Dark Matter Bound-State Formation at Higher Order: a Non-Equilibrium Quantum Field Theory Approach

T. Binder

Kavli IPMU (WPI), UTIAS, University of Tokyo,
Kashiwa, Chiba, Japan

B. Blobel

Mathematisches Institut, Göttingen

J. Harz

Physik Department T70, Technische Universität München, Garching

K. Mukaida

Deutsches Elektronen-Synchrotron DESY, Hamburg

Dark Matter bound-state formation at higher order: a non-equilibrium quantum field theory approach

Tobias Binder[a] **Burkhard Blobel**[b] **Julia Harz**[c] **Kyohei Mukaida**[d]

[a]*Kavli IPMU (WPI),*
 UTIAS, The University of Tokyo, Kashiwa, Chiba 277-8583, Japan

[b]*Mathematisches Institut,*
 Bunsenstr. 3-5, 37073 Göttingen, Germany

[c]*Physik Department T70, James-Franck-Straße, Technische Universität München, 85748 Garching, Germany*

[d]*Deutsches Elektronen-Synchrotron (DESY), Notkestraße 85, Hamburg, D-22607 Germany*

 E-mail: tobias.binder@ipmu.jp,
 Burkhard.Blobel@mathematik.uni-goettingen.de, julia.harz@tum.de,
 kyohei.mukaida@desy.de

ABSTRACT: The formation of meta-stable dark matter bound states in coannihilating scenarios could efficiently occur through the scattering with a variety of Standard Model bath particles, where light bosons during the electroweak cross over or even massless photons and gluons are exchanged in the t-channel. The amplitudes for those higher-order processes, however, are divergent in the collinear direction of the in- and out-going bath particles if the mediator is massless. To address the issue of collinear divergences, we derive the bound-state formation collision term in the framework of non-equilibrium quantum field theory. The main result is an expression for a more general cross section, which allows to compute higher-order bound-state formation processes inside the primordial plasma background in a comprehensive manner. Based on this result, we show that next-to-leading order contributions, including the bath-particle scattering, are i) collinear finite and ii) generically dominate over the on-shell emission for temperatures larger than the absolute value of the binding energy. Based on a simplified model, we demonstrate that the impact of these new effects on the thermal relic abundance is significant enough to make it worthwhile to study more realistic coannihilation scenarios.

Contents

1   Introduction

One of the leading dark matter (DM) candidates are Weakly Interacting Massive Particles (WIMPs) [1–3], which can account for all the present DM energy fraction [4] through their thermal production mechanism in the early Universe. While current upper bounds on the coupling strength to Standard Model (SM) particles put a variety of thermal dark matter candidates around the electroweak scale under tension, the TeV mass region and above remains less constrained and an attractive possibility. In such a high mass region, however, even the heaviest gauge bosons of the SM start to act as a long-range force between annihilating WIMP pairs, leading to a variety of quantum mechanical phenomena. These are important to include for predicting i) the thermal relic abundance precisely and ii) the flux of final state SM particles produced from DM annihilation in, e.g., the galactic center.

One famous example in the Minimal Supersymmetric extension of the SM (MSSM) is the traditional case of wino-like neutralino, for which the important role of quantum mechanical effects was pointed out in seminal works [5–8]. Already at the TeV mass region, the ladder exchange of the most massive SM gauge bosons enhances the probability that a slowly moving wino pair annihilates. This quantum mechanical effect is called *Sommerfeld enhancement* (**SE**) [9] or *Sakharov enhancement* [10] and lowers the predicted abundance by about 50 % compared to a tree-level computation in this case [8]. In turn, the enhancement in the cross section allows for a 30% larger wino mass to compensate for the effect. The flux of SM particles from the present neutralino annihilation in, e.g., the galactic center is sensitive in particular to the predicted mass, since small variations can lead to Sommerfeld resonances [6]. This example shows that it is required to predict the DM mass precisely once quantum mechanical effects can drastically change the observational outcome, in order to constrain the model or to estimate the required exposure time for fully testing the thermal case. Various follow-up works extended the studies of the Sommerfeld effect in the MSSM to more general cases, see, e.g., refs. [11–23] and [24–30] for formal aspects. For recently refined calculations of the present day SM flux in wino and higgsino models, see refs. [31–33].

Additional quantum mechanical effects caused by attractive long-range interactions are the existence of meta-stable bound states in the two-particle spectrum of WIMPs. Their formation and subsequent decay into SM particles can be seen as an additional channel depleting the relic density [34] and therefore allows for even larger dark matter masses. On the one hand, bound-state effects turn out to be negligibly small [35] in the traditional wino case within the current description of bound-state formation (**BSF**) via the emission of an on-shell mediator. On the other hand, large corrections to the predicted mass were found in other electroweak coannihilation scenarios, e.g., the quintuplet case in the context of minimal dark matter [35]. Similar strong effects were identified for coannihilation scenarios with colored charged particles [36–39]. In addition to the electroweak gauge boson, photon, and gluon induced bound states, also the Higgs boson [40, 41] can attractively contribute to confine DM into a meta-stable bound state. Self-interacting DM [42] with new light mediators [43–54], motivated from a bottom-up approach to alleviate the diversity problem in galactic rotation curves [55, 56] or in certain cases even simultaneously ameliorate the Hubble tension [57, 58], are further examples where long-range interactions can affect the dark matter relic abundance.

