$ 1 keV (0.6 keV),[3] which according to Eq. (7) results in an upper limit on the cutoff mass of

$$M^{\rm cut}\lesssim 2\times10^8\ M_\odot(10^9\ M_\odot). \tag{9}$$

Cutoff masses in the range 10^7 to 10^9 $M_\odot$ are indicated in Fig. 4 of Sec. IV, where $M^{\rm cut}$ 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: $\chi\bar\chi\to\phi\phi$ and $\chi\bar\chi\to l\bar l$. 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:

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

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

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

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

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

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 $(\sigma v_{\rm rel})_{0,i}^p\propto v_{\rm rel}^2$.

$S(v_{\rm rel})$ 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

---

[2] 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].

[3] The strong (the weak) limit is derived from the Lyman-α 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])

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

where the two dimensionless parameters are defined as

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

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

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

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

is fulfilled. The position of the nth "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

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

when ϵ_ϕ 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 $(\sigma v_{\rm rel})_V$ 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 $(\sigma v_{\rm rel})_V$ 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 χ 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_χ 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,

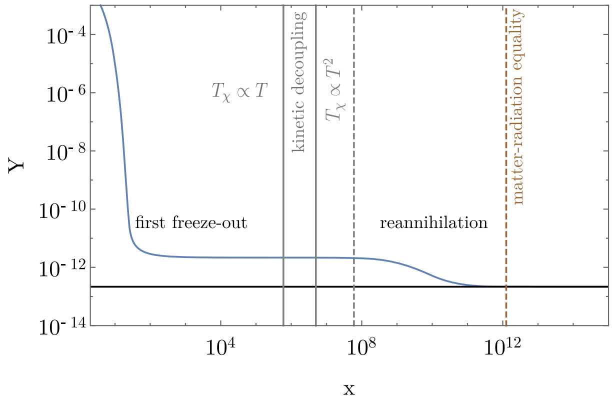


FIG. 1. Reannihilation process shown as a function of $x \equiv m_\chi/T_\gamma$, changing the DM comoving number density $Y \equiv n_\chi/s$ by 1 order of magnitude. The final abundance coincides with the correct value (black horizontal line). Here, we have chosen the parameters $m_\chi = 1$ TeV, $\alpha_\chi = 0.007$, $m_\phi \simeq 1$ GeV, and the resonance number $n \simeq 2$ (where m_ϕ 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,

$$\Gamma \equiv \frac{\langle \sigma v_{\rm rel} \rangle_{x'} Y}{H/s}, \tag{19}$$

where the dimensionless form of the DM number density n_χ is defined as

$$Y \equiv \frac{n_\chi}{s}. \tag{20}$$

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

$$x \equiv \frac{m_\chi}{T_\gamma}. \tag{21}$$

We follow the evolution of Γ 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_χ and can be written in the nonrelativistic limit as

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

where it is a function of

$$x' \equiv \frac{m_\chi}{T_\chi}. \tag{23}$$

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

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

$$x' = \begin{cases} \frac{m_\chi}{T_l} = \frac{x}{r} & \text{before kinetic decoupling,} \\ \frac{m_\chi T_l^{\rm kd}}{T_l^2} = \frac{x^2}{r^2 x_l^{\rm kd}} & \text{after kinetic decoupling.} \end{cases} \tag{24}$$

Here, $x_l^{\rm kd} \equiv m_\chi / T_l^{\rm kd}$ 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 Γ drops below 1, Γ scales as x^{-1} until the Sommerfeld factor (or the total s-wave annihilation cross section) starts to dominantly scale as $S \propto 1/v_{\rm rel}$. From this point to kinetic decoupling Γ further decreases in the phase of the $S \propto 1/v_{\rm rel}$ scaling, followed by a period where Γ stays constant. After kinetic decoupling, when T_χ starts to drop quickly, the $S \propto 1/v_{\rm rel}^2$ scaling dominates and leads to an increase of Γ as is highlighted by boldface $\boldsymbol{x}$ in the table. When Γ 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 $(\sigma v_{\rm rel}) \propto$ const and to a fast

TABLE I. Evolution of Γ at different cosmological epochs and for different DM annihilation cross sections $(\sigma v_{\rm rel})$. Only after kinetic decoupling and with a cross section scaling as $1/v^2$ can the ratio of annihilation over expansion rate, Γ, 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).

	$(\sqrt{g_{\rm eff}}/g_s) \times \Gamma \propto$			
	Radiation-dominated epoch		Matter-dominated epoch	
$(\sigma v_{\rm rel}) \propto$	*Before* kinetic decoupling	*After* kinetic decoupling	*Before* kinetic decoupling	*After* kinetic decoupling
Constant	x^{-1}	x^{-1}	$x^{-3/2}$	$x^{-3/2}$
$1/v_{\rm rel}$	$r^{-1/2}\, x^{-1/2}$	r^{-1}	$r^{-1/2} x^{-1}$	$r^{-1} x^{-1/2}$
$1/v_{\rm rel}^2$	r^{-1}	$r^{-2}\boldsymbol{x}$	$r^{-1} x^{-1/2}$	$r^{-2}\boldsymbol{x}^{1/2}$

decrease of Γ—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 Γ can vary depending on the model parameters.

An important quantity, used in the following sections, is the redshift $z_{\rm rea}$ defined by when the DM comoving number density changes first by 1% due to reannihilation. $z_{\rm rea}$ can be determined from the value of Γ 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

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

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 $z_{\rm rea} \ll z_{\rm eq}$, where $z_{\rm eq} \simeq 3400$ 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 $z_{\rm rea}$ 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 $T_\gamma^{\rm sat}$ is higher than the matter-radiation equality temperature $T_\gamma^{\rm eq} = 0.80$ eV. The saturation temperature $T_\gamma^{\rm sat}$ 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 $T_\gamma^{\rm sat}$ is given by the minimum α_χ value that can give a resonance. This occurs when $n = 1$ in Eq. (17) and is given by

$$\alpha_\chi^{\rm min} = \frac{\pi^2}{6}\frac{m_\phi}{m_\chi}. \tag{26}$$

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

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

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

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

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 $(\sigma v_{\rm rel}) \propto v_{\rm rel}^{-1-\epsilon}$, 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:

$$\frac{{\rm d}Y}{{\rm d}x} = -\frac{s}{\tilde{H}x}\langle\sigma v_{\rm rel}\rangle_y Y^2, \tag{29}$$

$$\frac{{\rm d}y}{{\rm d}x} = -\frac{2\gamma}{\tilde{H}x}[y - y^{\rm eq}] + \frac{s}{\tilde{H}x} yY[\langle\sigma v_{\rm rel}\rangle_y - \langle\sigma v_{\rm rel}\rangle_{y,2}], \tag{30}$$

$$\frac{{\rm d}Y_l}{{\rm d}x} = -\frac{H}{\tilde{H}x}Y_l + \frac{s}{\tilde{H}x}\langle\sigma v_{\rm rel}\rangle_y Y^2, \tag{31}$$

where we have defined the dimensionless temperatures as

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

$$y^{\rm eq} \equiv \frac{m_\chi T_l}{s^{2/3}} = \frac{m_\chi r T_\gamma}{s^{2/3}}. \tag{33}$$

The energy density of the injected dark radiation, given by

$$Y_l \equiv \frac{\rho_l}{m_\chi s}, \tag{34}$$

is fully included in the Hubble expansion rate

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

where the total dark sector energy density is given by

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

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

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

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

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

$$m_\chi(\partial_t - H\mathbf{p}\cdot\nabla_{\mathbf{p}})f_\chi = C^{\mathrm{non-rel}}_{\chi\chi\to\phi\phi,ll}[f_\chi] + C^{\mathrm{non-rel}}_{\chi l\leftrightarrow\chi l}[f_\chi, f_l^{\mathrm{eq}}], \tag{38}$$

by taking the zeroth

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

and the second moment with respect to momentum

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

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 γ 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 ϕ into fermions l are included in the equations via the *total* averaged cross section $\langle\sigma v_{\mathrm{rel}}\rangle_y$. 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_χ 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, $\langle\sigma v_{\mathrm{rel}}\rangle$ and $\langle\sigma v_{\mathrm{rel}}\rangle_2$ defined as

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

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

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:

$$f_\chi = \frac{n_\chi(T)}{n_{\chi,\mathrm{eq}}(T_\chi)}\exp\left(-\frac{m_\chi + p^2/(2m_\chi)}{T_\chi}\right)\Bigg|_{T_\chi = ys^{2/3}/m_\chi}, \tag{43}$$

where the T_χ evolution is governed via Eq. (30). This ansatz leads to the final form of $\langle\sigma v_{\mathrm{rel}}\rangle_y$ given in Eq. (22), and for simplifying the momentum square weighted annihilation cross section $\langle\sigma v_{\mathrm{rel}}\rangle_{y,2}$ 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 $[\langle\sigma v_{\mathrm{rel}}\rangle_y - \langle\sigma v_{\mathrm{rel}}\rangle_{y,2}]$. For Sommerfeld enhanced cross sections, this difference is always positive since $\langle\sigma v_{\mathrm{rel}}\rangle_{y,2}$ 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 $x \sim 10^6$.

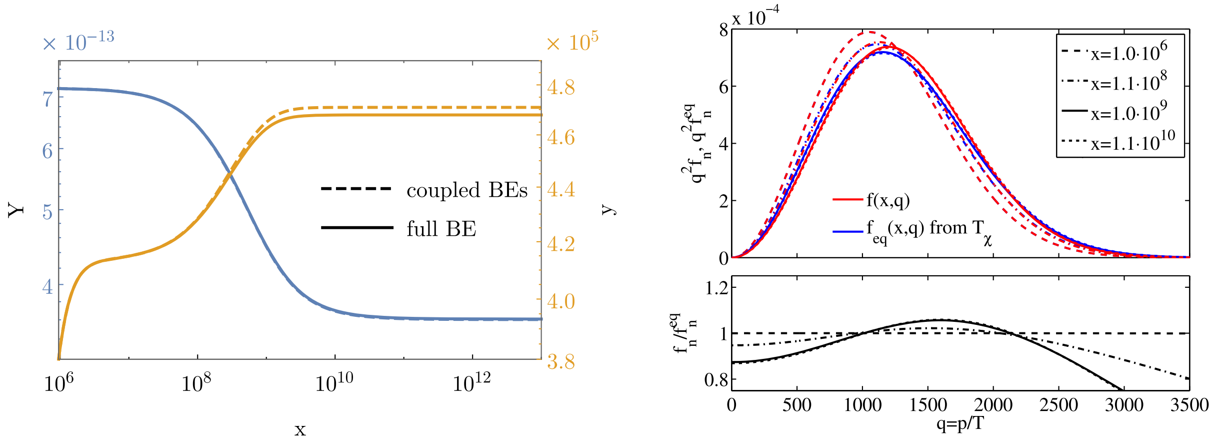


FIG. 2. Time evolution of the DM abundance $Y \equiv n_\chi/s$, its temperature $y \equiv \frac{m_\chi}{s^{2/3}} T_\chi$, and its phase-space density $f(q)$ with $q = p/T_\gamma$. *Left panel:* The evolution of Y (blue line) and y (yellow line) in the case of strongly self-interacting DM (dotted lines) 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 $f_n^{\mathrm{eq}}(q)$ with the same "temperature" T_χ (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 $f_n(q)/f_n^{\mathrm{eq}}(q)$. The DM model is $m_\chi = 600$ GeV, $m_\phi \simeq 1$ GeV, and α_χ 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

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

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

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

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

where θ 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 $\langle \sigma v_{\mathrm{rel}} \rangle$ 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 $f_\chi^{\mathrm{eq}}(q)$ (blue curves) that have the same T_χ, 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_χ falls below the corresponding thermal distribution $f_\chi^{\rm eq}$ at larger q does not have any practical implications—as the number density in the high momentum tail is negligible. The reason for this $f_\chi/f_\chi^{\rm eq}$ suppression is a spurious effect from annihilation at low momenta. The alteration of f_χ at low momenta leads to a best-fit thermal distribution $f_\chi^{\rm eq}$ 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 $f_\chi^{\rm eq}$.

IV. PARAMETER SCAN

Reannihilation leads to the fact that multiple values of α_χ can give the observed DM abundance for fixed model parameters m_χ and m_ϕ. An initial DM overabundance from the first freeze-out due to lower α_χ values can be compensated by a second period of annihilation. More precisely this is possible if the three free parameters[4] 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 α_χ for fixed values of m_χ and m_ϕ. 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_χ and m_ϕ values. The ability to lower Ω_χ 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 α_χ in the vicinity of the exact resonance point that reproduces the measured relic density. In fact, there are two α_χ possibilities for each of these resonances since $\Omega_\chi h^2$ is a smooth function of α_χ that coincides with the off-resonant result between the resonances. Larger values of α_χ than those shown in the figure do not lead to the correct abundance. To conclude, for given m_χ and m_ϕ 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 α_χ.

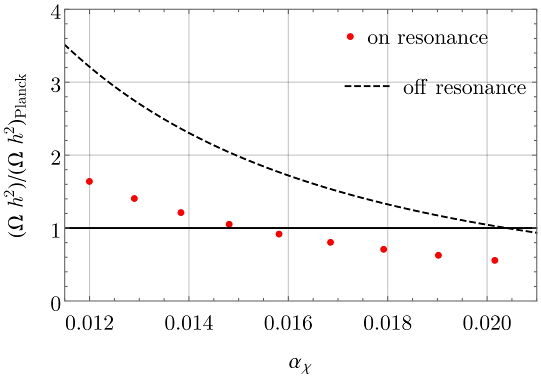


FIG. 3. Relic abundance ratio shown vs the coupling α_χ for fixed $m_\chi = 1$ TeV and $m_\phi = 10$ MeV. Dashed black curve is the result for only taking the standard thermal freeze-out into account (labeled as off resonance). The red dots present points where the parametric resonance condition is exactly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

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_χ, m_ϕ)-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 α_χ 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 α_χ values.

The region of multiple α_χ values is bordered from below by the grey shaded region, where no resonances leading to the correct relic density exist because $\epsilon_\phi^{\rm min} \equiv \frac{m_\phi}{\alpha_\chi^{\rm max} m_\chi} > \frac{6}{1^2 \pi^2}$, where $\alpha_\chi^{\rm max}$ 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

---

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

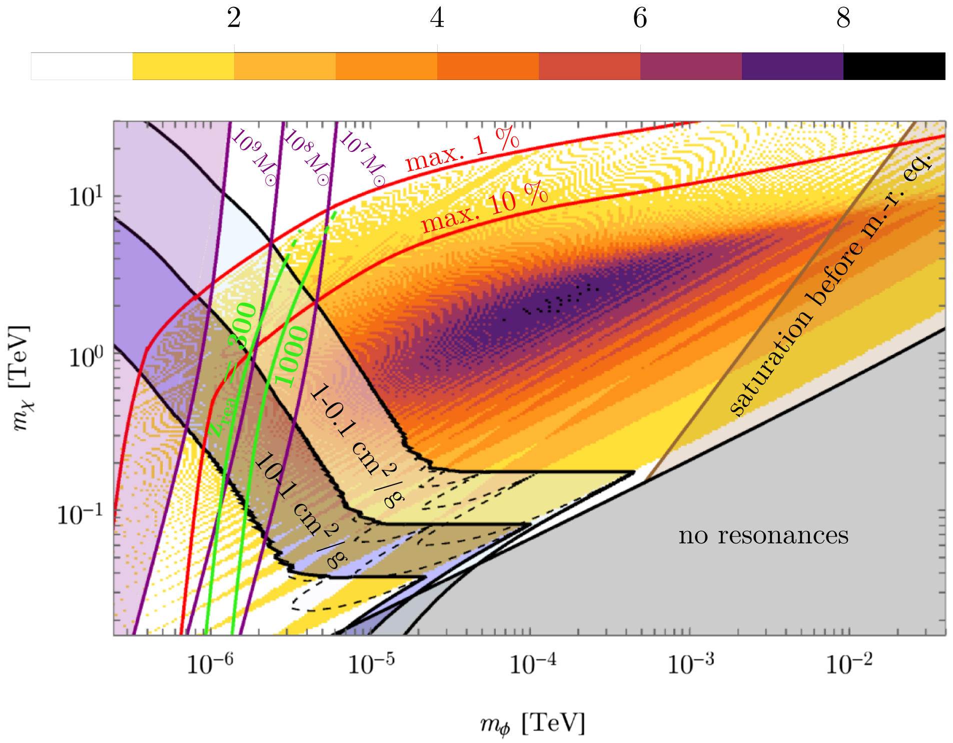


FIG. 4. Number of Sommerfeld resonances, color coded as given in the top panel, leading to the correct relic density today and changing the comoving DM number density by at least 1% during the epoch of reannihilation. The red solid lines show our analytic estimates (see Appendix C 2) of the border where 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: $(\sigma_T)_{30\ \mathrm{km/s}}/m_\chi \in [0.1, 10]\ \mathrm{cm}^2\ \mathrm{g}^{-1}$. 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, α_χ is adjusted in the computation of σ_T such that for given m_χ and m_ϕ 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 α_χ 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 ($m_\phi \simeq 1$ MeV, $m_\chi \simeq$ TeV), saturation of the Sommerfeld enhancement before today prevents the abundance to change. In the region where the red curves bend a second time ($m_\phi \simeq 10$ MeV, $m_\chi \simeq 10$ 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 $r_{\mathrm{BBN}} \neq 0.5$ 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 α_χ 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., $T_\gamma^{\mathrm{sat,min}} = 0.80$ 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 $(\sigma_T)_{30\ \mathrm{km/s}}/m_\chi \in [0.1, 10]\ \mathrm{cm}^2\,\mathrm{g}^{-1}$ 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_{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 $(\sigma_T)_{30\,\mathrm{km/s}}$ needs some further consideration in the parameter region where reannihilation happens. The multiple α_χ values leading to the correct relic abundance would also lead to multiple values of $(\sigma_T)_{30\,\mathrm{km/s}}$ for fixed m_ϕ and m_χ. For the blue self-scattering band in Fig. 4 we take the nearly-on-resonant α_χ value that is closest to the off-resonant α_χ 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 $(\sigma_T)_{30\,\mathrm{km/s}}$ does not change significantly when choosing an off-resonant or the closest on-resonant value of α_χ. However, in the quantum-resonant regime it makes a significant difference from using an off-resonant value when computing $(\sigma_T)_{30\,\mathrm{km/s}}$ as in, e.g., Refs. [30,33,54,55], which results in the dashed black curves (where α_χ is uniquely set by the standard relic density constraint, taking no reannihilation into account). In the Born regime ($\epsilon_\phi \gg 1$) we chose α_χ 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_{rea} contours given by the green lines in Fig. 4. Our calculations show that a reannihilation process at $z_{\mathrm{rea}} \simeq 300$ 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 $z_{\mathrm{rea}} \simeq 300$ as a lower critical value above which ~5% change in the abundance can be achieved. The reannihilation process starting at $z_{\mathrm{rea}} \simeq 1000$ with ~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 $z_{\mathrm{rea}} \simeq 1000$ 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_{rea}. However, this equation is strictly speaking only valid in the regime where $z_{\mathrm{rea}} \ll z_{\mathrm{eq}}$ and therefore a not good approximation in the case of $z_{\mathrm{rea}} \simeq 1000$. 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 ΛCDM cosmology. Since this process is time dependent, the naive constraints on extra relativistic degrees of freedom ΔN_{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 θ_* at $z = z_*$, where z_* is defined as the redshift where the optical depth τ 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]:

$$100\theta_* = 1.04105 \pm 0.00046, \tag{47}$$

along with

$$z_* = 1090.09 \pm 0.42. \tag{48}$$

From Ref. [90] we have

$$100\theta_* = 100 \times r_s(z_*)/D_{\mathrm{A}}(z_*). \tag{49}$$

The sound horizon r_s and angular diameter distance D_{A} are given by

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

$$D_{\mathrm{A}}(z) = \int_{1/(1+z)}^{1} \frac{da}{a^2 H}, \tag{51}$$

where

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

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

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

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 ΛCDM cosmology. When including reannihilation we will replace the standard CDM energy density ρ_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_ν entering the parametrization of energy density ρ_ν in Eq. (53). However, we do not consider this possibility here and fix m_ν 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 ΛCDM 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 redshifting of the injected dark radiation, which delays the return to the standard Hubble expansion rate.

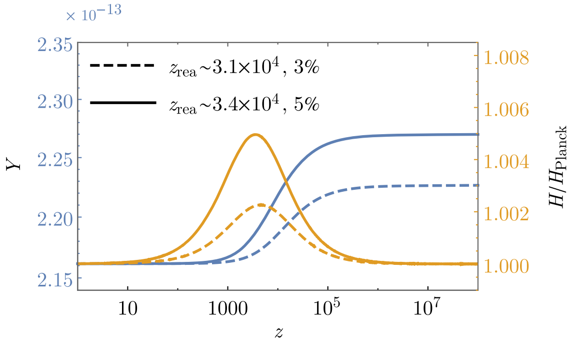


FIG. 5. Evolution of the DM number density $Y = n_\chi/s$ (blue line) and the corresponding expansion rate H (yellow line) shown as a function of the redshift. The onset (1% change in Y) of reannihilation for the dashed and solid curves is around $z \simeq 3 \times 10^4$ and the DM abundance is initially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

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_{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_ϕ around $\mathcal{O}(10)$ MeV and adjusted α_χ 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_{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 α_χ values.

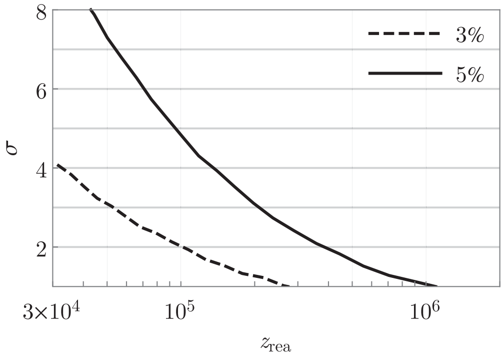


FIG. 6. Number of standard deviations from the $(100\theta_*)_{\mathrm{Planck}}$ measurement vs the redshift of reannihilation onset z_{rea} (i.e., when the comoving DM abundance first changes by more than 1%). The two curves refer to 3% and 5% total change in the relic abundance where the final value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

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, $z_{\rm rea} \lesssim z_*$, 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 compatibleness 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 ΛCDM model. Several works have pointed out the so-called H_0 tension; a discrepancy within the ΛCDM model between the measured value of the Hubble constant using CMB data [1], $H_0 = 67.31 \pm 0.96\ {\rm km\,s^{-1}\,Mpc^{-1}}$ (68% C.L.), and local measurements using only low redshift data, $H_0 = 73.24 \pm 1.74\ {\rm km\,s^{-1}\,Mpc^{-1}}$ (68% C.L.) [39]. Another tension concerns large-scale structure data and the value of the matter fluctuation amplitude on scales of $8h^{-1}$ Mpc, σ_8. This issue is related to the H_0 tension, as the Hubble parameter correlates with the matter density Ω_m and σ_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 σ_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, $r_s(z_*) = r_s^{\rm Planck}(z_*)$, since it is a distance derived from integrating $H(a)$ from $a = 0$ to the redshift of recombination. Depending on $z_{\rm rea}$, 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 ΛCDM 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 ρ_Λ 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 ρ_Λ 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 $z_{\rm rea} = 300$ and 700. To be in the special SIDM region we have used the parameters $m_\chi = 700$ GeV, $m_\phi \in [1.6, 2.0]$ MeV, and α_χ 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 $z_{\rm rea}$ 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 ΛCDM 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 $H/H_{\rm Planck}$ ratio increases at low redshifts below $z \lesssim 1$ because Ω_Λ 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 ΛCDM cosmology.

The reduction of Ω_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 ΛCDM [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 ΛCDM 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 χ 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

$$\dot\rho_M + 3H\rho_M = -\Gamma\rho_M, \tag{54}$$

$$\dot\rho_D + 4H\rho_D = \Gamma\rho_M, \tag{55}$$

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

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

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

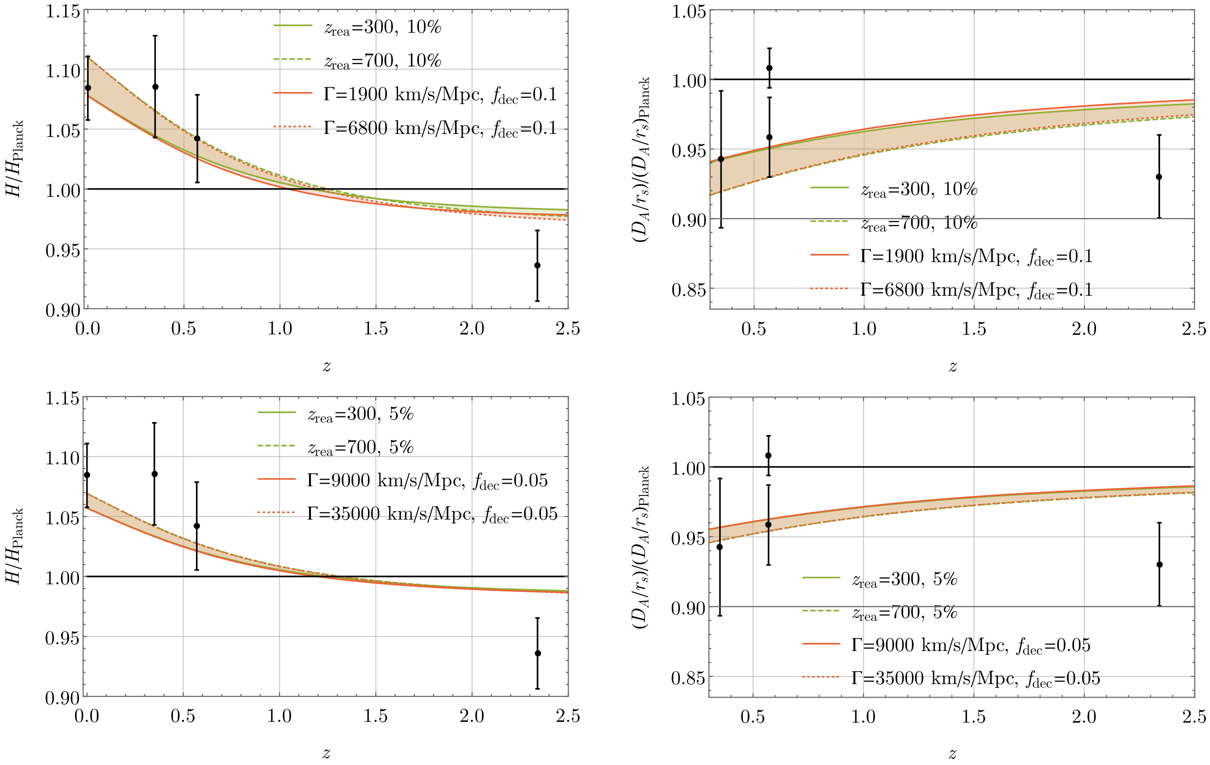


FIG. 7. Hubble expansion rate (left) and angular diameter distance (right) ratio vs redshift for reannihilating (green lines) and decaying DM (red lines). In both DM scenarios the abundance is changed by 10% (top) and 5% (bottom) after recombination, and the results are almost identical in this redshift 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 ΛCDM 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 ρ_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 ΛCDM 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 σ_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 σ_T/m_χ 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 δ_ℓ 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]

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

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

$$T(v_{\rm rel}) \simeq -\frac{1}{2\epsilon_v}(H(\alpha_+) + H(\alpha_-) + H(-\alpha_+) + H(-\alpha_-) - \{p \to p_0\}), \tag{B2}$$

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

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

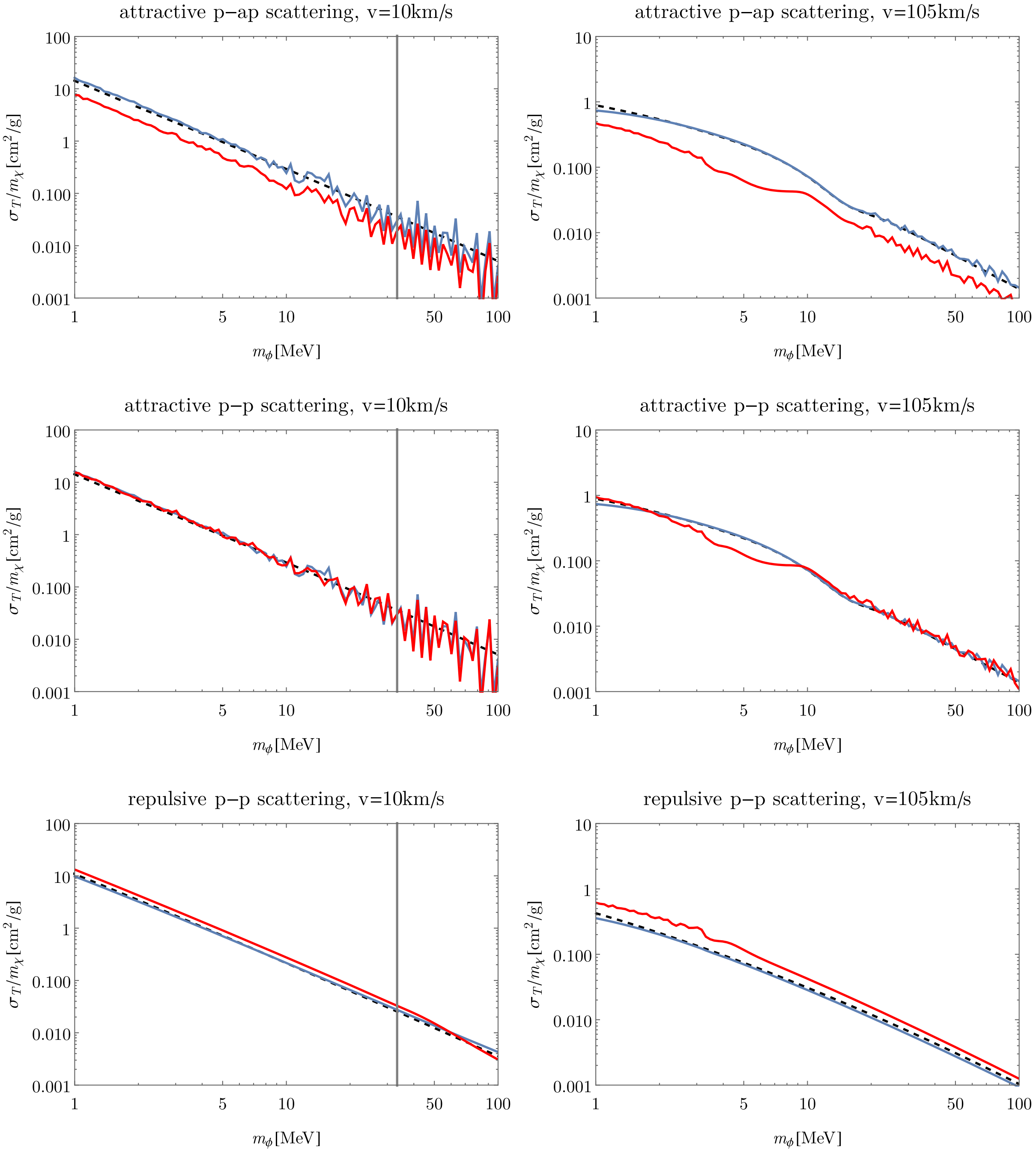


FIG. 8. Self-scattering transfer cross sections vs mediator mass m_ϕ from numerical solutions of the Schrödinger equation [33,54,55] (red and blue lines) compared to the ETHOS [24] fitting functions (black dashed lines). The particle model is fixed to $m_\chi = 1$ TeV and $\alpha_\chi = 0.033$. *Left:* 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 δ_ℓ up to $\ell = 150$ (left) and $\ell = 225$ (right). The red curve corresponds to the computation of σ_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 $1 - |\cos\theta|$.

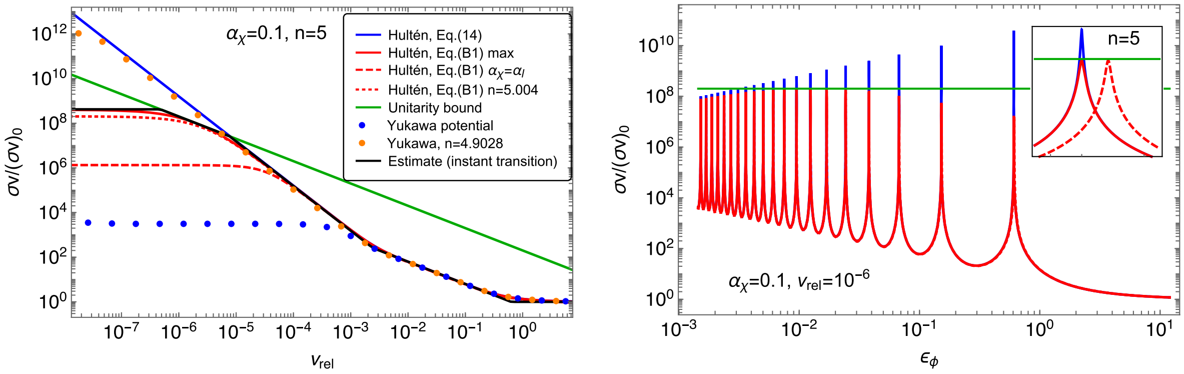


FIG. 9. *Left:* Comparison between various approximations of the s-wave Sommerfeld factor $(\sigma v_{\rm rel})_{\rm ann}/(\sigma v_{\rm rel})_{\rm ann,0}$ for a model with $\alpha_\chi = \alpha_l = 0.1$. The regulated Hulthén potential solution (red line) with $\sigma_{\rm sc,0} = (\mu/4\pi)(\sigma v_{\rm rel})^2_{\rm ann,0}$ 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 $\sigma_{\rm sc,0} = 3\alpha_\chi^2\pi/(4m_\chi^2)$ 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 ϵ_ϕ. 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 ϵ_ϕ and 4 orders of magnitude range in $(\sigma v_{\rm rel})_{\rm ann}$, showing the choice of $\sigma_{\rm sc,0}$ only shifts the resonance slightly while leaving the maximal enhancement unaffected.

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

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

$$\sigma_{\rm sc,0} = \frac{3\alpha_\chi^2\pi}{4m_\chi^2}, \tag{B5}$$

which can be obtained from the s-channel diagram of nonrelativistic particle-antiparticle scattering. In our computation of $T(v_{\rm rel})$ we drop the matching term $\{p \to p_0\}$, 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 $(\sigma v_{\rm rel})_{\rm ann,0}$ and $\sigma_{\rm sc,0}$ 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 $\sigma_{\rm sc,0} = \frac{\mu}{4\pi}(\sigma v_{\rm rel})^2_{\rm ann,0}$ 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 $\sigma_{\rm sc,0}$ 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 $(\sigma v_{\rm rel})_{\rm ann}$ 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])

$$\sigma_{\rm max} = \frac{\pi}{\mu^2 v_{\rm rel}^2}, \tag{B6}$$

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 $\sigma_{\rm max} = (\sigma v_{\rm rel})_{\rm ann,0} S(v_{\rm rel})/v_{\rm rel}$, namely,

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

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. C 1 we find approximate expressions of the Sommerfeld enhancement that we then use to estimate $\langle \sigma v_{\rm rel} \rangle_{x'}$. The latter quantity is used in Sec. C 2 to find analytical expressions for Γ, 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. C 3.

1. Sommerfeld enhanced annihilation on a resonance

In the following, we first estimate $S(v)$ and then $\langle\sigma v_{\mathrm{rel}}\rangle_{x'}$ in the case where the parameter ϵ_ϕ 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:

$$v_{1/v} = 2\pi\alpha_\chi, \tag{C1}$$

$$v_{1/v^2} = \frac{m_\phi}{m_\chi}, \tag{C2}$$

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

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

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 ϵ_ϕ and found in all the cases a similarly good result. The value of v_{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_{sat} we consider that S follows the scaling of the partial-wave unitarity bound, i.e., $1/v$ scaling. v_{max} is the velocity where S reaches the maximal value:

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

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

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

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

A similar instantaneous transitions description will next be used to estimate the temperature evolution of the thermally averaged cross section $\langle\sigma v_{\mathrm{rel}}\rangle_{x'}$ defined in Eq. (22). In the following we will drop the index x' to shorten the notation. $\langle\sigma v_{\mathrm{rel}}\rangle$ 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

$$x'_{1/v} = \frac{c_{1/v}}{v^2_{1/v}}, \tag{C6}$$

$$x'_{1/v^2} = \frac{c_{1/v^2}}{v^2_{1/v^2}}, \tag{C7}$$

$$x'_{\mathrm{sat}} = \frac{c_{\mathrm{sat}}}{v^2_{\mathrm{sat}}}, \tag{C8}$$

$$x'_{\mathrm{max}} = \frac{c_{\mathrm{max}}}{v^2_{\mathrm{max}}}, \tag{C9}$$

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

$$c_{\mathrm{max}} = \frac{c_{1/v}c_{1/v^2}}{c_{\mathrm{sat}}}. \tag{C10}$$

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

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_{kd}) and today (x_0) can be obtained from the standard solution of the Boltzmann equation [Eq. (29)], given by

$$\frac{Y(x_{\mathrm{kd}})}{Y(x_0)} = 1 + \int_{x_{\mathrm{kd}}}^{x_0} \mathrm{d}x\frac{\Gamma}{x}. \tag{C11}$$

Here, Γ is defined as in Eq. (19), but with the replacement $Y(x) \to Y(x_{\mathrm{kd}})$. 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 Γ is maximal. In the following, we first derive simple power-law expressions of the maximal value of Γ and second show how to approximate the time integration in various cases.

The usual order of chemical before kinetic decoupling ($x_{\rm cd} \lesssim x_{\rm kd}$) and kinetic decoupling before matter radiation equality ($x_{\rm kd} \lesssim x_0$) to have an adequate structure formation history implies the following time order: $x_{\rm cd} \lesssim x_{\rm kd} \lesssim x_{\rm eq} \lesssim x_0$, 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 $x_{\rm sat}$ [set by Eqs. (C6), (C7), and (C8) after converting x' into x via Eq. (24)] in between the fixed time order $x_{\rm cd} \lesssim x_{\rm kd} \lesssim x_{\rm eq} \lesssim x_0$. It turns out that only five different cases (time alignments) are relevant for us, and those are summarized as follows:

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

The two options given in each case lead to the same result in the final form of Γ 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: $x_0 \lesssim x_{\rm sat}$. This implies that Γ 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 $x_{1/v} \lesssim x_{\rm cd}$.

From here on we are always assuming that we are exactly on a Sommerfeld resonance point. The maximal value of Γ as a function of the free parameters in these five different cases can be obtained as follows. We define $x_{\rm cd}$ as the time when $\Gamma = 1$. Requiring $Y(x_{\rm kd})$ to coincide with the value of Y corresponding to the correct relic density, we can determine $x_{\rm cd}$ as a function of m_χ only. For the m_χ range between 10 GeV and 40 TeV we find that $x_{\rm cd}$ 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 $x_{\rm cd}$, the evolution of Γ in all five cases directly follows from the entries of Table I and the results of the previous section. For example, the estimate of Γ_1 is found to be

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

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 Γ is given by

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

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

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

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]

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

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

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

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

We now simplify the time integration of Γ to obtain the change in DM abundance due to reannihilation. A simple case is case 1 where $x_0 \lesssim x_{\mathrm{sat}}$. Γ takes the maximal value of Γ_1 today, and hence the integration can be simplified as

$$\int_{x_{\mathrm{kd}}}^{x_0} \mathrm{d}x \frac{\Gamma}{x} \approx \int_{x_{\mathrm{eq}}}^{x_0} \mathrm{d}x \frac{\Gamma}{x} \simeq 2 \times \Gamma_1. \tag{C18}$$

Inserting this result into Eq. (C11) and solving for m_χ for given m_ϕ 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_{cd} and α_χ, where the latter quantity is chosen such that $Y(x_{\mathrm{kd}})$ gives the correct relic density. Another simple case is case 5 where $x_{\mathrm{sat}} \lesssim x_{\mathrm{eq}}$. Here, the maximum value of Γ is given by the saturation temperature in the radiation dominated epoch, leading to the simplification

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

In the last approximation we assumed that $T_\gamma^{\mathrm{sat}} \ll T_\gamma^{1/v^2}$ and $T_\gamma^{\mathrm{max}} \gg T_\gamma^{\mathrm{eq}}$. The temperature ratio appearing in the latter equation is a constant and is given by $T_\gamma^{\mathrm{sat}}/T_\gamma^{\mathrm{max}} \simeq 34$, 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 Γ 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:

$$r_0 = 0.36, \tag{C20}$$

$$T_\gamma^{\mathrm{eq}} = 0.80\ \mathrm{eV}, \tag{C21}$$

$$T_\gamma^0 = 2.34 \times 10^{-4}\ \mathrm{eV}, \tag{C22}$$

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

3. Redshift of reannihilation onset

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

$$0.01 = \int_{x_{\mathrm{kd}}}^{x_{\mathrm{rea}}} \mathrm{d}x \frac{\Gamma}{x}, \tag{C24}$$

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

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

The time integration can be approximated as

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

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

$$\int_{x_{\mathrm{kd}}}^{x_{\mathrm{rea}}} \mathrm{d}x \frac{\Gamma}{x} \approx 2\Gamma(x_{\mathrm{rea}}). \tag{C27}$$

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

In regions where reannihilation can only change the DM abundances by less than 1%, z_{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 $z_{\rm rea}$ 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 ρ_c by $\rho_{\rm dark}$ via Eq. (36), and when studying decaying DM we replace ρ_c via Eq. (56). In all cases, we take an effective neutrino mass m_ν into account in the time evolution of ρ_ν. We introduce a single massive eigenstate (minimum-mass normal hierarchy) such that the SM neutrino energy density evolves according to

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

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

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

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

$$\Omega_\gamma h^2 = 2.4728 \times 10^{-5}, \tag{D3}$$

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

$$m_\nu = 0.06\ \text{eV}, \tag{D4}$$

$$N_{\rm eff} = 3.046. \tag{D5}$$

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

$$\Omega_c h^2 = 0.1197 \pm 0.0022, \tag{D6}$$

$$\Omega_b h^2 = 0.02222 \pm 0.00023, \tag{D7}$$

and the derived parameters from the same analysis are given by

$$\Omega_\Lambda = 0.685 \pm 0.013, \tag{D8}$$

$$h = 0.6731 \pm 0.0096, \tag{D9}$$

$$z_* = 1090.09 \pm 0.42, \tag{D10}$$

$$z_{\rm drag} = 1059.57 \pm 0.46. \tag{D11}$$

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 $r_s(z_{\rm drag})$,

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

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

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

---

[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:astroph/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:astroph/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:astroph/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.

Chapter 5

Long-range interactions at finite temperature

The number density equations introduced so far have been constructed from a classical on-shell Boltzmann equation. All quantities entering the collision term describe the particle interactions in an idealized way. The transition amplitudes of certain processes are computed by means of standard QFT methods assuming the particle reactions are taking place under perfect vacuum conditions. The phase-space densities in the Boltzmann formalism are identical to the occupation densities of a noninteracting gas and the energy integration was carried out by treating all particles as on-shell, with a trivial vacuum dispersion relation.

Thermal quantum field theory allows to describe interacting particle systems beyond this idealized Boltzmann formalism. There is increasing interest in the application of thermal field theory techniques to study temperature-dependent corrections to the annihilation and production rates of dark matter in the early Universe. Especially the description of bound states inside a hot and dense plasma background is an interesting topic, since bound-state contributions can revive certain DM models by lifting the upper bound on the DM mass.

In this chapter, different approaches to describe models with attractive long-range interaction are compared. Section 5.1 first introduces a simplified model to study, which features the main thermal effects. Section 5.2 and 5.3 is a literature review of existing approaches. The main body of this thesis is presented in the article reprinted in section 5.4. Section 5.5 and 5.6 contain supplementary material to the article.

5.1 Simplified model

In the framework of thermal quantum field theory it is well known that gauge bosons like the photon or gluon acquire inside a hot and dense QED or QCD plasma a temperature dependent mass, called Debye mass, which screens the interaction between, e.g., fermionic particles in the plasma (see, e.g., [108, 109]). This quasiparticle phenomena can not be explained within the Boltzmann formalism. In analogy, the same effect would also occur if DM self-interacts via a long-range force which is tightly coupled to, e.g., other charged particles potentially present during the chemical decoupling process. Naively, the Debye mass contributes to the exponential screening of the Yukawa potential entering the two-body Schrödinger equation. Strong screening can lead to the effect that bound states, contributing to the modification of the DM overclosure bound in the Boltzmann framework, do not *exist* inside a medium.

The plasma effect as described above is only a simplified picture of the situation. There are indeed other corrections which can contribute to the effective potential. The main part of this thesis, which is the article reprinted in section 5.4, is about a systematic derivation of *i)* an effective potential for a non-relativistic particle-antiparticle pair interacting via a long-range force inside a hot and dense plasma environment and *ii)* a differential equation describing the chemical non-equilibrium transition of DM including bound states and finite temperature corrections.

To address both issues simultaneously, the authors work in the Keldysh-Schwinger non-equilibrium formalism [110, 111]. In this formalism, the central objects are averaged expectation values of time-ordered operator products along a closed time-path contour. The main idea is to derive from the equation of motion of those abstract objects a more general number density equation for DM which allows to describe long-range force enhanced annihilation and bound-state decay inside a hot and dense plasma environment. The closed time-path formalism itself is introduced in the dedicated part "Real-time formalism prerequisites" in the article reprinted in section 5.4. This introduction has been written by the author of this thesis. In the following, the status of literature is reviewed (based on different methods) in order to make a comparison and highlight the improvements achieved in section 5.4 later on.