So far, the formation of dark matter bound states via the emission of an on-shell mediator was considered as the dominant process. However, it was pointed out recently that this does not necessarily have to be the dominant BSF channel during the thermal dark matter freeze-out [59]. The conversion between scattering and bound states in the radiation dominated epoch can also efficiently take place via bath-particle scattering, where a mediator is exchanged virtually between a DM two-body pair and a relativistic primordial plasma particle.

While in ref. [59], a mediator with a mass larger than the binding energy was investigated, we consider massless or lighter mediators in the present work. The masses of SM force-carries is temperature dependent during the electroweak cross over, which motivates to investigate these cases more carefully. The massless case is also relevant in coannihila-
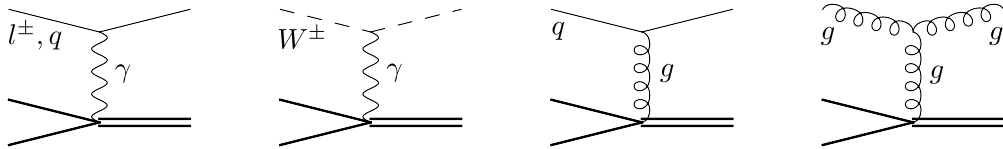
Figure 1. Graphic shows examples of bound-state formation processes via bath-particle scattering. The amplitudes for these processes are divergent in the forward scattering direction of the bath particles. Parallel black solid lines represent the bound state, while open lines correspond to the initial two-particle scattering state.

tion scenarios with electroweakly charged or colored particles. Examples with photons or gluons are illustrated diagrammatically in fig. 1, where bound states are formed via SM bath-particle scattering. Since they all come with a large multiplicity, one may ask an interesting question by how much these additional processes contribute to the previously studied case of BSF via the emission of an on-shell mediator only.

The problem is that the bound-state formation amplitudes diverge for massless mediators in the forward scattering direction of the in- and out-going bath particles in fig. 1 (see also ref. [59]). Since these processes are temperature dependent, the Kinoshita–Lee–Nauenberg (KLN) theorem [60, 61] is not directly applicable. Therefore this problem can not be addressed through the computation of higher-order amplitudes in the collision term of the Boltzmann equation in zero-temperature quantum field theory. Throughout this paper, we refer to this approach as the "conventional Boltzmann formalism". Thus, the question by how much these processes could additionally deplete the dark matter relic abundance can not be answered given the conventional methods. Naively, one might insert a Debye screening mass as a regulator. However, even this procedure is not justified for all temperatures. The energy flowing inside the mediator propagator from the inelastic BSF process is at least as large as the binding energy. Thus, hard thermal loop (HTL) effective field theory [62] would break down for temperatures about the binding energy or below.

The main purpose of this work is to develop a more general method of how to calculate dark matter bound-state formation processes inside the Early Universe plasma. One main focus is to address the problem of collinear divergences once massless mediators are involved and to refine the description for light mediators during the electroweak cross over. We manage to derive the BSF collision term in a more general framework of non-equilibrium quantum field theory, explicitly demonstrated for a vector mediator model in section 2. A more general BSF cross section is defined from the collision term, expressed in terms of thermal correlation functions. Based on this novel result, we demonstrate in section 3 that our BSF cross section correctly contains the on-shell mediator emission at the perturbation order of free correlation functions. In section 4, a full next-to-leading order computation for the direct capture into the ground state is presented, cancellation of forward scattering divergences for bath-particle scattering and the off-shell decay is proven in a model independent way, as well as the renormalization of the ultraviolet vacuum divergence are discussed. The impact on the thermal relic density from all these new next-to-leading order effects is shown in section 5 and our findings discussed in section 6. We concluded in section 7.

2 Generalized bound-state formation cross section

In order to study the infrared divergence structure of higher order bound-state formation processes within the primordial plasma environment for the case of a massless mediator, we consider for simplicity a QED-like model:

$$\mathcal{L} \supset -g\bar{\chi}\gamma^\mu \chi A_\mu + \mathcal{L}_{\text{env}}, \tag{2.1}$$