To compare to the literature, the following simplified model is considered:

$$\mathcal{L} = \bar{\chi}\left(i\partial\!\!\!/ - M\right)\chi + g_\chi \bar{\chi}\gamma^\mu \chi A_\mu + \bar{\psi}\left(i\partial\!\!\!/ - m\right)\psi + g_\psi \bar{\psi}\gamma^\mu \psi A_\mu - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{5.1}$$

This model will be also adopted in the paper reprinted in section 5.4. Heavy fermionic dark matter is denoted by χ, the long-range force is mediated by an abelian vector boson A_μ, and light fermionic particles are denoted by ψ. All fermionic particles are of Dirac type. The vector boson and the light fermionic particles build the dense relativistic plasma background with temperature T. Inside this system, DM is assumed to be non-relativistic and sufficiently dilute, i.e. $T \ll E_\chi \sim m_\chi$ (as it is the case during the freeze-out process). A further property valid at the freeze-out is that the energy density of DM is much smaller compared to the background energy density. This simplifies the situation since one can neglect the impact of DM on the environment and considers only DM placed in an environment which does not depend on the DM subsystem.

The next two sections are a literature review on this subject. For comparison, results are translated into the model given in Eq. (5.1).

5.2 Two-particle effective in-medium potential

The system described in previous section is physically similar to a heavy quark-antiquark pair in a relativistic quark-gluon plasma produced in heavy ion collisions at colliders like the LHC. The bound states of heavy quark-antiquarks, like the J/ψ ($c\bar{c}$), can annihilate into, e.g., hard dileptons. The latter can escape from the quark-gluon plasma and are then measured in the

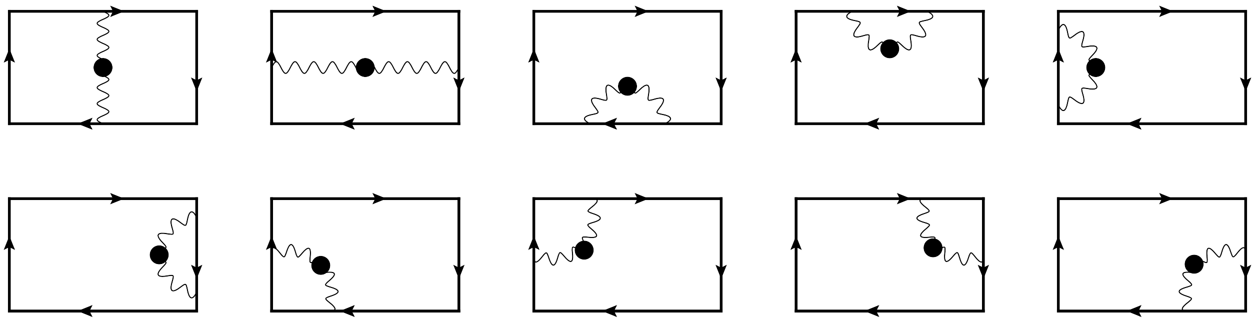


Fig. 4: The graphic illustrates the Wilson correlator C_E to order g_χ^2. Solid lines represent DM or Wilson lines and dots represent interacting "dark photon" correlators. Plot style resembles ref. [112].

detector. The amount of hard dileptons produced at a certain energy contains information of the thermal modification of the heavy quark pairs inside the medium.

In 1986, it was predicted in ref. [113] that the production of J/ψ is suppressed in nuclear collisions for certain energies due to the electric and colour Debye screening of the forces inside the plasma environment. The computation of the effective potential is based on a thermal two Polyakov loop (Wilson loop) correlator. However, it has been shown in the literature [114–117] that the potential at asymptotically large distances fails to give an expected Debye-screened potential (see also [112]).

In a seminal work by M. Laine et al. [112] in 2007, a static potential for heavy quark-antiquark in a quark gluon plasma has been derived based on a four line Wilson loop[5] correlator in Euclidean spacetime, given by

$$C_E(\tau,\mathbf{r}) \equiv \langle W[(0,\mathbf{r});(\tau,\mathbf{r})]W[(\tau,\mathbf{r});(\tau,\mathbf{0})]W[(\tau,\mathbf{0});(0,\mathbf{0})]W[(0,\mathbf{0}),(\tau,\mathbf{r})]\rangle, \tag{5.2}$$

where $\langle ...\rangle \equiv \mathrm{Tr}[\hat{\rho}\ (...)]$ denotes expectation value with density matrix $\hat{\rho} \propto e^{-\beta H}$ and H is the Hamiltonian of the system. The Euclidean Wilson line is given by[6]

$$W(\tilde{x},\tilde{y}) = \exp\left(ig_\chi \int_{\tilde{x}}^{\tilde{y}} A_\mu(\tilde{z})\mathrm{d}\tilde{z}_\mu\right). \tag{5.3}$$

The analytic continuation $\tau \to it$ of the Euclidean correlator C_E has a relation to another Minkowski time correlator. The latter is the actual quantity, which has a direct connection to the differential production rate of, e.g., di-muons (see [112]) from heavy quark pair annihilation. Here, we are only interested in the evaluation of C_E to extract the effective potential between heavy quarks.

For comparison, the results of the original work in ref. [112] are here translated into the simplified model considered in Eq. (5.1) and we have replaced gluons by "dark photon" A_μ, heavy quarks by DM χ, light quarks by ψ and adjust colour factors etc. The leading order term in the perturbative evaluation of Eq. (5.2) is of the order $\mathcal{O}(g_\chi^2)$ since $\langle 1\rangle = 1$ and a single insertion of A_μ has no contribution. All diagrams contributing at the leading order are shown in fig. 4. The dark photon two-point correlators are fully interacting and can be written in the imaginary time formalism in a resummed form as (see, e.g., [108, 109]):

---

[5] For introduction into Minkowski Wilson lines and their gauge invariance, see, e.g., [118].

[6] Euclidean coordinates and their Fourier transform follow the same convention as given in ref. [112].

$$\langle A_\mu(\tilde{x})A_\nu(\tilde{y})\rangle = \sum_{\tilde{q}_0}\int_{\mathbf{q}} e^{i\tilde{Q}(\tilde{x}-\tilde{y})}\left[\frac{P^T_{\mu\nu}(\tilde{Q})}{\tilde{Q}^2+\Pi_T(\tilde{Q})}+\frac{P^L_{\mu\nu}(\tilde{Q})}{\tilde{Q}^2+\Pi_L(\tilde{Q})}+\xi\frac{\tilde{q}_\mu\tilde{q}_\nu}{(\tilde{Q}^2)^2}\right]. \tag{5.4}$$

The sum over $\tilde{q}_0$ indicates summation over bosonic Matsubara frequencies. P, Π and ξ are projectors, one-loop self-energies and gauge parameter, respectively. Since DM (heavy quarks) are assumed to not influence the plasma, only the ψ particles (light quarks) contribute to the one-loop self-energies. Now, the important observation made in ref. [112] is that the gauge parameter ξ explicitly cancels out when summing over all contributions represented by the diagrams in fig. 4, leading to a gauge independent result. To further simplify the result, the self-energies are evaluated in the Hard Thermal Loop (HTL) approximation [119], which results in an effective theory for energies and momenta smaller compared to temperature. The effective potential in Minkowski spacetime is defined as $i\partial_t C(t,r) \equiv V(t,r)C(t,r)$, where C is the analytic continuation of C_E. To order $\mathcal{O}(g_\chi^2)$ this finally leads to the following static potential [112]:

$$\lim_{t\to\infty} V^{(2)}(t,r) = -\alpha_\chi\left[m_D+\frac{1}{r}e^{-m_D r}\right] - i\alpha_\chi T\phi(m_D r). \tag{5.5}$$

The effective in-medium potential for particle-antiparticle DM shows a real constant energy shift, an attractive screened Yukawa potential and an imaginary part which is related to Landau damping of a space-like dark photon inside the ψ plasma. The Debye screening mass for the DM model is here given by $m_D^2 = g_\psi^2 T^2/3$. The dimensionless function $\phi(x)$ has limiting values $\phi(0) = 0$ and $\phi(\infty) = 1$ and is monotonically increasing for increasing x. This potential has also been reproduced in other works [120, 121] later on.

The imaginary part introduces a non-trivial prediction regarding the energy spectra of, e.g., dilepton production from heavy quarkonium annihilation. The bound states inside a plasma acquire a thermal width which would lead to a smearing of the energy spectra of the dileptons. For high enough detector resolution this phenomena might be observable (see, e.g., [112]). Moreover, the imaginary part leads also to the fact, that bound states do not abruptly disappear from the two-particle spectrum for increasing temperature (as it would be the case for Debye screened potential only) but rather leads to a continuous *melting* of the bound-states poles [122].

5.3 Chemical equilibration rate

The effective in-medium potential introduced in the previous section determines the modification of the long-range force between heavy particles in a relativistic plasma background. Now the question is considered, how to supplement the Boltzmann equation by the modified in-medium annihilation rate. This is by far a non-trivial question since the imaginary part, which contributes to the potential, leads to "ill-defined" bound states due to the overlap with the continuum (see spectral functions in, e.g., [122]). A related difficulty is that in the heavy quark case it is assumed that they are in chemical equilibrium inside the quark gluon plasma. To describe the DM freeze-out, however, one has to find a differential equation for the number density which allows the DM to develop a finite chemical potential.

It has been argued in the literature that the standard Boltzmann equation for DM could be replaced by the following one [123]:

$$\dot{n}_\eta + 3Hn_\eta = -\frac{\Gamma_{\text{chem}}}{2n_\eta^{\text{eq}}}\left[n_\eta^2 - (n_\eta^{\text{eq}})^2\right]. \tag{5.6}$$

The quantity Γ_{chem} is called the chemical equilibration rate and is defined by

$$\Gamma_{\text{chem}} = \frac{\Omega_{\text{chem}}}{(4\chi_\eta/\beta)M^2}, \tag{5.7}$$

where Ω_{chem} is a transport coefficient and χ_η is the heavy DM number susceptibility, given in the DM dilute limit by $\chi_\eta = \beta n_\eta^{\text{eq}}$. The transport coefficient is related to a two-particle (DM-antiDM) spectral function ρ as:

$$\Omega_{\text{chem.}}/M^2 \simeq 16(\sigma v_{\text{rel}})_0 e^{-2\beta M}\int_{-\infty}^{\infty}\frac{\mathrm{d}^3\boldsymbol{P}}{(2\pi)^3}e^{-\beta\frac{\boldsymbol{P}^2}{4M}}\int_{-\infty}^{\infty}\frac{\mathrm{d}E}{(2\pi)}e^{-\beta E}\rho(E). \tag{5.8}$$

Here, the results of ref. [123, 124] have been translated to s-wave annihilation of the model considered in Eq. (5.1). The s-wave tree-level annihilation cross-section $(\sigma v_{\text{rel}})_0$ is a constant. It has been shown that the spectral function is related to a certain two-particle Green's function by $\rho(E) = 2\Im\left[i\lim_{\mathbf{r},\mathbf{r}'\to 0} G(\mathbf{0},\mathbf{0};E)|_{l=0}\right]$, where the Green's function value at the origin follows from the solution of a "Schrödinger-like" equation:

$$\left[\frac{\boldsymbol{\nabla}_{\mathbf{r}}^2}{M} + E - V^{(2)}(r)\right]G(\mathbf{r},\mathbf{r}';E) = 2i\delta^3(\mathbf{r}-\mathbf{r}'). \tag{5.9}$$

Furthermore, it was found in [122] that the static potential $V^{(2)}(r) \equiv \lim\limits_{t\to\infty} V^{(2)}(t,r)$ is the effective in-medium potential as introduced in the previous section. A nonperturbative solution of Eq. (5.9) for a complex in-medium potential was developed in [122] which is based on earlier findings, related to top quark pair production at the threshold [125].

All these results together set a new theoretical framework which allows to describe the chemical evolution of DM beyond the standard on-shell Boltzmann equation. Various effects of long-range forces can be taken into account within this formalism by only evaluating the two-particle spectral function. First of all, the resummation of Coulomb divergent ladder diagrams is taken into account by the non-perturbative solution of Eq. (5.9). This means that the Sommerfeld effect, leading to an enhanced annihilation cross-section, is contained in the solution of $\rho(E)$. Second, if the mass of the mediator is lighter than the inverse Bohr radius, the spectral function features a peak structure at certain negative energy values of E (binding energy), representing the bound-state solutions. And finally, the formalism allows to simultaneously include thermal corrections to the spectrum. Various DM models were found to be described within smaller modifications of this formalism [126–131].

It is mentioned that the formalism itself has limitations and is intrinsically only valid for energies $|E|$ smaller than temperature. This is a consequence of taking the HTL effective theory, which is only valid for energies and momenta (flowing inside the dark photon propagator) smaller compared to temperature as mentioned in the previous section.

Moreover, when taking the vacuum limit ($T \to 0$) in the effective potential $V^{(2)}$ one might expect that one should recover the coupled system of Boltzmann equations, provided in section 5.5. This seems somehow not to be the case if bound-state solutions exist in the spectral function.[7]

Bound states contribute to the spectral function as delta peaks, e.g. the spectral function is $\propto \delta(E - E_B)$ for a certain bound state in the vacuum limit. In the regime where the binding energy $|E_B|$ of a bound state is larger compared to temperature, the annihilation rate $\frac{\Gamma_{\text{chem}}}{2n_\eta^{\text{eq}}} sY$ would dominantly scale with temperature as

$$\frac{\Omega_{\text{chem.}}}{n_\eta^{\text{eq}} n_\eta^{\text{eq}} \beta^3} \propto \beta^{-3/2} e^{\beta |E_B|}. \tag{5.10}$$

This shows that the rate increases exponentially for decreasing temperature (not expected from the Boltzmann formalism). It might be due to the fact that the chemical equilibration rate in Eq. (5.8) is derived starting from a linearised Langevin equation [123] and is then inserted into the non-linear Boltzmann Eq. (5.6) “by hand”. The Langevin equation is a classical equation and the connection to quantum correlators is made via the correspondence principle (see pp. 199 in ref. [132]). It also has a noise term which is absent in the Boltzmann equation.

In the article reprinted in the next section, the authors derive the differential equation for the DM number density in a different way, starting from equation of motion of non-equilibrium Green’s function.

---

[7] However, in the *linear regime close the chemical equilibrium* the equation resulting from the vacuum limit is consistent.

5.4 *Publication:* Dark matter Sommerfeld-enhanced annihilation and bound-state decay at finite temperature

Reprinted article with permission from the authors:
Tobias Binder, Laura Covi, and Kyohei Mukaida.
Physical Review D **98**, 115023 (2018)
https://doi.org/10.1103/PhysRevD.98.115023
Copyright 2018 by the American Physical Society.

Author contributions L. C. had the idea of investigating finite temperature corrections to long-range force enhanced DM annihilation and founded the project. T. B. proposed as some part of the solution strategy to use non-equilibrium Green's function methods and the truncation procedure of the correlation function hierarchy. T. B. and K. M. did all analytic computations, partially in parallel using different methods while other parts were developed in strong collaboration. T. B. did all numerical computations as well as graphical illustrations independently. T. B. pointed out the importance of the degree of ionization leading to a non-trivial structure of the number density equation. Most of the article was written by T. B. and K. M. . All authors discussed the interpretation of the results and contributed to the revision of the manuscript.

PHYSICAL REVIEW D **98,** 115023 (2018)

Dark matter Sommerfeld-enhanced annihilation and bound-state decay at finite temperature

Tobias Binder,[1,*] Laura Covi,[1,†] and Kyohei Mukaida[2,‡]

[1] *Institute for Theoretical Physics, Georg-August University Göttingen, Friedrich-Hund-Platz 1, Göttingen, D-37077 Germany*

[2] *Deutsches Elektronen-Synchrotron (DESY), Notkestraße 85, Hamburg, D-22607 Germany*

(Received 5 September 2018; published 17 December 2018)

Traditional computations of the dark matter (DM) relic abundance, for models where attractive self-interactions are mediated by light force-carriers and bound states exist, rely on the solution of a coupled system of classical on-shell Boltzmann equations. This idealized description misses important thermal effects caused by the tight coupling among force-carriers and other charged relativistic species potentially present during the chemical decoupling process. We develop for the first time a comprehensive *ab initio* derivation for the description of DM long-range interactions in the presence of a hot and dense plasma background directly from nonequilibrium quantum field theory. Our results clarify a few conceptional aspects of the derivation and show that under certain conditions the finite temperature effects can lead to sizable modifications of DM Sommerfeld-enhanced annihilation and bound-state decay rates. In particular, the scattering and bound states get strongly mixed in the thermal plasma environment, representing a characteristic difference from a pure vacuum theory computation. The main result of this work is a novel differential equation for the DM number density, written down in a form which is manifestly independent under the choice of what one would interpret as a bound or a scattering state at finite temperature. The collision term, unifying the description of annihilation and bound-state decay, turns out to have in general a nonquadratic dependence on the DM number density. This generalizes the form of the conventional Lee-Weinberg equation which is typically adopted to describe the freeze-out process. We prove that our number density equation is consistent with previous literature results under certain limits. In the limit of vanishing finite temperature corrections our central equation is fully compatible with the classical on-shell Boltzmann equation treatment. So far, finite temperature corrected annihilation rates for long-range force systems have been estimated from a method relying on linear response theory. We prove consistency between the latter and our method in the linear regime close to chemical equilibrium.

DOI: 10.1103/PhysRevD.98.115023

I. INTRODUCTION

The cosmological standard model successfully describes the evolution of the large-scale structure of our Universe. It requires the existence of a cold and collisionless matter component, called dark matter (DM), which dominates over the baryon content in the matter dominated era of our Universe. The Planck satellite measurements of the cosmic microwave background (CMB) temperature anisotropies have nowadays determined the amount of dark matter to an unprecedented precision, reaching the level of subpercentage accuracy in the observational determination of the abundance when combining CMB and external data [1,2], e.g., measurements of the baryon acoustic oscillation.

Interestingly, astrophysical observation and structure formation on subgalactic scales might point toward the nature of dark matter as velocity-dependent self-interacting elementary particles. On the one hand, observations of galaxy cluster systems, where typical rotational velocities are of the order $v_0 \sim 1000$ km/s, set the most stringent bounds on the self-scattering cross section to be less than $\sigma/m_{\mathrm{DM}} \lesssim 0.7(0.1)\ \mathrm{cm}^2/\mathrm{g}$ in the bullet cluster [3] (in order to guarantee the production of elliptical halos [4,5]). On the other hand, a DM self-scattering cross section of the order $\sigma/m_{\mathrm{DM}} \sim 1\ \mathrm{cm}^2/\mathrm{g}$ on dwarf-galactic scales, where velocities are of the order $v_0 \sim 30$–100 km/s, would lead to a compelling solution of the cusp-core and diversity problem without strongly relying on uncertain assumptions of modeling the barionic feedbacks in simulations. This velocity dependence of the self-scattering cross section can naturally be realized in models where a light mediator

---

[*] tobias.binder@theorie.physik.uni-goettingen.de

[†] laura.covi@theorie.physik.uni-goettingen.de

[‡] kyohei.mukaida@desy.de

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP^3.

acts as a long-range force between the dark matter particles. For a recent review on self-interacting DM, see [6].

Generically, long-range forces can lead to sizable modifications of the DM tree-level annihilation cross section in the regime where the annihilating particles are slow. For the most appealing DM candidates, known as weakly interacting massive particles (WIMP) dark matter [7–9], such that the relic abundance in the early Universe is set by the thermal freeze-out mechanism when the DM is nonrelativistic, these effects can be sizable already at the time of chemical decoupling. Then the inclusion of the long-range force modification in the computation of the relic abundance is necessary to reach the required level of the accuracy set by the Planck precision measurement [1,2]. If the light mediators induce an attractive force between the annihilating DM particles, the total cross section is typically enhanced [10,11] which is often referred to as the Sommerfeld enhancement [12] or Sakharov enhancement [13]. Additionally, such attractive forces can lead to the existence of DM bound-state solutions [14–16]. This opens the possibility for conversion processes between scattering and bound states via radiative processes, influencing the evolution of the abundance of the stable scattering states during the DM thermal history. DM scenarios with Sommerfeld enhancement or with bound states have been widely studied in the literature [17–59] and it has been shown that the main effect of such corrections is to shift in the parameter space the upper bounds on the DM mass, otherwise the theoretically predicted DM density would be too large (overclosure bound).

Classic WIMP candidates with large corrections via Sommerfeld enhancement or bound states are particles charged under the electroweak interactions, like the Wino neutralino in supersymmetric models [60] or the first Kaluza-Klein excitation of the gauge boson in models with extra dimensions [61]. For the supersymmetric case it was realized very early on by [10,11] that the Sommerfeld effect reduces the Wino density up to 30% and pushes the mass of the Wino dark matter candidate to few TeVs in order to obtain the correct relic density. These studies have later been extended to the case of general components of the neutralino [29,30]. Similar and even stronger effects from the Sommerfeld enhancement and bound states were found in the case of coannihilation of the WIMP particle with charged or colored states [31–41]. If the electroweakly charged dark matter is sufficiently heavy, the Sommerfeld enhancement or the presence of bound states due to the exchange of electroweak gauge or Higgs bosons, see e.g., [42,43], are very generic as it was shown e.g., in the minimal dark matter model [40,44] and in Higgs portal models [45]. In these cases, long-range force effects play an important role also for the indirect detection limits [14,15,46–48] and especially for the Wino the Sommerfeld enhancement has lead to the exclusion of most parameter space [49–52]. Note that this effect can be important also when the dark matter is not itself a WIMP, but it is produced by WIMP decay out of equilibrium, like in the super-WIMP mechanism [62,63]. Indeed in such a scenario, the DM inherits part of the energy density of the mother particle and so any change in the latter freeze-out density is directly transferred to the superweakly interacting DM and can relax the BBN constraints on the mother particle [53].

While a lot of effort has been made to compute quantitatively the effects of a long-range force on the DM relic density employing the classical Boltzmann equation method, it is still unclear if that is a sufficient description. Indeed, considering the presence of a thermal plasma on the long-range force leads on one side to a possible screening by the presence of a thermal mass, or on the other to the issue if the coherence in the (in principle infinite) ladder diagram exchanges between the two slowly moving annihilating particles is guaranteed. Moreover, from a conceptional point of view, there is yet no consistent formulation in the existing literature of how to deal with long-range forces *at finite temperature*, especially if the dark matter is, or, enters an out-of-equilibrium state (already the standard freeze-out scenario is a transition from chemical equilibrium to out-of-chemical equilibrium). The main concern of our work will be to clarify conceptional aspects of the derivation and the solution of the number density equation for DM particles with attractive long-range force interactions in the presence of a hot and dense plasma background, starting from first principles. From the beginning, we work in the *real-time formalism*, which has a smooth connection to generic out-of-equilibrium phenomena.

The simplified DM system we would like to describe in the presence of a thermal environment is similar to heavy quarks in a hot quark gluon plasma. For this setup it has been shown that finite temperature effects can lead to a melting of the heavy-quark bound states which influences, e.g., the annihilation rate of the heavy quark pair into dilepton [64]. For DM or heavy quark systems, the Sommerfeld enhancement at finite temperature has been discussed in the literature, where the chemical equilibration rate is (i) estimated from *linear response theory* [37,65,66] and (ii) based on *classical* rate arguments [67], is then inserted into the *nonlinear* Boltzmann equation for the DM number density "by hand." Relying on those estimates, it has been shown that the overclosure bound of the DM mass can be strongly affected by the melting of bound states in a plasma environment [55–57]. However, strictly speaking, the linear response theory method is only applicable if the system is close to chemical equilibrium. Indeed the computation has been done in the *imaginary-time formalism* so far, capturing the physics of thermal equilibrium. One of our goals is to obtain a more general result directly from the *real-time* formalism, valid as well for systems way out of equilibrium.

Most of the necessary basics of the real-time method we use are provided in Sec. II as a short review of out-of-equilibrium quantum field theory. Within this mathematical framework, an exact expression for the DM number density equation of our system is derived in Sec. III, where the Sommerfeld-enhanced annihilation or decay rate at finite temperature can be computed from a certain component of a four-point correlation function. We derive the equation of motion for the four-point correlation function on the real-time contour in Sec. IV which becomes in its truncated form a Bethe-Salpeter type of equation. Since we close the correlation functions hierarchy by truncation, the system of coupled equations we have to solve contains only terms with DM two- and four-point functions. In Sec. V, we derive a simple semianalytical solution of the four-point correlator under certain assumptions valid for WIMP-like freeze-outs. Our result does not rely on linear response theory and it is therefore quite general applying also in case of large deviations from chemical equilibrium. The limit of vanishing finite temperature corrections is taken in Sec. VI, showing the consistency between our general results and the classical Boltzmann equation treatment. Here, we also compare to the linear response theory method and clarify the assumptions needed to reproduce those results. Our main numerical results for the finite temperature case are given in Sec. VII, both for a gauge theory and for a Yukawa potential, and discussed in detail in Sec. VIII. Finally, we conclude in Sec. IX.

II. REAL-TIME FORMALISM PREREQUISITES

The generalization of quantum field theory on the *closed-time-path* (CTP) contour, or *real-time formalism*, is a mathematical method which allows one to describe the dynamics of quantum systems out of equilibrium. Prominent applications are systems on curved space-time and/or systems having a finite temperature. In this work, we assume that the equilibration of DM in the early Universe is a fast process, and consequently, the initial memory effects before the freeze-out process can be ignored. This leads to the fact that the adiabatic assumption for such a system is an excellent approximation, motivating us to take the Keldysh-Schwinger prescription[1] of the CTP contour, as illustrated in Fig. 1. The time contour $\mathcal{C}$ in the Keldysh-Schwinger prescription consists of two branches denoted by τ^+ and τ^-. The upper time contour τ^+ ranges from the initial time $t_i = -\infty$ to $t = \infty$ while the lower contour τ^- is considered to go from ∞ back to $-\infty$. Therefore, times on the τ^- branch are said to be always later compared to the times on τ^+. The time ordering of operator products on $\mathcal{C}$ can generically be written as

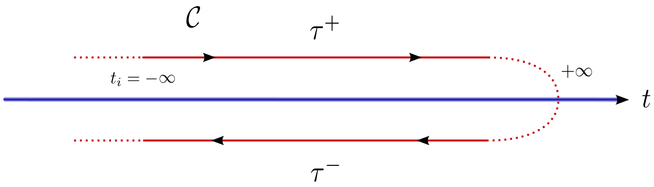


FIG. 1. Keldysh-Schwinger approximation of the closed-time-path contour $\mathcal{C}$, consisting of two time branches τ^+ and τ^-.

$$\begin{aligned} T_{\mathcal{C}}[O_1(t_1)\ldots O_n(t_n)] &\equiv \sum_P (-1)^{F(p)} \theta_{\mathcal{C}}(t_{p(1)}, t_{p(2)})\ldots\theta_{\mathcal{C}}(t_{p(n-1)}, t_{p(n)}) \\ &\quad\times O_{p(1)}(t_{p(1)})\ldots O_{p(n)}(t_{p(n)}), \end{aligned} \tag{2.1}$$

where the sum is over all the permutations P of the set of operators O_i and $F(p)$ is the number of permutations of fermionic operators. The unit step function and the delta distribution on the Keldysh-Schwinger contour is defined as

$$\begin{aligned} \theta_{\mathcal{C}}(t_1, t_2) &\equiv \begin{cases} \theta(t_1 - t_2) & \text{if } t_1, t_2 \in \tau^+ \\ \theta(t_2 - t_1) & \text{if } t_1, t_2 \in \tau^- \\ 1 & \text{if } t_1 \in \tau^-, t_2 \in \tau^+ \\ 0 & \text{if } t_1 \in \tau^+, t_2 \in \tau^- \end{cases} \\ &\hat{=} \begin{pmatrix} \theta(t_1 - t_2) & 0 \\ 1 & \theta(t_2 - t_1) \end{pmatrix}, \\ \delta_{\mathcal{C}}(t_1, t_2) &\hat{=} \begin{pmatrix} \delta(t_1 - t_2) & 0 \\ 0 & -\delta(t_1 - t_2) \end{pmatrix}. \end{aligned} \tag{2.2}$$

Correlation functions, i.e., contour $\mathcal{C}$-ordered operator products averaged over all states where the weight is the density matrix of the system denoted by $\hat{\rho}$, are defined by

$$\langle T_{\mathcal{C}} O(x_1, x_2, \ldots, x_n)\rangle \equiv \mathrm{Tr}[\hat{\rho} T_{\mathcal{C}} O(x_1, x_2, \ldots, x_n)]. \tag{2.3}$$

Let us introduce commonly used notations and properties of two-point correlation functions of fermionic or bosonic operator pairs relevant for this work. Because of the two-time structure, there are four possibilities to align the times x^0 and y^0 on $\mathcal{C}$ and hence four different components of a general two-point function denoted by $G(x, y)$, where in matrix notation it can be written as

$$\begin{aligned} G(x, y) &\equiv \langle T_{\mathcal{C}} \psi(x)\psi^\dagger(y)\rangle \\ &= \theta_{\mathcal{C}}(x^0, y^0)\langle\psi(x)\psi^\dagger(y)\rangle \mp \theta_{\mathcal{C}}(y^0, x^0)\langle\psi^\dagger(y)\psi(x)\rangle \\ &\hat{=} \begin{pmatrix} G^{++}(x, y) & G^{+-}(x, y) \\ G^{-+}(x, y) & G^{--}(x, y) \end{pmatrix}. \end{aligned} \tag{2.4}$$

---

[1] In ordinary QFT the initial vacuum state Ω appearing in correlation functions $\langle\Omega|_{\mathrm{in}} T[O(x, \ldots)]|\Omega\rangle_{\mathrm{in}}$ is equivalent up to a phase to the final vacuum state. For this special situation the operators are ordered along the "flat" time axis ranging from $t_{\mathrm{in}} = -\infty$ to $t_{\mathrm{out}} = \infty$. By means of Lehmann-Symanzik-Zimmermann (LSZ) reduction formula it is then possible to relate correlation functions to the S matrix and compute cross sections. This in-out formalism breaks down once, e.g., the initial vacuum is not equivalent to the final state vacuum. An expanding background or external sources can introduce such a time dependence. In our work, there are mainly two sources of breaking the time translation invariance. First, since we have a thermal population, we consider traces of time-ordered operator products, where the trace is taken over all possible states. The many particle states are in general time dependent. Second, we have a density matrix next to the time ordering. The CTP, or, in-in formalism we adopt in this work can be, pragmatically speaking, seen as just a mathematical way of how to deal with such more general expectation values. The Keldysh description of the CTP contour applies if initial correlations can be neglected and we refer for a more detailed discussion and limitation to [68].

Here, $G^{\sigma_x\sigma_y}$ means $x^0 \in \tau^{\sigma_x}$ and $y^0 \in \tau^{\sigma_y}$ with $\sigma_i = \pm$ for $i = x, y$ and the four different components of $G(x,y)$ are defined as

$$G^{-+}(x,y) \equiv \langle \psi(x)\psi^\dagger(y)\rangle, \tag{2.5}$$

$$G^{+-}(x,y) \equiv \mp \langle \psi^\dagger(y)\psi(x)\rangle, \tag{2.6}$$

$$G^{++}(x,y) \equiv \theta(x^0 - y^0)G^{-+}(x,y) + \theta(y^0 - x^0)G^{+-}(x,y), \tag{2.7}$$

$$G^{--}(x,y) \equiv \theta(x^0 - y^0)G^{+-}(x,y) + \theta(y^0 - x^0)G^{-+}(x,y), \tag{2.8}$$

where $-(+)$ on the r.h.s. of the second line applies for fermionic (bosonic) field operators. From these definitions one can recognize that not all components are independent, namely the following relation holds:

$$G^{++}(x,y) + G^{--}(x,y) = G^{+-}(x,y) + G^{-+}(x,y). \tag{2.9}$$

Furthermore, let us introduce *retarded* and *advanced* correlators defined by

$$\begin{aligned} G^R(x,y) &\equiv \theta(x^0 - y^0)[G^{-+}(x,y) - G^{+-}(x,y)] \\ &= G^{++}(x,y) - G^{+-}(x,y) \\ &= -G^{--}(x,y) + G^{-+}(x,y), \end{aligned} \tag{2.10}$$

$$\begin{aligned} G^A(x,y) &\equiv -\theta(y^0 - x^0)[G^{-+}(x,y) - G^{+-}(x,y)] \\ &= G^{++}(x,y) - G^{-+}(x,y) \\ &= -G^{--}(x,y) + G^{+-}(x,y), \end{aligned} \tag{2.11}$$

as well as the *spectral function* given by

$$G^\rho(x,y) \equiv G^R(x,y) - G^A(x,y) = G^{-+}(x,y) - G^{+-}(x,y). \tag{2.12}$$

From these definitions we can derive further useful properties:

$$\begin{aligned} G^A(x,y) &= -[G^R(y,x)]^\dagger, \\ G^{+-}(x,y) &= [G^{+-}(y,x)]^\dagger, \\ G^{-+}(x,y) &= [G^{-+}(y,x)]^\dagger. \end{aligned} \tag{2.13}$$

In the case of free (unperturbed) propagators G_0, the following *semigroup* property holds:

$$G_0^R(x,y) = \int \mathrm{d}^3z G_0^R(x,z)G_0^R(z,y), \quad \text{for } t_x > t_z > t_y. \tag{2.14}$$

This equality can be verified by noticing that for those time configurations the correlators are proportional to on-shell plane waves in Fourier space. Note that there is no time integration in the above equation. Together with the relations in Eq. (2.13) further semigroup properties can be derived and all important ones are summarized for the use in subsequent sections in Appendix A.

As an example, the time integration over the Schwinger-Keldysh contour $\mathcal{C}$ of products of correlators can be written as

$$\begin{aligned} C^{++}(x,y) &= \left[\int_{z^0\in\mathcal{C}} \mathrm{d}z^0 \int \mathrm{d}^3z A(x,z)B(z,y)\right]_{x^0,y^0\in\tau^+} \\ &= \left[\left(\int_{\tau^+} \mathrm{d}z^0 + \int_{\tau^-} \mathrm{d}z^0\right) \int \mathrm{d}^3z A(x,z)B(z,y)\right]_{x^0,y^0\in\tau^+} \\ &= \int_{-\infty}^{\infty} \mathrm{d}z^0 \int \mathrm{d}^3z A^{++}(x,z)B^{++}(z,y) + \int_{\infty}^{-\infty} \mathrm{d}z^0 \int \mathrm{d}^3z A^{+-}(x,z)B^{-+}(z,y) \\ &= \int_{-\infty}^{\infty} \mathrm{d}^4z (A^{++}(x,z)B^{++}(z,y) - A^{+-}(x,z)B^{-+}(z,y)). \end{aligned} \tag{2.15}$$

Equation (2.15) is called a *Lagereth rule* and it is straightforward to work out similar rules for, e.g., different components or double integrations of higher-order products of Keldysh-Schwinger correlators as they will appear later in this work. Let us move on and define *Wigner coordinates* according to

$$T = (x^0 + y^0)/2, \qquad t = x^0 - y^0, \tag{2.16}$$

$$\mathbf{R} = (\mathbf{x} + \mathbf{y})/2, \qquad \mathbf{r} = \mathbf{x} - \mathbf{y}. \tag{2.17}$$

In the second line all variables are 3-vectors. The Wigner-transformed correlators are defined as

$$\begin{aligned} \tilde{G}(t,\mathbf{r},\mathbf{R},T) &\equiv G(T + t/2, \mathbf{R} + \mathbf{r}/2, T - t/2, \mathbf{R} - \mathbf{r}/2) \\ &= G(x^0\mathbf{x}, y^0\mathbf{y}). \end{aligned} \tag{2.18}$$

In all computations, the tilde will be dropped such that we can write for the Fourier transformation of $G(x, y)$:

$$G(\omega, \mathbf{p}, \mathbf{R}, T) = \int \mathrm{d}t\mathrm{d}^3 r e^{i(\omega t - \mathbf{p}\cdot\mathbf{r})} G(t, \mathbf{r}, \mathbf{R}, T). \tag{2.19}$$

One of the great advantages of separating microscopic (t, $\mathbf{r}$) and macroscopic (T, $\mathbf{R}$) variables according to the Wigner transformation is that Fourier transformations of integral expressions can be considerably simplified by using the *gradient expansion*. For example, Eq. (2.15) in Fourier space can be written as

$$\begin{aligned} C^{++}(\omega, \mathbf{p}, \mathbf{R}, T) &= A^{++}(\omega, \mathbf{p}, \mathbf{R}, T)\mathcal{G}_{AB}B^{++}(\omega, \mathbf{p}, \mathbf{R}, T) - A^{+-}(\omega, \mathbf{p}, \mathbf{R}, T)\mathcal{G}_{AB}B^{-+}(\omega, \mathbf{p}, \mathbf{R}, T) \\ &\simeq A^{++}(\omega, \mathbf{p}, \mathbf{R}, T)B^{++}(\omega, \mathbf{p}, \mathbf{R}, T) - A^{+-}(\omega, \mathbf{p}, \mathbf{R}, T)B^{-+}(\omega, \mathbf{p}, \mathbf{R}, T), \end{aligned} \tag{2.20}$$

$$\mathcal{G}_{AB} \equiv e^{-i(\partial_T^A \partial_\omega^B - \partial_\omega^A \partial_T^B - \nabla_{\mathbf{R}}^A \cdot \nabla_{\mathbf{p}}^B + \nabla_{\mathbf{p}}^A \cdot \nabla_{\mathbf{R}}^B)/2} \simeq 1. \tag{2.21}$$

Throughout this work, such exponentials containing derivatives are always approximated as in the last line, defining the leading order term of the gradient expansion. For homogeneous and isotropic systems the correlators do not depend on R and thus for the spatial part the leading order term is exact. For a discussion of the validity of the leading order term of the temporal part we refer to [69]. Let us emphasize here, that for typical DM scenarios the leading order term is always assumed to be a very good approximation.

Next, important properties of two-point correlators in thermal equilibrium are provided. Under this assumption, different components of correlators become related which are otherwise independent. For a system being in equilibrium (here we consider kinetic as well as chemical equilibrium), the density matrix takes the form

$$\hat{\rho} \propto e^{-\beta H}, \tag{2.22}$$

where H is the Hamiltonian of the system and β factor is the inverse temperature T of the system. The density matrix in thermal equilibrium can be formally seen as a time evolution operator, where the inverse temperature is regarded as an evolution in the imaginary time direction. Making use of the cyclic property of the trace it can be shown that under the assumption of equilibrium the components are related via

$$G^{-+}(x^0 - y^0) = \mp\, G^{+-}(x^0 - y^0 + i\beta). \tag{2.23}$$

This important property is called the Kubo-Martin-Schwinger (KMS) relation, where $-$ ($+$) applies for two-point correlators of fermionic (bosonic) operators. Furthermore, in equilibrium the correlators should depend only on the difference of the time variables due to time translation invariance. Consequently, the KMS condition in Fourier space reads

$$G^{-+}(\omega, \mathbf{p}) = \mp e^{\beta\omega} G^{+-}(\omega, \mathbf{p}). \tag{2.24}$$

From this condition, various equilibrium relations follow:

$$\begin{aligned} G^{+-}(\omega, \mathbf{p}) &= \mp n_{\mathrm{F/B}}(\omega) G^{\rho}(\omega, \mathbf{p}), \\ G^{-+}(\omega, \mathbf{p}) &= [1 \mp n_{\mathrm{F/B}}(\omega)] G^{\rho}(\omega, \mathbf{p}), \end{aligned} \tag{2.25}$$

$$G^{++}(\omega, p) = \frac{G^R(\omega, \mathbf{p}) + G^A(\omega, \mathbf{p})}{2} + \left[\frac{1}{2} \mp n_{\mathrm{F/B}}(\omega)\right] G^{\rho}(\omega, \mathbf{p}), \tag{2.26}$$

where the Fermi-Dirac or Bose-Einstein phase-space densities are given by $n_{\mathrm{F/B}}(\omega) = 1/(e^{\beta\omega} \pm 1)$. Thus in equilibrium, all correlator components can be calculated from the retarded/advanced components, where the spectral function G^{ρ} is related to those via Eq. (2.12).

General out-of-equilibrium observables, like the dynamic of the number density or spectral information of the system, can be directly inferred from the equation of motions (EoM) of the corresponding correlators. Throughout this work, we derive the correlator EoM from the invariance principle of the path integral measure under infinitesimal perturbations of the fields. The equivalence of CTP correlators and the path integral formulation is given by

$$\langle T_{\mathcal{C}} O(x_1, x_2, \ldots, x_n) \rangle = \int \mathrm{d}\mu_i \rho(\mu_i) \int_{\mu_i} [\mathrm{d}\mu] O(x_1, x_2, \ldots, x_n) \times e^{i \int_{x \in \mathcal{C}} \mathcal{L}(x)} \tag{2.27}$$

and the action on the *CPT* contour is $S = \int_{x\in\mathcal{C}} \mathcal{L}(x)$. μ collectively represents the fields, and ρ stands for a state that could be either pure or mixed, as in Eq. (2.3). The second integral in Eq. (2.27) is a path integral with a boundary condition of μ_i at the initial time t_i that we take to $-\infty$ in the Schwinger-Keldysh prescription of the contour, and the first one takes the average of μ_i with the weight of $\rho(\mu_i)$. Now, to derive the EoM for two-point correlators, let us consider an infinitesimal perturbation $O'^{\dagger} = O^{\dagger} + \epsilon$, satisfying $\epsilon(t_i) = 0$. By relying on the measure-invariance principle under this transformation, one obtains the EoM of the two-point correlators from

$$\langle O'^{\dagger}(y)\rangle = \langle O^{\dagger}(y)\rangle + \epsilon(x) + i\int_{x\in\mathcal{C}} \epsilon(x)\left\langle T_{\mathcal{C}}\frac{\delta S}{\delta_{\mathrm{L}} O^{\dagger}(x)} O^{\dagger}(y)\right\rangle + \mathcal{O}(\epsilon^2) \tag{2.28}$$

$$\Rightarrow 0 = \delta_{\mathcal{C}}(x,y) + i\left\langle T_{\mathcal{C}}\frac{\delta S}{\delta_{\mathrm{L}} O^{\dagger}(x)} O^{\dagger}(y)\right\rangle, \tag{2.29}$$

where δ_{L} represents a derivative acting from the left. The same procedure can be applied to the case of O, as well as for deriving the EoM of higher correlation functions. The relation between the abstract EoM of correlators and differential equations for observables will be part of the next section. In general, a correlator EoM depends on higher and lower correlators which is called the Martin-Schwinger hierarchy. For systems where the coupling expansion is appropriate, it might be sufficient to work in the one-particle self-energy approximation, where the EoM are closed in terms of two-point functions, and the kinetic equations can systematically be obtained by expanding the DM self-energy perturbatively in the coupling constant. The kinetic equations of two-point functions in the self-energy approximation are also known as the Kadanoff-Baym equations. For example, at next-to-leading order (NLO) in the self-energy expansion of the two-point correlators the standard Boltzmann equation is recovered.

Finite temperature corrections to nonperturbative systems, e.g., Sommerfeld-enhanced DM annihilations or bound-state decays, where a subclass of higher-order diagrams becomes comparable to the leading order (LO) in vacuum, are less understood in the CTP formalism. The strategy in the next section will be to address this issue by going beyond the one-particle self-energy approximation [69]. More precisely, we derive the exact Martin-Schwinger hierarchy of our particle setup in the CTP formalism by using Eq. (2.29) and truncate the hierarchy at the six-point function level. The system of equations is then closed with respect to two- *and four-point* functions. This approach allows us to account for the resummation of the Coulomb diagrams and their finite temperature corrections at the same time. Furthermore, we show how to extract the DM number density equation from the EoM of two-point functions and that it depends on the four-point correlator. The complication is that the differential equation of the four-point correlator is coupled to the two-point correlator and in subsequent sections we solve this coupled systems of equations.

III. EQUATIONS OF MOTION IN REAL-TIME FORMALISM FOR NONRELATIVISTIC PARTICLES IN A RELATIVISTIC PLASMA ENVIRONMENT

Throughout this paper, we consider the following minimal scenario capturing important effects to study long-range force enhanced DM annihilations and bound states under the influence of a hot and dense plasma environment:

$$\mathcal{L} = \bar{\chi}(i\not{\partial} - M)\chi + g_{\chi}\bar{\chi}\gamma^{\mu}\chi A_{\mu} + \bar{\psi}(i\not{\partial} - m)\psi + g_{\psi}\bar{\psi}\gamma^{\mu}\psi A_{\mu} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{3.1}$$

The particles of the equilibrated plasma environment with temperature T are the Abelian mediators A_{μ} and the light fermionic particles ψ with mass $m \ll T$. Fermionic DM χ is assumed to be nonrelativistic, i.e., $M \gg T$. All fermionic particles are considered to be of Dirac type. We assume the mediator to be massless; however, we provide the final results also for the case of a massive A_{μ} with mass $m_V \ll M$.

Let us illustrate how we can get the DM nonrelativistic effective action in the thermal medium of light particles. It is obtained in two steps. First, hard modes of $p \gtrsim M$ are integrated out. In this limit, the DM four-component spinor χ splits into two parts, a term for the particle denoted by the two-component spinor η and a term for the antiparticle denoted by ξ. Second, we assume that DM does not influence the plasma environment during the freeze-out process. This is typically the case since the DM energy densities are smaller than that of light particles at this epoch. And thus, we may also integrate out the plasma fields by assuming they remain in thermal equilibrium. The resulting effective action on the CTP contour for particle η and anti-particle ξ DM is given by

$$S_{\mathrm{NR}}[\eta,\xi] = \int_{x\in\mathcal{C}} \eta^{\dagger}(x)\left[i\partial_t + \frac{\nabla^2}{2M}\right]\eta(x) + \xi^{\dagger}(x)\left[i\partial_t - \frac{\nabla^2}{2M}\right]\xi(x) + \int_{x,y\in\mathcal{C}} i\frac{g_{\chi}^2}{2}J(x)D(x,y)J(y) + iO_s^{\dagger}(x)\Gamma_s(x,y)O_s(y). \tag{3.2}$$

Dark matter long-range force interactions are encoded in the first term of the interactions in Eq. (3.2). This term includes the current and the full two-point correlator of the electric potential on the CTP contour which are defined by

$$\begin{aligned} J(x) &\equiv \eta^{\dagger}(x)\eta(x) + \xi^{\dagger}(x)\xi(x), \\ D(x,y) &\equiv \langle T_{\mathcal{C}}A_0(x)A_0(y)\rangle, \end{aligned} \tag{3.3}$$

respectively. The last term in Eq. (3.2) describes the annihilation part and we only keep the s-wave contribution

$$\begin{aligned} O_s(x) &\equiv \xi^{\dagger}(x)\eta(x), \\ \Gamma_s(x,y) &\equiv \frac{\pi(\alpha_{\chi}^2 + \alpha_{\chi}\alpha_{\psi})}{M^2}\begin{pmatrix} \delta^4(x-y) & 0 \\ 2\delta^4(x-y) & \delta^4(x-y) \end{pmatrix}, \end{aligned} \tag{3.4}$$

with the fine-structure constant being $\alpha_i \equiv g_i^2/4\pi$, and summation over the spin indices are implicit. Γ_s is shown in the matrix representation of the CTP formalism; see e.g., Eq. (2.4) in previous section. Hence the delta functions on the right-hand side are defined on the usual real-time axis.

Similar to the vacuum theory, the annihilation part Γ_s can be computed by cutting the box diagram (containing two A_μ) and the vacuum polarization diagram (containing one A_μ and a loop of light fermions ψ), where now all propagators are defined on the *CPT* contour. Finite temperature corrections to these hard processes in Γ_s are neglected[2] and for a derivation we refer to Appendix B.

In our effective action Eq. (3.2), we have discarded higher order terms in ∇/M (like magnetic interactions) and also interactions with ultrasoft gauge bosons,[3] since we focus on threshold singularities of annihilations at the leading order in the coupling g_χ and the velocity ∇/M. Furthermore, our effective field theory is non-Hermitian because we have integrated out (or traced out) hard and thermal degrees of freedom (d.o.f.). The first source of non-Hermite nature is the annihilation term which originates from the integration of hard d.o.f. A similar term would also be present in vacuum [10,11,46] and belongs to the ++ component of Γ. Thus, as a first result we have generalized the annihilation term towards the CTP contour. Another one stems from the gauge boson propagator D that encodes interactions with the thermal plasma. While the annihilation term containing Γ_s in our action breaks the number conservation of DM, the interaction term containing D can not. From this observation, one may anticipate that the non-Hermitian potential contributions of the gauge boson propagator never lead to a violation of the DM particle or antiparticle number conservation. Later, we will show this property directly from the EoM, respecting the global symmetries of our action.

In the next sections we proceed as in the following. First, we compute the finite temperature one-loop corrections contained in the potential term D explicitly. Since the number density of DM becomes Boltzmann suppressed in the nonrelativistic regime of the freeze-out process, the dominant thermal loop contributions arise from the relativistic species ψ. This implies that we can solve for D independently of the DM system since we assume DM does not modify the property of the plasma. The correction terms for the DM self-interactions are screening effects on the electric potential, as well as imaginary contributions arising from soft DM-ψ scatterings, derived and discussed in detail in Sec. III A. Second, in Sec. III B the kinetic equations for the DM correlators are derived. We show how to extract from these equations the number density equation, including finite temperature corrected processes for the negative energy spectrum (bound-state decays) as well as for the positive energy contribution (Sommerfeld-enhanced annihilation) in one single equation.

A. Thermal corrections to potential term

In this section, we briefly summarize how the electric component of the mediator propagator D gets modified by the thermal presence of ultrarelativistic ψ fields. The plasma environment is regarded to be perturbative and in the one-particle self-energy framework we can write down the Dyson equation on the Schwinger-Keldysh contour for the mediator:

$$D_{\mu\nu}(x,y)=D^0_{\mu\nu}(x,y)-i\int_{w,z\in\mathcal{C}}D^0_{\mu\alpha}(x,w)\Pi^{\alpha\beta}(w,z)D_{\beta\nu}(z,y), \tag{3.5}$$

$$\Pi^{\mu\nu}(x,y)=(-i)g_\psi^2(-1)\mathrm{Tr}[\gamma^\mu S(x,y)\gamma^\nu S(y,x)], \tag{3.6}$$

where $S(x,y)\equiv\langle T_{\mathcal{C}}\psi(x)\bar{\psi}(y)\rangle_0$ are unperturbed ψ correlators. A_μ and ψ are assumed to be in equilibrium and thus, according to the discussion in Sec. II, we only need to compute the retarded/advanced propagators. From those, we can construct all other components by using the KMS condition. The Dyson equation for retarded (advanced) mediator-correlator can be obtained by subtracting the $+-$ $(-+)$ component of Eq. (3.5) from the $++$ component of the same equation. In momentum space this results in

$$D^{R/A}_{\mu\nu}=(D^{R/A}_{\mu\nu})_0-i(D^{R/A}_{\mu\alpha})_0\Pi^{\alpha\beta}_{R/A}D^{R/A}_{\beta\nu}, \tag{3.7}$$

where the mediator's retarded self-energy is defined as

$$\begin{aligned}\Pi^{\mu\nu}_R(P)&=\Pi^{\mu\nu}_{++}-\Pi^{\mu\nu}_{+-}\\&=ig_\psi^2\int\frac{d^4K}{(2\pi)^4}(\mathrm{Tr}[\gamma^\mu S^{++}(K-P)\gamma^\nu S^{++}(K)]\\&\quad-\mathrm{Tr}[\gamma^\mu S^{+-}(K-P)\gamma^\nu S^{-+}(K)]).\end{aligned}\tag{3.8}$$

A sketch of efficiently calculating the thermal one-loop Eq. (3.8) is provided in Appendix C. In the computation we utilize the *hard thermal loop* (HTL) approximation [71] to extract leading thermal corrections.[4] In the HTL approximation we are allowed to resum the self-energy contributions of the retarded/advanced component and the result is gauge independent. We work in the noncovariant Coulomb gauge which is known to be fine at finite temperature since Lorentz invariance is anyway broken by the plasma temperature. We find for the dressed longitudinal component $\mu = \nu = 0$ of the mediator propagator in the HTL resummed self-energy approximation and in the Coulomb gauge the following:

---

[2] Assuming free plasma field correlators in the computation of Γ_s is a good approximation since the energy scale of the hard process is $\sim M$ which is much larger for nonrelativistic particles than typical finite temperature corrections being of the order $\sim g_\psi T$. Consequently, the dominant thermal corrections should be in the modification of the long-range force correlator D, where the typical DM momentum-exchange scale enters which is much lower compared to the annihilation scale.

[3] To fully study the out-of-equilibrium dynamics of the bound-state formation [21,22,42,70] at late times of the freeze-out process, it is necessary to include emission and absorption via ultrasoft gauge bosons, e.g., via an electric dipole operator. We drop for simplicity ultrasoft contributions and discuss in detail the limitation of our approach later in this work; see Sec. VI. Note here that at high enough temperature those processes are typically efficient, leading just to the ionization equilibrium among bounded and scattering states. As long as ionization equilibrium is maintained, the effective action we use is sufficient enough to describe Sommerfeld enhanced annihilation and bound-state decay at finite temperature. To estimate the time when the ionization equilibrium is violated concretely, we have to take into account these processes in the thermal plasma, which will be presented elsewhere. In vacuum, ionization equilibrium starts to become violated when the decay width of the lowest bound state exceeds the ionization rate.

$$D^{R,A}(\omega, \mathbf{p}) = \frac{-i}{\omega^2 - \mathbf{p}^2 + \Pi^{00}_{\mathrm{R,A}}(\omega, \mathbf{p}) \pm i\,\mathrm{sign}(\omega)\epsilon}, \tag{3.9}$$

$$\Pi^{00}_{\mathrm{R,A}}(\omega, \mathbf{p}) = -m_D^2 \left[1 - \frac{\omega}{2|\mathbf{p}|} \ln\left(\left|\frac{\omega + |\mathbf{p}|}{\omega - |\mathbf{p}|}\right|\right) \pm i \frac{\omega}{|\mathbf{p}|} \frac{\pi}{2} \theta(|\mathbf{p}|^2 - \omega^2)\right], \tag{3.10}$$

where we introduced the Debye screening mass $m_D^2 = g_\psi^2 T^2/3$. One can recognize that there is correction to the real part of the mediator propagator as well as a branch cut for spacelike exchange. Using the equilibrium relation Eq. (2.26), the $++$ component of D in the static limit reads

$$\begin{aligned}\lim_{\omega\to 0} D^{++}(\omega, \mathbf{p}) &= \lim_{\omega\to 0}\left[\frac{D^R(\omega, \mathbf{p}) + D^A(\omega, \mathbf{p})}{2} + \left(\frac{1}{2} + n_B(\omega)\right) D^\rho(\omega, \mathbf{p})\right] \\ &= \frac{i}{\mathbf{p}^2 + m_D^2} + \pi \frac{T}{|\mathbf{p}|} \frac{m_D^2}{(\mathbf{p}^2 + m_D^2)^2},\end{aligned} \tag{3.11}$$

while for a massive mediator we have simply

$$\lim_{\omega\to 0} D^{++}(\omega, \mathbf{p}) = \frac{i}{\mathbf{p}^2 + m_V^2 + m_D^2} + \pi \frac{T}{|\mathbf{p}|} \frac{m_D^2}{(\mathbf{p}^2 + m_V^2 + m_D^2)^2}. \tag{3.12}$$

The static D^{++} component is of special importance for describing DM long-range interactions in a plasma environment as we will see later in this work. The first term in Eqs. (3.11) and (3.12) will result in a screened Yukawa potential after Fourier transformation while the second terms will lead to purely imaginary contributions. Physically, the latter part originates from the scattering of the photon with plasma fermions, leading to a damping rate [72]. Indeed in the quasiparticle picture, the mediator has a limited propagation time within the plasma, which limits as well the coherence of the mediator exchange processes. For what regards the DM particles, this term will later give rise to DM-ψ scattering with zero energy transfer, leading also to a thermal width for the DM states. In the following sections, we try to keep generality and work in most of the computations with the unspecified form $D(x, y)$ and take just at the very end the static and HTL limit. Let us finally remark that the simple form of Eq. (3.12) allows us to achieve semianalytical results for the DM annihilation or decay rates in the presence of a thermal environment.

B. Exact DM number density equation from correlator equation of motion

The main purpose of this section is to derive the equation for the DM number density directly from the exact EoM of our nonrelativistic action. Defining the DM particle and antiparticle correlators as

$$G_\eta(x, y) \equiv \langle T_{\mathcal{C}} \eta(x) \eta^\dagger(y) \rangle, \tag{3.13}$$

$$G_\xi(x, y) \equiv \langle T_{\mathcal{C}} \xi(x) \xi^\dagger(y) \rangle, \tag{3.14}$$

we derive respective EoM from the path-integral formalism, as briefly explained at the end of Sec. II, for the nonrelativistic effective action S_{NR} given in Eq. (3.2):

$$\begin{aligned}&\left[i\partial_{x^0} + \frac{\nabla_x^2}{2M}\right] G_\eta(x, y) \\ &= i\delta_{\mathcal{C}}(x, y) - i g_\chi^2 \int_{z\in\mathcal{C}} D(x, z) \langle T_{\mathcal{C}} \eta(x) J(z) \eta^\dagger(y) \rangle \\ &\quad - i \int_{z\in\mathcal{C}} \Gamma_s(x, z) \langle T_{\mathcal{C}} \xi(x) \{\xi^\dagger(z) \eta(z)\} \eta^\dagger(y) \rangle,\end{aligned} \tag{3.15}$$

$$\begin{aligned}&\left[-i\partial_{y^0} + \frac{\nabla_y^2}{2M}\right] G_\eta(x, y) \\ &= i\delta_{\mathcal{C}}(x, y) - i g_\chi^2 \int_{z\in\mathcal{C}} D(y, z) \langle T_{\mathcal{C}} \eta(x) J(z) \eta^\dagger(y) \rangle \\ &\quad - i \int_{z\in\mathcal{C}} \Gamma_s(z, y) \langle T_{\mathcal{C}} \eta(x) \{\eta^\dagger(z) \xi(z)\} \xi^\dagger(y) \rangle,\end{aligned} \tag{3.16}$$

---

[4] Let us briefly summarize here the assumptions of the HTL approximation. First, we drop all vacuum contributions and only keep temperature dependent parts. Second, we assume the external energy P^0 and momentum $\mathbf{p}$ to be much smaller than typical loop momentum $\mathbf{k}$ which is of the order temperature (see Appendix C). The discussion of the validity of the HTL approximation depends on where the dressed mediator correlator is attached to. One can not naively argue for the case where one would attach it to the DM correlator that the external momentum $\mathbf{p}$ is the DM momentum which is of course much larger than temperature, thus invalidating the HTL approximation by this argumentation. For example, in our case the dressed mediator correlator enters the DM single-particle self-energy (see Fig. 2 and Sec. V C) and the dominant energy and momentum region in the loop diagram is where HTL effective theory is valid.

$$\left[i\partial_{x^0} - \frac{\nabla_x^2}{2M}\right] G_\xi(x,y)$$
$$= i\delta_{\mathcal{C}}(x,y) - ig_\chi^2 \int_{z\in\mathcal{C}} D(x,z)\langle T_{\mathcal{C}}\xi(x)J(z)\xi^\dagger(y)\rangle$$
$$- i\int_{z\in\mathcal{C}} \Gamma_s(x,z)\langle T_{\mathcal{C}}\eta(x)\{\eta^\dagger(z)\xi(z)\}\xi^\dagger(y)\rangle, \tag{3.17}$$

$$\left[-i\partial_{y^0} - \frac{\nabla_y^2}{2M}\right] G_\xi(x,y)$$
$$= i\delta_{\mathcal{C}}(x,y) - ig_\chi^2 \int_{z\in\mathcal{C}} D(y,z)\langle T_{\mathcal{C}}\xi(x)J(z)\xi^\dagger(y)\rangle$$
$$- i\int_{z\in\mathcal{C}} \Gamma_s(y,z)\langle T_{\mathcal{C}}\xi(x)\{\xi^\dagger(z)\eta(z)\}\eta^\dagger(y)\rangle. \tag{3.18}$$

The anticipated structure in Eqs. (3.15)–(3.18) shows the dependence of the two-point correlators on higher correlation functions. Here, the curly brackets stand for the summation over the spin indices and J is the current as already introduced in Eq. (3.3). It might be helpful to mention that we used a special property of two-point functions of Hermitian bosonic field operators: $D(x,y) = D(y,x)$. This exact property can be verified directly from the definition in Eq. (3.3).

In the following, the number density equation of particle and antiparticle DM is derived from this set of differential equations. First of all, we would like to clarify what is the number density in terms of fields appearing in S_{NR} in Eq. (3.2). For this purpose, let us switch off the annihilation term $\Gamma_s \to 0$ in S_{NR} and seek for conserved quantities. In this limit, the theory has the following global symmetries: $\eta \mapsto e^{i\theta_\eta}\eta$ and $\xi \mapsto e^{-i\theta_\xi}\xi$. The associated Noether currents for the DM particle and antiparticle are

$$J_\eta^\mu(x) = \left(\eta^\dagger(x)\eta(x), \frac{1}{2iM}\eta^\dagger(x)\overset{\leftrightarrow}{\nabla}\eta(x)\right),$$
$$J_\xi^\mu(x) = \left(\xi(x)\xi^\dagger(x), \frac{1}{2iM}\xi(x)\overset{\leftrightarrow}{\nabla}\xi^\dagger(x)\right). \tag{3.19}$$

The thermal-averaged zeroth component is the number density and the average over spatial component results in the current density. We obtain the differential equation for the two DM currents directly from the two-point function EoM, by subtracting Eqs. (3.16) from Eqs. (3.15) and Eqs. (3.18) from Eqs. (3.17), and by taking the spin-trace and the limit $y \to x$. For the particle DM, we obtain as an intermediate result after all these steps:

$$i\partial_\mu\langle J_\eta^\mu\rangle = -\left(i\partial_{x^0} + i\partial_{y^0} + \frac{\nabla_x^2}{2M} - \frac{\nabla_y^2}{2M}\right)\mathrm{Tr}G_\eta(x,y)\big|_{y\to x}$$
$$= +ig_\chi^2\int_{z\in\mathcal{C}} D(x,z)[\langle T_{\mathcal{C}}\{\eta(x)\eta^\dagger(x)\}J(z)\rangle - \langle T_{\mathcal{C}}\{\eta(x)\eta^\dagger(x)\}J(z)\rangle]$$
$$- i\int_{z\in\mathcal{C}} [\Gamma_s(x,z)\langle T_{\mathcal{C}}\{\eta(z)\xi^\dagger(z)\}\{\xi(x)\eta^\dagger(y)\}\rangle - \Gamma_s(z,y)\langle T_{\mathcal{C}}\{\eta(x)\xi^\dagger(y)\}\{\xi(z)\eta^\dagger(z)\}\rangle]_{y\to x}. \tag{3.20}$$

The trace and the curly brackets indicate the summation over the spin indices. It is important to note that the first line in the second equality cancels out, even in the case of a fully interacting correlator D including finite temperature corrections. Thus, we confirm from the EoM that, e.g., non-Hermitian potential corrections arising from soft thermal DM-ψ scatterings in the HTL approximation of D [see Eq. (3.11)], never violate the current conservation of each individual DM species. For a homogeneous and isotropic system (vanishing divergence of current density) this would mean that the individual number densities of particles and antiparticles do not change by self-scattering processes, real physical DM-ψ scatterings, soft DM-ψ scatterings or other finite temperature corrections leading to potential contributions in D.

It can be recognized, that the current conservation is only violated by the annihilation term Γ_s in the last line in Eq. (3.20), since this contribution does not cancel to zero. This term can be simplified by using Eq. (3.4) and by fixing the time component x^0 to either τ^+ or τ^-. We have explicitly checked that both choices of x^0 lead to the same final result. With the definition of the four-point correlator on the closed-time-path contour

$$G_{\eta\xi,s}(x,y,z,w) \equiv \langle T_{\mathcal{C}}\eta^i(x)\xi_i^\dagger(y)\xi^j(w)\eta_j^\dagger(z)\rangle, \tag{3.21}$$

we obtain our final form of the current equations:

$$\partial_\mu\langle J_\eta^\mu(x)\rangle = -2\frac{\pi(\alpha_\chi^2 + \alpha_\chi\alpha_\psi)}{M^2}G_{\eta\xi,s}^{++--}(x,x,x,x), \tag{3.22}$$

$$\partial_\mu\langle J_\xi^\mu(x)\rangle = -2\frac{\pi(\alpha_\chi^2 + \alpha_\chi\alpha_\psi)}{M^2}G_{\eta\xi,s}^{++--}(x,x,x,x). \tag{3.23}$$

The current conservation is only violated by contributions coming from Γ_s. This is consistent with the expectations from the symmetry properties of the action when annihilation is turned on. Namely, only a linear combination of both global transformations leaves the action invariant which leads to the conservation of $\partial_\mu(J_\eta^\mu - J_\xi^\mu) = 0$, which is nothing but the DM asymmetry current conservation. The conservation of the total DM number density, $\partial_\mu(J^\mu_\eta + J^\mu_\xi)$, is violated by the annihilation term.

Before we discuss the four-point correlator appearing in Eqs. (3.22) and (3.23) in detail, let us now assume a Friedman-Robertson-Walker universe and make the connection to the Boltzmann equation for the number density that is typically adopted in the literature when calculating the thermal history of the dark matter particles. First, the spatial divergence on the left-hand side of the current Eqs. (3.22) and (3.23) vanishes due to homogeneity and isotropy. Second, the adiabatic expansion of the background introduces a Hubble expansion term H. Third, it can be seen from the sign of the right-hand side of the current equations that only a DM loss term occurs. The production term is missing because we have assumed *a priori*, when deriving the nonrelativistic action, that the DM mass is much larger than the thermal plasma temperature. Within this mass-to-infinity limit the DM production term is set to zero in the computation of the annihilation term Γ_s and not expected to occur. Let us therefore add on the r.h.s. of the current equations *a posteriori* the production term of the DM via the assumption of detailed balance, resulting in the more familiar number density equations:

$$\dot{n}_\eta + 3Hn_\eta = -2(\sigma v_{\mathrm{rel}})[G^{++--}_{\eta\xi,s}(x,x,x,x) - G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}], \tag{3.24}$$

$$\dot{n}_\xi + 3Hn_\xi = -2(\sigma v_{\mathrm{rel}})[G^{++--}_{\eta\xi,s}(x,x,x,x) - G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}]. \tag{3.25}$$

The tree-level s-wave annihilation cross section of our system was defined as $(\sigma v_{\mathrm{rel}}) = \pi(\alpha_\chi^2 + \alpha_\chi\alpha_\psi)/M^2$ and $|_{\mathrm{eq}}$ in the last term means the evaluation at thermal equilibrium. Note that in the CTP formalism a cross section strictly speaking does not exist. The reason why this result is equal to the vacuum computation is because we computed the annihilation part Γ_s at the leading order, where it is expected that zero and finite temperature results should coincide. The correlation function $G^{++--}_{\eta\xi,s}$ however is fully interacting. We summarize with two concluding remarks on our main results:

(i) *Sommerfeld-enhancement factor at finite temperature:* One of our findings is that the Sommerfeld factor is contained in a certain component of the interacting four-point correlation function, namely $G^{++--}_{\eta\xi,s}$. This result is valid for a generic out-of-equilibrium state of the dark matter system. The remaining task is to find a solution for this four-point correlator. As we will see later, the solution can be obtained from the Bethe-Salpeter equation on the CTP contour, derived in the next section. For example, expanding the Bethe-Salpeter equation to zeroth order in the DM self-interactions $2G^{++--}_{\eta\xi,s}(x,x,x,x) \simeq -2G^{+-}_\eta(x,x) \times G^{+-}_\xi(x,x) = n_\eta n_\xi$ and inserting this into Eqs. (3.24) and (3.25) would result in a well-known expression for the number density equation of the DM particles with velocity-independent annihilation. As we will see later, higher terms in the interaction or a fully nonperturbative solution contain the finite temperature corrected negative and positive energy spectrum. In other words, $G^{++--}_{\eta\xi,s}$ contains both, the bound-state *and* scattering state contributions at the same time and at finite temperature they turn out to be not separable as it is sometimes done in vacuum computations. Bound state contributions will automatically change the cross section into a decay width and thus, n_η appearing on the l.h.s. of Eq. (3.24) is the *total* number of η particles including the ones in the bound states and similar interpretation for the antiparticle ξ.

(ii) *Particle number conservation:* In Sec. III A, we have seen that the thermal corrections to the mediator propagator D can contain, next to the real Debye mass, an imaginary contribution. It was shown that these non-Hermitian corrections to the potential never violate the particle number conservation due to the exact cancellation of the second line in Eq. (3.20). This was expected from the beginning, since, when switching off the annihilation $\Gamma_s \to 0$, the nonrelativistic action in Eq. (3.2) has two global symmetries $\eta \mapsto e^{i\theta_\eta}\eta$ and $\xi \mapsto e^{-i\theta_\xi}\xi$. The conserved quantities are the particle and antiparticle currents in Eqs. (3.22) and (3.23) in the limit $\Gamma_s \to 0$ (vanishing r.h.s.). When annihilation is included, the nonrelativistic action is only invariant if both global transformations are performed at the same time, resulting in the conserved asymmetry current $J_\eta - J_\xi$. We conclude that thermal corrections can never violate these symmetries, even not at higher loop level. On the other hand, the solution of the Sommerfeld factor is contained in $G^{++--}_{\eta\xi,s}$ and hence the annihilation rate will depend on the thermal loop corrected long-range mediator D, as we will see in the next section.

IV. TWO-TIME BETHE-SALPETER EQUATIONS

The exact number density Eq. (3.24) depends on the Keldysh four-point correlation function $G^{++--}_{\eta\xi,s}(x,x,x,x)$. In this section, we derive the system of closed equation of motion needed in order to obtain a solution for this four-point function, including the full resummation of Coulomb divergent ladder diagrams. The result will be a coupled set of two-time Bethe-Salpeter equations on the Keldysh contour as given by the end of this section, Eqs. (4.20)–(4.21). They apply in general for out-of-equilibrium situations and include in their nonperturbative form also the bound-state contributions if present. In order to arrive at those equations, a set of approximations and assumptions is needed. We therefore would like to start from the beginning in deriving those equations, which might lead to a better understanding of their limitation.

In the first simplification, we treat the annihilation term Γ_s as a perturbation and ignore it in the following computations, since the leading order term in the annihilation part is already contained in Eq. (3.24). The exact set of EoMs for two- and four-point functions in the limit $\Gamma_s \to 0$ are given by

$$\left[i\partial_{x^0} + \frac{\nabla_x^2}{2M}\right] G_\eta(x,y) = i\delta_{\mathcal{C}}(x,y) - ig_\chi^2 \int_{z\in\mathcal{C}} D(x,z) \times [G_{\eta\xi}(x,z,y,z) - G_{\eta\eta}(x,z,y,z)], \tag{4.1}$$

$$\left[i\partial_{x^0} + \frac{\nabla_x^2}{2M}\right] \bar{G}_\xi(x,y) = i\delta_{\mathcal{C}}(x,y) - ig_\chi^2 \int_{z\in\mathcal{C}} D(x,z) \times [G_{\eta\xi}(z,x,z,y) - G_{\xi\xi}(x,z,y,z)], \tag{4.2}$$

$$\begin{aligned}\left[i\partial_{x^0} + \frac{\nabla_x^2}{2M}\right] G_{\eta\xi}(x,y,z,w) &= i\delta_{\mathcal{C}}(x,z)\bar{G}_\xi(y,w) \\ &\quad - ig_\chi^2 \int_{\bar{x}\in\mathcal{C}} D(x,\bar{x}) \langle T_{\mathcal{C}} \eta(x) J(\bar{x}) \xi^\dagger(y) \xi(w) \eta^\dagger(z)\rangle,\end{aligned} \tag{4.3}$$

where we will work for the rest of this work with the conjugate antiparticle correlator $\bar{G}_\xi$ and, here, the spin-uncontracted correlators are defined as

$$\bar{G}_\xi(x,y) \equiv \langle T_{\mathcal{C}} \xi^\dagger(x) \xi(y)\rangle, \tag{4.4}$$

$$G_{\eta\xi}(x,y,z,w) \equiv \langle T_{\mathcal{C}} \eta(x) \xi^\dagger(y) \xi(w) \eta^\dagger(z)\rangle, \tag{4.5}$$

$$G_{\eta\eta}(x,y,z,w) \equiv \langle T_{\mathcal{C}} \eta(x) \eta(y) \eta^\dagger(w) \eta^\dagger(z)\rangle, \tag{4.6}$$

$$G_{\xi\xi}(x,y,z,w) \equiv \langle T_{\mathcal{C}} \xi^\dagger(x) \xi^\dagger(y) \xi(w) \xi(z)\rangle. \tag{4.7}$$

The EoMs for the two-point functions Eqs. (4.1) and (4.2) are equivalent to Eqs. (3.15) and (3.17) in the limit $\Gamma_s \to 0$, and we have just rewritten them in terms of the four-point correlators and conjugate antiparticle propagator, defined in Eqs. (4.4)–(4.7). In our notation, the spinor indices of the operators having equal space-time arguments in the four-point correlators are summed and J is the current as defined in Eq. (3.3). The different conventions for the $\eta\eta$ and $\xi\xi$ four-point correlators are because $\eta^\dagger$ and ξ are the creation operators. From the exact differential equation of the four-point correlator in Eq. (4.3), it can be seen that the correlator hierarchy is still not closed yet, since it depends on the six-point function. We close this hierarchy of correlators by truncating the six-point function at the leading order:

$$\begin{aligned}&\left[i\partial_{y^0} + \frac{\nabla_y^2}{2M}\right] \langle T_{\mathcal{C}} \eta(x) J(\bar{x}) \xi^\dagger(y) \xi(w) \eta^\dagger(z)\rangle \\ &\simeq i\delta_{\mathcal{C}}(y,w)[G_{\eta\xi}(x,\bar{x},z,\bar{x}) - G_{\eta\eta}(x,\bar{x},z,\bar{x})] \\ &\quad - i\delta_{\mathcal{C}}(\bar{x},y) G_{\eta\xi}(x,y,z,w),\end{aligned} \tag{4.8}$$

i.e., *only the integral kernel of the six-point function containing the eight-point function was dropped*. The set of correlator differential equations is now closed under this truncation procedure. A fully self-consistent solution requires in principle to solve the equations for the five correlators G_η, G_ξ, $G_{\eta\xi}, G_{\eta\eta}, G_{\xi\xi}$ simultaneously. This is beyond the scope of this work and we have to further approximate the system in order to obtain at least a simple semianalytical solution at the end.

The solution of our target component $G_{\eta\xi}$ can formally be decoupled from the solution of $G_{\eta\eta}$ and $G_{\xi\xi}$ by approximating the latter two quantities as the leading order contribution (dropping the integral kernel, known as Hartree-Fock approximation). Then, one can recognize that

$$\begin{aligned}&[G_{\eta\xi}(x,z,y,z) - G_{\eta\eta}(x,z,y,z)] \\ &\simeq [G_\eta(x,y)\bar{G}_\xi(z,z) - G_\eta(x,y)G_\eta(z,z) + G_\eta(x,z)G_\eta(z,y)] \\ &= G_\eta(x,z)G_\eta(z,y),\end{aligned} \tag{4.9}$$

$$\begin{aligned}&[G_{\eta\xi}(z,x,z,y) - G_{\xi\xi}(x,z,y,z)] \\ &\simeq [\bar{G}_\xi(x,y)G_\eta(z,z) - \bar{G}_\xi(x,y)\bar{G}_\xi(z,z) + \bar{G}_\xi(x,z)\bar{G}_\xi(z,y)] \\ &= \bar{G}_\xi(x,z)\bar{G}_\xi(z,y).\end{aligned} \tag{4.10}$$

In both equations, the last step is a strict equality only if the DM particle $G_\eta(z,z) = -n_\eta$ and antiparticle $\bar{G}_\xi(z,z) = -n_\xi$ number densities are equal. This is true if there is no DM asymmetry which we will assume throughout this work.

In the last approximation, we perform a coupling expansion of Eqs. (4.1)–(4.3). After inserting the results of the two-point functions into the equation of $G_{\eta\xi}$, by using the relations Eqs. (4.9) and (4.10) in the free limit, we obtain for the four-point correlator to the leading order in g_χ:

$$\begin{aligned}G_{\eta\xi}(x,y,z,w) &\simeq G_{\eta,0}(x,z)\bar{G}_{\xi,0}(y,w) + g_\chi^2 \int_{\bar{x},\bar{y}\in\mathcal{C}} G_{\eta,0}(x,\bar{x})\bar{G}_{\xi,0}(y,\bar{y}) D(\bar{x},\bar{y}) G_{\eta,0}(\bar{x},z)\bar{G}_{\xi,0}(\bar{y},w) \\ &\quad + \int_{\bar{x},\bar{y}\in\mathcal{C}} G_{\eta,0}(x,z)\bar{G}_{\xi,0}(y,\bar{y})(-i)\bar{\Sigma}_\xi(\bar{y},\bar{x})\bar{G}_{\xi,0}(\bar{x},w) + \int_{\bar{x},\bar{y}\in\mathcal{C}} G_{\eta,0}(x,\bar{x})\bar{G}_{\xi,0}(y,w)(-i)\Sigma_\eta(\bar{x},\bar{y})G_{\eta,0}(\bar{y},z),\end{aligned} \tag{4.11}$$

where in the last two terms, particle and antiparticle are disconnected and we have introduced the *single-particle self-energies* according to

$$\Sigma_\eta(x,y) \equiv -ig_\chi^2 D(x,y) G_{\eta,0}(x,y),$$
$$\bar{\Sigma}_\xi(x,y) \equiv -ig_\chi^2 D(x,y) \bar{G}_{\xi,0}(x,y). \tag{4.12}$$

The first integral term in Eq. (4.11) contains the ladder diagram exchange between particle and antiparticle. Similar equations for $G_{\eta\eta}$ and $G_{\xi\xi}$ can also be obtained by applying the same steps. In order to obtain the spectrum of bound-state solutions as well as a fully nonperturbative treatment of the Sommerfeld-enhancement we have to resumm Eq. (4.11) somehow.

We define our resummation scheme of the four-point correlator by resumming the ladder exchange, as well as the self-energy contributions in Eq. (4.11) on an equal footing.

In other words, we are resumming the leading order terms in the coupling expansion of the four-point correlator $G_{\eta\xi}$ and similar for $G_{\eta\eta}$ and $G_{\xi\xi}$. On the one hand, this is a critical point, since this procedure is not exact and we cannot guarantee that other contributions do not play an important role as well, e.g., one of the limitations are systems with large coupling constants or large DM density where we cannot decouple the solution of $G_{\eta\xi}$ from $G_{\eta\eta}$ and $G_{\xi\xi}$. The latter limitation might not be a problem since when focussing on the freeze-out the DM density becomes dilute. On the other hand, as we will discuss in detail later in this work, the final result based on this resummation scheme behaves physically, fulfils KMS condition in equilibrium,[5] reproduces the correct vacuum limit, enables us to study bound states and DM Sommerfeld-enhanced annihilation at finite temperature, and in some other limits we recover the literature results based on linear response theory. Furthermore, a combined resummation of one-particle self-energies and ladder-diagram exchanges seems to be necessary at finite temperature, since similar combinations would also occur when calculating the effective potential from Wilson-loop lines [73] guaranteeing the gauge invariance.

Before we can resum Eq. (4.11), we need some further exact rearrangements and manipulations of the four-point function components. Since we are interested in the solution of $G_{\eta\xi}^{++--}$ at equal times, it is sufficient to only consider two-time four-point functions, where we will adopt the short notation $G_{\eta\xi}(t,t') = G_{\eta\xi}(t\mathbf{xy}, t'\mathbf{zw})$. Certain combinations of the components of Eq. (4.11) turn out to be closed, e.g., let us define[6]

$$G_{\eta\xi}^R(t,t') \equiv \theta(t-t')[G_{\eta\xi}^{++--}(t,t') - G_{\eta\xi}^{+--+}(t,t') - G_{\eta\xi}^{-++-}(t,t') + G_{\eta\xi}^{--++}(t,t')], \tag{4.13}$$

$$G_{\eta\xi}^A(t,t') \equiv -\theta(t'-t)[G_{\eta\xi}^{++--}(t,t') - G_{\eta\xi}^{+--+}(t,t') - G_{\eta\xi}^{-++-}(t,t') + G_{\eta\xi}^{--++}(t,t')]. \tag{4.14}$$

In the following we will show that, by using the semigroup properties of free correlators, the following structure of the retarded equations can be achieved:

$$G_{\eta\xi}^R(t,t') = G_{\eta,0}^R(t,t')G_{\xi,0}^R(t,t') + g_\chi^2 \int \mathrm{d}t_1 G_{\eta,0}^R(t,t_1)G_{\xi,0}^R(t,t_1) \times \int \mathrm{d}t_2 \Sigma_{\eta\xi}^R(t_1,t_2) G_{\eta,0}^R(t_2,t')G_{\xi,0}^R(t_2,t'), \tag{4.15}$$

and similar for the advanced. The precise terms contained in the two-particle self-energy $\Sigma_{\eta\xi}$ will be given later. From the form of Eq. (4.15) it is clear that our resummation scheme as described above is just the replacement of the free two-point correlators at the end by $G_{\eta,0}^R(t_2,t')G_{\xi,0}^R(t_2,t') \to G_{\eta\xi}^R(t_2,t')$.

Let us in the following sketch the way to obtain this resummable structure. The first term on the r.h.s. in Eq. (4.15) can be obtained by using simple relations

$$G_{\eta,0}^{+-}G_{\xi,0}^{+-} - G_{\eta,0}^{+-}G_{\xi,0}^{-+} - G_{\eta,0}^{-+}G_{\xi,0}^{+-} + G_{\eta,0}^{-+}G_{\xi,0}^{-+} = G_{\eta,0}^{\rho}G_{\xi,0}^{\rho} = G_{\eta,0}^R G_{\xi,0}^R + G_{\eta,0}^A G_{\xi,0}^A. \tag{4.16}$$

---

[5] Another resummation scheme for $G_{\eta\xi}$ we tested can be obtained directly after the steps of the truncation in Eq. (4.8) and Hartree-Fock approximation in Eqs. (4.9)–(4.10). After some algebra, this would result in the Bethe-Salpeter equation $G_{\eta\xi}(x,y,z,w) = G_\eta(x,z)\bar{G}_\xi(y,w) + g_\chi^2 \int_{\bar{x},\bar{y}\in\mathcal{C}} G_\eta(x,\bar{x})\bar{G}_\xi(y,\bar{y})D(\bar{x},\bar{y})G_{\eta\xi}(\bar{x},\bar{y},z,w)$. Note that this equation would be equivalent to the original vacuum BS equation when naively extending the time integration in the latter equation towards the Keldysh contour $\mathcal{C}$. The main difference here compared to our resummation scheme is that the two-point functions are fully interacting. The l.h.s. of the $++--$ or $--++$ component of this equation fulfils the KMS condition in equilibrium, when taking the two-time limit. The right-hand side depends in general on three times. It can be reduced to only two times by assuming a static form for the mediator correlator $D(x,y) = \delta_C(t_x,t_y)V(\mathbf{x}-\mathbf{y})$. Integrating over Keldysh-contour delta function leads to the fact that also the r.h.s. fulfils the KMS condition. However, this is a strong assumption on the form of the mediator correlation function, sending the off-diagonal terms D^{+-} or D^{-+} to zero. The simplest possibility to take into account the off-diagonal terms and at the same time obtain a two-time structure also of the r.h.s. would be to assume that every component of D is proportional to a time delta function. A crucial observation we have made is that this type of approximation seems to violate the KMS condition through the off-diagonal terms, although the r.h.s. has a two-time structure. The reason might be in the resummation of different orders in the coupling, as caused by the interacting DM two-point correlators in the BS kernel. The coupling expansion and the resummation of terms of equal order in the coupling parameter (our scheme), seems to be essential in order to obtain our final BS Eq. (4.20), fulfilling the KMS condition.

[6] This combination is not obvious at first, but when subtracting the advanced component from the retarded it can be shown that the resulting spectral function has a similar completeness relation as the spectral function of two-point correlators [74].

In the last step we used $G^R G^A = 0$ for equal time products and when multiplying this with the unit step function as in the definition Eqs. (4.13) and (4.14) this just projects out the respective product. The integral term is more complicated. Let us consider for simplicity only the last integral term in Eq. (4.11) and perform the sum over the different contributions of the components for the retarded two-time four-point correlator, resulting in

$$I = \int \mathrm{d}t_1 \mathrm{d}t_2 G^R_{\eta,0}(t,t_1)\bar{G}^R_{\xi,0}(t,t')(-i)\Sigma^R_\eta(t_1,t_2)G^R_{\eta,0}(t_2,t'), \tag{4.17}$$

where the retarded one-particle self-energy is defined as $\Sigma^R = \Sigma^{++} - \Sigma^{+-}$. Since all propagators are free, we can use the semigroup property (see also Appendix A)

$$\bar{G}^R_{\xi,0}(t,t') = \bar{G}^R_{\xi,0}(t,t_1)\bar{G}^R_{\xi,0}(t_1,t_2)\bar{G}^R_{\xi,0}(t_2,t'), \quad \text{for } t < t_1 < t_2 < t'. \tag{4.18}$$

For brevity, we suppress the space integration here. This property can be used in Eq. (4.17) since the times satisfy the inequality due to the product of retarded correlators, resulting in

$$I = \int \mathrm{d}t_1 G^R_{\eta,0}(t,t_1)\bar{G}^R_{\xi,0}(t,t_1)\int \mathrm{d}t_2 [(-i)\Sigma^R_\eta(t_1,t_2)\bar{G}^R_{\xi,0}(t_1,t_2)] \times G^R_{\eta,0}(t_2,t')\bar{G}^R_{\xi,0}(t_2,t'). \tag{4.19}$$

Comparing with Eq. (4.15), we indeed find the anticipated structure. Applying similar steps to all integral terms in Eq. (4.11), we find the following *two-time Bethe-Salpeter equations*:[7]

$$\begin{aligned}
G^\Phi_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) = \mathcal{G}^\Phi_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) &+ \int\limits_{t_1,\mathbf{x}_1,\mathbf{x}_2} \mathcal{G}^R_{\eta\xi}(t\mathbf{x}\mathbf{y},t_1\mathbf{x}_1\mathbf{x}_2) \int\limits_{t_2,\mathbf{x}_3,\mathbf{x}_4} \Sigma^R_{\eta\xi}(t_1\mathbf{x}_1\mathbf{x}_2,t_2\mathbf{x}_3\mathbf{x}_4)G^\Phi_{\eta\xi}(t_2\mathbf{x}_3\mathbf{x}_4,t'\mathbf{z}\mathbf{w}) \\
&+ \int\limits_{t_1,\mathbf{x}_1,\mathbf{x}_2} \mathcal{G}^R_{\eta\xi}(t\mathbf{x}\mathbf{y},t_1\mathbf{x}_1\mathbf{x}_2) \int\limits_{t_2,\mathbf{x}_3,\mathbf{x}_4} \sigma^\Phi_{\eta\xi}(t_1\mathbf{x}_1\mathbf{x}_2,t_2\mathbf{x}_3\mathbf{x}_4)G^A_{\eta\xi}(t_2\mathbf{x}_3\mathbf{x}_4,t'\mathbf{z}\mathbf{w}) \\
&+ \int\limits_{t_1,\mathbf{x}_1,\mathbf{x}_2} \mathcal{G}^\Phi_{\eta\xi}(t\mathbf{x}\mathbf{y},t_1\mathbf{x}_1\mathbf{x}_2) \int\limits_{t_2,\mathbf{x}_3,\mathbf{x}_4} \Sigma^A_{\eta\xi}(t_1\mathbf{x}_1\mathbf{x}_2,t_2\mathbf{x}_3\mathbf{x}_4)G^A_{\eta\xi}(t_2\mathbf{x}_3\mathbf{x}_4,t'\mathbf{z}\mathbf{w})
\end{aligned} \tag{4.20}$$

with $\Phi = \{++--\};\{+--+\};\{-++-\};\{--++\}$, *and*

$$\begin{aligned}
G^{R/A}_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) = \mathcal{G}^{R/A}_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) &+ \int\limits_{t_1,\mathbf{x}_1,\mathbf{x}_2} \mathcal{G}^{R/A}_{\eta\xi}(t\mathbf{x}\mathbf{y},t_1\mathbf{x}_1\mathbf{x}_2) \\
&\times \int\limits_{t_2,\mathbf{x}_3,\mathbf{x}_4} \Sigma^{R/A}_{\eta\xi}(t_1\mathbf{x}_1\mathbf{x}_2,t_2\mathbf{x}_3\mathbf{x}_4) \\
&\times G^{R/A}_{\eta\xi}(t_2\mathbf{x}_3\mathbf{x}_4,t'\mathbf{z}\mathbf{w}).
\end{aligned} \tag{4.21}$$

The products of free correlators are defined as

$$\begin{aligned}
&\mathcal{G}^R_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) \equiv G^R_{\eta,0}(t\mathbf{x},t'\mathbf{z})\bar{G}^R_{\xi,0}(t\mathbf{y},t'\mathbf{w}), \\
\text{and}\quad &\mathcal{G}^A_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{z}\mathbf{w}) \equiv -G^A_{\eta,0}(t\mathbf{x},t'\mathbf{z})\bar{G}^A_{\xi,0}(t\mathbf{y},t'\mathbf{w}),
\end{aligned} \tag{4.22}$$

and similar for the other components, e.g., $\mathcal{G}^{++--}_{\eta\xi} = G^{+-}_{\eta,0}G^{+-}_{\xi,0}$. Furthermore, we introduced the *two-particle self-energies* according to

$$\begin{aligned}
\Sigma^R_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{w}\mathbf{z}) \equiv (-i)\Sigma^R_\eta(t\mathbf{x},t'\mathbf{w})\bar{G}^R_{\xi,0}(t\mathbf{y},t'\mathbf{z}) &+ (-i)\Sigma^R_\xi(t\mathbf{y},t'\mathbf{z})G^R_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2\bar{G}^R_{\xi,0}(t\mathbf{y},t'\mathbf{z})D^{-+}(t\mathbf{x},t'\mathbf{z})G^R_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2 G^R_{\eta,0}(t\mathbf{x},t'\mathbf{w})D^{+-}(t'\mathbf{w},t\mathbf{y})\bar{G}^R_{\xi,0}(t\mathbf{y},t'\mathbf{z}) \\
&+ g_\chi^2\bar{G}^{+-}_{\xi,0}(t\mathbf{y},t'\mathbf{z})D^R(t\mathbf{x},t'\mathbf{z})G^R_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2 G^{+-}_{\eta,0}(t\mathbf{x},t'\mathbf{w})D^A(t'\mathbf{w},t\mathbf{y})\bar{G}^R_{\xi,0}(t\mathbf{y},t'\mathbf{z}),
\end{aligned} \tag{4.23}$$

where the retarded single-particle self-energy is defined in terms of the components of the definition Eq. (5.20), namely $\Sigma^R_i \equiv \Sigma^{++}_i - \Sigma^{+-}_i$. The advanced two-particle self-energy is given by

$$\begin{aligned}
-\Sigma^A_{\eta\xi}(t\mathbf{x}\mathbf{y},t'\mathbf{w}\mathbf{z}) \equiv (-i)\Sigma^A_\eta(t\mathbf{x},t'\mathbf{w})\bar{G}^A_{\xi,0}(t\mathbf{y},t'\mathbf{z}) &+ (-i)\Sigma^A_\xi(t\mathbf{y},t'\mathbf{z})G^A_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2\bar{G}^A_{\xi,0}(t\mathbf{y},t'\mathbf{z})D^{-+}(t\mathbf{x},t'\mathbf{z})G^A_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2 G^A_{\eta,0}(t\mathbf{x},t'\mathbf{w})D^{+-}(t'\mathbf{w},t\mathbf{y})\bar{G}^A_{\xi,0}(t\mathbf{y},t'\mathbf{z}) \\
&+ g_\chi^2\bar{G}^{+-}_{\xi,0}(t\mathbf{y},t'\mathbf{z})D^A(t\mathbf{x},t'\mathbf{z})G^A_{\eta,0}(t\mathbf{x},t'\mathbf{w}) \\
&+ g_\chi^2 G^{+-}_{\eta,0}(t\mathbf{x},t'\mathbf{w})D^R(t'\mathbf{w},t\mathbf{y})\bar{G}^A_{\xi,0}(t\mathbf{y},t'\mathbf{z}),
\end{aligned} \tag{4.24}$$

---

[7] Similar equations where also obtained in [74], although the derivation, particle content, and potential is slightly different compared to ours. Nevertheless, further helpful steps for bringing the integral terms into a resumable form by using semigroup properties can be found in the Appendix of [74].

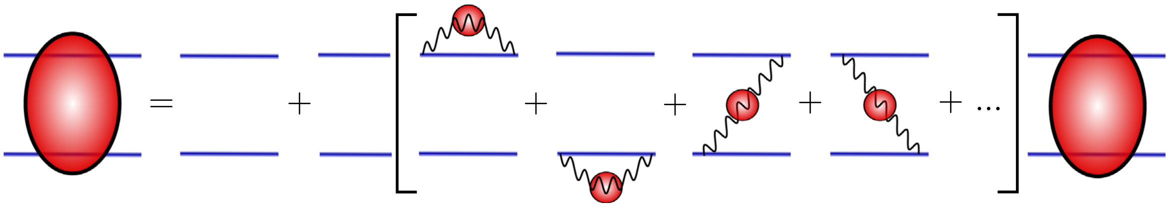


FIG. 2. Resummation scheme shown for the retarded four-point correlator Eq. (4.21). The terms in the brackets belong to one-particle self-energy contributions as well as the mediator exchange between particle and antiparticle. Dots represent terms containing the DM distribution in the two-particle self-energy. Due to the Boltzmann suppression at the freeze-out, those contributions will be dropped later (see DM dilute limit in Sec. V B).

and the most important statistical components are defined by

$$\begin{aligned}&\sigma_{\eta\xi}^{++--}(t\mathbf{xy},t'\mathbf{wz})\\&\equiv(-i)\Sigma_{\eta}^{+-}(t\mathbf{x},t'\mathbf{w})\bar{G}_{\xi,0}^{+-}(t\mathbf{y},t'\mathbf{z})\\&\quad+(-i)\Sigma_{\xi}^{+-}(t\mathbf{y},t'\mathbf{z})G_{\eta,0}^{+-}(t\mathbf{x},t'\mathbf{w})\\&\quad+g_{\chi}^{2}G_{\eta,0}^{+-}(t\mathbf{x},t'\mathbf{w})D^{-+}(t'\mathbf{w},t\mathbf{y})\bar{G}_{\xi,0}^{+-}(t\mathbf{y},t'\mathbf{z})\\&\quad+g_{\chi}^{2}G_{\eta,0}^{+-}(t\mathbf{x},t'\mathbf{w})D^{+-}(t\mathbf{x},t'\mathbf{z})\bar{G}_{\xi,0}^{+-}(t\mathbf{y},t'\mathbf{z}),\end{aligned}\tag{4.25}$$

$$\begin{aligned}&\sigma_{\eta\xi}^{--++}(t\mathbf{xy},t'\mathbf{wz})\\&\equiv(-i)\Sigma_{\eta}^{-+}(t\mathbf{x},t'\mathbf{w})\bar{G}_{\xi,0}^{-+}(t\mathbf{y},t'\mathbf{z})\\&\quad+(-i)\Sigma_{\xi}^{-+}(t\mathbf{y},t'\mathbf{z})G_{\eta,0}^{-+}(t\mathbf{x},t'\mathbf{w})\\&\quad+g_{\chi}^{2}G_{\eta,0}^{-+}(t\mathbf{x},t'\mathbf{w})D^{+-}(t'\mathbf{w},t\mathbf{y})\bar{G}_{\xi,0}^{-+}(t\mathbf{y},t'\mathbf{z})\\&\quad+g_{\chi}^{2}G_{\eta,0}^{-+}(t\mathbf{x},t'\mathbf{w})D^{+-}(t\mathbf{x},t'\mathbf{z})\bar{G}_{\xi,0}^{-+}(t\mathbf{y},t'\mathbf{z}).\end{aligned}\tag{4.26}$$

A graphical illustration of the resummation scheme for the retarded or advanced component is shown in Fig. 2. Below, we summarize the essential properties of our two-time BS equations, based on our resummation of one-particle self-energies as well as the mediator exchanges on an equal footing.

(i) *Two-time structure.*— The remarkable result of our resummation scheme is that we achieved a two-time dependence of the Bethe-Salpeter Eqs. (4.20) and (4.21) without assuming a static property of mediator correlator D or, more general, without assuming *any* particular form.

(ii) *KMS condition.*— In general, the two-time correlators $G_{\eta\xi}^{++--}(t,t')$ and $G_{\eta\xi}^{--++}(t,t')$ are related in equilibrium via the KMS condition: $G_{\eta\xi}^{--++}(t\mathbf{xy};t'\mathbf{zw})=G_{\eta\xi}^{++--}(t+i\beta,\mathbf{xy};t'\mathbf{zw})$ or $G_{\eta\xi}^{--++}(t\mathbf{xy};t'+i\beta,\mathbf{zw})=G_{\eta\xi}^{++--}(t\mathbf{xy};t'\mathbf{zw})$. Now, one of the great features of our resummation scheme is that it respects this KMS condition which is nontrivial since the equations are not exact. This means that the *left- and right-hand* sides of BS Eq. (4.20) for the components $G_{\eta\xi}^{++--}(t,t')$ and $G_{\eta\xi}^{--++}(t,t')$ transform respectively into each other in equilibrium. This can be seen, by Fourier transforming the kernels, needed for proper analytic continuation of time, and by using the fact that in equilibrium all quantities only depend on the difference of the time variable. Indeed one finds that always the statistical part in the three kernels of Eq. (4.20) transform into their counterparts. The solution of our target component $G_{\eta\xi}^{++--}$ becomes a very simple expression when utilizing the power of the KMS condition as we will see in the next section.

(iii) Similar to the two-point functions, two-time retarded and advanced components of the four-point correlator are related by complex conjugation $G_{\eta\xi}^{A}(t,t')=-[G_{\eta\xi}^{R}(t',t)]^{\dagger}$, as can be shown directly from the definition Eqs. (4.13) and (4.14).

(iv) Other components of four-point correlators not listed above can be constructed by the given ones [74].

(v) The full set of equations are able to describe Bose-Einstein condensation, e.g., relevant for fermionic systems where bound-state solutions exist and the density and chemical potential are in a critical regime. Since we focus on multi-TeV particles, the density will be always low enough to ignore those quantum-statistical effects.

V. TWO-PARTICLE SPECTRUM AT FINITE TEMPERATURE

For general out-of-equilibrium situations, the coupled system of two-time BS Eqs. (4.20)–(4.21) might require a fully numerical treatment. However, when relying on some well-motivated assumptions which are guaranteed for WIMP like freeze-outs, we show in this chapter that the coupled equations can by drastically simplified. The result will be a formal solution of our target component $G_{\eta\xi}^{++--}$, appearing in our main number density Eqs. (3.24)–(3.25), in terms of the DM *two-particle spectral function*. Furthermore, we fully provide the details for finding the explicit solution of the DM two-particle spectral function from a Schrödinger-like equation with an effective in-medium potential, including thermal corrections. For a better understanding of the limitation of this final solution we share, step by step, the approximations needed. All assumptions leading to this simple result will be made systematically and discussed separately in this section.

A. Formal solution in grand canonical ensemble

We assume the DM system to be in a grand canonical state where the density matrix takes the form

$$\hat{\rho} = \frac{1}{Z} e^{-\beta(\hat{H} - \mu_\eta \hat{N}_\eta - \mu_\xi \hat{N}_\xi)}; \qquad \hat{H} = \hat{H}_{\mathrm{NR}} + M(\hat{N}_\eta + \hat{N}_\xi),$$
$$\hat{N}_\bullet \equiv \int_{\boldsymbol{x}} j^0_\bullet, \tag{5.1}$$

where $\bullet = \eta$, ξ and we assume symmetric DM resulting in equal chemical potentials $\mu = \mu_\eta = \mu_\xi$. Further, we assume the Hamiltonian to commute with the number operators by treating the annihilation Γ_s as a perturbation. This is valid as long as the processes driving the system to a grand canonical state are much more efficient compared to annihilation or the decay of bound states. In this sense, the chemical potential is effectively time dependent. It is related to the total number density, appearing on the l.h.s. in Eqs. (3.24)–(3.25). The time dependence of the number density is set by the Hubble term and the production and loss terms appearing on the r.h.s.

Let us insert the grand canonical density matrix into the components $++--$ and $--++$ and show how they are related. Utilizing the commutation relations of $-[\hat{N}_\eta, \eta^{(\dagger)}] = (-)\eta^{(\dagger)}$, $[\hat{N}_\xi, \xi^{(\dagger)}] = (-)\xi^{(\dagger)}$, one can derive the KMS relations for the two-time four-point correlators in the presence of the chemical potentials. Recalling that the Hamiltonian is a generator of the time evolution, one finds the KMS condition for a grand canonical state:

$$G^{++--}_{\eta\xi,s}(t\boldsymbol{xy}; t'\boldsymbol{zw}) = e^{-2\beta(M-\mu)} G^{--++}_{\eta\xi,s}(t\boldsymbol{xy}; t' + i\beta, \boldsymbol{zw}). \tag{5.2}$$

Its Fourier transform reads

$$G^{++--}_{\eta\xi,s}(\boldsymbol{r}, \boldsymbol{r}'; \omega, \boldsymbol{P}) = e^{-\beta(\omega + 2M - 2\mu)} G^{--++}_{\eta\xi,s}(\boldsymbol{r}, \boldsymbol{r}'; \omega, \boldsymbol{P}), \tag{5.3}$$

where we introduced the Wigner-transformed four-point correlators according to

$$G^{\bullet\bullet\bullet\bullet}_{\eta\xi,s}\left(t, \boldsymbol{X} + \frac{\boldsymbol{r}}{2}, \boldsymbol{X} - \frac{\boldsymbol{r}}{2}; t', \boldsymbol{X}' + \frac{\boldsymbol{r}'}{2}, \boldsymbol{X}' - \frac{\boldsymbol{r}'}{2}\right)$$
$$\equiv \int_{\omega,\boldsymbol{P}} e^{-i\omega(t-t') + i\boldsymbol{P}\cdot(\boldsymbol{X}-\boldsymbol{X}')} G^{\bullet\bullet\bullet\bullet}_{\eta\xi,s}(\boldsymbol{r}, \boldsymbol{r}'; \omega, \boldsymbol{P}). \tag{5.4}$$

Here, we have used the fact that the operator $\hat{\rho}$ commutes with the Hamiltonian $\hat{H}$ and the translation operator $\hat{\boldsymbol{P}}$. Defining the two-particle spectral function as

$$G^{\rho}_{\eta\xi}(t\boldsymbol{xy}; t'\boldsymbol{zw}) \equiv G^{--++}_{\eta\xi,s}(t\boldsymbol{xy}; t'\boldsymbol{zw}) - G^{++--}_{\eta\xi,s}(t\boldsymbol{xy}; t'\boldsymbol{zw}), \tag{5.5}$$

our target component is formally solved in terms of the spectral function and chemical potential by utilizing the KMS relation for a grand canonical state:

$$G^{++--}_{\eta\xi,s}(x,x,x,x)$$
$$= \int_{\omega,\boldsymbol{P}} f_B(\omega + 2M - 2\mu) G^{\rho}_{\eta\xi}(\boldsymbol{0}, \boldsymbol{0}; \omega, \boldsymbol{P}) \tag{5.6}$$

$$\simeq \int_{\omega,\boldsymbol{P}} e^{-\beta(\omega + 2M - 2\mu)} G^{\rho}_{\eta\xi}(\boldsymbol{0}, \boldsymbol{0}; \omega, \boldsymbol{P}) \tag{5.7}$$

$$= e^{-2\beta(M-\mu)} \int_{-\infty}^{\infty} \frac{\mathrm{d}^3\boldsymbol{P}}{(2\pi)^3} e^{-\beta \boldsymbol{P}^2/4M}$$
$$\times \int_{-\infty}^{\infty} \frac{\mathrm{d}E}{(2\pi)} e^{-\beta E} G^{\rho}_{\eta\xi}(\boldsymbol{0}, \boldsymbol{0}; E). \tag{5.8}$$

In the second line we approximated the Bose-Einstein distribution f_B as Maxwell-Boltzmann, assuming the DM system to be dilute $T \ll (M - \mu)$. The same limit should also be taken in the explicit solution of the spectral function, as is done in the next section. In the last equality (5.8), we have used that the spectral function only depends on $E \equiv \omega - \boldsymbol{P}^2/4M$ (as we will explicitly see later) and adopted the loose notation $G^{\rho}_{\eta\xi}(\boldsymbol{0}, \boldsymbol{0}; \omega, \boldsymbol{P}) = G^{\rho}_{\eta\xi}(\boldsymbol{0}, \boldsymbol{0}; E)$. As a consequence of Fourier transformation, the energy integration in Eq. (5.8) ranges from minus infinity to plus infinity.

If the theory has bound-state solutions in the spectrum, the two-particle spectral function has strong contributions at particular negative values of E (binding energy). These contributions are further enhanced by the Boltzmann factor $e^{-\beta E}$ compared to the scattering solutions with positive energy, as can be directly seen from Eq. (5.8). This is expected from the assumption of a grand canonical system. All DM states of energy E must be populated with the Boltzmann factor $e^{-\beta E}$ for a given DM number density. As a result, the bound states are preferred compared to the scattering states because their energies are smaller.

Under the key assumption of the grand canonical ensemble, our main number density Eqs. (3.24)–(3.25) are formally closed. This is because the effective chemical potential appearing in Eq. (5.8) is related to the total number density by Legendre transformation. On the one hand, the validity of adopting a grand canonical state requires scattering processes to be efficient in order to keep DM in kinetic equilibrium with the plasma particles A_μ and ψ. For light mediators this is indeed guaranteed for times much later than the freeze-out. On the other hand, if the theory has bound-state solutions and is described by a grand canonical ensemble with a single chemical potential as we have introduced, there appears a hidden assumption on the internal chemical relation between scattering and bound-state contributions. As we will see later, it automatically implies the Saha condition for ionization equilibrium. Ionization equilibrium can only be achieved by efficient radiative processes like the ultrasoft emissions of mediators. In the description of our theory we have traced out from the beginning these contributions but can now formally include them by assuming ionization equilibrium. Thus, a grand canonical description of systems where bound states exist is only appropriate if the ionization equilibrium can be guaranteed. In Sec. VI D, we come back to this issue in detail, provide an explicit expression for the chemical potential and prove the implication of ionization equilibrium.

We would like to finally remark that once a grand canonical picture is appropriate, all finite temperature corrections to the annihilation or decay rate are contained in the solution of the two-particle spectral function in Eq. (5.8) through Eqs. (3.24) and (3.25). Indeed as we will see, the negative and positive energy solutions of the spectral function will merge continuously together if finite temperature effects are strong. In this case it turns out to be impossible to distinguish bound from scattering state solutions. Due to the form of Eq. (5.8) it is, however, not required to distinguish between these two contributions. Integrating the spectral function over the whole energy range automatically takes into account *all* contributions. In summary, for a grand canonical ensemble, the finite temperature corrections to Sommerfeld-enhanced annihilation and bound-state decay processes are contained in the two-particle spectral function and an explicit solution of the latter quantity is derived in following sections.

B. Two-particle spectral function in DM dilute limit

A key observation is that we have factored out in (5.8) the leading order contribution in the DM phase-space density and the remaining task is to find the two-particle spectral properties. For the computation of a two-particle spectral function we can now approximate the DM system to be *dilute* which is the limit $T \ll (M-\mu)$:

$$G^{+-}_{\eta/\xi}(x,y) \ll G^{-+}_{\eta/\xi}(x,y) \Rightarrow G^{-+}_{\eta/\xi}(x,y) \simeq G^{\rho}_{\eta/\xi}(x,y) \\ = G^{R}_{\eta/\xi}(x,y) - G^{A}_{\eta/\xi}(x,y), \qquad (5.9)$$

where G^ρ is the single particle spectral function as introduced in Sec. II. In the DM dilute limit, the two-particle spectral function is related to the imaginary part of the dilute solution of the retarded four-point correlator, where the result is given by:

$$\begin{aligned} G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E) &\simeq G^{--++}_{\eta\xi}(\mathbf{0},\mathbf{0};E) \\ &= G^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E) - G^{\mathrm{adv}}_{\eta\xi}(\mathbf{0},\mathbf{0};E) \\ &= 2\Im[iG^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E)]. \end{aligned} \qquad (5.10)$$

Here, we defined $G^{\mathrm{ret}}_{\eta\xi}$ and $G^{\mathrm{adv}}_{\eta\xi}$ as the DM dilute limit of the Eq. (4.21) for $G^{\mathrm{R}}_{\eta\xi}$ and $G^{\mathrm{A}}_{\eta\xi}$, respectively. For the rest of this section, the computation of the dilute limit of these equations is given, proving the claim $G^{--++}_{\eta\xi}(\mathbf{0},\mathbf{0};E) = 2\Im[iG^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E)]$.

Applying the DM dilute limit Eq. (5.9) to the two-time BS Eq. (4.21) for $G^{R}_{\eta\xi}$, one finds:

$$G^{\mathrm{ret}}_{\eta\xi} = G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0} + \int G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0}\Sigma^{\mathrm{ret}}_{\eta\xi}G^{\mathrm{ret}}_{\eta\xi}, \qquad (5.11)$$

$$\begin{aligned} \Sigma^{\mathrm{ret}}_{\eta\xi}(tt'\mathbf{xywz}) = {}& g_\chi^2 G^{R}_{\eta,0}(t\mathbf{x},t'\mathbf{w})\bar{G}^{R}_{\xi,0}(t\mathbf{y},t'\mathbf{z}) \\ &\times[-D^{-+}(t\mathbf{x},t'\mathbf{w}) - D^{-+}(t\mathbf{y},t'\mathbf{z}) \\ &+ D^{-+}(t\mathbf{x},t'\mathbf{z}) + D^{+-}(t'\mathbf{w},t\mathbf{y})], \end{aligned} \qquad (5.12)$$

where $\Sigma^{\mathrm{ret}}_{\eta\xi}$ is the dilute limit of the two-particle self-energy $\Sigma^{R}_{\eta\xi}$ in Eq. (4.23). All space-time dependences remain the same as in Eq. (4.21) but we suppress them hereafter for simplicity. Similar limit can be taken for the advanced component. Important to note is that due to the DM dilute limit, it can be recognized that the retarded Eq. (5.11) and hence the two-particle spectral function are *independent* of the DM number density. For the freeze-out process the DM dilute limit is an excellent approximation. Now to continue with the proof of $G^{--++}_{\eta\xi}(\mathbf{0},\mathbf{0};E) = 2\Im[iG^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E)]$, the equation for the $G^{--++}_{\eta\xi}$ [see Eq. (4.20)] in the DM dilute limit is

$$\begin{aligned} G^{--++}_{\eta\xi} = {}& G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0} + G^{A}_{\eta,0}\bar{G}^{A}_{\xi,0} \\ &+ \int G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0}\Sigma^{\mathrm{ret}}_{\eta\xi}[G^{\mathrm{ret}}_{\eta\xi} - G^{\mathrm{adv}}_{\eta\xi}] \\ &+ G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0}\sigma^{--++}_{\eta\xi,\mathrm{dil}}G^{\mathrm{adv}}_{\eta\xi} \\ &+ [G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0} + G^{A}_{\eta,0}\bar{G}^{A}_{\xi,0}]\Sigma^{\mathrm{adv}}_{\eta\xi}G^{\mathrm{adv}}_{\eta\xi} \end{aligned} \qquad (5.13)$$

$$\begin{aligned} = {}& G^{\mathrm{ret}}_{\eta\xi} - G^{\mathrm{adv}}_{\eta\xi} \\ &+ \int G^{R}_{\eta,0}\bar{G}^{R}_{\xi,0}[-\Sigma^{\mathrm{ret}}_{\eta\xi} + \Sigma^{\mathrm{adv}}_{\eta\xi} + \sigma^{--++}_{\eta\xi,\mathrm{dil}}]G^{\mathrm{adv}}_{\eta\xi}, \end{aligned} \qquad (5.14)$$

where in the step to the last equality we used two-time BS Eq. (4.21) backward. The statistical two-particle self-energy in the dilute limit is given by

$$\begin{aligned} &\sigma^{--++}_{\eta\xi,\mathrm{dil}}(tt'\mathbf{xywz}) \\ &= g_\chi^2[G^{R}_{\eta,0}(t\mathbf{x},t'\mathbf{w})\bar{G}^{R}_{\xi,0}(t\mathbf{y},t'\mathbf{z}) + G^{A}_{\eta,0}(t\mathbf{x},t'\mathbf{w})\bar{G}^{A}_{\xi,0}(t\mathbf{y},t'\mathbf{z})] \\ &\quad\times[-D^{-+}(t\mathbf{x},t'\mathbf{w}) - D^{-+}(t\mathbf{y},t'\mathbf{z}) \\ &\quad+ D^{-+}(t\mathbf{x},t'\mathbf{z}) + D^{+-}(t'\mathbf{w},t\mathbf{y})], \end{aligned} \qquad (5.15)$$

where we have used

$$\begin{aligned} G^{-+}_{\eta,0}(t,t')\bar{G}^{-+}_{\xi,0}(t,t') &\simeq G^{\rho}_{\eta,0}(t,t')\bar{G}^{\rho}_{\xi,0}(t,t') \\ &= G^{R}_{\eta,0}(t,t')\bar{G}^{R}_{\xi,0}(t,t') \\ &\quad + G^{A}_{\eta,0}(t,t')\bar{G}^{A}_{\xi,0}(t,t'). \end{aligned} \tag{5.16}$$

The first similarity is a consequence of the dilute limit. In the last equality we used $G^R G^A = 0$ for equal time products. The integral term in Eq. (5.14) vanishes by noting that in the dilute limit we have indeed $-\Sigma^{\rm ret}_{\eta\xi} + \Sigma^{\rm adv}_{\eta\xi} + \sigma^{--++}_{\eta\xi,\rm dil} = 0$ which finally proves the claim $G^{--++}_{\eta\xi}(\mathbf{0},\mathbf{0};E) = 2\Im iG^{\rm ret}_{\eta\xi}(\mathbf{0},\mathbf{0};E)$.

C. Retarded equation in static potential limit

In the previous section, we related the two-particle spectrum to the solution of the retarded equation in the DM dilute limit: $G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E) = 2\Im[iG^{\rm ret}_{\eta\xi}(\mathbf{0},\mathbf{0};E)]$. As a final step, we further simplify the two-time Bethe-Salpeter Eq. (5.11) for $G^{\rm ret}_{\eta\xi}$ by taking the proper static limit of the mediator correlator D, resulting finally in a simple Schrödinger-like equation. We start by acting with the inverse two-particle propagator from the left on Eq. (5.11), arriving at

$$\begin{aligned} &[\mathcal{G}^R_{\eta\xi}]^{-1} G^{\rm ret}_{\eta\xi}(t\mathbf{xy},t'\mathbf{zw}) \\ &= i\delta(t-t')\delta^3(\mathbf{x}-\mathbf{z})\delta^3(\mathbf{y}-\mathbf{w}) \\ &\quad + i\int_{t_2,\mathbf{x}_3,\mathbf{x}_4} \Sigma^{\rm ret}_{\eta\xi}(t\mathbf{xy},t_2\mathbf{x}_3\mathbf{x}_4) G^{\rm ret}_{\eta\xi}(t_2\mathbf{x}_3\mathbf{x}_4,t'\mathbf{zw}). \end{aligned} \tag{5.17}$$

Here, we suppress spin-indices for simplicity but quote the final result in full form later. Let us simplify the interaction kernel in Fourier space, where we take Wigner transform in time $\tau \equiv t - t'$:

$$\begin{aligned} &G^{\rm ret}_{\eta\xi}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}'_1,\mathbf{p}'_2,\omega) \\ &= \int \mathrm{d}^3x\mathrm{d}^3y\mathrm{d}^3z\mathrm{d}^3w\mathrm{d}\tau e^{i(\omega\tau-\mathbf{p}_1\cdot\mathbf{x}-\mathbf{p}_2\cdot\mathbf{y}+\mathbf{p}'_1\cdot\mathbf{z}+\mathbf{p}'_2\cdot\mathbf{w})} \\ &\quad \times G^{\rm ret}_{\eta\xi}(\mathbf{xyzw},\tau). \end{aligned} \tag{5.18}$$

Taking this Fourier transform of the kernel leads to two distinct parts:

$$\begin{aligned} \widehat{\int \Sigma^{\rm ret}_{\eta\xi} G^{\rm ret}_{\eta\xi}} &= g_\chi^2 \int \frac{\mathrm{d}^3q}{(2\pi)^3} [I_1(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\omega) \\ &\quad \times G^{\rm ret}_{\eta\xi}(\mathbf{p}_1,\mathbf{p}_2,\mathbf{p}'_1,\mathbf{p}'_2,\omega) \\ &\quad + I_2(\mathbf{p}_1,\mathbf{p}_2,\mathbf{q},\omega) G^{\rm ret}_{\eta\xi}(\mathbf{p}_1-\mathbf{q},\mathbf{p}_2+\mathbf{q},\mathbf{p}'_1,\mathbf{p}'_2,\omega)], \end{aligned} \tag{5.19}$$

where I_1 results from the sum of the two one-particle self-energy contributions, whereas I_2 originates from the exchange term between particle and antiparticle (see Fig. 2):

$$I_1 = (-i)\int \frac{\mathrm{d}\omega_1\mathrm{d}\omega_2\mathrm{d}\omega_3}{(2\pi)^3} \frac{G^{\rho}_{\eta,0}(\omega_1,\mathbf{p}_1-\mathbf{q})D^{-+}(\omega_3,\mathbf{q})\bar{G}^{\rho}_{\xi,0}(\omega_2,\mathbf{p}_2) + G^{\rho}_{\eta,0}(\omega_1,\mathbf{p}_1)D^{-+}(\omega_3,\mathbf{q})\bar{G}^{\rho}_{\xi,0}(\omega_2,\mathbf{p}_2-\mathbf{q})}{\omega-\omega_1-\omega_2-\omega_3+i\epsilon}, \tag{5.20}$$

$$I_2 = i\int \frac{\mathrm{d}\omega_1\mathrm{d}\omega_2\mathrm{d}\omega_3}{(2\pi)^3} \frac{G^{\rho}_{\eta,0}(\omega_1,\mathbf{p}_1-\mathbf{q})D^{-+}(\omega_3,\mathbf{q})\bar{G}^{\rho}_{\xi,0}(\omega_2,\mathbf{p}_2) + G^{\rho}_{\eta,0}(\omega_1,\mathbf{p}_1)D^{+-}(-\omega_3,\mathbf{q})\bar{G}^{\rho}_{\xi,0}(\omega_2,\mathbf{p}_2+\mathbf{q})}{\omega-\omega_1-\omega_2-\omega_3+i\epsilon}. \tag{5.21}$$

We can perform the two ω integrations over one-particle spectral function, where in the free limit they are given by

$$\begin{aligned} G^{\rho}_{\eta,0}(\omega,\mathbf{p}) &= (2\pi)\delta\left(\omega - \frac{\mathbf{p}^2}{2M}\right), \\ \bar{G}^{\rho}_{\xi,0}(\omega,\mathbf{p}) &= (2\pi)\delta\left(\omega - \frac{\mathbf{p}^2}{2M}\right). \end{aligned} \tag{5.22}$$

Now the integrals are reduced to

$$I_1 = (-i)\int \frac{\mathrm{d}\bar{\omega}}{(2\pi)} \frac{D^{-+}(\bar{\omega},\mathbf{q}) + D^{+-}(-\bar{\omega},\mathbf{q})}{\Omega_1 - \bar{\omega} + i\epsilon}, \tag{5.23}$$

$$I_2 = i\int \frac{\mathrm{d}\bar{\omega}}{(2\pi)} \frac{D^{-+}(\bar{\omega},\mathbf{q}) + D^{+-}(-\bar{\omega},\mathbf{q})}{\Omega_2 - \bar{\omega} + i\epsilon}, \tag{5.24}$$

where Ω_i contain the respective on-shell energies. Noticing that the Fourier transform of $D^{++}(t,r) = \theta(t)D^{-+}(t,r) + \theta(-t)D^{+-}(t,r)$ is given by

$$D^{++}(\Omega_i,\mathbf{q}) = i\int \frac{\mathrm{d}\bar{\omega}}{(2\pi)} \frac{D^{-+}(\bar{\omega},\mathbf{q}) + D^{+-}(-\bar{\omega},\mathbf{q})}{\Omega_i - \bar{\omega} + i\epsilon}, \tag{5.25}$$

we can now take the proper static limit $\Omega_i \to 0$ which results in

$$
\begin{aligned}
& i \widehat{\int \Sigma^{\mathrm{ret}}_{\eta\xi} G^{\mathrm{ret}}_{\eta\xi}} \\
& \quad = (-i) g_\chi^2 \int \frac{\mathrm{d}^3 q}{(2\pi)^3} [D^{++}(0,\mathbf{q}) G^{\mathrm{ret}}_{\eta\xi}(\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_1', \mathbf{p}_2', \omega) \\
& \qquad - D^{++}(0,\mathbf{q}) G^{\mathrm{ret}}_{\eta\xi}(\mathbf{p}_1 - \mathbf{q}, \mathbf{p}_2 + \mathbf{q}, \mathbf{p}_1', \mathbf{p}_2', \omega)]. \qquad (5.26)
\end{aligned}
$$

Introducing Wigner-momenta and Fourier transforming back with respect to the difference variables

$$
\begin{aligned}
& \mathbf{P} = (\mathbf{p}_1 + \mathbf{p}_2) - (\mathbf{p}_1' + \mathbf{p}_2'), \\
& \mathbf{p} = (\mathbf{p}_1 - \mathbf{p}_2)/2, \qquad \mathbf{p}' = (\mathbf{p}_1' - \mathbf{p}_2')/2, \qquad (5.27)
\end{aligned}
$$

$$
G^{\mathrm{ret}}_{\eta\xi}(\mathbf{r}, \mathbf{r}', \mathbf{P}, \omega) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{\mathrm{d}^3 p'}{(2\pi)^3} e^{i(\mathbf{p}\cdot\mathbf{r} - \mathbf{p}'\cdot\mathbf{r}')} G^{\mathrm{ret}}_{\eta\xi}(\mathbf{p}, \mathbf{p}', \mathbf{P}, \omega), \qquad (5.28)
$$

we finally end up with the Schrödinger-like equation for the retarded four-point correlator in the static limit of the potential:

$$
\begin{aligned}
& \left[\omega - \frac{\mathbf{P}^2}{4M} + \frac{\nabla^2_{\mathbf{r}}}{M} + i\epsilon - V_{\mathrm{eff}}(\mathbf{r})\right] G^{\mathrm{ret}}_{\eta\xi}(\mathbf{r}, \mathbf{r}'; \mathbf{P}, \omega) \\
& \quad = \mathrm{Tr}[\mathbf{1}_{2\times 2}] i \delta^3(\mathbf{r} - \mathbf{r}'). \qquad (5.29)
\end{aligned}
$$

Now we see that the retarded and hence also the spectral function only depends on $E \equiv \omega - \mathbf{P}^2/(4M)$.

The retarded two-time BS equation in static and dilute limit can be written as:

$$
\left[\frac{\nabla^2_{\mathbf{r}}}{M} + E + i\epsilon - V_{\mathrm{eff}}(\mathbf{r})\right] G^{\mathrm{ret}}_{\eta\xi}(\mathbf{r}, \mathbf{r}'; E) = \mathrm{Tr}[\mathbf{1}_{2\times 2}] i \delta^3(\mathbf{r} - \mathbf{r}'), \qquad (5.30)
$$

where the effective in-medium potential is defined as

$$
V_{\mathrm{eff}}(\mathbf{r}) \equiv -i g_\chi^2 \int_{-\infty}^{+\infty} \frac{\mathrm{d}^3 q}{(2\pi)^3} (1 - e^{i\mathbf{q}\cdot\mathbf{r}}) D^{++}(0, \mathbf{q}). \qquad (5.31)
$$

The spectral function, we would actually like to compute, is obtained from the solution of this equation according to the relation $G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E) = 2\Im[iG^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E)]$, as proven in the previous section. In the next section, we will derive the explicit solution of the retarded equation where we will further approximate the static mediator correlator D^{++} in the hard thermal loop limit, as already given in Eq. (3.11). The first term in the effective potential in Eq. (5.31) originates from the sum of the two single-particle self-energies, while the second term accounts for the mediator exchange between particle and antiparticle. The trace in the Schrödinger-like Eq. (5.30) takes the correct spin summation into account, which we have suppressed in this section for simplicity.

D. Explicit solution in static HTL approximation

Taking the static HTL approximation of the massless mediator as derived in Eq. (3.11), the effective potential according to Eq. (5.31) results in

$$
V_{\mathrm{eff}}(\mathbf{r}) = -\alpha_\chi m_D - \frac{\alpha_\chi}{r} e^{-m_D r} - i\alpha_\chi T \phi(m_D r), \qquad (5.32)
$$

$$
\phi(x) = 2 \int_0^\infty \mathrm{d}z \frac{z}{(z+1)^2} \left(1 - \frac{\sin(zx)}{zx}\right), \qquad (5.33)
$$

and $\phi(0) = 0$ and $\phi(\infty) = 1$. One can recognize the real part of the potential is corrected by the Debye mass as expected. It shifts the energy by $\alpha_\chi m_D$ (twice the Salpeter correction of single particle self-energies) and screens the Coulomb potential. At the same time, the effective potential contains an imaginary part. The physical meaning of this term is scatterings of DM with light particles in the thermal plasma. If particle and antiparticle are far away, the imaginary part must be solely determined by scattering with the thermal plasma without the Yukawa force. One can also see that this is actually the case, since the imaginary part becomes twice the thermal width of single particles $-i\alpha_\chi T$ for $r \to \infty$ (see single particle corrections in Appendix E 2, as well as Salpeter correction in Appendix E 3). This property follows from our resummation scheme, treating the DM self-energy on an equal footing with the ladder diagram exchange. Since the finite temperature corrections introduce a constant imaginary term for large distances, we can drop the $i\epsilon$ term in the following derivation of the explicit solution of Eq. (5.30). This solution will be general and contains also the correct vacuum limit, where $i\epsilon$ has to be carefully taken into account. The effective potential in Eq. (5.32) was first obtained in [73] and reproduced subsequently by other methods [70,75]. Let us remark that we derived it independently, i.e., for the first time starting from a set of two-time Bethe-Salpeter equations on the Keldysh contour.

To derive the solution, we expand $G^{\mathrm{ret}}_{\eta\xi}$ in terms of partial waves

$$
G^{\mathrm{ret}}_{\eta\xi}(\boldsymbol{r}, \boldsymbol{r}'; E) = \sum_l \frac{2l+1}{4\pi} P_l(\cos\theta_{\boldsymbol{r},\boldsymbol{r}'}) (-i) G^{\mathrm{ret}}_{\eta\xi,l}(r, r'; E), \qquad (5.34)
$$

leading to the $l = 0$ (s-wave) equation:

$$
\begin{aligned}
& \left[-\frac{1}{M}\frac{1}{r^2}\partial_r(r^2\partial_r) - E + V_{\mathrm{eff}}(r)\right] G^{\mathrm{ret}}_{\eta\xi,s}(r, r'; E) \\
& \quad = \mathrm{Tr}[\mathbf{1}_{2\times 2}] \frac{1}{rr'} \delta(r - r'). \qquad (5.35)
\end{aligned}
$$

The physically relevant boundary conditions we impose on $G^{\mathrm{ret}}_{\eta\xi,s}$ are listed below.

(i) $G^{\mathrm{ret}}_{\eta\xi,s}(r,r';E)$ is finite $\forall\, r,\, r'$.
(ii) For $|r-r'|\to\infty$, $G^{\mathrm{ret}}_{\eta\xi,s}$ decays exponentially (as a consequence of constant imaginary potential).

Note here that, since we are working in the dilute limit, the Feynman propagator and the retarded function are the same. These requirements set the form of the solution uniquely (see also [46,76]):

$$G^{\mathrm{ret}}_{\eta\xi,s}(r,r';E)=\mathrm{Tr}[\mathbf{1}_{2\times2}]\frac{M}{rr'}[\theta(r-r')g_>(r)g_<(r') + \theta(r'-r)g_<(r)g_>(r')], \tag{5.36}$$

where $g_{>/<}$ are the solutions of the homogeneous differential equation

$$\left[-\frac{1}{M}\partial_r^2-E+V_{\mathrm{eff}}(r)\right]g_{>/<}(r)=0, \tag{5.37}$$

whose boundary conditions are given as follows:
(i) $g_<(0)=0$ and $g'_<(0)=1$;
(ii) $g_>(0)=1$ and $g_>(r)$ decays exponentially for $r\to\infty$.

One can explicitly check that the solution of the form (5.36) with the boundary conditions for $g_{>/<}$ fulfils the requirements on $G^{\mathrm{ret}}_{\eta\xi,s}$ as listed above.

The two-particle spectral function, needed to compute the annihilation or decay rate according to Eqs. (5.8) and (5.10), can be written in terms of the homogeneous solution as

$$\begin{aligned}G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{l=0}&=2\Im\left[iG^{\mathrm{ret}}_{\eta\xi}(\mathbf{0},\mathbf{0};E)\Big|_{l=0}\right]\\&=\frac{1}{2\pi}\Im\left[\lim_{r,r'\to0}G^{\mathrm{ret}}_{\eta\xi,s}(r,r';E)\right]\\&=\frac{1}{2\pi}\mathrm{Tr}[\mathbf{1}_{2\times2}]M\Im[g'_>(0)].\end{aligned} \tag{5.38}$$

In the last term, prime stands for the derivative with respect to r. Formally this solution is correct but might be troublesome in the numerical evaluation, since the real part of $g'_>(0)$ has a singularity due to the $1/r$ behavior of the effective potential. To resolve this issue, we rewrite the imaginary part of $g'_>(0)$ by means of a different solution. We closely follow the discussion given in Ref. [76]. Let us define another singular solution g_s whose boundary conditions are given by $g_s(0)=1$ and $\Im[g'_s(0)]=0$. The outgoing solution for $r\to\infty$ can be expressed by a linear combination of

$$g_>=g_s+Bg_<. \tag{5.39}$$

By definition, we have $\Im[g'_>(0)]=\Im[B]$. The fact that $g_>$ is outgoing forces it to decay for $r\to\infty$, which determines B. And thus, one finds

$$\Im[g'_>(0)]=\Im[B]=-\Im\left[\lim_{r\to\infty}\left[\frac{g_s(r)}{g_<(r)}\right]\right]. \tag{5.40}$$

The physical quantity $\Im[B]$ we would like to compute does not depend on the choice of $\Re[g'_s(0)]$. This is why we can get the correct result without handling the divergence of $\Re[g'_s(0)]$. The final step is to rewrite the singular solution as

$$g_s(r)=-g_<(r)\int_0^r \mathrm{d}r'\frac{1}{g^2_<(r')}. \tag{5.41}$$

One may also write down the expression for $g_>$ by using Eq. (5.41):

$$g_>(r)=g_<(r)\int_r^\infty \mathrm{d}r'\frac{1}{g^2_<(r')}. \tag{5.42}$$

One can check that it fulfills the boundary conditions $g_s(0)=g_>(0)=\lim_{r\to0}r(1/r)=1$ and $\Im[g'_s(0)]=\Im[g_>(0)]-\Im[B]=0$. Plugging Eq. (5.41) into Eq. (5.40) and recalling the relation Eq. (5.38), we finally arrive at the convenient form to evaluate the imaginary part (see also [64]):

$$\begin{aligned}\Im[G^{\mathrm{ret}}_{\eta\xi,s}(0,0;E)]&=\mathrm{Tr}[\mathbf{1}_{2\times2}]M\Im[g'_>(0)]\\&=\mathrm{Tr}[\mathbf{1}_{2\times2}]M\lim_{\delta\to0}\int_\delta^\infty \mathrm{d}r\Im\left(\frac{1}{g_<(r)}\right)^2.\end{aligned} \tag{5.43}$$

The great advantage of this general solution is that it applies for the whole two-particle energy spectrum of our theory, i.e., for negative and positive E. Here, we have introduced the tolerance $\delta\ll1$ as initial value for numerical studies. Let us finally introduce dimensionless variables, expressing distances in terms of the Bohr radius $x\equiv\alpha_\chi Mr$, and summarize the equations in terms of these units. The s-wave part of two-particle spectral function reads:

$$G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{l=0} =\frac{1}{2\pi}\mathrm{Tr}[\mathbf{1}_{2\times2}]\alpha_\chi M^2\lim_{x'\to0}\int_{x'}^\infty \mathrm{d}x\Im\left[\frac{1}{g_<(x)g_<(x)}\right]. \tag{5.44}$$

The homogeneous equations for a massless and massive mediator are given by:

$$g''_<(x)+\left[\frac{E}{\alpha_\chi^2M}+\frac{m_D}{\alpha_\chi M}+\frac{1}{x}e^{-\frac{m_D}{\alpha_\chi M}x}+i\frac{T}{\alpha_\chi M}\phi\left(\frac{m_D}{\alpha_\chi M}x\right)\right]g_<(x)=0, \tag{5.45}$$

$$g''_{<}(x) + \left[\frac{E}{\alpha_\chi^2 M} + \frac{\sqrt{m_V^2 + m_D^2} - m_V}{\alpha_\chi M} + \frac{1}{x} e^{-\frac{\sqrt{m_V^2+m_D^2}}{\alpha_\chi M} x} + i \frac{T}{\alpha_\chi M} \frac{1}{\sqrt{1 + m_V^2/m_D^2}} \phi\left(\frac{\sqrt{m_V^2 + m_D^2}}{\alpha_\chi M} x\right)\right] g_{<}(x) = 0. \tag{5.46}$$

In practice, we use the following initial conditions which can be obtained by power series approach:

$$g_{<}(x) = x - x^2/2 + i\gamma_i x^5, \tag{5.47}$$

$$\gamma_C = -\frac{1}{40} \frac{T}{\alpha_\chi M} \left[\frac{m_D}{\alpha_\chi M}\right]^2, \tag{5.48}$$

$$\gamma_Y = -\frac{1}{40} \frac{T}{\alpha_\chi M} \frac{1}{\sqrt{1 + (m_V/m_D)^2}} \left[\frac{\sqrt{m_D^2 + m_V^2}}{\alpha_\chi M}\right]^2, \tag{5.49}$$

where γ_C applies for the Coulomb case Eq. (5.45) while γ_Y can be taken for the Yukawa case Eq. (5.46). To efficiently deal with the sometimes highly oscillatory integrand in Eq. (5.44), local adaptive integrating methods are useful. In finding the initial power in x for the imaginary part we have assumed that $\phi(x) \sim \frac{1}{2}x^2$ for small x [64] which is only approximately true. In general, one has to carefully check if the correct value of the potential is close enough to this approximation at the initial value which we have done for the numerical results presented in subsequent sections.

VI. DM NUMBER DENSITY EQUATION IN GRAND CANONICAL ENSEMBLE

In the previous section, we have obtained a formal solution of the out-of-equilibrium term $G^{++--}_{\eta\xi,s}$, entering our main number density Eq. (3.24). This was achieved by assuming the DM system is in a grand canonical state, formally solving $G^{++--}_{\eta\xi,s}$ in terms of the chemical potential and two-particle spectral function by KMS relation. Inserting this formal solution given in Eq. (5.8) into the main number density Eq. (3.24) results in our master formula for the DM number density equation in a grand canonical ensemble:

$$\dot{n}_\eta + 3Hn_\eta = -2(\sigma v_{\rm rel}) G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\rm eq} (e^{\beta 2\mu_\eta[n_\eta]} - 1), \tag{6.1}$$

where a symmetric plasma $2\mu_\eta = \mu_\eta + \mu_\xi$ is assumed and $(\sigma v_{\rm rel})$ is the s-wave tree-level annihilation cross section. The latter quantity is averaged over initial internal d.o.f. (spin) and summed over final.

The chemical potential $\mu_\eta[n_\eta]$ is a function of the total number density n_η as it appears on the left-hand side of our master formula. The term $G^{++--}_{\eta\xi,s}|_{\rm eq}$ is the chemical equilibrium limit $\mu \to 0$ of Eq. (5.8), given by

$$G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\rm eq} = e^{-\beta 2M} \int_{-\infty}^{\infty} \frac{{\rm d}^3 \boldsymbol{P}}{(2\pi)^3} e^{-\beta \boldsymbol{P}^2/4M} \times \int_{-\infty}^{\infty} \frac{{\rm d}E}{(2\pi)} e^{-\beta E} G^{\rho}_{\eta\xi}(\boldsymbol{0},\boldsymbol{0};E)|_{l=0}. \tag{6.2}$$

We presented a general method in the previous section of how to compute the in-medium two-particle spectral function $G^{\rho}_{\eta\xi}$ explicitly. It contains finite temperature corrections to the Sommerfeld enhancement and bound-state decay. The only parameter left in Eq. (6.1) is the chemical potential, which has not yet been explicitly solved. The chemical potential $\mu_\eta[n_\eta]$ can be obtained in two steps as demonstrated in the following. First, the total number density as a function of the chemical potential is computed. For a grand canonical ensemble this follows from basic relations of quantum statistical mechanics and is given by

$$n_\eta[\mu_\eta] = \frac{\partial p}{\partial \mu_\eta}\bigg|_{T,\Omega}, \tag{6.3}$$

where the total pressure is

$$p\Omega = T \ln Z_{\rm gr}(\Omega, T, \mu) = T \ln {\rm Tr}[e^{-\beta(H - \mu_\eta N_\eta - \mu_\xi N_\xi)}]. \tag{6.4}$$

Here, Ω is the volume and $Z_{\rm gr}$ is the grand canonical partition function. Second, by inversion of Eq. (6.3) one obtains the chemical potential as a function of the total number density. The functional dependence of $\mu[n_\eta]$ on the total number density can be nontrivial especially for the case if bound-state solutions exist as we will see later. The n_η on the l.h.s. of Eq. (6.3) is equivalent to n_η appearing on the l.h.s. of our master Eq. (6.1).

In subsequent sections of this chapter we demonstrate how powerful our master Eq. (6.1) is. We self-consistently compute the component $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\rm eq}$ and the chemical potential μ. This means in both terms the same approximations should be made to obtain a well-behaved number density equation.

For a better understanding, we would like to start in the next Sec. VI A with the simplest case of our theory by taking the zero self-interaction and zero finite temperature correction limit. This means we take $g_\chi \to 0$ and $g_\psi \to 0$ in the effective in-medium potential Eq. (5.32) and compute the spectral function. The same limits are applied to the Hamiltonian entering in Eq. (6.4) to compute the total pressure. Under these limits, our master Eq. (6.1) reduces to the conventional Lee-Weinberg equation, describing constant s-wave annihilation of DM.

As a next step, we allow for long-range self-interactions but neglect finite temperature corrections. This corresponds to the limit $g_\psi \to 0$, leading to the fact that the remaining term in the effective in-medium potential is the standard Coulomb or unscreened Yukawa potential. In Sec. VI B, the two-particle spectrum for this simple case of our theory is shown. We stress the point that only in this limit there is a direct relation between spectral function and standard expressions for the Sommerfeld-enhancement factor or the decay width of the bound states. Section VI C completes the results by computing the chemical potential for the same limit. Combining the analytic expressions for the spectral function and chemical potential, we prove our master formula Eq. (6.1) to be consistent with the classical on-shell Boltzmann equation treatment for vanishing thermal corrections. We also point out that adopting a grand canonical ensemble with one single time-dependent chemical potential as in our master formula implies ionization equilibrium between the scattering and bound states. A detailed discussion is given on the validity of ionization equilibrium during the freeze-out process. If no bound-state solutions exist, the only limitation of our master formula is essentially kinetic equilibrium [77,78].

We relax the assumption of zero finite temperature corrections in Sec. VI D. This brings us to another central result of this work: a DM number density equation, generalizing the conventional Lee-Weinberg equation and classical on-shell Boltzmann equation treatment as a consequence of accounting simultaneously for DM annihilation and bound-state decay at finite temperature. However, it should be noted that this equation in Sec. VI D strictly speaking only applies to the narrow thermal width case and is therefore less general compared to our master Eq. (6.1). This means we have neglected in Sec. VI D imaginary-part corrections to the effective in-medium potential for the computation of the chemical potential. While we can fully account for these non-Hermite corrections in the computation of $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$, it remains an open question of this work of how to consistently compute the chemical potential for the broad thermal width case. The broad thermal width case for $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$ we compute numerically later in this work (see Sec. VII). Nevertheless, we demonstrate that the chemical potential and the two-particle spectral function entering the number density equation in Sec. VI D can be evaluated self-consistently in the narrow thermal width limit. This approach, taking leading finite temperature real-part corrections into account, is already more general of what has been computed so far in the literature. In principle, it is possible to take a nonconsistent approach and compute $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$ including imaginary parts in the potential while only including real-part corrections to the chemical potential. However, some care must be taken when doing so. This is because the chemical potential corrects the functional form of the number density dependence in our master equation. We discuss in more detail the possibility of taking a non-self-consistent approach by the end of Sec. VI D.

Finally in Sec. VI E, we compare our master Eq. (6.1) to the previous literature, relying on the method of linear response theory. Consistency is proven in the linear regime close to chemical equilibrium.

A. Recovering the Lee-Weinberg equation

We take the limit of zero self-interactions $\alpha_\chi \to 0$ while keeping the annihilation term Γ_s as a perturbation. It should be emphasized again that we have to approximate the spectral function and the chemical potential both in the same limit in order to obtain a self-consistent solution. The free spectral function without self-interactions and the ideal pressure are given by

$$G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{l=0} = \theta(E)\frac{1}{2\pi}\mathrm{Tr}[\mathbf{1}_{2\times 2}]M^{3/2}E^{1/2}, \tag{6.5}$$

$$p_0\Omega = T\ln\mathrm{Tr}[e^{-\beta(H_0-\mu_\eta N_\eta-\mu_\xi N_\xi)}]. \tag{6.6}$$

Here, H_0 is the free Hamiltonian and for a derivation of this result for the two-particle spectral function directly starting from the general expression Eq. (5.44) can be found in Appendix D. The number density can be obtained from Eq. (6.3) by using the ideal pressure:

$$\begin{aligned} n_\eta[\mu_\eta] &= \frac{\partial p_0}{\partial \mu_\eta}\bigg|_{T,\Omega} = G^{+-}_{\eta,0}(x,x) \\ &= e^{\beta\mu}2\int\frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}e^{-\beta(M+\mathbf{p}^2/2M)} = e^{\beta\mu}n^{\mathrm{eq}}_{\eta,0}, \end{aligned} \tag{6.7}$$

$$n^{\mathrm{eq}}_{\eta,0} = 2\left(\frac{MT}{2\pi}\right)^{3/2}e^{-\beta M}. \tag{6.8}$$

In the second equality of the first line, we find the relation between ideal number density and the noninteracting correlator $G^{+-}_{\eta,0}(x,x)$. The latter quantity has to be evaluated in a grand canonical ensemble, which we have done in the third equality by using KMS condition and the DM dilute limit (see Appendix E 1). The DM dilute limit should be taken in the computation of $G^{+-}_{\eta,0}(x,x)$ to be consistent with the computation of the spectral function. For the latter quantity we have seen in Sec. V B only in the DM dilute limit it is independent of the DM number density and our general solution Eq. (5.44) relies on this assumption. In the last equality of the first line we defined the conventional chemical equilibrium number density of ideal particles. Finally, we obtain from the last equality the noninteracting (ideal) chemical potential by inversion: $\beta\mu = \ln[n_\eta/n^{\mathrm{eq}}_{\eta,0}]$. Note that this inversion can only be done analytically if one approximates the Fermi-Dirac distribution as Maxwell-Boltzmann (which is our DM dilute limit). Entering these results of the spectral function and the chemical potential into our master formula for the DM number density Eq. (6.1), leads to the conventional Lee-Weinberg equation for DM particles with zero self-interactions:

$$\dot{n}_\eta + 3Hn_\eta = -\langle \sigma v_{\mathrm{rel}} \rangle [n_\eta^2 - (n_{\eta,0}^{\mathrm{eq}})^2]. \tag{6.9}$$

Here, we have recovered the standard thermal averaged cross section by using the simple substitution $E = Mv_{\mathrm{rel}}^2/4$ for the positive energy spectrum:

$$\langle \sigma v_{\mathrm{rel}} \rangle = \frac{(M/T)^{3/2}}{2\sqrt{\pi}} \int_0^\infty \mathrm{d}v_{\mathrm{rel}} e^{-\frac{v_{\mathrm{rel}}^2 M}{4T}} v_{\mathrm{rel}}^2 (\sigma v_{\mathrm{rel}}) \tag{6.10}$$

$$= (\sigma v_{\mathrm{rel}}). \tag{6.11}$$

The last equality holds for constant s-wave annihilation cross section $(\sigma v_{\mathrm{rel}})$ as it is the case for our model.

B. Spectral function, Sommerfeld enhancement factor, and decay width for vanishing thermal corrections

We turn now to the interacting case $\alpha_\chi \neq 0$ and compute the two-particle spectral function. The s-wave two-particle spectral function is numerically solved according to Eq. (5.44) in the limit of vanishing finite temperature corrections and the results are shown in Fig. 3. Poles in the negative energy spectrum represent the bound states, while the spectrum is continuous for the scattering states at positive energy. In the vacuum limit one clearly sees that the scattering states can be separated from the bound-state contribution at $E = 0$. Due to this separation, the solution of the spectral function is directly related to the Sommerfeld enhancement factor $S(v_{\mathrm{rel}})$ and bound-state decay width Γ_n. These relations are given below and now it becomes clear that the two-particle spectral function as shown in Fig. 3 is, for the vacuum case, just a convenient way of presenting all contributions simultaneously.

The relations between two-particle spectral function $G_{\eta\xi}^{\rho}$, Sommerfeld enhancement factor S, and decay width Γ_n in the limit of vanishing finite temperature corrections are given by:

$$(\sigma v_{\mathrm{rel}}) G_{\eta\xi}^{\rho}(\mathbf{0},\mathbf{0};E)|_{E>0,l=0} = \frac{1}{4\pi} \mathrm{Tr}[\mathbf{1}_{2\times 2}] M^2 v_{\mathrm{rel}} (\sigma v_{\mathrm{rel}}) S(v_{\mathrm{rel}}), \tag{6.12}$$

$$(\sigma v_{\mathrm{rel}}) G_{\eta\xi}^{\rho}(\mathbf{0},\mathbf{0};E)|_{E<0,l=0} = \frac{\pi}{2} \mathrm{Tr}[\mathbf{1}_{2\times 2}] \sum_n \delta(E - E_{B_n}) \Gamma_n, \tag{6.13}$$

where $E = Mv_{\mathrm{rel}}^2/4$ for the scattering states, E_{B_n} is the (negative) binding energy for the bound states, and $(\sigma v_{\mathrm{rel}})$ is the tree-level s-wave annihilation cross section. Γ_n is decay width of the bound state (not to be confused with our annihilation term Γ_s at the beginning of this work).

These relations can be proven directly from our general solution Eq. (5.44), see Appendix D for a derivation. On a first look, the spectral function in the vacuum case seems just a nice way of presentation. Instead one should emphasize that the notion of spectral function is more general and unifies the picture of scattering state annihilation and bound-state decay. This observation becomes important for the finite temperature case discussed in Sec. VII, where it is impossible to separate or distinguish between annihilation and decay. The spectrum includes both. Only in the absolute vacuum case a clear distinction between annihilation and decay can be made.

Coming back to Fig. 3 and now keeping in mind the relations Eqs. (6.12)–(6.13). There is an infinite number of exited S-bound states for the Coulomb case (left plot) with binding energy $E_{B_n} = -\alpha_\chi^2 M/(4n^2)$, where n is the number of the exited states and $n = 1$ is the ground state with lowest binding energy, shown as the pole most to the left. At small positive energies, where $v_{\mathrm{rel}} \lesssim \alpha_\chi$, the spectral function is constant, resulting in the familiar scaling $S(v_{\mathrm{rel}}) \propto v_{\mathrm{rel}}^{-1}$ according to Eq. (6.12).

In the Yukawa potential case, shown in the right plot of Fig. 3, there is a finite number of bound-state solutions. For certain ratios of $\epsilon_\phi \equiv m_V/(\alpha_\chi M)$ there exist a bound-state solution with zero binding energy ($E = 0$). For those special cases the Sommerfeld enhancement factor scales as $S(v_{\mathrm{rel}}) \propto v_{\mathrm{rel}}^{-2}$ for $v_{\mathrm{rel}} \lesssim m_V/M \ll \alpha_\chi$, called on-resonance regime, leading to an interesting observational impact on cosmology at very late times [79,80]. Roughly, those poles where the spectral function would diverge are at the on-resonance condition $\epsilon_\phi = 6/(m^2\pi^2)$ where m is integer.[8] The on-resonance divergences give rise to partial wave unitarity violation of the total cross section, as can be seen in the right plot of Fig. 3 at $E = 0$. It has been pointed out in [54] that once the imaginary contribution of the annihilation part (proportional to our Γ_s) is included self-consistently in the solution of the Schrödinger-like Eq. (5.30), then the Sommerfeld enhancement starts to saturate below the unitarity limit.[9] This means that for some small velocity there is a transition from the divergent

---

[8] This is exactly true for the Hulthén potential case which is analytically solvable, while for the Yukawa case this resonance condition slightly deviates.

[9] Since we have treated the annihilation term Γ_s as a perturbation, where the leading order contribution was found to contribute to the change in the number density, the contribution of Γ_s does not occur in Eq. (5.30). We will see later that at finite temperature the imaginary parts in the potential will dominate over the imaginary contribution from the annihilation term and thus at sufficiently finite temperature we will always get a saturation below the unitarity bound.

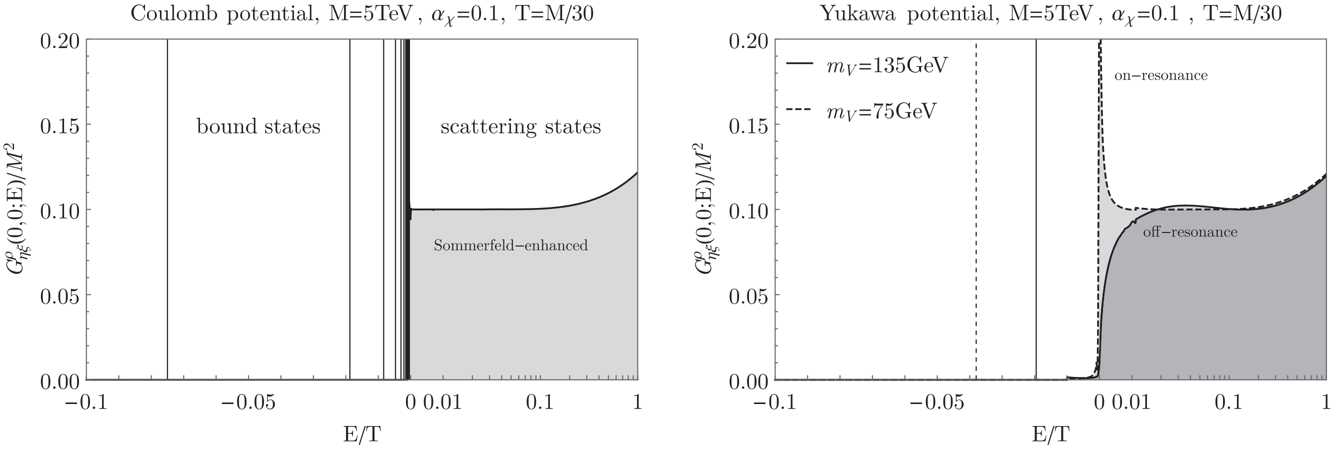


FIG. 3. S-wave two-particle spectral function vs the energy E in units of typical freeze-out temperature shown for a standard Coulomb (left) and Yukawa (right) potential. The two-particle spectral function enters directly our master formula and is weighted by the Boltzmann factor for all the energy range.

scaling $S(v_{\rm rel}) \propto v_{\rm rel}^{-2}$ to $S(v_{\rm rel}) \propto$ const which results in zero spectrum at zero energy. The saturation is always present if the on-resonance condition is not exactly fulfilled. The other extreme case is if ϵ_ϕ is taken exactly in between neighboring on-resonance values, called off-resonance. Then, $S(v_{\rm rel})$ never scales stronger than $S(v_{\rm rel}) \propto v_{\rm rel}^{-1}$ and at some small velocity of the order $v_{\rm rel} \lesssim m_V/M$ the Sommerfeld enhancement factor starts to saturate and the spectral function approaches zero.

C. DM number density equation for vanishing thermal corrections

In the previous section we have proven in the limit of vanishing finite temperature corrections a relation between spectral function, standard expression of Sommerfeld enhancement factor and the bound-state decay width. Inserting these relations Eqs. (6.12)–(6.13) into our master Eq. (6.1) leads to the following differential equation for the total number density n_η:

$$\dot{n}_\eta + 3Hn_\eta = -\langle\sigma v_{\rm rel}\rangle[(n^{\rm eq}_{\eta,0})^2 e^{\beta 2\mu} - (n^{\rm eq}_{\eta,0})^2] - \sum_i \Gamma_i [n^{\rm eq}_{B_i,0} e^{\beta 2\mu} - n^{\rm eq}_{B_i,0}]. \tag{6.14}$$

In the limit of zero chemical potential, defining chemical equilibrium, the r.h.s. vanishes as expected. Here, we recovered the thermal averaged Sommerfeld enhancement factor:

$$\langle\sigma v_{\rm rel}\rangle = \frac{(M/T)^{3/2}}{2\sqrt{\pi}} \int_0^\infty {\rm d}v_{\rm rel}\, e^{-\frac{v_{\rm rel}^2 M}{4T}} v_{\rm rel}^2 (\sigma v_{\rm rel}) S(v_{\rm rel}). \tag{6.15}$$

The chemical equilibrium number density for the scattering states $n^{\rm eq}_{\eta,0}$ was coming out as already defined in Eq. (6.8). This outcome is fully consistent with the result one would get from integrating the Maxwell-Boltzmann equilibrium phase-space density (multiplied by spin factor 2) of non-relativistic particles. The chemical equilibrium number density of the bound-states was defined as

$$n^{\rm eq}_{B_i,0} = \left(\frac{2MT}{2\pi}\right)^{3/2} e^{-\beta M_{B_i}}, \tag{6.16}$$

where the mass of bound-state i is $M_{B_i} = 2M - |E^i_B|$ and the subscript 0 stands for ideal bound states, respectively. The term in front of the exponential in the bound-state number density Eq. (6.16) needs some further explanation. Since we have only considered s-wave contributions to the spectral function, $n^{\rm eq}_{B_i,0}$ is the equilibrium number density of the ith exited para-WIMPonium. The decay, as well as the annihilation, of ortho-WIMPonium into three A_μ would be a p-wave process. To form para-WIMPonium there is only one spin option while for ortho-WIMPonium there are 3, consistent with the picture of having in total 4 spin d.o.f. Therefore, the spin factor 1 in Eq. (6.16) comes out correctly. When carefully looking at the term in front of the exponential in Eq. (6.16), it can be seen that the normalization of the distribution came out as like integrating the phase space density:

$$n^{\rm eq}_{B_i,0} = \int \frac{{\rm d}^3\mathbf{P}}{(2\pi)^3} e^{-\beta(M_{B_i} + \mathbf{P}/4M)}. \tag{6.17}$$

The kinetic term $\mathbf{P}/4M$ of the bound state misses the correction coming from the binding energy, because the conventional normalization would give $(n^{\rm eq}_{B_i,0})_{(c)} = (M_{B_i}/2\pi)^{3/2} e^{-\beta M_{B_i}}$. The reason why this correction of the order $\mathcal{O}(E^i_B/M)$ does not come out as in the conventional case can be explained by how we have approximated the

Martin-Schwinger hierarchy. In our paper, we expand the equation for the four-point correlator around the product of two free propagators. This is why the spectral function only depends on $E = \omega - \mathbf{P}/4M$ as we have shown in Eq. (5.29). If we would iterate the solution, e.g., correcting the free correlators and inserting them again into the solution of the four-point correlator, we could obtain the conventional result. However, note that the correction is small for perturbative systems, since typically $\mathcal{O}(E^i_B/M) \sim \mathcal{O}(\alpha^2)$.

We are turning to the discussion of the chemical potential μ in Eq. (6.14). The chemical potential can be obtained by inverting the number density as a function of the chemical potential. For an *ideal gas* description it is known that the total number density of η-particles n_η, as it appears on the l.h.s. of Eq. (6.14), is in general just given by the sum of scattering and bound-state contributions:

$$n_\eta = n_{\eta,0} + \sum_i n_{B_i,0} \tag{6.18}$$

$$n_{\eta,0} = n^{\mathrm{eq}}_{\eta,0} e^{\beta\mu_\eta}, \tag{6.19}$$

$$n_{\xi,0} = n^{\mathrm{eq}}_{\xi,0} e^{\beta\mu_\xi}, \tag{6.20}$$

$$n_{B_i,0} = n^{\mathrm{eq}}_{B_i,0} e^{\beta\mu_{B_i}}. \tag{6.21}$$

Since we have imposed a grand canonical state with only one single chemical potential, the chemical potentials for the scattering and bound states are related and therefore the number densities are not independent quantities. Assuming a grand canonical ensemble with only one time dependent chemical potential μ implies $2\mu = 2\mu_\eta = 2\mu_\xi = \mu_{B_i}$, which leads to the relation

$$n_{B_i,0} = 2\left(\frac{\pi}{MT}\right)^{3/2} n^2_{\eta,0} e^{-\beta|E^i_B|}. \tag{6.22}$$

This is nothing but the *Saha ionization equilibrium condition*. To see it explicitly, let us insert Eq. (6.22) into Eq. (6.18), leading to a quadratic equation for the number density of free scattering states:

$$n_\eta = n_{\eta,0} + K^{\mathrm{id}}(T) n_{\eta,0} n_{\eta,0}, \quad K^{\mathrm{id}}(T) = \frac{\sum_i n_{B_i,0}}{n_{\eta,0} n_{\eta,0}}. \tag{6.23}$$

$K^{\mathrm{id}}(T)$ is according to Eq. (6.22) independent of the chemical potential. This quadratic Eq. (6.23) can be solved, leading to the degree of ionization α^{id} for ideal gases:

$$\frac{n_{\eta,0}}{n_\eta} = \alpha^{\mathrm{id}}(n_\eta K^{\mathrm{id}}(T)), \qquad \alpha^{\mathrm{id}}(x) = \frac{1}{2x}(\sqrt{1+4x} - 1). \tag{6.24}$$

The chemical potential can now be obtained from this equation by using $n_{\eta,0} = n^{\mathrm{eq}}_{\eta,0} e^{\beta\mu}$ resulting in

$$\beta\mu = \ln\left[\frac{\alpha^{\mathrm{id}} n_\eta}{n^{\mathrm{eq}}_{\eta,0}}\right]. \tag{6.25}$$

Inserting this chemical potential into Eq. (6.14) we finally end up with the Boltzmann equation for vanishing finite temperature corrections, ideal gas approximation, and the system in a grand canonical state:

$$\begin{aligned}\dot{n}_\eta + 3Hn_\eta = &-\langle\sigma v_{\mathrm{rel}}\rangle[(\alpha^{\mathrm{id}} n_\eta)^2 - (n^{\mathrm{eq}}_{\eta,0})^2] \\ &- \sum_i \Gamma_i n^{\mathrm{eq}}_{B_i,0}\left[\left(\frac{\alpha^{\mathrm{id}} n_\eta}{n^{\mathrm{eq}}_{\eta,0}}\right)^2 - 1\right].\end{aligned} \tag{6.26}$$

The equation is closed in terms of the total number density n_η. Before we discuss this result in detail, let us consider the case where we would have treated all bound and scattering states to be independent. This could have been realized in Eq. (6.14) by assigning different chemical potentials to scattering and bound states. Then we would have ended up with decoupled equations:

$$\begin{aligned}\dot{n}_{\eta,0} + 3Hn_{\eta,0} &= -\langle\sigma v_{\mathrm{rel}}\rangle[n^2_{\eta,0} - (n^{\mathrm{eq}}_{\eta,0})^2], \\ \sum_i \dot{n}_{B_i,0} + 3Hn_{B_i,0} &= -\sum_i \Gamma_i[n_{B_i,0} - n^{\mathrm{eq}}_{B_i,0}].\end{aligned} \tag{6.27}$$

These are the standard equations if bound and scattering states are decoupled. They might be helpful to understand Eq. (6.26) better. Namely, when adding the well-known Boltzmann Eqs. (6.27) and *imposing* ionization equilibrium, one would end up with Eq. (6.26). We summarize and discuss the main findings of this section below.

(i) The differential Eq. (6.26) describes the out-of-chemical equilibrium evolution of the total number density, including the reactions $\eta\xi \rightleftharpoons \{AA, \psi\psi\}$ (Sommerfeld-enhanced annihilation and production) and $(\eta\xi)_{\mathrm{B}} \rightleftharpoons \{AA, \psi\psi\}$ (bound-state decay and production) under the constraint of ionization equilibrium for all times. The total number density n_η counts both: free particles as well as particles in the bound state. Equation (6.26) is equivalent to the coupled set of Boltzmann equations including soft emissions and absorptions [21] in the limit of ionization equilibrium. The equations are independent of the bound-state formation or ionization cross section since the rates are, by assumption, balanced. One can also see it from a different perspective. Via Eq. (6.26) it is very elegant to include bound-state formation or dissociation processes without calculating the cross sections or solving a coupled system of differential equations. Note that when comparing our single equation to the coupled set of Boltzmann equations in Refs. [21], the last term in Eq. (6.26) accounting for the direct production of bound states from two photon annihilation (inverse decay) was dropped. It might have little impact since bound-state formation from radiative processes can be much more efficient around the freeze-out.

(ii) Let us discuss some asymptotic regimes of Eq. (6.26), assuming the system has bound-state contributions. For this case, the ionization degree α^{id} has a nontrivial dependence on the total number density n_η, as can be seen from Eq. (6.24). This leads to the fact that the collision term of the number density Eq. (6.26) has in general neither a linear nor a quadratic dependence on n_η, as it is for the decoupled conventional Eqs. (6.27). However, we recover correctly the quadratic and linear form in some regimes discussed now. Close to the freeze-out, the temperature is much larger than the binding energy of bound states. As a consequence, the bound-state contribution in the ionization degree is subdominant and can be neglected. This can be seen directly from Eq. (6.23). In the high temperature regime $x \ll 1$, leading to $\alpha^{\mathrm{id}} \sim 1$ (fully ionized DM plasma), the dominant part of the r.h.s. of the number density Eq. (6.26) is *quadratic* in n_η. At late times, roughly when temperature becomes of the order the binding energy, the contribution of the bound states becomes strongly enhanced due to Boltzmann factor $\propto e^{\beta|E_B|}$ in $K^{\mathrm{id}}(T)$. For this low temperature regime $x \gg 1$, leading to $\alpha^{\mathrm{id}} \simeq 1/\sqrt{n_\eta K^{\mathrm{id}}(T)}$ which is much smaller than unity. This behavior is expected since bound states with energy $2M - |E_B^i|$ are thermodynamically more favored compared to the scattering states with energy $2M$ for $T < |E_B^i|$. Therefore, the equilibrium limit for low temperature is that most of the particles are contained in the lowest bound state (assuming for a moment that such state is stable). Inserting the low temperature behavior of $\alpha^{\mathrm{id}} \simeq 1/\sqrt{n_\eta K^{\mathrm{id}}(T)}$ into Eq. (6.26) results in the fact that at late times the r.h.s. the of number density equation is *linear* in n_η and the proportionality factor is effectively the bound-state decay rate. Since typically the decay rates are much larger compared to the Hubble rate and the number density equation is linear in n_η, the total number density gets depleted exponentially fast in time. This means at late times, if ionization equilibrium is assumed, also the free DM particle number decreases simply because it must follow the exponential fast decay of the bound states in order to maintain the imposed ionization equilibrium. In summary, if bound-state solutions are present and one would integrate numerically Eq. (6.26) until today, effectively no DM would remain as a consequence of the imposed ionization equilibrium. However, we know from the full coupled set of classical Boltzmann equations [21] that ionization equilibrium is *not* maintained for all times during the DM depletion phase caused by bound-state decays. This is because, once the decay rate exceeds the ionization rate, the system leaves the ionization equilibrium. The temperature dependence of the bound-state formation (BSF) and ionization determines by how much the stable components are depleted during this critical epoch. Once the BSF rate drops below the cosmic expansion rate H the depletion stops.

(iv) If we instead would have treated bound and scattering states separately, with independent chemical potentials, we would have obtained Eq. (6.27). In this equation one can see that the bound states are independent of the scattering states. Since the differential equation of the former is linear in the number density on the r.h.s., at some point all bound states start to decay away. The coupling between bound and scattering states via radiative processes are not included in our theory and therefore do not appear. As we have learned, in our theory treating bound states as composite particles there is only one chemical potential. Therefore, by naively just giving bound and scattering states different chemical potentials would lead to the fact that we describe the system not with a grand canonical ensemble. Moreover, the KMS condition does not hold for this case. In order to be able to introduce different chemical potentials might require to rewrite our theory in terms of effective operators, creating only scattering or only bound states respectively. Then, there might be for every operator a conserved charge and one can associate individually different chemical potentials. We will see later that this more “phenomenological” procedure is definitely not applicable for finite temperature case. There it becomes impossible to introduce such operators since the eigenvalues might be not well defined when non-Hermitian thermal corrections are included.

(v) For going beyond the ionization equilibrium it is required to include ultrasoft terms from the beginning and re-derive the DM correlator EoM including those corrections. This we leave for future work and we restrict our equations to be valid as long as ionization equilibrium can be maintained. If no bound-state solution exists, our equations presented here only assume kinetic equilibrium [77,78]. If bound states exist, the validity of Eq. (6.26) can be estimated. Once the decay rate exceeds the ionization rate at late times, the regime of out-of-ionization equilibrium starts (see also [21]). Until then, our description is valid. In the next section, we will generalize this equation for the finite temperature case. We will assume that the thermal width in the effective potential is small. Already in this case, the number of bound states is dynamical since the screening as well as the constant real part term in the effective potential are temperature dependent. They lead for decreasing temperature to an abrupt occurrence of bound-state poles in the spectral function (see also Sec. VII). Therefore, at finite temperature the description via Eq. (6.26) is insufficient. The more general description can be obtained from our formalism by going back to our master equation, expressing the annihilation or decays in terms of the two-particle spectral function. The spectral function automatically decides for a given temperature which part of the spectrum contributes to the continuum and which part is bounded, as we will see later. Moreover, as we discuss in the next section, it is required at finite temperature to include nonideal contributions to the idealized Saha equation, leading to finite temperature corrections of the chemical potential. Thus, for a consistent treatment we need to compute both: long-range modified annihilation and the chemical potential to the same order of approximation.

D. Chemical potential in narrow thermal width approximation

In the previous section, we established what is the outcome of our master equation in the limit of vanishing thermal corrections. In this section, we compute the chemical potential for finite g_χ and g_ψ. The only assumption we make is that the thermal width in the mediator correlator D is subleading. Only the real-part correction in Eq. (5.32) is kept. This is called the narrow width (or quasiparticle) approximation. It leads to the fact that the correlator takes the simple form

$$D(x,y) = \delta_{\mathcal{C}}(t_x,t_y)(-i)V(\mathbf{x}-\mathbf{y}), \tag{6.29}$$

where V is the screened Yukawa potential. To now evaluate the pressure explicitly, we make use of the structure of the Hamiltonian for this potential:

$$\begin{aligned} H = H_0 &+ M(N_\eta + N_\xi) \\ &+ \frac{g_\chi^2}{2}\int_{\mathbf{x},\mathbf{y}} V(\mathbf{x}-\mathbf{y})[\eta^\dagger(\mathbf{x})\eta^\dagger(\mathbf{y})\eta(\mathbf{y})\eta(\mathbf{x}) \\ &+ \xi(\mathbf{x})\xi(\mathbf{y})\xi^\dagger(\mathbf{y})\xi^\dagger(\mathbf{x}) - 2\eta^\dagger(\mathbf{x})\xi(\mathbf{y})\xi^\dagger(\mathbf{y})\eta(\mathbf{x})]. \end{aligned} \tag{6.30}$$

Substituting $g_\chi^2 \to \lambda g_\chi^2$ and taking the partial derivative of the partition function with respect to λ we arrive at the convenient form

$$p\Omega - p_0\Omega = T\int_0^1 \mathrm{d}\lambda \frac{1}{Z_{\mathrm{gr}}}\partial_\lambda Z_{\mathrm{gr}}(\Omega,T,\mu) \tag{6.31}$$

$$\begin{aligned} &= T\int_0^1 \mathrm{d}\lambda \int_{\mathbf{x},\mathbf{y}} \frac{g_\chi^2}{2} V(\mathbf{x}-\mathbf{y}) \\ &\quad \times [2G_{\eta\xi}^{++--}(\mathbf{x},\mathbf{y},\mathbf{x},\mathbf{y};t-t',\lambda)|_{t=t'} \\ &\quad - G_{\eta\eta}^{++--} - G_{\xi\xi}^{++--}]. \end{aligned} \tag{6.32}$$

Spin indices are summed over equal arguments and $G_{\eta\eta}^{++--}$, $G_{\xi\xi}^{++--}$ carry the same arguments as $G_{\eta\xi}^{++--}$. Applying the KMS condition, and expressing G^{++--} in terms of spectral function we arrive, after introducing Wigner coordinates and Fourier transformation, at

$$\begin{aligned} \beta p = \beta p_0 &+ e^{-\beta 2M}\int \frac{\mathrm{d}^3\mathbf{P}}{(2\pi)^3} e^{-\beta \mathbf{P}^2/4M} \int \frac{\mathrm{d}E}{(2\pi)} e^{-\beta E}\int_0^1 \mathrm{d}\lambda \\ &\times \int_{\mathbf{r}} \frac{g_\chi^2}{2} V(\mathbf{r})\Big[2e^{\beta(\mu_\eta+\mu_\xi)}G^\rho_{\eta\xi}(\mathbf{r},\mathbf{r};E,\lambda) \\ &- e^{\beta 2\mu_\eta}G^\rho_{\eta\eta} - e^{\beta 2\mu_\xi}G^\rho_{\xi\xi}\Big]. \end{aligned} \tag{6.33}$$

This equation tells us that the total pressure is equal to the ideal pressure, defined as

$$p_0\Omega = T\ln \mathrm{Tr}[e^{-\beta(H_0-\mu_\eta N_\eta - \mu_\xi N_\xi)}], \tag{6.34}$$

plus nonideal contributions arising from DM long-range self-interactions. It is possible to eliminate the λ integration in Eq. (6.33) by partial integration and using the BS equation backwards. The final result can be expressed in terms of bound-state contributions and the change of the scattering phase with respect to the energy for the scattering states. The equation is then known as Beth-Uhlenbeck formula. For the following discussion, it is however not required to explicitly give those expressions and therefore we just refer to the result in standard text books for nonideal plasmas; see [81].[10] Let us remark that one can also directly solve Eq. (6.33) numerically via the solution of the two-particle spectral function. This is the power of Eq. (6.33). One can solve self-consistently for $G_{\eta\xi,s}^{++--}(x,x,x,x)|_{\mathrm{eq}}$ and the chemical potential (see below) entering our master equation just by evaluating the two-particle spectral function *without* specifying what is a bound or a scattering state. The two-particle spectral function automatically takes into account everything. This is because Eq. (6.33) is *exact* in the narrow thermal width limit. We have just reduced the problem of evaluating the total pressure to the evaluation of a four-point correlation function.

For the moment, only the structure of Eq. (6.33) is important. Namely, when differentiating the total pressure equation with respect to the chemical potential we obtain, according to Eq. (6.3), for the total DM number densities

$$n_\eta = n_{\eta,0} + K(T)n_{\eta,0}n_{\xi,0}, \tag{6.35}$$

$$n_\xi = n_{\xi,0} + K(T)n_{\eta,0}n_{\xi,0}, \tag{6.36}$$

$$n_{\eta,0} = n_{\eta,0}^{\mathrm{eq}}e^{\beta\mu_\eta}, \tag{6.37}$$

$$n_{\xi,0} = n_{\xi,0}^{\mathrm{eq}}e^{\beta\mu_\xi}, \tag{6.38}$$

---

[10] In nonideal plasma literature, the nonideal contribution in Eq. (6.33) is referred to as the second-viral coefficient. It might be also interesting to note that the number of bound states are related to scattering phases according to the Levinson theorem.

$$n^{\mathrm{eq}}_{\eta,0} = n^{\mathrm{eq}}_{\xi,0} = 2\left(\frac{MT}{2\pi}\right)^{3/2} e^{-\beta M}. \tag{6.39}$$

The subscript 0 labels ideal number densities, i.e., $n_{\eta,0}$ is a number density for a free DM without gauge interactions as in the previous section. Superscript "eq" stands for chemical equilibrium and a symmetric plasma is assumed $\mu_\eta = \mu_\xi$. We stress that $n_{\eta,0}$ should not be confused with number densities of one quasiparticle excitation of DM in thermal plasma. This equation just tells us that the exact total DM number density, n_η and n_ξ, including corrections from thermal plasma and bound states, can be simply expressed as Eqs. (6.35) and (6.36) by means of the ideal number densities given in Eqs. (6.37)–(6.39). All the effects from interactions are encoded in $K(T)$. An explicit expression for $K(T)$ can be obtained by comparing Eq. (6.35) to (6.33). Again, it is only important to know that $K(T)$ includes bound-state contributions as well as scattering parts which can be seen from the integration of spectral function in the whole energy range. One can see that in the limit of $K(T) \to 0$ the number densities of free DM are recovered. If we would send finite temperature corrections to zero and only include the bound-state contribution, then $K(T)$ just reproduces the ideal gas case from the previous section; see Eqs. (6.23) and (6.22). In this sense, *standard Boltzmann equations (see also previous section) typically used in numerical codes solve for the DM number density non-self-consistently.* This is because they miss the (small) nonideal corrections coming from the scattering contributions in $K(T)$ and evaluate the chemical potential in the ideal gas approximation.

For a symmetric plasma $n_{\eta,0} = n_{\xi,0}$, Eq. (6.35) is a quadratic equation in $n_{\eta,0}$ and the solution is given by

$$\frac{n_{\eta,0}}{n_\eta} = \alpha(n_\eta K(T)), \quad \alpha(x) = \frac{1}{2x}(\sqrt{1+4x}-1), \tag{6.40}$$

where α gives the ratio between the ideal number density of free particles (no bound states and no interactions) and the total number density n_η, including bound states, nonideal and thermal corrections. This is the main difference compared to the ideal definition in the previous section. The generalized Saha Eq. (6.40) accounts for nonideal contributions like self-interactions as well as for finite temperature corrections. Important to note is that we do not have to define what is a bound or a scattering contribution to the total number density. The spectral function in Eq. (6.33) does the job automatically and takes all contributions into account when integrating over the whole energy range.

Since $n_{\eta,0} = n^{\mathrm{eq}}_{\eta,0} e^{\beta\mu_\eta}$, we can finally determine the chemical potential μ_η as a function of total number density from Eq. (6.40).

Assuming a grand canonical state for a system having bound-state solutions in the spectrum automatically implies the Saha ionization equilibrium. Under this assumption, the chemical potential is set by

$$\beta\mu_\eta = \ln\left[\frac{\alpha n_\eta}{n^{\mathrm{eq}}_{\eta,0}}\right], \tag{6.41}$$

and our master formula for the total number density Eq. (6.1) can be written in a fully closed form as

$$\dot{n}_\eta + 3Hn_\eta = -\frac{2(\sigma v_{\mathrm{rel}}) G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}}{n^{\mathrm{eq}}_{\eta,0}(T) n^{\mathrm{eq}}_{\eta,0}(T)} \times [\alpha^2(n_\eta K(T)) n_\eta n_\eta - n^{\mathrm{eq}}_{\eta,0}(T) n^{\mathrm{eq}}_{\eta,0}(T)], \tag{6.42}$$

where

$$G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}} = e^{-\beta 2M} \int_{-\infty}^{\infty} \frac{\mathrm{d}^3\boldsymbol{P}}{(2\pi)^3} e^{-\beta \boldsymbol{P}^2/4M} \times \int_{-\infty}^{\infty} \frac{\mathrm{d}E}{(2\pi)} e^{-\beta E} G^{\rho}_{\eta\xi}(\boldsymbol{0},\boldsymbol{0};E)|_{l=0}. \tag{6.43}$$

Importantly, the obtained number density Eq. (6.42) is in general *not quadratic in* n_η contrary to the conventional Lee-Weinberg equation because the generalized ionization fraction, $\alpha(n_\eta K(T))$, exhibits a nontrivial dependence on n_η. As we have discussed in the previous section, this equation contains all the number violating processes of DM, i.e., annihilations and bound-state decay. The process which dominates the decrease of DM number density is determined by how α evolves in time as we have discussed at the end of the previous section.

The insights we gained in this section are important for the understanding of our work. Let us put below these results more into context.

(i) Equation (6.42) describes the out-of-chemical equilibrium (finite μ) evolution of the total number density under the constraint of ionization equilibrium and is one of our main results. It is a generalization of the idealized vacuum Eq. (6.26), accounting for (i) finite temperature corrections to the annihilation/decay rates and (ii) nonideal corrections to the chemical potential. The nonideal corrections to the chemical potential consist of finite temperature corrections as well as scattering contributions. The main advantage of the Eq. (6.42) is that we do not have to define what is a bound or scattering state contribution since at finite temperature this is meaningless to do. All expressions needed in order to solve our generalized number density equation numerically can be obtained by evaluating the two-particle spectral function. Since one has to integrate over the whole energy spectrum, the result manifestly takes all contributions into account, without the need of differentiating between bound and scattering states. Later in

Sec. VII, we show the results for the two-particle spectral function $G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{l=0}$ in Eq. (6.42) including finite temperature corrections.

(ii) Let us discuss a bit more in detail the generalized Saha Eq. (6.40). For vanishing finite temperature corrections and the ideal gas limit, we have emphasized that it reduces to the standard expression Eq. (6.24). Since (6.40) looks the same as the one in the ideal gas limit aside from how $K(T)$ depends on T, one may understand how α evolves from the discussion at the end of the previous section. However, at finite temperature it is more complicated to precisely estimate since the number of bound states is dynamical. Furthermore, it is also more complicated to discuss the case when ionization equilibrium is broken at finite temperature due to this reason.

Finite temperature effects might extend the period where the ionization rate is much larger compared to the decay rate, since ψ particles can efficiently destroy the bound state. Thus it might be true that the validity of our equations holds longer compared to the vacuum case. Another difficulty at finite temperature is, only in the limit of narrow thermal width it might be possible to estimate the validity of (6.42). This is because in order to estimate the rates one has to define what is the decay width of the bound state, which becomes hard to define beyond the narrow thermal width limit. At late times of DM freeze-out, however, we naively expect that the thermal width becomes less important and one may estimate the rates by just including the real-part corrections (but still in this case the number of highly excited bound states can be dynamical).

(iii) According to previous discussions, we would like to emphasize again that our number density Eq. (6.42) is not applicable to the regime where bound-state decay rates exceed the ionization rates at late times causing an out-of-ionization equilibrium state (now disregarding the issue of how we can precisely estimate those rates at finite temperature). However, our more general equations Eqs. (4.20)–(4.21) do not assume ionization equilibrium and can be applied to any out-of-equilibrium state. The point is, since we have dropped for simplicity soft emissions from the beginning, there are no processes like BSF via the emission of a mediator relating the number of bound and free particles. Consequently, if one would solve the general Eqs. (4.20)–(4.21) numerically, with an initial out-of-ionization equilibrium state, the system would remain for all times in out-of-ionization equilibrium. It is required for future work to include soft emissions via e.g., an electric dipole operator in thermal plasma, to account for a correct description of the DM thermal history at late times. We will see, however, if ionization equilibrium can be guaranteed, our description accounts for sizable finite temperature corrections during the early phase of the freeze-out which can not be captured by the classical on-shell Boltzmann equation treatment as in [21]. Thus these different approaches are in some sense complementary. Furthermore, the approach in [21] uses only the main contribution of the ground state 1S, while via Eqs. (6.42) and (5.44) it is very elegant and efficient to include *all* (here only s-wave) bound-state contributions (but under the assumption of ionization equilibrium).

(iv) From this section we learned that the standard Boltzmann equations at zero temperature are a non-self-consistent set of equations. They miss scattering contributions in $K(T)$ which can now be fully accounted for. Bearing in mind that these contributions might be small, it might also be acceptable to adopt a non-self-consistent solution of Eq. (6.42). By this we mean one can in principle compute the chemical potential in the narrow width approximation; however, in the computation of $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\rm eq}$, the two-particle spectral function can be solved in its most general form including finite temperature width (as we present in Sec. VII). We emphasize again that a non-self-consistent solution might cause some troubles and care should be taken. In Appendix E 3, we discuss other non-self-consistent computations of the chemical potential and point out their failure, especially for late times if bound-state solutions exist.

E. Comparison to linear-response-theory method

Our dark matter system is similar to heavy-quark pair annihilation in a thermal quark gluon plasma produced in heavy ion collisions. In literature, the annihilation rate of the heavy-quark pair into dileptons is estimated from linear response theory. Let us just quote their results in the following without deeply diving into the details. For comparison we translate the expressions for $SU(3)$ to our U(1) by adjusting color and flavor factors and call the heavy (anti)quark dark matter (ξ) χ. The linearized Boltzmann equation around chemical equilibrium is given by

$$\dot{n}_\eta + 3Hn_\eta = -\Gamma_{\rm chem}(n_\eta - n_{\eta,{\rm eq}}) + \mathcal{O}(n_\eta - n_{\eta,{\rm eq}})^2, \qquad (6.44)$$

where $\Gamma_{\rm chem}$ is called the chemical equilibration rate. $\Gamma_{\rm chem}$ can be extracted from linear response theory, assuming thermal equilibrium and a perturbation linear around chemical equilibrium. It is defined as [64–67]

$$\Gamma_{\rm chem} \equiv \frac{\Omega_{\rm chem}}{(4\chi_\eta/\beta)M^2}, \qquad (6.45)$$

where $\Omega_{\rm chem}$ is a transport coefficient and χ_η is the heavy DM number susceptibility. The transport coefficient is quoted as

$$\Omega_{\rm chem}/M^2 \simeq 16(\sigma v_{\rm rel}) \int_{\omega,\boldsymbol{P}} e^{-\beta(2M+\omega)} \rho_{\eta\xi}(\omega,\boldsymbol{P}), \qquad (6.46)$$

where we consider always only s-wave contributions and the tree-level annihilation cross section is defined as in our case, while $\rho_{\eta\xi}$ is also called a spectral function but might be defined slightly differently than ours. This expression looks similar to our term in chemical equilibrium

$$2(\sigma v_{\mathrm{rel}})G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}} \\ = 2(\sigma v_{\mathrm{rel}})\int_{\omega,\boldsymbol{P}} f_B(2M+\omega)G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};\omega,\boldsymbol{P}). \quad (6.47)$$

Indeed, we find both spectral functions in Eqs. (6.46) and (6.47) give identical results if we take the DM dilute limit of our equation.

The number susceptibility is defined as the response of the total number density with respect to infinitesimal variation of the chemical potential:

$$\chi_\eta \equiv \left.\frac{\partial n_\eta}{\partial \mu_\eta}\right|_{\mu_\eta=0}. \quad (6.48)$$

In the dilute limit, using Eq. (6.35), we have $\chi_\eta = \beta n_\eta^{\mathrm{eq}}$. So we find in total for the linearised equation using linear response theory:

$$\dot{n}_\eta + 3Hn_\eta = -\frac{4(\sigma v_{\mathrm{rel}})\int_{\omega,\boldsymbol{P}} e^{-\beta(2M+\omega)}\rho_{\eta\xi}(\omega,\boldsymbol{P})}{n_\eta^{\mathrm{eq}} n_\eta^{\mathrm{eq}}} \\ \times [n_\eta n_\eta^{\mathrm{eq}} - n_\eta^{\mathrm{eq}} n_\eta^{\mathrm{eq}}]. \quad (6.49)$$

In the following, we compare this equation based on linear response theory with our Eq. (6.42). We start from our more general expression and reproduce Eq. (6.49) in the linear regime around chemical equilibrium while clarifying the underlying approximations. For this purpose, we have to linearize our equation around $n_\eta \sim n_\eta^{\mathrm{eq}}$, where “eq” labels chemical equilibrium. On the one hand, as we have shown in Sec. V, the spectral function, $G^{\rho}_{\eta\xi}$, does not depend on n_η as long as the DM number densities are dilute. On the other hand, the generalized ionization fraction, α, does depend on n_η and hence it must also be approximated as to be close to chemical equilibrium. One may easily see this by looking at the definition of α given in Eq. (6.40). From this expression it is clear that if the DM number densities are close to chemical equilibrium, the ionization fraction is always close to one, i.e., $\alpha_{\mathrm{eq}} \simeq 1$, because $n_{\eta,0}^{\mathrm{eq}} \propto e^{-\beta M}$ and $K(T) \propto e^{\beta|E_B|}$. Hence, in the linear regime, the total DM number densities are fully ionized and can be approximated as the free scattering contributions: $n_\eta^{\mathrm{eq}} \simeq n_{\eta,0}^{\mathrm{eq}}$. From these observations, we reproduce Eq. (6.49) from our Eq. (6.42) *in the linear regime around chemical equilibrium.*

The linear response theory method applies, by definition, only to the linear regime around chemical equilibrium where $n_\eta \sim n_\eta^{\mathrm{eq}}$. It is not possible via this method to see what is the correct form of the underlying Lee-Weinberg equation describing the nonlinear regime where $n_\eta \gg n_\eta^{\mathrm{eq}}$. Nevertheless, one might be tempted to use it by replacing the right-hand side of Eq. (6.44) with $\frac{\Gamma_{\mathrm{chem}}}{2n_\eta^{\mathrm{eq}}}(n_\eta^2 - (n_\eta^{\mathrm{eq}})^2)$ as done in [37,55–57,66]. Most important to note is that the collision term obtained from this replacement matches our equation only if the generalized ionization fraction is close to *chemical equilibrium*, i.e., $\alpha_{\mathrm{eq}} \simeq 1$, which is no longer true at late times. We can see how α depends on time from Eq. (6.40). As we have already discussed in detail at the end of Secs. VI C and VI D, at late times of DM freeze-out the total comoving number density approaches a constant value while $K(T)$ starts to grow rapidly for $T < |E_B|$. As a result, we find the generalized ionization fraction to be $\alpha \simeq 1/\sqrt{n_\eta K(T)} \ll 1$ at late times, which invalidates the naive replacement above. One can understand this behavior intuitively because the bound states of energy $2M - |E_B|$ are thermodynamically favored compared to the scattering states of energy $2M$ at $T < |E_B|$. For a fixed total number of DM, the bound states dominate over the scattering states at some point as in the case of the recombination. There is also another issue when correcting n_η^{eq} in $\frac{\Gamma_{\mathrm{chem}}}{2n_\eta^{\mathrm{eq}}}(n_\eta^2 - (n_\eta^{\mathrm{eq}})^2)$ only by the Salpeter term. Since this discussion requires some detailed knowledge about thermal corrections, we share it in Appendix E 3. In summary, we would like to emphasize that care must be taken if one matches the equation obtained by linear response theory to nonlinear differential equations.

Instead, one may match the equation obtained by linear response theory to our corrected form of the Lee-Weinberg equation in the nonlinear regime. By this procedure it is now also clear what the limitation exactly is. As we have discussed in detail in the previous section, our master formula for the number density equation is valid as long as ionization equilibrium can be maintained. Ionization equilibrium is broken if the decay rate exceeds the ionization rate. The temperature where this happens can be estimated for finite temperature systems, at least in the narrow width case.

VII. NUMERICAL RESULTS FOR TWO-PARTICLE SPECTRAL FUNCTION AT FINITE TEMPERATURE

We turn to the numerical solution of the two-particle spectral function for the full in-medium potential. The effects of the finite temperature corrections can be simplest understood for the case of the Coulomb potential as given in Eq. (5.32). The first correction is a real constant term that shifts effectively only the energy by $\alpha_\chi m_D$. When only taking this correction into account one would thus expect that the infinite number of bound states in the spectral function of the Coulomb case just move to lower binding energies and similar shift to the threshold as well as to the positive energy spectrum. The second real-part correction is an exponential screening of the Coulomb potential with radius m_D. This introduces another effect. It leads to a disappearance of the bound states closest to the threshold since Yukawa potentials have only a finite number of bound states. The disappearance of bound states wins against the move of the poles towards lower energies at the Mott transition, where all bound states disappear and the spectrum is exclusively continuous. Additionally, we have imaginary-part corrections coming from the soft DM-ψ scatterings, leading to a finite thermal width of the bound states. Once the thermal width is comparable to the binding energy, the bound-state poles are strongly broadened.

The combination of all effects are shown in Fig. 4, where we present the numerical solution of the two-particle spectral function for the full in-medium potential according to Eq. (5.44). We show the case of the Coulomb and Yukawa potential. In this figure, we have fixed the temperature to $T = M/30$ (slightly below the typical DM freeze-out temperature) and varied the Debye mass where the maximal value shown corresponds to the equal charge case $g_\psi = g_\chi$ of our minimal model: $m_D^2 = g_\chi^2 T^2/3$. The mass of the DM is fixed to 5 TeV and the coupling $\alpha_\chi = 0.1$ very roughly chosen to account for the correct order of the abundance.

All finite temperature effects together lead to a continuous *melting* of the bound-state poles. As can be seen, the melting of the bound states leads to the fact that even at negative energies, the spectrum is continuous at finite temperature. The reshuffling of the spectrum towards lower energies affects the rates *exponentially* according to Eq. (5.8). This is because the integrand (the spectral function) has more support at negative energies which is, due to Boltzmann factor, exponentially preferred in kinetic equilibrium. It now becomes clear that the notion of spectral function is more general compared to the vacuum case where one could separate the spectrum for bound and scattering states. Here, it is evident that such a distinction is impossible. It is also not necessary to do so since the integration of spectral function times Boltzmann factor takes already all contributions into account.

In all cases we study in this section, the integral has a fast convergence at negative energies well inside the validity region of HTL resummed effective theory $|E| \lesssim T$. An intuitive reason why our treatment breaks down for a large negative energy $-E \gg T$ is the following. In this regime dark matter and antidark matter are tightly coupled. As a result, typical scatterings with momentum exchange of T cannot probe inside the dark matter and antidark matter pair. Thus, we expect the imaginary part in the effective potential for $|E| \gg T$ to be suppressed by an additional Boltzmann factor. In a word, for large enough $|E|$, the situation should revert to the vacuum case and we no longer expect thermal corrections for the bound states. Indeed, for vanishing imaginary part in the potential the two-particle spectral function has no support for negative energies below the ground state energy.

Nevertheless, let us discuss how only the positive energy spectrum is affected. According to the theorem of Levinson, the scattering phases (and hence the wave function at the origin) depend on the amount or properties of the bound states. This means that thermal modifications of the bound states automatically affect also the positive energy spectrum. The impact on the positive energy spectrum depends on the melting status of the bound states. In general, there can be both, a suppression or further enhancement of the positive energy spectrum as can be seen by carefully looking at the value around $E = 0$ in Fig. 4. We would like to stress that a suppression of the positive spectrum does not imply that the total rate is less. For the computation of the rate one has to integrate the spectral function over the whole energy range, where $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\rm eq}$ becomes due to the reshuffling towards lower energies exponentially enhanced.

While for the Coulomb case, the impact of the melting on the positive energy spectrum is only very little (which does not mean that the overall effect is small), the impact for the Yukawa potential case can be much larger. In Fig. 5, we compare the positive energy solution of the Yukawa spectrum at zero and finite temperature, as a function of the mediator mass m_V. The vacuum line (dashed) is obtained by

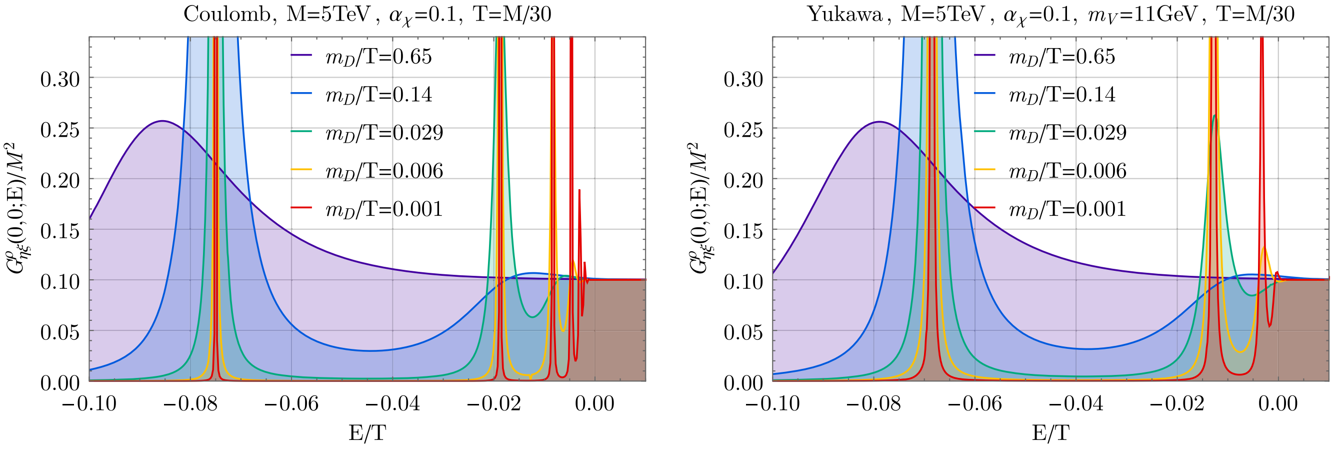


FIG. 4. Two-particle spectral function at finite temperature shown vs the energy in units of typical freeze-out temperature. The violet line corresponds to the equal charge case $g_\psi = g_\chi$ of our minimal model and hence $m_D^2 = g_\chi^2 T^2/3$.

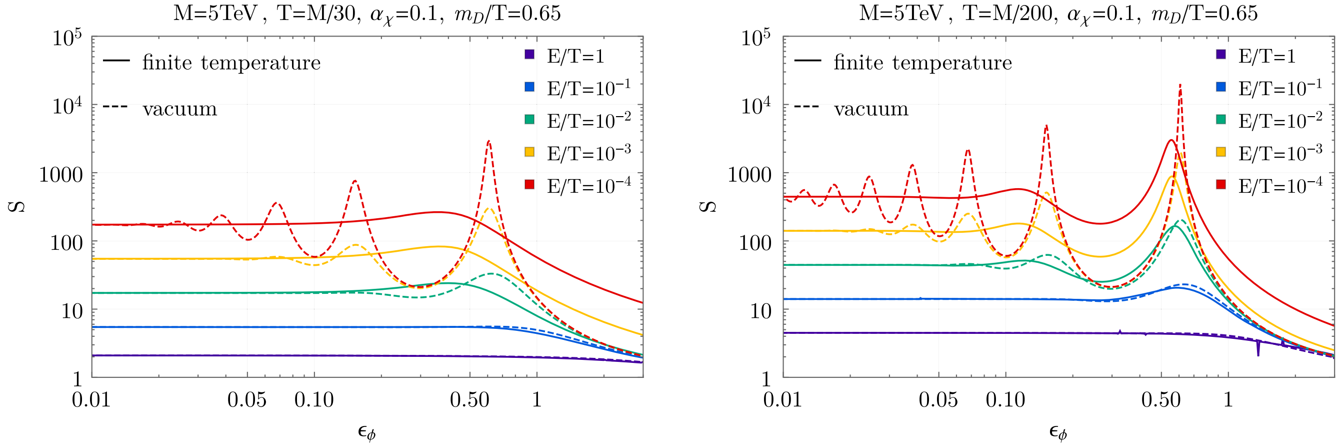


FIG. 5. Sommerfeld enhancement factor for a Yukawa potential shown vs the mediator mass in units of the Bohr radius $\epsilon_\phi \equiv m_V/(\alpha_\chi M)$. Dashed and solid lines correspond to the vacuum limit and the full in-medium potential, respectively. Effects at different temperatures are compared. Left side for typical freeze-out temperature and right plot at a typical temperature where the annihilation rate would be much smaller compared to the Hubble rate.

solving the spectral function in the limit of vanishing finite temperature corrections. The Sommerfeld factor for this case can be obtained from Eq. (6.12) according to

$$S(v_{\mathrm{rel}}) = \frac{2\pi}{M^2 v_{\mathrm{rel}}} G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{E>0,l=0} \qquad (7.1)$$

$$\Leftrightarrow S(E) = \frac{\pi}{M\sqrt{ME}} G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{E>0,l=0}. \qquad (7.2)$$

In the second line we used $E = M v_{\mathrm{rel}}^2/4$ for on-shell particles. At finite temperature this kinetic energy relation does not hold. We therefore use the second Eq. (7.2) to define the Sommerfeld enhancement factor at finite temperature as shown in Fig. 5. The enhancement or suppression of the Sommerfeld enhancement factor due to the thermal effects is largest if the ground state is close to the threshold of $E = 0$ (around the first peak from the right, here $\epsilon_\phi \sim 0.6$). For our minimal model, it is also shown in the right plot of Fig. 5 that the whole temperature range of a typical freeze-out process can be affected. In the limit $m_V \to 0$ the Coulomb limit is recovered. Again, this does not mean that it is sufficient to just take the standard expression of the Sommerfeld enhancement factor of the Coulomb potential to describe the DM freeze-out. There is also a contribution from the negative energy spectrum. Therefore, one has to be careful in interpreting Fig. 5. On the one hand side the positive energy solution (Sommerfeld enhancement factor) can be suppressed or equal compared to the vacuum case, but on the other side the total $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$ entering our master formula Eq. (6.1), which requires the integration over the whole energy spectrum, can be enhanced.

As an extreme example of this situation, let us discuss the case where there are no bound states (e.g., Yukawa potential in the Born regime $\epsilon_\phi \gg 1$). The finite temperature spectral function for this case is shown in the right plot of Fig. 6. Indeed, the positive energy spectrum is

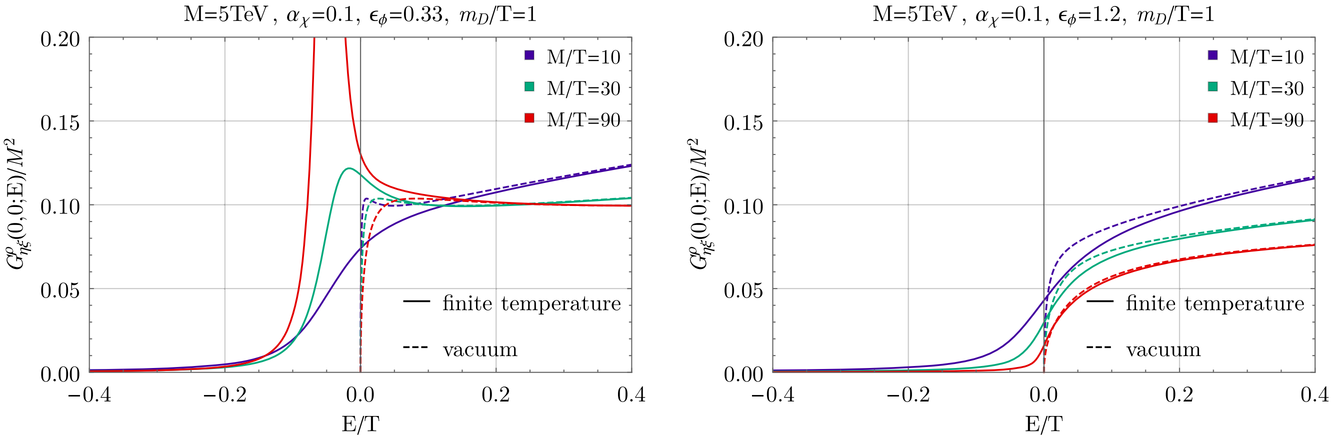


FIG. 6. Comparison between extreme examples, where only the ground state exists and is close to the threshold (left) and where no bound states exist (right). Here, the vacuum curve is defined as not including delta-peaked bound-state contributions.

dominantly suppressed compared to the vacuum case but when integrating the spectral function over the Boltzmann factor in the whole energy range there is still an enhancement of order 1% compared to the vacuum case. Another extreme example, where the corrections to the positive spectrum are strongest, is the case where only the ground state exists and is close to threshold [37,55,56]. This example is shown in the left plot of Fig. 6. We find in this case, the value of the integration of the spectral function times the Boltzmann factor is by up to 10% (30%) larger compared to the vacuum case *without* bound-state peak at the typical freeze-out temperature $T = M/30$ ($T = M/90$). The correction increases for lower temperature due to the Boltzmann factor.

The total rate is only proportional to the integration of the spectral function times the Boltzmann factor. There are additional finite temperature corrections to the chemical potential (see previous section) which can be also obtained from the spectral function. We have not explicitly computed those nonideal corrections yet but leave it for future work once we have included ultrasoft emissions in our system description.

VIII. DISCUSSION

A self-interacting DM system, where long-range forces and bound-state solutions exist, is in general a complex ensemble where many processes with different rates are taking place at the same time during the DM thermal history. Essentially, there are three quite different approaches in the literature with distinct motivation to describe the evolution of the abundance of the stable components for such systems:

(1) The first approach is based on a coupled set of classical on-shell Boltzmann equations. If bound-state solutions are absent, the description of the DM freeze-out acquires dominantly corrections from the Sommerfeld-enhanced annihilation of free DM particles [10,11]. If the two-particle spectrum has support at negative energies, the free DM particles can form a bound state via radiative processes [21,82]. The reverse process can also happen, called ionization. If there are several bound-state solutions present, further processes like excitation or deexcitation can happen [22,42]. All those processes are in general coupled, and as we see, the list of Boltzmann equations needed to describe such systems can be long. When relying on those classical Boltzmann equation computations, treating e.g., the number density of free particles and bound states separately and as idealized, potential strong modifications arising from higher-order plasma interactions might be missed. In this approach, however, it is always guaranteed that the nonlinearity of out-of-chemical equilibrium reactions are accurately described. And there can be in general many such out-of-equilibrium reactions as listed above.

(2) The second approach starts from the EoM of correlation functions on the Keldysh contour and takes into account some finite temperature corrections. The major difference to our work is that in [69] it is assumed that the correlator hierarchy can by truncated at the lowest order, resulting in closed equations for the two-point functions in terms of the one-particle self-energy only. One of the equations are the so called kinetic equations, being the differential equations for describing the evolution of observables in terms of the macroscopic Wigner coordinates. In the one-particle self-energy approximation they are also known as Kadanoff-Baym equations. Expanding the self-energy in terms of the coupling to NLO results in the standard Boltzmann equation. At NNLO first finite temperature corrections enter. The advantage of a fixed order calculation is that infrared divergences, arising at NNLO cancel [69]. At NNLO in the self-energy expansion of the kinetic equations, the thermal corrections turn out to be strongly suppressed, i.e., to high power in T/M, compared to the NLO result. One should, however, keep in mind that there are next to the kinetic equations also the equations for the microscopic Wigner coordinates, called mass-shell equations accounting for, e.g., thermal corrections to the dispersion relation. Kinetic and mass-shell equations are in general coupled. Therefore, a self-consistent solution in principle requires to take account of corrections also from the mass-shell equation. In any case, the problem within this systematic approach is that a fixed order calculation can never account for correctly describing the Sommerfeld enhancement beyond the Born regime and also bound-state solutions will never appear.

(3) The third approach addresses the description of long-range force systems at finite temperature in a nonperturbative sense, i.e., by resummation of the Coulomb divergent ladder diagrams including thermal corrections. Clearly, first attempts were made in the literature of heavy quark pair annihilation in a quark gluon plasma [64], produced in heavy-ion collisions at the LHC. More recently, some of these authors have applied the same techniques also to the DM freeze-out [37]. The method is based on *linear response theory* [65–67], estimating the DM Sommerfeld-enhanced annihilation and bound-state decay from a spectral function including finite temperature corrections. It has been shown in Ref. [55–57] that the DM overclosure bound, computed by this method, can be strongly affected by finite temperature effects if bound-state solutions exist. Compared to a fixed order calculation as in approach 2, the finite temperature corrections are larger. The reason is because the mass-shell equations are solved by resummation of the Hard thermal loop contribution. Albeit there are potentially strong effects, the linear response theory is strictly speaking valid only for systems close to thermal equilibrium, e.g., $n \sim n_{\mathrm{eq}}$. At finite temperature the spectral function can in general depend on the DM density. Therefore, it is *a priory* not clear if the transport coefficients extracted from linear response theory can be inserted into a nonlinear Boltzmann equation describing the DM

freeze-out in a nonlinear regime where $n \gg n^{\mathrm{eq}}$. From vacuum computations it is known that the Sommerfeld effect can still be efficient in such a regime. This is because the transition $n \sim n_{\mathrm{eq}}$ to $n \gg n^{\mathrm{eq}}$ happens in a short time, since $n_{\mathrm{eq}} \propto e^{-M/T}$ decreases rapidly. To the best of our knowledge this method, inserting transport coefficients obtained from linear response theory into a nonlinear Boltzmann equation, has not been tested so far by using other treatments applying for generic out-of equilibrium situations.

Our formalism, presented in this work, aims towards a first step in unifying the approach 1 and 3, by generalizing the approach 2 for long-range force systems. In other words, we derived from the EoM of Keldysh correlation functions the number density equation for DM including finite temperature corrections and accounting for the full resummation of Coulomb divergent ladder diagrams. This allows to study the finite temperature corrected Sommerfeld-enhanced annihilation as well as bound-state decay. Moreover, our master Eq. (6.1) is able to describe the correct nonlinear transition to out-of-chemical equilibrium, i.e., the freeze-out process.

Although we have derived all equations on the Keldysh contour, and therefore they should be valid for any out-of-equilibrium situation of the system, the reader should be reminded what precisely our system is. While it remains true that we can describe correctly Sommerfeld-enhanced annihilation and bound-state decay in the presence of a relativistic plasma background for out-of equilibrium situations, we have dropped from the beginning, when deriving our nonrelativistic effective action, ultrasoft contributions of the fully relativistic action. Hence, Eqs. (4.20)–(4.21) are missing ultrasoft contributions leading to bound-state formation and ionization processes via the emission or absorption of a mediator, as well as contributions to excitation or de-excitation processes if multiple bound states exist. Once ultrasoft terms are included in the system of equations, we expect the final equations, if finite temperature effects are neglected, to coincide with the full set of equations of approach 1. Moreover, the inclusion of emission and absorption in the Keldysh formalism might lead to new insights in the production rate of dileptons or photons, produced from heavy-quark pair annihilation in a quark gluon plasma.

In the second half of the work, we have indirectly included all bound-state formation, ionization, excitation, and deexcitation processes. This was achieved by assuming our system is in a grand canonical state with one single time dependent chemical potential as in our master Eq. (6.1). The important observation that adopting a grand canonical picture automatically implies ionization equilibrium if bound states are present was by far not obvious to us. This key observation brought us to the conclusion that our equations in the limit of vanishing thermal corrections are equivalent to the coupled system of classical Boltzmann equations in the limit of ionization equilibrium. Thus we have shown that under certain assumptions our approach and approach 1 consistently fall together. Another important point based on this observation was that, since the ionization fraction at chemical equilibrium is close to unity, our and approach 3 are equivalent to approach 1 in the regime linear near chemical equilibrium.

Important to recognize was that our approach and approach 3 give different results if the transport coefficients extracted from linear response theory is inserted into a nonlinear Boltzmann equation just by replacing $\Gamma_{\mathrm{chem}}(n_\eta - n_\eta^{\mathrm{eq}})$ with $\frac{\Gamma_{\mathrm{chem}}}{2n_\eta^{\mathrm{eq}}}(n_\eta^2 - (n_\eta^{\mathrm{eq}})^2)$ [55–57]. This is because the ionization fraction depends on n_η where another nonlinearity comes in, and in particular, the ionization fraction will be much smaller than unity at late times. This is intuitively because the bound states are exponentially favored compared to the scattering states for $T < |E_B|$. Furthermore, the ionization fraction counteracts against the exponential grow of Γ_{chem} or of our $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$ for late limes if bound-state solutions exist. In a word, while the spectral function is identical between the linear response and ours in the DM dilute limit, the ionization fraction makes the difference. This effect is non-negligible when at late times the DM gets depleted by bound-state formation effects.

Our master Eq. (6.1) cannot be used at very late times where ionization equilibrium is not maintained. Therefore, one has to be careful in relying on our so far simplified treatment for all times during the DM thermal history. More generally, when using our equations, it has to be ensured that the rates driving the system to kinetic and ionization equilibrium are much faster than any other rates leading to a potential out-of-kinetic or -ionization equilibrium state. In the case of no ψ particles (no finite temperature corrections), it was shown in Refs. [21,22,42] that the decay of the bound state becomes faster than the ionization via emission and absorption processes by an electric dipole operator at some point, which breaks the ionization equilibrium at late times. Later, when the bound-state formation becomes inefficient compared to the cosmic expansion, the dark matter number freezes out completely. Estimating the valid regime of our approach in the presence of finite temperature corrections is a more complicated task. To draw a definitive conclusion in our case, one has to estimate these processes, including emission and absorption of ultrasoft gauge bosons, in the presence of the thermal plasma. As we have discussed already in detail, this might only be realizable if the thermal width is negligible compared to the real-part corrections. Furthermore, one has to keep in mind that the number of existing bound-state solutions is temperature dependent when already only real-part corrections are taken into account.

After this warning, we now would like to discuss the case where a grand canonical description with one single chemical potential is justified. As we have in detail presented in this work, all finite temperature corrections are then totally encoded in the solution of the DM two-particle spectral function. In the presence of ultrarelativistic fermionic particles in the background, the hard thermal loop resummed corrections to the DM system can be classified into three contributions [see effective in-medium potential Eq. (5.32)]. The first two contributions are real-part corrections to the DM effective in-medium potential. The first one leads to an energy shift (Salpeter correction) in the DM two-particle spectral function by an amount of $\alpha_\chi m_D$ towards lower energies compared to the vacuum case. Second, the Debye mass m_D leads to a screening of the Coulomb potential, resulting in a temperature dependent Yukawa potential with screening radius m_D. The third contribution to the effective potential is purely imaginary and originates from soft DM scattering processes with the ultrarelativistic fermions in the hot and dense background.

Let us first discuss the two real-part corrections and their implication on the two-particle spectral function for the case of a Coulomb potential. In the limit of vanishing real-part corrections, it is well known that a Coulomb system has infinitely many bound-state solutions. Furthermore, the bound-state solutions and the scattering states can clearly be separated sharply at the energy $E = 0$ in the two-particle spectral function (see Fig. 3). If the real-part finite-temperature corrections are included, bound states close to the threshold $E = 0$ disappear in the spectrum. This is simple to understand. First of all, by the real-part corrections the Coulomb potential transforms into a Yukawa potential. Yukawa potentials have only a finite number of bound states. Secondly, due to the energy shift caused by the other real-part correction, the threshold is lowered. The combination of these two effects causes the disappearance of highly excited bound states close to the threshold and the spectrum is continuous instead of discrete. The effect gets stronger with increasing Debye mass leading at the Mott transition to a total disappearance of all bound states.

Already when only real-part corrections to the in-medium potential are included, the number of bound states as well as their binding energies are temperature dependent according to the discussion above. It implies that a sharp definition of bound and scattering states can not be made for all times during the DM thermal history. However, for our total number density equation, it is NOT required at all to distinguish between bound and scattering states. For example, the computation of the ionization fraction via the generalized Saha Eq. (6.40) can always be performed without specifying what is a bound or scattering contribution. Another example is the spectral function entering the total number density in the production term. Also here, the integration of the spectral function automatically takes into account all contributions from the spectrum. Only in the absolute vacuum limit, it is possible to separate contributions. At finite temperature everything is mixed into one single object. A separation would only cause problems like unphysical jumps in projected thermodynamical quantities when bound states abruptly disappear (if only real-part corrections are included).

The mixing between scattering and bound-state solutions becomes even stronger once the imaginary contributions to the effective in-medium potential are included. As we have seen in Sec. VII, these corrections lead to a thermal width of the peaks of the bound states and a continuous melting of the poles for increasing Debye mass, as illustrated in Fig. 4. Note that instead of changing the independent coupling in the Debye mass, we could have also increased the number of generations of ultrarelativistic particles in the plasma. The Debye mass is proportional to the square root of number of generations, and thus, if most of the particle content of the Standard Model would run in the thermal loop, we can have a large Debye mass although the coupling is still small. The broadening of the peak and the shift towards lower binding energies increases the annihilation or decay rate exponentially (again its not necessary to distinguish between these two) due to the integration of the product of two-particle spectral function and the Boltzmann factor, as in Eq. (5.8).

Although, we focused on a simple U(1) like theory and s-wave contributions in the present work, most of the equations for higher gauge theories, scalar mediators, or higher partial waves will change in the expected way. Let us already mention some major changes. The mediator self-energy would acquire further contributions from self-interactions as well as different colour or flavor prefactors. The definition of the effective in-medium potential in Eq. (5.31) remains the same and is computed from the specific dressed mediator correlator. The r.h.s. of the Schrödinger-like Eq. (5.30) will be proportional to the number of colours. Our number density Eq. (3.24) in the case of velocity dependent tree-level annihilation cross sections (like in p-wave case) will have a space derivative on the r.h.s. As expected, our formalism breaks down for temperatures around the confining scale of confining theories.

IX. SUMMARY AND CONCLUSION

Traditional computations of DM Sommerfeld-enhanced annihilation and bound-state decay rates rely on the assumption that reactions of such processes are taking place under perfect vacuum conditions. In this work we developed a comprehensive derivation of a more general description, taking into account nonideal contributions arising from simultaneous interactions with the hot and dense plasma environment in the early Universe. We have derived the evolution equation for the DM number density which is applicable to the case where scattering and bound states get strongly mixed due to the influence of the thermal plasma surrounding. Our master Eq. (6.1) for the total DM density simultaneously accounts for annihilation and bound-state decay and hence its collision term is in general not quadratic in the DM number density. We showed that finite temperature effects can lead to strong modifications of the shape of the two-particle spectrum, which in turn modifies the DM annihilation or decay rates.

The Keldysh formalism we adopted throughout this work applies for the description of the dynamics of generic out-of-equilibrium states. Within this mathematical framework, we derived in the first part of this work directly from our nonrelativistic effective action the exact equation of motion of the DM two-point correlation functions. We extracted for the first time from those EoM the differential equation for the DM number density [see Eqs. (3.24) and (3.25)], which turns out to only depend on a special component of the DM four-point function on the Keldysh contour, namely $G^{++--}_{\eta\xi}$. Let us emphasize again that this equation for the number density is exact within our nonrelativistic effective action; however, it is not closed since it depends on the solution of this four-point correlation function. The long-range force enhanced annihilations, the decay of bounded particles as well as the finite temperature corrections are all contained in the solution of this one single four-point correlator.

In the second part of this work, we derived the EoM for the DM four-point function on the Keldysh contour. We developed the approximations needed in order to close the hierarchy of correlators but at the same time keep the resummation of Coulomb divergent ladder diagrams as well as the finite temperature corrections. Based on our approximation and resummation scheme, the final form of the equation for our target component $G^{++--}_{\eta\xi}$ is physically sound and maintains important relations like the KMS condition in equilibrium. The coupled system of equations is general enough to apply for the description of DM out-of-chemical equilibrium states.

In the third part, we explored further approximations needed in order to obtain a simple solution to our target component and to reproduce from our general equations the results in the literature, based on different assumptions. So far existing literature has estimated transport coefficients from linear response theory and entered those into a nonlinear Boltzmann equation by classical rate arguments [37,55–57,64–67]. We have proven that our master Eq. (6.1) is equivalent to the method of linear response only in the linear regime close to chemical equilibrium. Finally, we must point out that the Lee-Weinberg equation, adopted in [37,55–57,66] to rederive the DM overclosure bound in the nonlinear regime, is not the correct form of the number density equation to use if bound-state solutions exist in the spectrum. The ionization fraction causes the difference as discussed in great detail in our work.

When taking the vacuum limit, our master equation reduces correctly to the coupled system of classical Boltzmann equations for ideal number densities of bound and scattering states *in the limit of ionization equilibrium*. In our method, it came out as a consequence of assuming the system is in a grand canonical state. Namely, we have proven that the assumption of a grand canonical state automatically implies the Saha ionization equilibrium if bound-state solution exist. One has to take the assumption of ionization equilibrium to be fulfilled for all times with a grain of salt for the following reason. From the vacuum treatment it is known that the duration of Saha ionization equilibrium is limited. Therefore, when using our master equation one has to carefully check that this condition is satisfied for a sufficiently long period. And especially when the assumption of ionization equilibrium is not justified, one has to make sure that at least the abundance of the stable scattering states are not affected by out-of-ionization equilibrium effects which might be model dependent.

The reason why in our Keldysh formalism we cannot resolve this issue at the moment lies in one particular approximation, made from the beginning. Ultrasoft emissions and absorptions were dropped for simplicity. We leave the inclusion of those quantities for future work, but expect once they are included we can fully recover the general set of coupled classical Boltzmann equations in the vacuum limit of our (future) updated equations. Moreover, this would allow us to describe Sommerfeld-enhanced annihilation and bound-state decay at finite temperature for the first time beyond the ionization equilibrium.

In the regime where ionization equilibrium is maintained, we have shown that finite temperature effects strongly mix bound and scattering states and the effects are all encoded in the solution of the two-particle spectral function. Let us remark that the numerical results for the spectral function obtained in Sec. VII are compatible with the linear response theory approach [37,55–57,64–67], although we started from a completely different method. The component $G^{++--}_{\eta\xi,s}|_{\rm eq}$ in our master Eq. (6.1) can be enhanced by much more than 10%. In addition, our master equation is applicable to the nonlinear regime beyond the limitation of linear response if, at least, the ionization equilibrium is maintained. These results make it definitely worthwhile to further generalize our Keldysh description in order to correctly describe the out-of-ionization equilibrium transition at late times by including contributions from the ultrasoft scale.

ACKNOWLEDGMENTS

We are thankful to Kalliopi Petraki, Andrzej Hryczuk, Mikko Laine, and Shigeki Matsumoto for very useful comments on our manuscript. T. B. thanks the DESY Theory Group and Kavli IPMU for hospitality during an early phase of this project. This work is supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Science, Sports, and Culture (MEXT), Japan, World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan, and the JSPS Research Fellowships for Young Scientists (K. M.). T. B. and L. C. received funding from the German Research Foundation

[Deutsche Forschungsgemeinschaft (DFG)] through the Institutional Strategy of the University of Göttingen (RTG 1493), the European Union's Horizon 2020 research and innovation program InvisiblesPlus RISE under the Marie Sklodowska-Curie Grant Agreement No. 690575, and from the European Union's Horizon 2020 research and innovation program Elusives ITN under the Marie Sklodowska-Curie Grant Agreement No. 674896.

APPENDIX A: SEMIGROUP PROPERTY OF FREE CORRELATORS

Free correlators G_0 fulfil *semigroup properties*:

$$G_0^R(x,y) = +\int \mathrm{d}^3z\, G_0^R(x,z)G_0^R(z,y), \quad \text{for } t_x > t_z > t_y, \tag{A1}$$

$$G_0^R(x,y) = +\int \mathrm{d}^3z\,\mathrm{d}^3w\, G_0^R(x,w)G_0^R(w,z)G_0^R(z,y), \quad \text{for } t_x > t_w > t_z > t_y, \tag{A2}$$

$$G_0^A(x,y) = -\int \mathrm{d}^3z\, G_0^A(x,z)G_0^A(z,y), \quad \text{for } t_x < t_z < t_y, \tag{A3}$$

$$G_0^A(x,y) = +\int \mathrm{d}^3z\,\mathrm{d}^3w\, G_0^A(x,w)G_0^A(w,z)G_0^A(z,y), \quad \text{for } t_x < t_w < t_z < t_y, \tag{A4}$$

$$G_0^{+-/-+}(x,y) = +\int \mathrm{d}^3z\, G_0^R(x,z)G_0^{+-/-+}(z,y), \quad \text{for } t_x > t_z, \tag{A5}$$

$$G_0^{+-/-+}(x,y) = -\int \mathrm{d}^3z\, G_0^{+-/-+}(x,z)G_0^A(z,y), \quad \text{for } t_z < t_y, \tag{A6}$$

$$G_0^{+-/-+}(x,y) = -\int \mathrm{d}^3z\,\mathrm{d}^3w\, G_0^R(x,w)G_0^{+-/-+}(w,z)G_0^A(z,y), \quad \text{for } t_x > t_w \text{ and } t_z < t_y. \tag{A7}$$

Note, there is no time integration here. All relations follow from the first Eq. (A1) by using Eq. (2.13). It might be helpful to prove the first equation from definition, where we have

$$G_0^R(x,y) = \theta(t_x - t_y)[G_0^{-+}(x,y) - G_0^{+-}(x,y)] = G_0^\rho(x,y), \quad \text{for } t_x > t_y. \tag{A8}$$

The free spectral function for nonrelativistic particles is in Fourier space given by $G_0^\rho(\omega,\mathbf{p}) = (2\pi)\delta(\omega - \frac{\mathbf{p}^2}{2m})$. Then it follows:

$$\int \mathrm{d}^3z\, G_0^R(x,z)G_0^R(z,y) \tag{A9}$$

$$= \int \mathrm{d}^3z\, G_0^\rho(x,z)G_0^\rho(z,y), \quad \text{for } t_x > t_z > t_y \tag{A10}$$

$$\begin{aligned} &= \int \mathrm{d}^3z \int \frac{\mathrm{d}^3p}{(2\pi)^3}\frac{\mathrm{d}\omega}{(2\pi)}(2\pi)\delta\left(\omega - \frac{\mathbf{p}^2}{2m}\right) \\ &\quad \times e^{i(\omega(t_x - t_z) - \mathbf{p}\cdot(\mathbf{x}-\mathbf{z}))}\int \frac{\mathrm{d}^3p'}{(2\pi)^3}\frac{\mathrm{d}\omega'}{(2\pi)}(2\pi) \\ &\quad \times \delta\left(\omega' - \frac{(\mathbf{p}')^2}{2m}\right)e^{i(\omega'(t_z - t_y) - \mathbf{p}'\cdot(\mathbf{z}-\mathbf{y}))} \end{aligned} \tag{A11}$$

$$= \int \frac{\mathrm{d}^3p}{(2\pi)^3} e^{i(\frac{\mathbf{p}^2}{2m}(t_x - t_y) - \mathbf{p}\cdot(\mathbf{x}-\mathbf{y}))} \tag{A12}$$

$$= G_0^R(x,y), \quad \text{for } t_x > t_y. \square \tag{A13}$$

APPENDIX B: ANNIHILATION TERM

Although one can directly compute $\Gamma_s(x,y)$ defined on the closed-time-path contour, it is instructive to see how one can recover it from the more common computation of annihilations. Usually, we compute the matrix element by means of Feynman correlators so as to evaluate annihilations, which means that both x and y are on the $\mathcal{C}_+$ contour. And thus, it yields the upper left component of Γ_s, i.e., Γ_s^{++}. The question is how to recover the remaining three components. To answer it, let us go back one step further, namely before integrating out hard products of the annihilation. Suppose that the interaction with them takes the following form: $O_s(x)O_H[\chi(x)]e^{-2iMx_0} + \mathrm{H.c.}$, with χ being the products of the annihilation. Here O_s is defined as Eq. (3.4) and O_H represents hard d.o.f. Then, Γ_s is obtained from the cuttings of $\langle T_\mathcal{C} O_H^\dagger[\chi(x)]O_H[\chi(y)]\rangle e^{2iM(x_0 - y_0)}$. Since we assume that the background plasma is in thermal equilibrium and does not change by η and ξ reactions, the two-point correlator of O_H only depends on the space-time difference $x - y$:

$$G_{O_H}(x-y) \equiv \langle T_\mathcal{C} O_H^\dagger[\chi(x)]O_H[\chi(y)]\rangle. \tag{B1}$$

By definition, incoming energy/momentum from O_s is much smaller than M, which justifies the following approximation:

$$\begin{aligned} &\int_{x,y\in\mathcal{C}_+} O_s^\dagger(x)O_s(y)iG_{O_H}^{++}(x-y)e^{2iM(x_0-y_0)} \\ &\simeq \int_{x,y\in\mathcal{C}_+} O_s^\dagger(x)O_s(y)iG_{O_H}^{++}(2M,\mathbf{0})\delta(x-y). \end{aligned} \tag{B2}$$

Taking the imaginary part and comparing it with Γ_s^{++}, one can see that the usual computation corresponds to $\Im iG_{O_H}^{++}(2M,\mathbf{0})$. As the background plasma is assumed to be close to equilibrium, we can use the Kubo-Martin-

Schwinger (KMS) relation, which essentially connects all the other combinations, $G^{-+}_{O_H}$, $G^{+-}_{O_H}$, $G^{--}_{O_H}$, with this one. Moreover, because of $M \gg T$, one can safely neglect $G^{+-}_{O_H}$. As a result, we end up with

$$\Im i G^{++}_{O_H}(2M, \mathbf{0}) = \frac{\pi(\alpha_\chi^2 + \alpha_\chi \alpha_\psi)}{M^2}, \tag{B3}$$

$$\Im i G^{+-}_{O_H}(2M, \mathbf{0}) = 0, \tag{B4}$$

$$\begin{aligned} \Im i G^{-+}_{O_H}(2M, \mathbf{0}) &= \Im i G^{\mathrm{ret}}_{O_H}(2M, \mathbf{0}) - \Im i G^{\mathrm{adv}}_{O_H}(2M, \mathbf{0}) \\ &= 2\Im i G^{++}_{O_H}(2M, \mathbf{0}), \end{aligned} \tag{B5}$$

$$\begin{aligned} \Im i G^{--}_{O_H}(2M, \mathbf{0}) &= -\Im i G^{\mathrm{ret}}_{O_H}(2M, \mathbf{0}) + \Im i G^{-+}_{O_H}(2M, \mathbf{0}) \\ &= \Im i G^{++}_{O_H}(2M, \mathbf{0}), \end{aligned} \tag{B6}$$

which results in Eq. (3.4). Here, several properties of equilibrium correlators were used, as given in Sec. II.

APPENDIX C: HARD THERMAL LOOP APPROXIMATION

The Fourier transform of the ideal ψ two-point function is in the Keldysh representation given by

$$\begin{aligned} S(P) = (\not{P} + m)\Bigg[&\begin{pmatrix} \frac{i}{P^2 - m^2 + i\epsilon} & 0 \\ 0 & \frac{-i}{P^2 - m^2 - i\epsilon} \end{pmatrix} - 2\pi\delta(P^2 - m^2) \\ &\times \begin{pmatrix} n_F(|p^0|) & -\theta(-p^0) + n_F(|p^0|) \\ -\theta(p^0) + n_F(|p^0|) & n_F(|p^0|) \end{pmatrix}\Bigg]. \end{aligned} \tag{C1}$$

Combining different components, the following retarded, advanced, and symmetric propagator can be obtained:

$$S^{R/A}(P) = \frac{i(\not{P} + m)}{P^2 - m^2 \pm i\,\mathrm{sign}(p^0)\epsilon}, \tag{C2}$$

$$\begin{aligned} S^s(P) &\equiv S^{++}(P) + S^{--}(P) \\ &= 2\pi(\not{P} + m)[1 - 2n_F(|p^0|)]\delta(P^2 - m^2). \end{aligned} \tag{C3}$$

The inverse relations are given by

$$\begin{aligned} S^{++} &= \frac{1}{2}(S^s + S^R + S^A), \\ S^{+-} &= \frac{1}{2}(S^s - S^R + S^A), \\ S^{-+} &= \frac{1}{2}(S^s + S^R - S^A). \end{aligned} \tag{C4}$$

By using these relations, the one-loop expression of the retarded mediator correlator can be simplified as

$$\begin{aligned} &\mathrm{Tr}[\gamma^\mu S^{++}(x,y)\gamma^\nu S^{++}(y,x)] - \mathrm{Tr}[\gamma^\mu S^{+-}(x,y)\gamma^\nu S^{-+}(y,x)] \\ &= \frac{1}{2}\mathrm{Tr}[\gamma^\mu S^R(x,y)\gamma^\nu S^S(y,x)] + \frac{1}{2}\mathrm{Tr}[\gamma^\mu S^S(x,y)\gamma^\nu S^A(y,x)]. \end{aligned} \tag{C5}$$

Let us take the limit $m \ll T$, where m is the ψ mass, leading in Fourier space to the following retarded self-energy of the mediator:

$$\Pi^{00}_R(P) = -g_\psi^2 8\pi \int \frac{\mathrm{d}^4 K}{(2\pi)^4}[(k^0 - p^0)k^0 + (\mathbf{k} - \mathbf{p})\cdot\mathbf{k}][1 - 2n_F(|k^0|)]\delta(K^2)\frac{1}{(K - P)^2 - i\,\mathrm{sgn}(k^0 - p^0)\epsilon}. \tag{C6}$$

Dropping the vacuum part and integrating over k^0 one obtains:

$$\Pi^{00}_R = g_\psi^2 8\pi \int \frac{\mathrm{d}^3 k}{(2\pi)^4}\frac{n_F(|\mathbf{k}|)}{|\mathbf{k}|}\left[\frac{(|\mathbf{k}| - p^0)|\mathbf{k}| + (\mathbf{k} - \mathbf{p})\cdot\mathbf{k}}{(|\mathbf{k}| - p^0)^2 - (\mathbf{k} - \mathbf{p})^2 - i\,\mathrm{sgn}(|\mathbf{k}| - p^0)\epsilon} + \frac{(|\mathbf{k}| + p^0)|\mathbf{k}| + (\mathbf{k} - \mathbf{p})\cdot\mathbf{k}}{(|\mathbf{k}| + p^0)^2 - (\mathbf{k} - \mathbf{p})^2 + i\,\mathrm{sgn}(|\mathbf{k}| + p^0)\epsilon}\right]. \tag{C7}$$

The result so far is exact up to the fact that we neglected the ψ mass and ignored the vacuum contribution. Now, the *hard-thermal-loop approximation* [71] assumes that the external energy p^0 and momentum $|\mathbf{p}|$ are smaller compared to the typical loop momentum $|\mathbf{k}|$ which is of the order temperature (hard), since the integrand contains n_F. Expanding the term in the brackets to leading order in $p^0/|\mathbf{k}|$ and $|\mathbf{p}|/|\mathbf{k}|$, all remaining integrals can be performed analytically leading to the finite result:

$$\Pi^{00}_{\mathrm{R,A}}(P) \simeq -m_D^2\left[1 - \frac{p^0}{2|\mathbf{p}|}\ln\left(\frac{p^0 + |\mathbf{p}| \pm i\epsilon}{p^0 - |\mathbf{p}| \pm i\epsilon}\right)\right], \tag{C8}$$

where the Debye mass is defined as $m_D^2 = g_\psi^2 T^2/3$. This result coincides with the result obtained from the imaginary time formalism, where instead one has to perform a sum over Matsubara frequencies.

APPENDIX D: VACUUM LIMIT OF TWO-PARTICLE SPECTRAL FUNCTION

Here we will derive Eq. (6.12) starting from Eq. (5.44) in the limit of vanishing finite temperature corrections. As we have briefly mentioned in Sec. V D, for the vacuum limit one has to carefully take into account the imaginary part $i\epsilon$ in the retarded equation, representing the small width which will be taken to be zero in the end. We will see that the result does not depend on this $i\epsilon$ prescription as long as ϵ is small enough.

Suppose that the potential almost vanishes for a large enough r. For a Yukawa type potential, this is true for $m_V r \gg 1$. The ϵ parameter should be much smaller than this mass parameter, namely $\epsilon \ll m_V$. In the case of the Coulomb potential one may introduce another small mass parameter to the gauge boson. In the end of the computation one can take it to be zero while keeping $m_V \gg \epsilon$. For $m_V r \gg 1$, the homogeneous solutions, $g_{>,<}$, can be well approximated by the plane wave:

$$g_>(r) \to C_> e^{i\sqrt{ME}r - \epsilon r}, \tag{D1}$$

$$g_<(r) \to \frac{1}{2}\left(C_< e^{i\sqrt{ME}r-\epsilon r} + C_<^\dagger e^{-i\sqrt{ME}r+\epsilon r}\right). \tag{D2}$$

The Wronskian tells us that there exists a nontrivial relation between two coefficients $C_{>,<}$. Since the Wronskian, $W(r) = g_>(r)g'_<(r) - g'_>(r)g_<(r)$, does not depend on r, one may equate it at $r = 0$ and $r \to \infty$, which yields

$$1 = W(0) = W(\infty) = -i\sqrt{ME}C_>C_<^\dagger \leftrightarrow C_> = \frac{i}{\sqrt{ME}}\frac{1}{C_<^\dagger}. \tag{D3}$$

We have taken ϵ to be zero in the end of the computation.

Let us evaluate the integral given in Eq. (5.44) by means of Eq. (D2). The first observation is that there is no imaginary part for $g_<$ if ϵ is zero from the beginning. Thus, the integrand becomes relevant only after $r \gtrsim \epsilon^{-1} \gg m_V^{-1}$. As a result one may evaluate the integral in Eq. (5.44) by substituting Eq. (D2):

$$G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{E>0,l=0} = \frac{1}{2\pi}\mathrm{Tr}[\mathbf{1}_{2\times2}]M\frac{1}{|C_<|^2}\lim_{\epsilon\to0}\Im \times \int_{m_V^{-1}}^{\infty} \mathrm{d}r \frac{1}{\cos^2[(\sqrt{ME}+i\epsilon)r + \delta_{C_<}]} \tag{D4}$$

$$= \frac{1}{2\pi}\mathrm{Tr}[\mathbf{1}_{2\times2}]M\frac{1}{|C_<|^2}\frac{1}{\sqrt{ME}}. \tag{D5}$$

Finally, substituting Eq. (D3) into this equation, we arrive at

$$G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{E>0,l=0} = \frac{1}{4\pi}\mathrm{Tr}[\mathbf{1}_{2\times2}]M^2 v_{\mathrm{rel}}|C_>|^2, \tag{D6}$$

where we have used $E = Mv_{\mathrm{rel}}^2/4$.

Now we are in a position to discuss its relation to the conventional definition of the Sommerfeld enhancement factor. In the limit of $\epsilon \to 0$, the wave function propagates to infinity. Then one may obtain the Sommerfeld enhancement factor by extracting the amplitude of the wave function at the infinity, which is nothing but the relation $S(v_{\mathrm{rel}}) = |C_>|^2$; see Ref. [44] for instance. Utilizing this relation, we finally get Eq. (6.12). For conventional reason, let us give the s-wave Sommerfeld enhancement factor for the Coulomb case consistent with our equations:

$$S(v_{\mathrm{rel}}) = \frac{2\pi\alpha_\chi}{v_{\mathrm{rel}}}\frac{1}{1-e^{-\frac{2\pi\alpha_\chi}{v_{\mathrm{rel}}}}}. \tag{D7}$$

For the bound state, it is much easier to solve the equation directly rather starting from Eq. (5.44). One may express the spectral function by means of the wave functions for the bound states [37]:

$$G^{\rho}_{\eta\xi}(\mathbf{0},\mathbf{0};E)|_{E<0,l=0} = 2\pi\mathrm{Tr}[\mathbf{1}_{2\times2}]\sum_n \delta(E-E_B^n)|\psi_n^{(B)}(0)|^2, \tag{D8}$$

where $\psi_n^{(B)}$ represents the normalized wave function for the nth bound state. For instance, the wave function for the lowest energy state $n = 1$ is given by

$$\psi_1^{(B)} = \frac{1}{\sqrt{\pi}}\left(\frac{\alpha_\chi M}{2}\right)^{3/2} e^{-\alpha_\chi Mr/2}. \tag{D9}$$

The decay rate of the bound state is related to its wave function at the origin. For the lowest state, one can easily show this from Eq. (D9):

$$(\sigma v_{\mathrm{rel}})|\psi_1^{(B)}(0)|^2 = \frac{1}{4}\Gamma_{{}^1S_0}, \tag{D10}$$

where the decay rate of the lowest bound state is given by

$$\Gamma_{{}^1S_0} = (\alpha_\chi^2 + \alpha_\chi\alpha_\psi)\alpha_\chi^3\frac{M}{2}. \tag{D11}$$

Similar calculation holds in the limit of negligible thermal width but finite real-part corrections. For this case one should substitute the kinetic energy $E \to E - \Re V_{\mathrm{eff}}(\infty) = Mv_{\mathrm{rel}}^2/4$ and similar for the bound-state energy. Also one has to take ϵ smaller than m_D and m_V.

APPENDIX E: NUMBER DENSITY AND CHEMICAL POTENTIAL IN GRAND CANONICAL ENSEMBLE

In this section, we present an alternative way of how to derive the chemical potential as a function of the total number density directly from the EoM. It is convenient to write the EoM in integral form as

$$G_\eta(x,y) = G_{\eta,0}(x,y) - g_\chi^2 \int_{w,z\in\mathcal{C}} G_{\eta,0}(x,w) D(w,z) \times [G_{\eta\xi}(w,z,y,z) - G_{\eta\eta}(w,z,y,z)], \tag{E1}$$

$$\bar{G}_\xi(x,y) = G_{\xi,0}(x,y) - g_\chi^2 \int_{w,z\in\mathcal{C}} G_{\xi,0}(x,w) D(w,z) \times [G_{\eta\xi}(z,w,z,y) - G_{\xi\xi}(w,z,y,z)]. \tag{E2}$$

The number density is given by the $+-$ component, namely in Wigner coordinates it reads $n_\eta(\mathbf{R},T) = \mathrm{Tr}[G_\eta^{+-}(x,x)] = \mathrm{Tr}[G_\eta^{+-}(\mathbf{0},0;\mathbf{R},T)]$. Now one can already clearly see the structure that the total number density is given by the ideal number density plus interactions. These equations are exact within our nonrelativistic effective action. Expanding D around the narrow width limit and assuming a grand canonical state, one should recover Eq. (6.35) for the number density since the only assumption entering there is the narrow width approximation. Equation (6.35) is the correct thermodynamic definition of the number density which should coincide with the result obtained by solving for the number density from these integral EoM when assuming a thermodynamic picture like the grand canonical ensemble. It requires a rigorous proof of this claim, which we would like to give somewhere else.

Instead, we would like to give some approximations in order to obtain an analytic solution of Eq. (E1). We restrict the discussion by assuming the system is in a grand canonical state. Utilizing the KMS condition for finite chemical potential, one can formally solve G_η^{+-} in terms of spectral function:

$$\begin{aligned} n_\eta &= \mathrm{Tr}[G_\eta^{+-}(\mathbf{0},0)] \\ &= \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}\frac{\mathrm{d}\omega}{(2\pi)}\frac{1}{e^{\beta(M+\omega-\mu_\eta)}+1}\mathrm{Tr}[G_\eta^\rho(\mathbf{p},\omega)] \\ &\simeq \int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}\frac{\mathrm{d}\omega}{(2\pi)} e^{-\beta(M+\omega-\mu_\eta)}\mathrm{Tr}[G_\eta^\rho(\mathbf{p},\omega)]. \end{aligned} \tag{E3}$$

In the last equality we assumed the DM gas to be dilute and approximated the Fermi-Dirac distribution as a Maxwellian. By KMS relation we have formally solved for the number density in terms of chemical potential and spectral function. The spectral function can be computed from the retarded component G_η^R, according to $G_{\eta,0}^\rho(\mathbf{p},\omega) = G_{\eta,0}^R(\mathbf{p},\omega) - G_{\eta,0}^A(\mathbf{p},\omega)$ (see Sec. II). The EoM of the retarded correlator can be obtained from Eq. (E1) by subtracting $+-$ from the $++$ component, e.g., $G_\eta^R = G_\eta^{++} - G_\eta^{+-}$. In subsequent sections we solve the retarded equation in various approximations, compute the spectral function and finally evaluate Eq. (E3).

1. Ideal gas approximation

The ideal gas approximation can be defined as the zeroth order contribution in Eq. (E1). This means we have to know the solution of the free retarded correlator. It can be obtained from the differential form of the retarded equations. In Fourier space of the microscopic Wigner coordinates it is given by $G_{\eta,0}^R(\mathbf{p},\omega) = i\delta_{ij}/(\omega - \mathbf{p}^2/(2M) + i\epsilon)$, leading to $G_{\eta,0}^\rho(\mathbf{p},\omega) = \delta_{ij}(2\pi)\delta(\omega - \mathbf{p}^2/2M)$. Inserting this into Eq. (E3) results in the ideal gas approximation of the free number density:

$$n_{\eta,0} = 2\int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} e^{-\beta(M+\frac{\mathbf{p}^2}{2M}-\mu_\eta)}. \tag{E4}$$

We can invert this relation to obtain the chemical potential as a function of the number density in the ideal gas approximation:

$$\beta\mu_\eta^{\mathrm{id}} = \ln\left[\frac{n_{\eta,0}}{n_{\eta,0}^{\mathrm{eq}}}\right], \quad \text{where } n_{\eta,0}^{\mathrm{eq}} = 2\int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3} e^{-\beta(M+\mathbf{p}^2/2M)}. \tag{E5}$$

And similar expressions for antiparticle ξ.

2. Hartree-Fock approximation

The Hartree-Fock approximation is the zero order approximation of the four-point correlator. Using Eqs. (4.9)–(4.10), one obtains

$$[G_{\eta\xi}(x,z,y,z) - G_{\eta\eta}(x,z,y,z)] \simeq G_{\eta,0}(x,z)G_{\eta,0}(z,y), \tag{E6}$$

$$[G_{\eta\xi}(z,x,z,y) - G_{\xi\xi}(x,z,y,z)] \simeq \bar{G}_{\xi,0}(x,z)\bar{G}_{\xi,0}(z,y), \tag{E7}$$

valid for symmetric DM. Inserting this into Eqs. (E1)–(E2) and subtracting the components $G_\eta^{++} - G_\eta^{+-}$, one obtains for the retarded equations:

$$G_\eta^R(x,y) = G_{\eta,0}^R(x,y) + \int \mathrm{d}^4z\mathrm{d}^4w G_{\eta,0}^R(x,z)(-i) \times \Sigma_\eta^R(z,w)G_{\eta,0}^R(w,y), \tag{E8}$$

$$\bar{G}_\xi^R(x,y) = \bar{G}_{\xi,0}^R(x,y) + \int \mathrm{d}^4z\mathrm{d}^4w \bar{G}_{\xi,0}^R(x,z)(-i) \times \bar{\Sigma}_\xi^R(z,w)\bar{G}_{\xi,0}^R(w,y). \tag{E9}$$

The single particle self-energies are defined on the CTP contour as we have introduced in Eq. (5.20). The retarded component is $\Sigma^R = \Sigma^{++} - \Sigma^{+-}$ and can be written as

$$\begin{aligned}
&\Sigma^R(x,y)/(-ig_\chi^2)\\
&\quad = D^{++}(x,y)G_0^{++}(x,y) - D^{+-}(x,y)G_0^{+-}(x,y) &\text{(E10)}\\
&\quad = D^{++}(x,y)G_0^{R}(x,y) - D^{R}(x,y)G_0^{+-}(x,y) &\text{(E11)}\\
&\quad = D^{-+}(x,y)G_0^{R}(x,y) - D^{R}(x,y)G_0^{+-}(x,y). &\text{(E12)}
\end{aligned}$$

In the last step, the definition $D^{++}(t,t') = \theta(t-t')D^{-+}(t,t') + \theta(t'-t)D^{+-}(t,t')$ was used and the fact that the retarded function G_0^R projects out only the $D^{-+}(t,t')$ contribution due to equal times in D and G_0^R. A key observation is that the self-energy can depend on the DM number density and chemical potential due to the G_0^{+-} contribution. This might lead to a nonlinear dependence of the total number density on $e^{\beta\mu}$ when inserting the spectral function obtained from the retarded equation into Eq. (E3). In the following we would like to perturbatively resum Eqs. (E8) and (E9) which brings us later to the widely used Salpeter correction. Let us therefore drop the dependence of the self-energy on $G_0^{+-}(x,y)$. Then the equation is closed in terms of the retarded correlators G_0^R. Within this approximation, the self-energy can be written in Fourier space as (to leading order gradient expansion):

$$\begin{aligned}
&\Sigma^R_\eta(\mathbf{p},\omega)\\
&\quad = g_\chi^2\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}\frac{d\omega_1}{(2\pi)}\frac{d\omega_2}{(2\pi)}\frac{G^\rho_{\eta,0}(\omega_1,\mathbf{p}-\mathbf{p}_1)D^{-+}(\omega_2,\mathbf{p}_1)}{\omega-\omega_1-\omega_2+i\epsilon} &\text{(E13)}\\
&\quad = g_\chi^2\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}\frac{d\omega_1}{(2\pi)}\frac{D^{-+}(\omega_1,\mathbf{p}_1)}{\Omega-\omega_1+i\epsilon} &\text{(E14)}\\
&\quad = g_\chi^2\frac{1}{2}\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}\frac{d\omega_1}{(2\pi)}\frac{D^{-+}(\omega_1,\mathbf{p}_1)+D^{-+}(\omega_1,\mathbf{p}_1)}{\Omega-\omega_1+i\epsilon} &\text{(E15)}\\
&\quad = g_\chi^2\frac{1}{2}\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}\frac{d\omega_1}{(2\pi)}\frac{D^{-+}(\omega_1,\mathbf{p}_1)+D^{+-}(-\omega_1,\mathbf{p}_1)}{\Omega-\omega_1+i\epsilon} &\text{(E16)}\\
&\quad = -ig_\chi^2\frac{1}{2}\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}D^{++}(\Omega,\mathbf{p}_1), &\text{(E17)}
\end{aligned}$$

where $\Omega = \omega - (\mathbf{p}-\mathbf{p}_1)^2/2M$. The final form is convenient for inserting the static HTL approximation of D^{++} as given in Eq. (3.11). Performing the one-loop calculation results in

$$\begin{aligned}
\lim_{\Omega\to 0}\Sigma^R(\mathbf{p},\omega) &= -ig_\chi^2\frac{1}{2}\int\frac{d^3\mathbf{p}_1}{(2\pi)^3}\lim_{\Omega\to 0}D^{++}(\Omega,\mathbf{p}_1)\\
&= -\frac{1}{2}(\alpha_\chi m_D + i\alpha_\chi T). \qquad \text{(E18)}
\end{aligned}$$

One can recognize that this result is exactly half the effective in-medium potential for two particles at large distance; see Eq. (5.32). Now, for perturbative resummation we replace the retarded correlators at the end of Eqs. (E8) and (E9) by the fully dressed one: $G^R_{\eta,0}(w,y)\to G^R_\eta(w,y)$, $\bar{G}^R_{\xi,0}(w,y)\to G^R_\xi(w,y)$. Performing Wigner and Fourier transformation of the equation leads at the leading order in gradient expansion to

$$G^R_\eta(\mathbf{p},\omega) = G^R_{\eta,0}(\mathbf{p},\omega) + G^R_{\eta,0}(\mathbf{p},\omega)(-i)\Sigma^R_\eta(\mathbf{p},\omega)G^R_\eta(\mathbf{p},\omega), \tag{E19}$$

and similar equation for the antiparticle. Then, by using geometric series one ends up with the HTL single particle correlators:

$$\begin{aligned}
G^R_\eta(\mathbf{p},\omega) &= \frac{i\delta_{ij}}{\omega-\mathbf{p}^2/2M-\Sigma^R_\eta(\mathbf{p},\omega)+i\epsilon},\\
\bar{G}^R_\xi(\mathbf{p},\omega) &= \frac{i\delta_{ij}}{\omega-p^2/2M-\Sigma^R_\xi(\mathbf{p},\omega)+i\epsilon}. \qquad \text{(E20)}
\end{aligned}$$

Computing the spectral function from the difference of retarded and advanced correlators results in a Breit-Wigner shape:

$$G^\rho_\eta(\mathbf{p},\omega) = \delta_{ij}\frac{\Gamma_\eta}{(\omega-\mathbf{p}^2/2M-\Re(\Sigma^R_\eta))^2+(\Gamma_\eta/2)^2}, \tag{E21}$$

where the particle width is defined by $\Gamma\equiv 2\Im(\Sigma^R)$. In summary, we evaluated the four-point correlator in the Hartree-Fock (HF) approximation and formally solved for ρ as a function of the self-energy in static HTL approximation, shifting the energy by $\alpha_\chi m_D/2$ and broadening the peak via imaginary contributions. Finally, let us quote the chemical potential in HF and static HTL approximation:

$$\begin{aligned}
&\beta\mu^{\mathrm{HF}}_\eta = \ln\left[\frac{n_\eta}{n^{\mathrm{eq}}_{\eta,\mathrm{HF}}}\right], \quad \text{where}\\
&n^{\mathrm{eq}}_{\eta,\mathrm{HF}} = 2\int\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{d\omega}{(2\pi)}e^{-\beta(M+\omega)}\\
&\qquad\times\frac{\Gamma}{(\omega-\mathbf{p}^2/2M-\Re(\Sigma^R_\eta))^2+(\Gamma/2)^2}. \qquad \text{(E22)}
\end{aligned}$$

This approximation might be already good enough if there are no bound states but the two-particle spectral function has support at negative energies due to thermal width. We also see that one-particle spectral function can have spectral support at negative energies.

3. Salpeter correction

Taking the limit $\Gamma \to 0$ in the spectral function Eq. (E21) results in

$$\lim_{\Gamma\to 0} G^{\rho}_{\eta}(p,\omega) = \delta_{ij}(2\pi)\delta(\omega - p^2/2M - \Re(\Sigma^R_\eta)). \tag{E23}$$

This is called the *narrow width* or *quasiparticle* approximation, taking only the real-part correction into account. Inserting the spectral function into Eq. (E3) leads to the chemical potential and equilibrium number density:

$$\beta\mu^{\mathrm{SP}}_{\eta} = \ln\left[\frac{n_\eta}{n^{\mathrm{eq}}_{\eta,\mathrm{SP}}}\right], \quad \text{where}$$
$$n^{\mathrm{eq}}_{\eta,\mathrm{SP}} = 2\int \frac{\mathrm{d}^3\mathbf{p}}{(2\pi)^3}\, e^{-\beta(M + p^2/2M + \Re[\Sigma^R_\eta])}. \tag{E24}$$

For the self-energy in the static HTL approximation, the real-part correction is $\Re(\Sigma^R_\eta) = -\alpha_\chi m_D/2$, according to Eq. (E18). This is the well-known *Salpeter correction* to the equilibrium distribution.

The Salpeter correction is a simple first order approximation for the description of quasi particles in a plasma. As we have seen in Sec. VI D, it might be however required for our number density equation to compute both, the annihilation/decay rates *and* the chemical potential, to the same level of approximation in the four-point correlator. Especially when bound-state solutions exist it is required to solve the four-point correlator nonperturbatively (by resummation). The Salpeter correction was obtained by approximating the four-point correlator as a product of free particles without self-interactions; see Eqs. (E6) and (E8). Thus, in this approximation of the chemical potential, bound-state contributions never appear. It might lead to inconsistencies in the number density equation, like exponentially growing terms for late times.

Let us illustrate why the Salpeter correction is not enough to correctly describe the freeze-out at late times. We plug the chemical potential μ^{SP}_η in Salpeter approximation into our master formula for the number density and obtain

$$\dot{n}_\eta + 3Hn_\eta = -2(\sigma v_{\mathrm{rel}}) G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}[e^{\beta 2\mu^{\mathrm{SP}}_\eta} - 1] \tag{E25}$$

$$= -\frac{2(\sigma v_{\mathrm{rel}}) G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}}{n^{\mathrm{eq}}_{\eta,\mathrm{SP}}(T)\, n^{\mathrm{eq}}_{\eta,\mathrm{SP}}(T)}[n_\eta n_\eta - n^{\mathrm{eq}}_{\eta,\mathrm{SP}}(T)\, n^{\mathrm{eq}}_{\eta,\mathrm{SP}}(T)]. \tag{E26}$$

There are two reasons why this description fails at late times if bound-state solutions exist. First of all, since no bound states are included in the computation of the chemical potential in Salpeter approximation, the ionization degree would always be approximated as 1. Second, if one would compute the spectral function in $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}}$ nonperturbatively, this would cause an exponential growing term at late times, caused by the bound states: $G^{++--}_{\eta\xi,s}(x,x,x,x)|_{\mathrm{eq}} \propto e^{\beta|E_B|}$. In the Salpeter approximation, the denominator $n^2_{\eta,\mathrm{SP}}$ cannot kill this unphysical behavior. Note however, if one computes the chemical potential to the same level as the annihilation rates, the degree of ionization exhibits a term leading to a cancellation of the exponential growing term in $G^{++--}_{\eta\xi,s}$. Our main number density equation automatically incorporates this because it evaluates the chemical potential and number density to the same level of approximation. The Salpeter correction together with solving the spectral function nonperturbatively was used in Refs. [55–57].

---

[1] P. A. R. Ade *et al.* (Planck Collaboration), Astron. Astrophys. **594**, A13 (2016).

[2] N. Aghanim *et al.* (Planck Collaboration), arXiv:1807.06209.

[3] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, Astrophys. J. **679**, 1173 (2008).

[4] J. Miralda-Escude, Astrophys. J. **564**, 60 (2002).

[5] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, Mon. Not. R. Astron. Soc. **430**, 105 (2013).

[6] S. Tulin and H.-B. Yu, Phys. Rep. **730**, 1 (2018).

[7] B. W. Lee and S. Weinberg, Phys. Rev. Lett. **39**, 165 (1977).

[8] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, Nucl. Phys. **B238**, 453 (1984).

[9] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre, S. Profumo, and F. S. Queiroz, Eur. Phys. J. C **78**, 203 (2018).

[10] J. Hisano, S. Matsumoto, and M. M. Nojiri, Phys. Rev. Lett. **92**, 031303 (2004).

[11] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, Phys. Lett. B **646**, 34 (2007).

[12] A. Sommerfeld, Ann. Phys. (Amsterdam) **403**, 257.

[13] A. D. Sakharov, Zh. Eksp. Teor. Fiz. **18**, 631 (1948) Usp. Fiz. Nauk **161**, 29 (1991).

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

[15] J. D. March-Russell and S. M. West, Phys. Lett. B **676**, 133 (2009).

[16] W. Shepherd, T. M. P. Tait, and G. Zaharijas, Phys. Rev. D **79**, 055022 (2009).

[17] K. Belotsky, D. Fargion, M. Khlopov, and R. V. Konoplich, Phys. At. Nucl. **71**, 147 (2008).

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

[19] T. R. Slatyer, J. Cosmol. Astropart. Phys. 02 (2010) 028.
[20] J. L. Feng, M. Kaplinghat, and H.-B. Yu, Phys. Rev. D **82**, 083525 (2010).
[21] B. von Harling and K. Petraki, J. Cosmol. Astropart. Phys. 12 (2014) 033.
[22] K. Petraki, M. Postma, and M. Wiechers, J. High Energy Phys. 06 (2015) 128.
[23] H. An, M. B. Wise, and Y. Zhang, Phys. Rev. D **93**, 115020 (2016).
[24] P. Asadi, M. Baumgart, P. J. Fitzpatrick, E. Krupczak, and T. R. Slatyer, J. Cosmol. Astropart. Phys. 02 (2017) 005.
[25] E. Johnson, E. Braaten, and H. Zhang, Proc. Sci., ICHEP2016 (**2016**) 210 [arXiv:1611.06212].
[26] E. Braaten, E. Johnson, and H. Zhang, J. High Energy Phys. 11 (2017) 108.
[27] E. Braaten, E. Johnson, and H. Zhang, J. High Energy Phys. 05 (2018) 062.
[28] E. Braaten, E. Johnson, and H. Zhang, J. High Energy Phys. 02 (2018) 150.
[29] A. Hryczuk, R. Iengo, and P. Ullio, J. High Energy Phys. 03 (2011) 069.
[30] M. Beneke, A. Bharucha, F. Dighera, C. Hellmann, A. Hryczuk, S. Recksiegel, and P. Ruiz-Femenia, J. High Energy Phys. 03 (2016) 119.
[31] A. Freitas, Phys. Lett. B **652**, 280 (2007).
[32] A. Hryczuk, Phys. Lett. B **699**, 271 (2011).
[33] K. Harigaya, K. Kaneta, and S. Matsumoto, Phys. Rev. D **89**, 115021 (2014).
[34] J. Harz, B. Herrmann, M. Klasen, K. Kovarik, and M. Meinecke, Phys. Rev. D **91**, 034012 (2015).
[35] M. Beneke, C. Hellmann, and P. Ruiz-Femenia, J. High Energy Phys. 05 (2015) 115.
[36] J. Ellis, J. L. Evans, F. Luo, and K. A. Olive, J. High Energy Phys. 02 (2016) 071.
[37] S. Kim and M. Laine, J. Cosmol. Astropart. Phys. 01 (2017) 013.
[38] S. El Hedri, A. Kaminska, and M. de Vries, Eur. Phys. J. C **77**, 622 (2017).
[39] S. P. Liew and F. Luo, J. High Energy Phys. 02 (2017) 091.
[40] A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, J. Cosmol. Astropart. Phys. 05 (2017) 006.
[41] J. Harz and K. Petraki, J. High Energy Phys. 07 (2018) 096.
[42] K. Petraki, M. Postma, and J. de Vries, J. High Energy Phys. 04 (2017) 077.
[43] J. Harz and K. Petraki, Phys. Rev. D **97**, 075041 (2018).
[44] M. Cirelli, A. Strumia, and M. Tamburini, Nucl. Phys. **B787**, 152 (2007).
[45] J. March-Russell, S. M. West, D. Cumberbatch, and D. Hooper, J. High Energy Phys. 07 (2008) 058.
[46] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, Phys. Rev. D **71**, 063528 (2005).
[47] M. Cirelli, R. Franceschini, and A. Strumia, Nucl. Phys. **B800**, 204 (2008).
[48] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, J. Cosmol. Astropart. Phys. 05 (2017) 036.
[49] J. Fan and M. Reece, J. High Energy Phys. 10 (2013) 124.
[50] B. Bhattacherjee, M. Ibe, K. Ichikawa, S. Matsumoto, and K. Nishiyama, J. High Energy Phys. 07 (2014) 080.
[51] M. Ibe, S. Matsumoto, S. Shirai, and T. T. Yanagida, Phys. Rev. D **91**, 111701 (2015).
[52] M. Beneke, A. Bharucha, A. Hryczuk, S. Recksiegel, and P. Ruiz-Femenia, J. High Energy Phys. 01 (2017) 002.
[53] C. F. Berger, L. Covi, S. Kraml, and F. Palorini, J. Cosmol. Astropart. Phys. 10 (2008) 005.
[54] K. Blum, R. Sato, and T. R. Slatyer, J. Cosmol. Astropart. Phys. 06 (2016) 021.
[55] S. Biondini and M. Laine, J. High Energy Phys. 08 (2017) 047.
[56] S. Biondini and M. Laine, J. High Energy Phys. 04 (2018) 072.
[57] S. Biondini, J. High Energy Phys. 06 (2018) 104.
[58] I. Baldes, M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, SciPost Phys. **4**, 041 (2018).
[59] I. Baldes and K. Petraki, J. Cosmol. Astropart. Phys. 09 (2017) 028.
[60] G. Jungman, M. Kamionkowski, and K. Griest, Phys. Rep. **267**, 195 (1996).
[61] G. Servant and T. M. P. Tait, Nucl. Phys. **B650**, 391 (2003).
[62] L. Covi, J. E. Kim, and L. Roszkowski, Phys. Rev. Lett. **82**, 4180 (1999).
[63] J. L. Feng, A. Rajaraman, and F. Takayama, Phys. Rev. Lett. **91**, 011302 (2003).
[64] Y. Burnier, M. Laine, and M. Vepsalainen, J. High Energy Phys. 01 (2008) 043.
[65] D. Bodeker and M. Laine, J. High Energy Phys. 01 (2013) 037.
[66] S. Kim and M. Laine, J. High Energy Phys. 07 (2016) 143.
[67] D. Bodeker and M. Laine, J. High Energy Phys. 07 (2012) 130.
[68] G. Stefanucci and R. van Leeuwen, *Nonequilibrium Many-Body Theory of Quantum Systems: A Modern Introduction* (Cambridge University Press, Cambridge, England, 2013).
[69] M. Beneke, F. Dighera, and A. Hryczuk, J. High Energy Phys. 10 (2014) 045; 07 (2016) 106(E).
[70] N. Brambilla, J. Ghiglieri, A. Vairo, and P. Petreczky, Phys. Rev. D **78**, 014017 (2008).
[71] E. Braaten and R. D. Pisarski, Nucl. Phys. **B337**, 569 (1990).
[72] M. L. Bellac, *Thermal Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, England, 2011).
[73] M. Laine, O. Philipsen, P. Romatschke, and M. Tassler, J. High Energy Phys. 03 (2007) 054.
[74] T. Bornath, D. Kremp, and M. Schlanges, Phys. Rev. A **61**, 032309 (2000).
[75] A. Beraudo, J. P. Blaizot, and C. Ratti, Nucl. Phys. **A806**, 312 (2008).
[76] M. J. Strassler and M. E. Peskin, Phys. Rev. D **43**, 1500 (1991).
[77] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, Phys. Rev. D **85**, 123512 (2012).
[78] T. Binder, T. Bringmann, M. Gustafsson, and A. Hryczuk, Phys. Rev. D **96**, 115010 (2017).
[79] T. Binder, M. Gustafsson, A. Kamada, S. M. R. Sandner, and M. Wiesner, Phys. Rev. D **97**, 123004 (2018).
[80] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, Phys. Rev. D **98**, 023543 (2018).
[81] W. Ebeling, V. E. Fortov, and V. Filinov, *Quantum Statistics of Dense Gases and Nonideal Plasmas* (Springer, New York, 2017).
[82] R. Oncala and K. Petraki, arXiv:1808.04854.

5.5 Boltzmann equation for ionization equilibrium

In the article reprinted in the previous section, it was said that in ionization equilibrium and for vanishing temperature corrections of the potential, the derived number density equation is consistent with the coupled Boltzmann equations. Here, we show it explicitly by starting from the Boltzmann equations and imposing ionization equilibrium.

Long-range forces can lead to the existence of bound-state solutions, which contribute to the chemical reaction network. For a single attractive Yukawa potential this occurs if the mediator mass is smaller compared to the inverse Bohr radius. Like in the hydrogen recombination case, two colliding dark matter particles can emit mediators, loose thereby angular momentum and energy, and fall into a bound state. Direct capture into the ground state is often dominated by one mediator emission [24]. The reverse process is called ionization. Furthermore, energy level transitions couple the number densities of different (excited) bound states among each other. These processes are called excitation and deexcitation [25, 26]. For the case where particle and anti-particle DM can annihilate, the bound states are unstable and can decay. Collecting all processes, one can formally write down the following coupled system of Boltzmann equations (see, e.g., [27]):

$$\begin{aligned}
\dot{n}_\eta + 3Hn_\eta = &- \langle(\sigma v_{\rm rel})_{\rm an}\rangle \left[n_\eta n_\xi - n_\eta^{\rm eq} n_\xi^{\rm eq}\right] \\
&- \sum_i \langle(\sigma v_{\rm rel})_{B_i}\rangle \left[n_\eta n_\xi - \frac{n_{B_i}}{n_{B_i}^{\rm eq}} n_\eta^{\rm eq} n_\xi^{\rm eq}\right] , \qquad (5.11)\\
\dot{n}_{B_i} + 3Hn_{B_i} = &- \Gamma_{B_i} \left[n_{B_i} - n_{B_i}^{\rm eq}\right] \\
&+ \langle(\sigma v_{\rm rel})_{B_i}\rangle \left[n_\eta n_\xi - \frac{n_{B_i}}{n_{B_i}^{\rm eq}} n_\eta^{\rm eq} n_\xi^{\rm eq}\right] \\
&- \sum_j \Gamma_{B_i \to B_j} \left[n_{B_i} - \frac{n_{B_j}}{n_{B_j}^{\rm eq}} n_{B_i}^{\rm eq}\right] . \qquad (5.12)
\end{aligned}$$

As in the article reprinted in the previous section, η and ξ denote particle and anti-particle, respectively. Bound states B are made of a particle and an anti-particle. The first equation describes the evolution of free particles and the second equation describes the i-th bound state, where in this simplified notation the label i represents certain bound-state quantum numbers. The first term on the right hand side of Eq. (5.11) describes the Sommerfeld enhanced annihilation of particle η and anti-particle ξ into, e.g., SM particles X and Y, as well as the reverse process. The Sommerfeld enhanced annihilation cross-section $(\sigma v_{\rm rel})_{\rm an}$ should be understood in this notation as a sum over all partial wave contributions, where the s-wave part would be $(\sigma v_{\rm rel})_{\rm an} \simeq (\sigma v_{\rm rel})_0 S$. The second term in the same equation describes the formation of a bound state i via emission of a mediator, as well as the reverse process. The first to the third term in Eq. (5.12) describes bound-state decay, bound-state formation and deexcitation into a lower state j, respectively, as well as all reverse processes.

When adding Eq. (5.12) to Eq. (5.11) and summing over all possible bound states, the bound-state formation and the level transition terms must cancel exactly. This is because the total η number is conserved in latter processes. The only remaining terms, which are η number violating, are annihilation and bound-state decay. In these terms, however, bound and scattering state number densities occur separately, which means if there is no relation among the number density of free and bound particles, one has to solve the evolution of the full chemical network Eq. (5.12) and Eq. (5.11).

Ionization equilibrium

The coupled system of Boltzmann equations can be simplified by assuming that the annihilation and decay rates are less efficient than bound-state formation and level transitions. In particular, it is assumed that the second term on the right hand side in Eq. (5.11) and Eq. (5.12) vanishes (rates are balanced), which implies the following relation between the number density of free DM particles and the bound states:

$$\frac{n_\eta n_\xi}{n_\eta^{\text{eq}} n_\xi^{\text{eq}}} = \frac{n_{B_i}}{n_{B_i}^{\text{eq}}},\ \forall i. \tag{5.13}$$

This also means that $n_{B_i}/n_{B_i}^{\text{eq}} = n_{B_j}/n_{B_j}^{\text{eq}}$ for $i \neq j$, which is indeed the attractor solution of the third line in Eq. (5.12). The system is said to be in ionization equilibrium, if the number density relations in Eq. (5.13) are fulfilled.

In the coupled Boltzmann equations, it is assumed that DM is dilute. Dilute number densities can be written in terms of their chemical potential as: $n_X = n_X^{\text{eq}} e^{\beta \mu_X}$. Expressing Eq. (5.13) in terms of chemical potentials and assuming symmetric dark matter (here $n_\eta/n_\eta^{\text{eq}} = n_\xi/n_\xi^{\text{eq}}$), one obtains the following relations $2\mu \equiv 2\mu_\eta = \mu_{B_i}, \forall i$. The remaining task is to determine μ as a function of the total η number density, defined as n. The latter is the sum over all the individual contributions:

$$n = n_\eta^{\text{eq}} e^{\beta\mu} + \sum_i n_{B_i}^{\text{eq}} e^{\beta 2\mu}. \tag{5.14}$$

This is a quadratic equation for $e^{\beta\mu}$, which has the solution:

$$\beta\mu = \ln\left[\frac{\alpha(nK(T))n}{n_\eta^{\text{eq}}}\right], \tag{5.15}$$

$$\alpha(x) = \frac{\sqrt{1+4x}-1}{2x}, \tag{5.16}$$

$$K(T) = \sum_i \frac{n_{B_i}^{\text{eq}}}{n_\eta^{\text{eq}} n_\eta^{\text{eq}}}. \tag{5.17}$$

Adding Eq. (5.12) to Eq. (5.11), summing over all possible bound states, imposing ionization equilibrium, and inserting μ results in a differential equation for the total η number density:

$$\dot{n} + 3Hn = -\langle(\sigma v_{\text{rel}})_{\text{an}}\rangle \left[\alpha^2 n^2 - n_\eta^{\text{eq}} n_\xi^{\text{eq}}\right] - \sum_i \Gamma_{B_i} n_{B_i}^{\text{eq}} \left[\left(\frac{\alpha n}{n_\eta^{\text{eq}}}\right)^2 - 1\right]. \tag{5.18}$$

The system is closed, since the degree of ionization α depends only on n and temperature T. Limitation of this formula and properties of α in asymptotic regimes were given in the article reprinted in section 5.4.

5.6 Comparison to coannihilation

For comparison to the previous section, the traditional case of coannihilation [133, 134] of supersymmetric particles in the framework of the Minimal Supersymmetric extension of the SM (MSSM) is considered. For simplicity, we write down first the chemical network for two particles and generalize it to N afterwards. The lightest neutralino (DM) is labelled by 1 and can annihilate into SM particles. Additionally, the neutralino can coannihilate with another slightly heavier supersymmetric particle, denoted by 2, into SM particles. The slightly heavier particle can decay into the lightest neutralino and a SM particle. The latter process is called conversion via decay. And last, there is also conversion via scattering. Collecting the processes, the Boltzmann equations can be written as (see, e.g., [133, 134]):

$$\begin{aligned}\dot{n}_1 + 3Hn_1 = &- \langle(\sigma v_{\mathrm{rel}})_{11}\rangle \left[n_1 n_1 - n_1^{\mathrm{eq}} n_1^{\mathrm{eq}}\right] - \langle(\sigma v_{\mathrm{rel}})_{12}\rangle \left[n_1 n_2 - n_1^{\mathrm{eq}} n_2^{\mathrm{eq}}\right] \\ &- \langle(\sigma v_{\mathrm{rel}})_{1X}\rangle \left[n_1 n_X^{\mathrm{eq}} - \frac{n_2}{n_2^{\mathrm{eq}}} n_1^{\mathrm{eq}} n_X^{\mathrm{eq}}\right] - \Gamma_{2\to 1Z} \left[\frac{n_1}{n_1^{\mathrm{eq}}} n_2^{\mathrm{eq}} - n_2\right], \end{aligned} \tag{5.19}$$

$$\begin{aligned}\dot{n}_2 + 3Hn_2 = &- \langle(\sigma v_{\mathrm{rel}})_{21}\rangle \left[n_2 n_1 - n_2^{\mathrm{eq}} n_1^{\mathrm{eq}}\right] - \langle(\sigma v_{\mathrm{rel}})_{22}\rangle \left[n_2 n_2 - n_2^{\mathrm{eq}} n_2^{\mathrm{eq}}\right] \\ &+ \langle(\sigma v_{\mathrm{rel}})_{1X}\rangle \left[n_1 n_X^{\mathrm{eq}} - \frac{n_2}{n_2^{\mathrm{eq}}} n_1^{\mathrm{eq}} n_X^{\mathrm{eq}}\right] + \Gamma_{2\to 1Z} \left[\frac{n_1}{n_1^{\mathrm{eq}}} n_2^{\mathrm{eq}} - n_2\right]. \end{aligned} \tag{5.20}$$

The first line on the right hand side in Eq. (5.19) and Eq. (5.20) describes the annihilation processes, while the second lines contain the conversion via scattering and via decay, respectively. X and Z label SM particles. Assuming conversion equilibrium (second lines vanish), relates the number densities as

$$\frac{n_i}{n_i^{\mathrm{eq}}} = \frac{n_j}{n_j^{\mathrm{eq}}}, \tag{5.21}$$

or in terms of chemical potential $\mu \equiv \mu_i = \mu_j$, where now N particles with $m_1 \leq m_2 \leq ... \leq m_N$ are considered. The chemical potential as a function of the total number of supersymmetric particles $n = \sum_i n_i$ follows from inversion of

$$n = \sum_i n_i^{\mathrm{eq}} e^{\beta\mu}. \tag{5.22}$$

Compared to the case of ionization equilibrium as in the previous section, this time the equation is linear in $e^{\beta\mu}$. Defining $n^{\mathrm{eq}} = \sum_i n_i^{\mathrm{eq}}$, adding all individual equations, and inserting the chemical potential $\beta\mu = \ln[n/n^{\mathrm{eq}}]$ into

$$\dot{n} + 3Hn = -\sum_{ij} \langle(\sigma v_{\mathrm{rel}})_{ij}\rangle \left[n_i n_j - n_i^{\mathrm{eq}} n_j^{\mathrm{eq}}\right], \tag{5.23}$$

results in a single number density equation for the total number of supersymmetric particles [134]:

$$\dot{n} + 3Hn = -\langle(\sigma v_{\rm rel})_{\rm eff}\rangle \left[n^2 - n_{\rm eq}^2\right] . \tag{5.24}$$

The effective cross-section is defined as

$$\langle(\sigma v_{\rm rel})_{\rm eff}\rangle \equiv \sum_{ij} \langle(\sigma v_{\rm rel})_{ij}\rangle \frac{n_i^{\rm eq} n_j^{\rm eq}}{n_{\rm eq}^2} . \tag{5.25}$$

The traditional number density Eq. (5.24) allows for efficient integration and is fully implemented in, e.g., DarkSUSY [3, 4]. The assumption of balanced conversion holds generically in the MSSM. This is a difference compared to the previous section, since ionization equilibrium will be violated once the decay rate exceeds the radiative transitions. There can be also exceptions in coannihilation scenarios, where the freeze-out is conversion driven [135, 136].

Chapter 6

Summary and Conclusion

On the basis of the results of this thesis, it can be concluded that the thermal production mechanism is still an appealing possibility to generate the relic abundance of dark matter. A new exception was found, a non-equilibrium formulation of the number density equation for long-range forces including finite temperature effects was presented, and new phenomenological aspects of light mediator models were discovered.

In the article reprinted in section 3.2, the standard description of the thermal freeze-out process was revisited. First of all, it was concluded that kinetic equilibrium during chemical decoupling is not maintained in general. It was shown that kinetic decoupling can indeed happen so early that the standard description via the independent Eqs. (3.1) and (3.2) fails. As a main result, two more general methods were developed to deal with the situation where chemical and kinetic decoupling essentially overlap. The first one is based on a direct solution of the Boltzmann equation on phase-space density level and the second one consists of a generalized coupled system of leading momentum moments. The author of this thesis contributed with analytic computations to the latter method. Both methods include relativistic corrections and were presented in a model independent way, i.e., in terms of particle physics input quantities like cross-section and momentum transfer rate. They were applied to the resonantly enhanced region of the Scalar Singlet model and the impact of early kinetic decoupling on the number density evolution was quantified. The relic density was found to deviate from the standard prediction by up to one order of magnitude for parameter points close to the excluded region by the invisible decay of the SM Higgs boson (see, e.g., [137]). Quantitatively, a larger coupling value of the Scalar Singlet to the Higgs was found to be needed from the DM abundance bound point of view, implying that the model in this region is generically more constrained compared to previous expectations using the standard treatment.

On the basis of the results of the model analysis, it is recommended for future work to check the assumption of kinetic equilibrium especially for portal DM models, where resonances generically occur. The authors plan for future work to make the phase-space density code public available, including further examples were DM kinetically decouples early.

In the article reprinted in section 5.4, long-range force enhanced pair-annihilation of heavy particles in the presence of a hot and dense plasma environment was formulated in the framework of non-equilibrium quantum field theory, based on a simplified model. The main result is a novel differential equation for the number density of dark matter, expressed in terms of thermal correlation functions. The more general form allows to describe DM chemical decoupling in a self-consistent way, taking into account the Sommerfeld effect, bound-state contributions if they exist, as well as finite temperature corrections to the effective two-particle potential arising from collective phenomena of the environment like charge screening and Landau damping. Moreover, it includes bound-state formation and level transitions under the assumption of detailed balance, called in this case ionization equilibrium. Consistency checks of the generalized number density equation were presented. The authors recovered under certain limits the effective two-particle in-medium potential given in Eq. (5.5), as originally formulated for heavy quark pair annihilation in a quark gluon plasma. While in the latter case, an Euclidean Wilson loop approach was taken, we have shown independently that the same static potential can be obtained starting from perturbative resummation of non-equilibrium correlators. Furthermore, the result correctly reduces to the Boltzmann Eq. (5.18) in the vacuum limit of the effective potential. And finally, it was shown that in the linear regime close to chemical equilibrium, the equation is consistent with the Langevin approach introduced in section 5.3.

The presented number density equation has limitations. The first one is the assumption of a grand canonical state, which implies ionization equilibrium if bound-state solution exist. The second one is the validity region of the hard thermal loop approximation which enters the effective in-medium potential.

For practical application and to deal with the large chemical network once bound-state solutions exist, it can be recommended to use ionization equilibrium at earlier times of the freeze-out process where the assumption of a grand canonical state is typically justified. This allows to include all bound-state formation and level transition processes in a balanced way and finite temperature corrections on top. The advantage is that the number density equation in this case is a single one and not coupled. This allows to overcome the known issue of numerical stiffness of number density equations at the freeze-out by using well-tested, accurate and efficient integration methods. The implicit trapezoidal algorithm with adaptive step size controller as implemented in DarkSUSY [3, 4] for coannihilation would be particularly suited. At later time, when ionization equilibrium starts to become broken, it is recommended to change to the coupled system of classical on-shell Boltzmann equations with initial conditions obtained from the single equation. Based on the fact that we have not yet obtained a complete theory for the whole thermal history, further theoretical developments are needed to achieve a more complete description for long-range force systems including finite temperature corrections.

In agreement with the findings from other literature results, we first of all conclude that radiative processes and bound-state decay can significantly contribute to a depletion of the relic abundance starting when temperature is comparable to the binding energy of the ground state. In addition to the Sommerfeld enhancement, bound-state effects can further push the allowed DM mass to larger values, which is the most interesting aspect for more complete models. This thesis has contributed to set a basis for investigating the impact of finite temperature corrections on the relic abundance prediction for such systems. The most important observation was that thermal corrections are contained in the spectral function, as well as in the chemical potential. Therefore, we conclude that a self-consistent computation of both quantities is required for studying the overall effect. A precise quantification of thermal corrections to the relic abundance prediction is left for future work.

In the article reprinted in section 4.3, a broader applicable collision term was used to study late kinetic decoupling of light mediator models. The t-averaging of the scattering amplitude was an important aspect to identify new types of mediators, which allow for late kinetic decoupling. It was shown that there are models, which can have a similar damping effect on the matter power spectrum at the dwarf galaxy scale in a similar parameter region. The identified type of mediators are complementary to the results of a comprehensive classification of DM-DR models featuring late kinetic decoupling [138], released soon after our work. The description of the DM perturbations, which collaborators in the reprinted article in section 4.3 have derived and implemented, are also complementary to the deep non-relativistic treatment in the ETHOS framework [139, 140]. The latter framework has a better description of the dark radiation perturbations, whereas in section 4.3 they were implemented to first order without collision term (without back reaction).

Light mediator models might not only be an attractive alternative solution of the "small scale crisis", but can also lead to the existence of very loosely bounded two-particle DM states. For such a case, the annihilation cross-section is strongly velocity-dependent can induce a second epoch where DM efficiently annihilates into dark radiation after kinetic decoupling [141], which we called reannihilation. In section 4.4, the detailed phenomenology and gravitational impact on cosmological observables of reannihilation were explored for the first time. A detailed analysis was presented quantifying when reannihilation can happen, for how long it can last, and by how much the relic abundance is allowed to change before and after SM hydrogen recombination. Surprisingly, it was found that there is a parameter region of the model where the "small scale crisis" and tensions between CMB and astrophysical measurements could be alleviated simultaneously. The analysis which brought us to this conclusion was based on estimates. A recent study confirmed the result of our estimates from a full CMB and power spectrum analysis [142], regarding the potential solution of tensions between CMB and astrophysical measurements.

Albeit this is an interesting observation that such a simple model features many non-trivial phenomena, one critical aspect is that an enormous fine tuning of the Sommerfeld resonance was necessary in order that reannihilation is effective at late times. Furthermore, in the golden parameter region of the model, there exist many bound-state solutions that were all neglected in the computation of reannihilation or in the relic abundance. A future direction would be to clarify how bound-state formation processes or level transitions contribute after kinetic decoupling. This is not only an issue to understand for light mediator models but certainly also a relevant aspect for a refined description of long-range force systems in general.

Bibliography

[1] P. A. R. Ade *et al.* (Planck), "Planck 2015 results. XIII. Cosmological parameters," Astron. Astrophys. **594**, A13 (2016), arXiv:1502.01589 [astro-ph.CO] .

[2] N. Aghanim *et al.* (Planck), "Planck 2018 results. VI. Cosmological parameters," (2018), arXiv:1807.06209 [astro-ph.CO] .

[3] P. Gondolo, J. Edsjö, P. Ullio, L. Bergström, M. Schelke, and E. A. Baltz, "DarkSUSY: Computing supersymmetric dark matter properties numerically," JCAP **0407**, 008 (2004), arXiv:astro-ph/0406204 [astro-ph] .

[4] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio, and L. Bergström, "DarkSUSY 6 : An Advanced Tool to Compute Dark Matter Properties Numerically," JCAP **1807**, 033 (2018), arXiv:1802.03399 [hep-ph] .

[5] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, "MicrOMEGAs: A Program for calculating the relic density in the MSSM," Comput. Phys. Commun. **149**, 103 (2002), arXiv:hep-ph/0112278 [hep-ph] .

[6] G. Belanger, F. Boudjema, A. Pukhov, and A. Semenov, "micrOMEGAs_3: A program for calculating dark matter observables," Comput. Phys. Commun. **185**, 960 (2014), arXiv:1305.0237 [hep-ph] .

[7] T. Bringmann, "Particle Models and the Small-Scale Structure of Dark Matter," New J. Phys. **11**, 105027 (2009), arXiv:0903.0189 [astro-ph.CO] .

[8] R. Huo, M. Kaplinghat, Z. Pan, and H.-B. Yu, "Signatures of Self-Interacting Dark Matter in the Matter Power Spectrum and the CMB," Phys. Lett. **B783**, 76 (2018), arXiv:1709.09717 [hep-ph] .

[9] S. Bose, M. Vogelsberger, J. Zavala, C. Pfrommer, F.-Y. Cyr-Racine, S. Bohr, and T. Bringmann, "ETHOS - an Effective Theory of Structure Formation: detecting dark matter interactions through the Lyman-α forest," (2018), arXiv:1811.10630 [astro-ph.CO] .

[10] R. A. Ibata, G. F. Lewis, and M. J. Irwin, "Uncovering cdm halo substructure with tidal streams," Mon. Not. Roy. Astron. Soc. **332**, 915 (2002), arXiv:astro-ph/0110690 [astro-ph] .

[11] J. H. Yoon, K. V. Johnston, and D. W. Hogg, "Clumpy Streams from Clumpy Halos: Detecting Missing Satellites with Cold Stellar Structures," Astrophys. J. **731**, 58 (2011), arXiv:1012.2884 [astro-ph.GA] .

[12] J. S. Bullock and M. Boylan-Kolchin, "Small-Scale Challenges to the ΛCDM Paradigm," Ann. Rev. Astron. Astrophys. **55**, 343 (2017), arXiv:1707.04256 [astro-ph.CO] .

[13] A. G. Riess *et al.*, "Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant," Astrophys. J. **861**, 126 (2018), arXiv:1804.10655 [astro-ph.CO] .

[14] A. G. Riess *et al.*, "A 2.4% Determination of the Local Value of the Hubble Constant," Astrophys. J. **826**, 56 (2016), arXiv:1604.01424 [astro-ph.CO] .

[15] N. MacCrann, J. Zuntz, S. Bridle, B. Jain, and M. R. Becker, "Cosmic Discordance: Are Planck CMB and CFHTLenS weak lensing measurements out of tune?" Mon. Not. Roy. Astron. Soc. **451**, 2877 (2015), arXiv:1408.4742 [astro-ph.CO] .

[16] H. Hildebrandt *et al.*, "KiDS-450: Cosmological parameter constraints from tomographic weak gravitational lensing," Mon. Not. Roy. Astron. Soc. **465**, 1454 (2017), arXiv:1606.05338 [astro-ph.CO] .

[17] S. Joudaki *et al.*, "KiDS-450 + 2dFLenS: Cosmological parameter constraints from weak gravitational lensing tomography and overlapping redshift-space galaxy clustering," Mon. Not. Roy. Astron. Soc. **474**, 4894 (2018), arXiv:1707.06627 [astro-ph.CO] .

[18] L. G. van den Aarssen, T. Bringmann, and C. Pfrommer, "Is dark matter with long-range interactions a solution to all small-scale problems of ΛCDM cosmology?" Phys. Rev. Lett. **109**, 231301 (2012), arXiv:1205.5809 [astro-ph.CO] .

[19] S. Tulin, H.-B. Yu, and K. M. Zurek, "Resonant Dark Forces and Small Scale Structure," Phys. Rev. Lett. **110**, 111301 (2013), arXiv:1210.0900 [hep-ph] .

[20] S. Tulin, H.-B. Yu, and K. M. Zurek, "Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure," Phys. Rev. **D87**, 115007 (2013), arXiv:1302.3898 [hep-ph] .

[21] J. Hisano, S. Matsumoto, and M. M. Nojiri, "Explosive dark matter annihilation," Phys. Rev. Lett. **92**, 031303 (2004), arXiv:hep-ph/0307216 [hep-ph] .

[22] A. Sommerfeld, "Über die Beugung und Bremsung der Elektronen," Annalen der Physik **403**, 257, https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19314030302 .

[23] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, "Non-perturbative effect on thermal relic abundance of dark matter," Phys. Lett. **B646**, 34 (2007), arXiv:hep-ph/0610249 [hep-ph] .

[24] B. von Harling and K. Petraki, "Bound-state formation for thermal relic dark matter and unitarity," JCAP **1412**, 033 (2014), arXiv:1407.7874 [hep-ph] .

[25] K. Petraki, M. Postma, and M. Wiechers, "Dark-matter bound states from Feynman diagrams," JHEP **06**, 128 (2015), arXiv:1505.00109 [hep-ph] .

[26] K. Petraki, M. Postma, and J. de Vries, "Radiative bound-state-formation cross-sections for dark matter interacting via a Yukawa potential," JHEP **04**, 077 (2017), arXiv:1611.01394 [hep-ph] .

[27] A. Mitridate, M. Redi, J. Smirnov, and A. Strumia, "Cosmological Implications of Dark Matter Bound States," JCAP **1705**, 006 (2017), arXiv:1702.01141 [hep-ph] .

[28] J. Harz and K. Petraki, "Higgs-mediated bound states in dark-matter models," (2019), arXiv:1901.10030 [hep-ph] .

[29] E. W. Kolb and M. S. Turner, "The Early Universe," Front. Phys. **69**, 1 (1990).

[30] B. W. Lee and S. Weinberg, "Cosmological Lower Bound on Heavy Neutrino Masses," Phys. Rev. Lett. **39**, 165 (1977).

[31] P. Gondolo and G. Gelmini, "Cosmic abundances of stable particles: Improved analysis," Nucl. Phys. **B360**, 145 (1991).

[32] Y. Hochberg, E. Kuflik, H. Murayama, T. Volansky, and J. G. Wacker, "Model for Thermal Relic Dark Matter of Strongly Interacting Massive Particles," Phys. Rev. Lett. **115**, 021301 (2015), arXiv:1411.3727 [hep-ph] .

[33] D. J. Fixsen, "The Temperature of the Cosmic Microwave Background," Astrophys. J. **707**, 916 (2009), arXiv:0911.1955 [astro-ph.CO] .

[34] R. J. Scherrer and M. S. Turner, "On the Relic, Cosmic Abundance of Stable Weakly Interacting Massive Particles," Phys. Rev. **D33**, 1585 (1986), [Erratum: Phys. Rev.D34,3263(1986)].

[35] E. Bertschinger, "The Effects of Cold Dark Matter Decoupling and Pair Annihilation on Cosmological Perturbations," Phys. Rev. **D74**, 063509 (2006), arXiv:astro-ph/0607319 [astro-ph] .

[36] T. Bringmann and S. Hofmann, "Thermal decoupling of WIMPs from first principles," JCAP **0704**, 016 (2007), [Erratum: JCAP1603,no.03,E02(2016)], arXiv:hep-ph/0612238 [hep-ph] .

[37] J. Kasahara, *Neutralino dark matter: the mass of the smallest halo and the golden region*, Ph.D. thesis, The University of Utah (2009).

[38] P. Gondolo, J. Hisano, and K. Kadota, "The Effect of quark interactions on dark matter kinetic decoupling and the mass of the smallest dark halos," Phys. Rev. **D86**, 083523 (2012), arXiv:1205.1914 [hep-ph] .

[39] C. Boehm, P. Fayet, and R. Schaeffer, "Constraining dark matter candidates from structure formation," Phys. Lett. **B518**, 8 (2001), arXiv:astro-ph/0012504 [astro-ph] .

[40] S. Hofmann, D. J. Schwarz, and H. Stoecker, "Damping scales of neutralino cold dark matter," Phys. Rev. **D64**, 083507 (2001), arXiv:astro-ph/0104173 [astro-ph] .

[41] A. M. Green, S. Hofmann, and D. J. Schwarz, "The power spectrum of SUSY - CDM on sub-galactic scales," Mon. Not. Roy. Astron. Soc. **353**, L23 (2004), arXiv:astro-ph/0309621 [astro-ph] .

[42] A. M. Green, S. Hofmann, and D. J. Schwarz, "The First wimpy halos," JCAP **0508**, 003 (2005), arXiv:astro-ph/0503387 [astro-ph] .

[43] A. Loeb and M. Zaldarriaga, "The Small-scale power spectrum of cold dark matter," Phys. Rev. **D71**, 103520 (2005), arXiv:astro-ph/0504112 [astro-ph] .

[44] M. Drees, F. Hajkarim, and E. R. Schmitz, "The Effects of QCD Equation of State on the Relic Density of WIMP Dark Matter," JCAP **1506**, 025 (2015), arXiv:1503.03513 [hep-ph] .

[45] J. F. Navarro, C. S. Frenk, and S. D. M. White, "A Universal density profile from hierarchical clustering," Astrophys. J. **490**, 493 (1997), arXiv:astro-ph/9611107 [astro-ph] .

[46] R. A. Flores and J. R. Primack, "Observational and theoretical constraints on singular dark matter halos," Astrophys. J. **427**, L1 (1994), arXiv:astro-ph/9402004 [astro-ph] .

[47] B. Moore, "Evidence against dissipationless dark matter from observations of galaxy haloes," Nature **370**, 629 (1994).

[48] B. Moore, T. R. Quinn, F. Governato, J. Stadel, and G. Lake, "Cold collapse and the core catastrophe," Mon. Not. Roy. Astron. Soc. **310**, 1147 (1999), arXiv:astro-ph/9903164 [astro-ph] .

[49] F. Donato, G. Gentile, P. Salucci, C. F. Martins, M. I. Wilkinson, G. Gilmore, E. K. Grebel, A. Koch, and R. Wyse, "A constant dark matter halo surface density in galaxies," Mon. Not. Roy. Astron. Soc. **397**, 1169 (2009), arXiv:0904.4054 [astro-ph.CO] .

[50] W. J. G. de Blok, "The Core-Cusp Problem," Adv. Astron. **2010**, 789293 (2010), arXiv:0910.3538 [astro-ph.CO] .

[51] K. A. Oman *et al.*, "The unexpected diversity of dwarf galaxy rotation curves," Mon. Not. Roy. Astron. Soc. **452**, 3650 (2015), arXiv:1504.01437 [astro-ph.GA] .

[52] A. A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada, "Where are the missing Galactic satellites?" Astrophys. J. **522**, 82 (1999), arXiv:astro-ph/9901240 [astro-ph] .

[53] B. Moore, S. Ghigna, F. Governato, G. Lake, T. R. Quinn, J. Stadel, and P. Tozzi, "Dark matter substructure within galactic halos," Astrophys. J. **524**, L19 (1999), arXiv:astro-ph/9907411 [astro-ph] .

[54] A. V. Kravtsov, "Dark matter substructure and dwarf galactic satellites," Adv. Astron. **2010**, 281913 (2010), arXiv:0906.3295 [astro-ph.CO] .

[55] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, "Too big to fail? The puzzling darkness of massive Milky Way subhaloes," Mon. Not. Roy. Astron. Soc. **415**, L40 (2011), arXiv:1103.0007 [astro-ph.CO] .

[56] M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat, "The Milky Way's bright satellites as an apparent failure of LCDM," Mon. Not. Roy. Astron. Soc. **422**, 1203 (2012), arXiv:1111.2048 [astro-ph.CO] .

[57] T. Sawala *et al.*, "The APOSTLE simulations: solutions to the Local Group's cosmic puzzles," Mon. Not. Roy. Astron. Soc. **457**, 1931 (2016), arXiv:1511.01098 [astro-ph.GA] .

[58] A. A. Dutton, A. V. Macciò, J. Frings, L. Wang, G. S. Stinson, C. Penzo, and X. Kang, "NIHAO V: too big does not fail – reconciling the conflict between ΛCDM predictions and the circular velocities of nearby field galaxies," Mon. Not. Roy. Astron. Soc. **457**, L74 (2016), arXiv:1512.00453 [astro-ph.GA] .

[59] A. R. Wetzel, P. F. Hopkins, J.-h. Kim, C.-A. Faucher-Giguere, D. Keres, and E. Quataert, "Reconciling dwarf galaxies with ΛCDM cosmology: Simulating a realistic population of satellites around a Milky Way-mass galaxy," Astrophys. J. **827**, L23 (2016), arXiv:1602.05957 [astro-ph.GA] .

[60] A. Fattahi, J. F. Navarro, T. Sawala, C. S. Frenk, L. V. Sales, K. Oman, M. Schaller, and J. Wang, "The cold dark matter content of Galactic dwarf spheroidals: no cores, no failures, no problem," (2016), arXiv:1607.06479 [astro-ph.GA] .

[61] I. M. Santos-Santos, A. Di Cintio, C. B. Brook, A. Macciò, A. Dutton, and R. Domínguez-Tenreiro, "NIHAO - XIV. Reproducing the observed diversity of dwarf galaxy rotation curve shapes in ΛCDM," Mon. Not. Roy. Astron. Soc. **473**, 4392 (2018), arXiv:1706.04202 .

[62] S. Y. Kim, A. H. G. Peter, and J. R. Hargis, "There is No Missing Satellites Problem," (2017), arXiv:1711.06267 [astro-ph.CO] .

[63] J. Schaye *et al.*, "The EAGLE project: Simulating the evolution and assembly of galaxies and their environments," Mon. Not. Roy. Astron. Soc. **446**, 521 (2015), arXiv:1407.7040 [astro-ph.GA] .

[64] V. Springel *et al.*, "First results from the IllustrisTNG simulations: matter and galaxy clustering," Mon. Not. Roy. Astron. Soc. **475**, 676 (2018), arXiv:1707.03397 [astro-ph.GA] .

[65] P. F. Hopkins *et al.*, "FIRE-2 Simulations: Physics versus Numerics in Galaxy Formation," Mon. Not. Roy. Astron. Soc. **480**, 800 (2018), arXiv:1702.06148 [astro-ph.GA] .

[66] A. Fitts *et al.*, "FIRE in the Field: Simulating the Threshold of Galaxy Formation," Mon. Not. Roy. Astron. Soc. **471**, 3547 (2017), arXiv:1611.02281 [astro-ph.GA] .

[67] P. Bode, J. P. Ostriker, and N. Turok, "Halo formation in warm dark matter models," Astrophys. J. **556**, 93 (2001), arXiv:astro-ph/0010389 [astro-ph] .

[68] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, "Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest," Phys. Rev. **D71**, 063534 (2005), arXiv:astro-ph/0501562 [astro-ph] .

[69] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns, J. Wang, D. M. White, A. Boyarsky, and O. Ruchayskiy, "The Haloes of Bright Satellite Galaxies in a Warm Dark Matter Universe," Mon. Not. Roy. Astron. Soc. **420**, 2318 (2012), arXiv:1104.2929 [astro-ph.CO] .

[70] M. R. Lovell, C. S. Frenk, V. R. Eke, A. Jenkins, L. Gao, and T. Theuns, "The properties of warm dark matter haloes," Mon. Not. Roy. Astron. Soc. **439**, 300 (2014), arXiv:1308.1399 [astro-ph.CO] .

[71] W. Hu, R. Barkana, and A. Gruzinov, "Cold and fuzzy dark matter," Phys. Rev. Lett. **85**, 1158 (2000), arXiv:astro-ph/0003365 [astro-ph] .

[72] D. J. E. Marsh and J. Silk, "A Model For Halo Formation With Axion Mixed Dark Matter," Mon. Not. Roy. Astron. Soc. **437**, 2652 (2014), arXiv:1307.1705 [astro-ph.CO] .

[73] A. V. Maccio, S. Paduroiu, D. Anderhalden, A. Schneider, and B. Moore, "Cores in warm dark matter haloes: a Catch 22 problem," Mon. Not. Roy. Astron. Soc. **424**, 1105 (2012), arXiv:1202.1282 [astro-ph.CO] .

[74] B. Schwabe, *Simulating Structure Formation with Ultra-light Bosonic Dark Matter*, Ph.D. thesis, Georg-August-University Göttingen, `http://hdl.handle.net/11858/00-1735-0000-002E-E55C-D` (2019).

[75] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, "Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data," Phys. Rev. **D88**, 043502 (2013), arXiv:1306.2314 [astro-ph.CO] .

[76] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville, and M. Viel, "Lyman-alpha Forests cool Warm Dark Matter," *SDSS-IV Collaboration Meeting, July 20-23, 2015*, JCAP **1608**, 012 (2016), arXiv:1512.01981 [astro-ph.CO] .

[77] V. Iršič *et al.*, “New Constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data,” Phys. Rev. **D96**, 023522 (2017), arXiv:1702.01764 [astro-ph.CO] .

[78] H. Deng, M. P. Hertzberg, M. H. Namjoo, and A. Masoumi, “Can Light Dark Matter Solve the Core-Cusp Problem?” Phys. Rev. **D98**, 023513 (2018), arXiv:1804.05921 [astro-ph.CO] .

[79] D. N. Spergel and P. J. Steinhardt, “Observational evidence for selfinteracting cold dark matter,” Phys. Rev. Lett. **84**, 3760 (2000), arXiv:astro-ph/9909386 [astro-ph] .

[80] S. Tulin and H.-B. Yu, “Dark Matter Self-interactions and Small Scale Structure,” Phys. Rept. **730**, 1 (2018), arXiv:1705.02358 [hep-ph] .

[81] N. Yoshida, V. Springel, S. D. M. White, and G. Tormen, “Weakly self-interacting dark matter and the structure of dark halos,” Astrophys. J. **544**, L87 (2000), arXiv:astro-ph/0006134 [astro-ph] .

[82] R. Dave, D. N. Spergel, P. J. Steinhardt, and B. D. Wandelt, “Halo properties in cosmological simulations of selfinteracting cold dark matter,” Astrophys. J. **547**, 574 (2001), arXiv:astro-ph/0006218 [astro-ph] .

[83] P. Colin, V. Avila-Reese, O. Valenzuela, and C. Firmani, “Structure and subhalo population of halos in a selfinteracting dark matter cosmology,” Astrophys. J. **581**, 777 (2002), arXiv:astro-ph/0205322 [astro-ph] .

[84] M. Rocha, A. H. G. Peter, J. S. Bullock, M. Kaplinghat, S. Garrison-Kimmel, J. Onorbe, and L. A. Moustakas, “Cosmological Simulations with Self-Interacting Dark Matter I: Constant Density Cores and Substructure,” Mon. Not. Roy. Astron. Soc. **430**, 81 (2013), arXiv:1208.3025 [astro-ph.CO] .

[85] J. Zavala, M. Vogelsberger, and M. G. Walker, “Constraining Self-Interacting Dark Matter with the Milky Way’s dwarf spheroidals,” Mon. Not. Roy. Astron. Soc. **431**, L20 (2013), arXiv:1211.6426 [astro-ph.CO] .

[86] M. Vogelsberger, J. Zavala, C. Simpson, and A. Jenkins, “Dwarf galaxies in CDM and SIDM with baryons: observational probes of the nature of dark matter,” Mon. Not. Roy. Astron. Soc. **444**, 3684 (2014), arXiv:1405.5216 [astro-ph.CO] .

[87] M. Kaplinghat, S. Tulin, and H.-B. Yu, “Dark Matter Halos as Particle Colliders: Unified Solution to Small-Scale Structure Puzzles from Dwarfs to Clusters,” Phys. Rev. Lett. **116**, 041302 (2016), arXiv:1508.03339 [astro-ph.CO] .

[88] M. Vogelsberger, J. Zavala, and A. Loeb, “Subhaloes in Self-Interacting Galactic Dark Matter Haloes,” Mon. Not. Roy. Astron. Soc. **423**, 3740 (2012), arXiv:1201.5892 [astro-ph.CO] .

[89] O. D. Elbert, J. S. Bullock, S. Garrison-Kimmel, M. Rocha, J. Oñorbe, and A. H. G. Peter, "Core formation in dwarf haloes with self-interacting dark matter: no fine-tuning necessary," Mon. Not. Roy. Astron. Soc. **453**, 29 (2015), arXiv:1412.1477 [astro-ph.GA] .

[90] M. Valli and H.-B. Yu, "Dark matter self-interactions from the internal dynamics of dwarf spheroidals," Nat. Astron. **2**, 907 (2018), arXiv:1711.03502 [astro-ph.GA] .

[91] A. Kamada, M. Kaplinghat, A. B. Pace, and H.-B. Yu, "How the Self-Interacting Dark Matter Model Explains the Diverse Galactic Rotation Curves," Phys. Rev. Lett. **119**, 111102 (2017), arXiv:1611.02716 [astro-ph.GA] .

[92] T. Ren, A. Kwa, M. Kaplinghat, and H.-B. Yu, "Reconciling the Diversity and Uniformity of Galactic Rotation Curves with Self-Interacting Dark Matter," (2018), arXiv:1808.05695 [astro-ph.GA] .

[93] P. Creasey, O. Sameie, L. V. Sales, H.-B. Yu, M. Vogelsberger, and J. Zavala, "Spreading out and staying sharp - Creating diverse rotation curves via baryonic and self-interaction effects," Mon. Not. Roy. Astron. Soc. **468**, 2283 (2017), arXiv:1612.03903 [astro-ph.GA] .

[94] A. Robertson *et al.*, "The diverse density profiles of galaxy clusters with self-interacting dark matter plus baryons," Mon. Not. Roy. Astron. Soc. **476**, L20 (2018), arXiv:1711.09096 [astro-ph.CO] .

[95] S. W. Randall, M. Markevitch, D. Clowe, A. H. Gonzalez, and M. Bradac, "Constraints on the Self-Interaction Cross-Section of Dark Matter from Numerical Simulations of the Merging Galaxy Cluster 1E 0657-56," Astrophys. J. **679**, 1173 (2008), arXiv:0704.0261 [astro-ph] .

[96] A. Robertson, R. Massey, and V. Eke, "What does the Bullet Cluster tell us about self-interacting dark matter?" Mon. Not. Roy. Astron. Soc. **465**, 569 (2017), arXiv:1605.04307 [astro-ph.CO] .

[97] J. M. Cline, Z. Liu, G. Moore, and W. Xue, "Scattering properties of dark atoms and molecules," Phys. Rev. **D89**, 043514 (2014), arXiv:1311.6468 [hep-ph] .

[98] J. Miralda-Escude, "A test of the collisional dark matter hypothesis from cluster lensing," Astrophys. J. **564**, 60 (2002), arXiv:astro-ph/0002050 [astro-ph] .

[99] A. H. G. Peter, M. Rocha, J. S. Bullock, and M. Kaplinghat, "Cosmological Simulations with Self-Interacting Dark Matter II: Halo Shapes vs. Observations," Mon. Not. Roy. Astron. Soc. **430**, 105 (2013), arXiv:1208.3026 [astro-ph.CO] .

[100] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, "A Theory of Dark Matter," Phys. Rev. **D79**, 015014 (2009), arXiv:0810.0713 [hep-ph] .

[101] R. Iengo, “Sommerfeld enhancement: General results from field theory diagrams,” JHEP **05**, 024 (2009), arXiv:0902.0688 [hep-ph] .

[102] A. D. Sakharov, “Interaction of an Electron and Positron in Pair Production,” Zh. Eksp. Teor. Fiz. **18**, 631 (1948), [Usp. Fiz. Nauk161,no.5,29(1991)].

[103] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, “Strong constraints on self-interacting dark matter with light mediators,” Phys. Rev. Lett. **118**, 141802 (2017), arXiv:1612.00845 [hep-ph] .

[104] M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, “Dark Matter’s secret liaisons: phenomenology of a dark U(1) sector with bound states,” JCAP **1705**, 036 (2017), arXiv:1612.07295 [hep-ph] .

[105] I. Baldes and K. Petraki, “Asymmetric thermal-relic dark matter: Sommerfeld-enhanced freeze-out, annihilation signals and unitarity bounds,” JCAP **1709**, 028 (2017), arXiv:1703.00478 [hep-ph] .

[106] I. Baldes, M. Cirelli, P. Panci, K. Petraki, F. Sala, and M. Taoso, “Asymmetric dark matter: residual annihilations and self-interactions,” SciPost Phys. **4**, 041 (2018), arXiv:1712.07489 [hep-ph] .

[107] A. Lewis, A. Challinor, and A. Lasenby, “Efficient computation of CMB anisotropies in closed FRW models,” Astrophys. J. **538**, 473 (2000), arXiv:astro-ph/9911177 [astro-ph] .

[108] M. L. Bellac, *Thermal Field Theory*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2011).

[109] J. I. Kapusta and C. Gale, *Finite-temperature field theory: Principles and applications*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, 2011).

[110] J. S. Schwinger, “Brownian motion of a quantum oscillator,” J. Math. Phys. **2**, 407 (1961).

[111] L. V. Keldysh, “Diagram technique for nonequilibrium processes,” Zh. Eksp. Teor. Fiz. **47**, 1515 (1964), [Sov. Phys. JETP20,1018(1965)].

[112] M. Laine, O. Philipsen, P. Romatschke, and M. Tassler, “Real-time static potential in hot QCD,” JHEP **03**, 054 (2007), arXiv:hep-ph/0611300 [hep-ph] .

[113] T. Matsui and H. Satz, “J/ψ Suppression by Quark-Gluon Plasma Formation,” Phys. Lett. **B178**, 416 (1986).

[114] S. Nadkarni, “Nonabelian Debye Screening. 1. The Color Averaged Potential,” Phys. Rev. **D33**, 3738 (1986).

[115] A. K. Rebhan, “NonAbelian Debye screening in one loop resummed perturbation theory,” Nucl. Phys. **B430**, 319 (1994), arXiv:hep-ph/9408262 [hep-ph] .

[116] E. Braaten and A. Nieto, “Asymptotic behavior of the correlator for Polyakov loops,” Phys. Rev. Lett. **74**, 3530 (1995), arXiv:hep-ph/9410218 [hep-ph] .

[117] P. B. Arnold and L. G. Yaffe, “The NonAbelian Debye screening length beyond leading order,” Phys. Rev. **D52**, 7208 (1995), arXiv:hep-ph/9508280 [hep-ph] .

[118] M. D. Schwartz, *Quantum Field Theory and the Standard Model* (Cambridge University Press, 2014).

[119] E. Braaten and R. D. Pisarski, “Soft Amplitudes in Hot Gauge Theories: A General Analysis,” Nucl. Phys. **B337**, 569 (1990).

[120] A. Beraudo, J. P. Blaizot, and C. Ratti, “Real and imaginary-time Q anti-Q correlators in a thermal medium,” Nucl. Phys. **A806**, 312 (2008), arXiv:0712.4394 [nucl-th] .

[121] N. Brambilla, J. Ghiglieri, A. Vairo, and P. Petreczky, “Static quark-antiquark pairs at finite temperature,” Phys. Rev. **D78**, 014017 (2008), arXiv:0804.0993 [hep-ph] .

[122] Y. Burnier, M. Laine, and M. Vepsalainen, “Heavy quarkonium in any channel in resummed hot QCD,” JHEP **01**, 043 (2008), arXiv:0711.1743 [hep-ph] .

[123] D. Bodeker and M. Laine, “Heavy quark chemical equilibration rate as a transport coefficient,” JHEP **07**, 130 (2012), arXiv:1205.4987 [hep-ph] .

[124] D. Bodeker and M. Laine, “Sommerfeld effect in heavy quark chemical equilibration,” JHEP **01**, 037 (2013), arXiv:1210.6153 [hep-ph] .

[125] M. J. Strassler and M. E. Peskin, “The Heavy top quark threshold: QCD and the Higgs,” Phys. Rev. **D43**, 1500 (1991).

[126] S. Kim and M. Laine, “Rapid thermal co-annihilation through bound states in QCD,” JHEP **07**, 143 (2016), arXiv:1602.08105 [hep-ph] .

[127] S. Kim and M. Laine, “On thermal corrections to near-threshold annihilation,” JCAP **1701**, 013 (2017), arXiv:1609.00474 [hep-ph] .

[128] S. Biondini and M. Laine, “Re-derived overclosure bound for the inert doublet model,” JHEP **08**, 047 (2017), arXiv:1706.01894 [hep-ph] .

[129] S. Biondini and M. Laine, “Thermal dark matter co-annihilating with a strongly interacting scalar,” JHEP **04**, 072 (2018), arXiv:1801.05821 [hep-ph] .

[130] S. Biondini, “Bound-state effects for dark matter with Higgs-like mediators,” JHEP **06**, 104 (2018), arXiv:1805.00353 [hep-ph] .

[131] S. Biondini and S. Vogl, “Coloured coannihilations: Dark matter phenomenology meets non-relativistic EFTs,” (2018), arXiv:1811.02581 [hep-ph] .

[132] M. Laine and A. Vuorinen, “Basics of Thermal Field Theory,” Lect. Notes Phys. **925**, pp.1 (2016), arXiv:1701.01554 [hep-ph] .

[133] K. Griest and D. Seckel, "Three exceptions in the calculation of relic abundances," Phys. Rev. **D43**, 3191 (1991).

[134] J. Edsjo and P. Gondolo, "Neutralino relic density including coannihilations," Phys. Rev. **D56**, 1879 (1997), arXiv:hep-ph/9704361 [hep-ph] .

[135] R. T. D'Agnolo, D. Pappadopulo, and J. T. Ruderman, "Fourth Exception in the Calculation of Relic Abundances," Phys. Rev. Lett. **119**, 061102 (2017), arXiv:1705.08450 [hep-ph] .

[136] M. Garny, J. Heisig, B. Lülf, and S. Vogl, "Coannihilation without chemical equilibrium," Phys. Rev. **D96**, 103521 (2017), arXiv:1705.09292 [hep-ph] .

[137] J. M. Cline, K. Kainulainen, P. Scott, and C. Weniger, "Update on scalar singlet dark matter," Phys. Rev. **D88**, 055025 (2013), [Erratum: Phys. Rev.D92,no.3,039906(2015)], arXiv:1306.4710 [hep-ph] .

[138] T. Bringmann, H. T. Ihle, J. Kersten, and P. Walia, "Suppressing structure formation at dwarf galaxy scales and below: late kinetic decoupling as a compelling alternative to warm dark matter," Phys. Rev. **D94**, 103529 (2016), arXiv:1603.04884 [hep-ph] .

[139] F.-Y. Cyr-Racine, K. Sigurdson, J. Zavala, T. Bringmann, M. Vogelsberger, and C. Pfrommer, "ETHOS—an effective theory of structure formation: From dark particle physics to the matter distribution of the Universe," Phys. Rev. **D93**, 123527 (2016), arXiv:1512.05344 [astro-ph.CO] .

[140] M. Vogelsberger, J. Zavala, F.-Y. Cyr-Racine, C. Pfrommer, T. Bringmann, and K. Sigurdson, "ETHOS – an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems," Mon. Not. Roy. Astron. Soc. **460**, 1399 (2016), arXiv:1512.05349 [astro-ph.CO] .

[141] L. G. van den Aarssen, T. Bringmann, and Y. C. Goedecke, "Thermal decoupling and the smallest subhalo mass in dark matter models with Sommerfeld-enhanced annihilation rates," Phys. Rev. **D85**, 123512 (2012), arXiv:1202.5456 [hep-ph] .

[142] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg, and P. Walia, "Converting nonrelativistic dark matter to radiation," Phys. Rev. **D98**, 023543 (2018), arXiv:1803.03644 [astro-ph.CO] .

Acknowledgments

First of all, I would like to express particular gratitude to my supervisor Prof. Laura Covi. I appreciated your scientific advice and the reliable guidance over all the years. She provided me various opportunities to develop a network and I also thank her for the support to attend numerous conferences, workshops, schools and research visits. I am grateful for her trust that I could take the responsibility for teaching and advisory of students in her group. Furthermore, I would like to thank Prof. Rehren for his encouragement and Prof. Schumann for always having an open door for helpful conversation. I acknowledge financial support from the Research Trainee Group 1493 and also thank Prof. Dorothea Bahns for the fellowship.

I thank all colleagues from the Institute for a nice working atmosphere, especially Michael, Simone, Fiona and Riccardo. Many thanks to the members of the Göttinger Uni Big Band and Cloud 6 for various concerts.

Thanks to all my Göttinger friends, the members of the Theoretikerstammtisch, Bodo, Burgi, Fabse, Thomas, and former flatmates Markus and Marcel. Last, I am especially grateful to my parents, sister and of course Hanna. You were my most important backing and mental supporters.

Curriculum vitae

Tobias Binder

Institute for Theoretical Physics
Friedrich-Hund-Platz 1, Göttingen, D-37077 Germany
✉ *tobias.binder@theorie.physik.uni-goettingen.de*
March, 2019

Personal information

Born May 16th, 1989, in Heilbronn-Neckargartach (Germany)
Nationality German

Academic education

October 2014 - March 2019 **Ph.D. in Physics**, *Institute for Theoretical Physics*, Georg-August-University Göttingen.
Thesis title:
“Refining the chemical and kinetic decoupling description of thermally produced dark matter”.
Date of thesis defense: March 13th, 2019.
Supervisor: Prof. Dr. Laura Covi.

2012-2014 **Master of Science in Physics**, *Institute for Theoretical Physics*.
Thesis title: “A particle physics view on the dark matter thermal history”.
Supervisor: Dr. Riccardo Catena.

2009-2012 **Bachelor of Science in Physics**, *2. Institute*, Georg-August-University Göttingen.
Thesis title: “The jet-multiplicity distribution in top-quark pair production at the LHC”.
Supervisor: Jun. Prof. Dr. Steffen Schumann.

2008 **Abitur**, *Hölderlin-Gymnasium Lauffen*.

Fellowships

2016, 2017 **Exchange fellowship**, *IPMU*, Tokyo University, Japan.
Grant support from the InvisiblesPlus exchange RISE project (total exchange duration: 4 month).

2014-2017 **Doctoral research fellowship**, *Mathematical Institute*, Göttingen, Germany.
Research Training Group 1493 “Mathematical Structures in Modern Quantum Physics”.

Work and teaching experience

2018 **Co-advisor of Bachelor’s thesis**, *Institute for Theoretical Physics, Göttingen*.
“Galactic Rotation Curves in Self-Interacting Dark Matter Models” by Robin Diedrichs.

2016 **Co-advisor of Bachelor’s thesis**, *Institute for Theoretical Physics, Göttingen*.
“Phenomenology of dark matter with a second freeze-out” by Max Wiesner,
“Dark matter scenarios with velocity dependent interactions” by Stefan Marinus Rodrigues Sandner.

2012 **Summer student**, *CERN, Geneva*.
Search for supersymmetric particles in the ATLAS SUSY team.

2010-2017 **Teaching assistant**, *Faculty of Physics, Göttingen*.
During the Ph.D.: Quantum Mechanics, Advanced Quantum Mechanics, Quantum Field Theory.
As an undergraduate: Electrodynamics, Scientific computing for physicists.

2008-2009 **Military service**, *Berlin*, Germany.
Musician in the Air Force Band “Luftwaffenmusikkorps 4”.

Workshops and Ph.D. schools

2018	Workshop “Beyond the beyond the standard model”, Ikaho, Japan.
2017	DESY Theory Workshop “Fundamental physics in the cosmos: The early, the large and the dark Universe”, DESY Hamburg, Germany.
2017	SIDM Workshop, Copenhagen, Denmark.
2017	Open Questions in Particle Physics and Cosmology, Göttingen, Germany.
2016	Foundations and Constructive Aspects of QFT, Göttingen, Germany.
2016	12th International Dark Side of the Universe Workshop, Bergen University, Norway.
2015	DESY Theory Workshop “Physics at the LHC and beyond”, DESY Hamburg, Germany.
2015	47th Ph.D. Autumn School in high-energy physics, Maria Laach Abbey, Germany.
2015	Invisibles15 School, Madrid, Spain.
2015	10th Kosmologietag Workshop, Bielefeld University, Germany.
2014	7th Winter School on Theoretical Physics, CP3-Origins in Odense, Denmark.