where χ is a heavy Dirac fermion, (e.g. coannihilating chargino partners) and the Lagrange density of the plasma environment contains all SM fields, e.g. the interaction with light fermions $\mathcal{L}_{\text{env.}} \supset -g\bar{e}\gamma^\mu e A_\mu$. The final form of the derived cross section in this section will be independent of the underlying particle content in the environment and therefore $\mathcal{L}_{\text{env}}$ can be chosen later. Temperatures around the typical chemical decoupling and below are considered, where the χ fields can be assumed to be non-relativistic to a good approximation. We commence from the potential non-relativistic (pNR) effective theory [63], which allows for naturally writing the two-body bound and scattering states in terms of wave function fields. In this framework, the transition between different two-body states are described by an "electric" dipole operator $\mathbf{r} \cdot \mathbf{E}(\mathbf{x}, t)$, where $\mathbf{E}(\mathbf{x}, t) = -\boldsymbol{\nabla} A^0(x) - \partial_t \mathbf{A}(x)$ in terms of the gauge field. In the Hamiltonian formalism, those dipole (dip) interactions are given by (see, e.g., ref. [63]):[1]

$$H_{\text{dip}}(t) = -\sum_{\text{Spin}} \int \mathrm{d}^3x \mathrm{d}^3r \, O_{sr}^\dagger(\mathbf{x}, \mathbf{r}, t) \left[g\, \mathbf{r} \cdot \mathbf{E}(\mathbf{x}, t) \right] O_{sr}(\mathbf{x}, \mathbf{r}, t), \text{ where} \tag{2.2}$$

$$O_{sr}(\mathbf{x}, \mathbf{r}, t) = $$
$$\int \frac{\mathrm{d}^3K}{(2\pi)^3} \left\{ \sum_{\mathcal{B}} e^{-i(\mathcal{E}_{\mathcal{B}}t - \mathbf{K}\cdot\mathbf{x})} \psi_{\mathcal{B}}(\mathbf{r}) \hat{c}_{\mathcal{B},\mathbf{K}}^{sr} + \int \frac{\mathrm{d}^3k}{(2\pi)^3} e^{-i(\mathcal{E}_{\mathbf{k}}t - \mathbf{K}\cdot\mathbf{x})} \psi_{\mathbf{k}}(\mathbf{r}) \hat{a}_{\mathbf{K}/2+\mathbf{k}}^s \hat{b}_{\mathbf{K}/2-\mathbf{k}}^r \right\}. \tag{2.3}$$

The dipole interaction Hamiltonian contains all possible types of two-body conversion processes, i.e., scattering-scattering, bound-bound, and scattering-bound state transitions. The general two-body field operator $O_{sr}(\mathbf{x}, \mathbf{r}, t)$ consists of the whole energy spectrum of χ-field pairs, whose individual components are converted into each other through dipole interactions. Spin configurations are conserved in the transitions, represented by s and r for particle and anti-particle, respectively, and summed over. The components with negative energy eigenvalues $\mathcal{E}_{\mathcal{B}}$ (binding energy) are created by fundamental bound-state operators $\hat{c}^\dagger$ with quantum number $\mathcal{B} = \{nlm\}$. The positive energy spectrum with kinetic energy $\mathcal{E}_{\mathbf{k}} = \mathbf{k}^2/(2\mu) = \mu v_{\text{rel}}^2/2$ and reduced mass μ is created by the particle operator $\hat{a}^\dagger$ and anti-particle operator $\hat{b}^\dagger$. The expansion coefficients involve the bound state $\psi_{\mathcal{B}}(\mathbf{r})$ and scattering state $\psi_{\mathbf{k}}(\mathbf{r})$ wave functions, which are the solution of the time-independent Schrödinger equation with a static Coulomb potential. The wave function dependence on the spin is neglected, and $\mathbf{r}$, $\mathbf{x}$ denote relative and center-of-mass coordinates, respectively.

---

[1]At this stage we would already like to remark that scattering and bound states in Eq. (2.3) are assumed to be separable, which is however not valid once the bound-state energy spectrum starts to overlap with the continuum in the high temperature regime for large thermal widths. This phenomena depends on what is contained in $\mathcal{L}_{\text{env}}$, see discussion in Sec. 6 and Fig. 10 for more details about the limitation of our formalism.

In the following, the time evolution of the particle number density is derived in the density matrix formalism, relating the collision term to the dipole Hamiltonian. For the sake of completeness, we first clarify the underlying assumptions. We take an unperturbed Hamiltonian so that the number densities for the scattering and bound states are conserved and compute the BSF and dissociation rates perturbatively. Let $\hat{H}$, $\hat{H}_0$, and $\hat{H}_{\rm int}$ being the full, free, and interaction Hamiltonian, namely $\hat{H} = \hat{H}_0 + \hat{H}_{\rm int}$. The unperturbed Hamiltonian is given by the summation of that for the scattering state, the bound state, and the thermal environment, $\hat{H}_0 = \hat{H}_{\mathcal{S}} + \hat{H}_{\mathcal{B}} + \hat{H}_{\rm env}$. We take the initial density matrix which is factorized into a tensor product of the scattering state, the bound state, and the thermal environment: $\hat{\rho}(t = 0) = \hat{\rho}_{\mathcal{S}} \otimes \hat{\rho}_{\mathcal{B}} \otimes \hat{\rho}_{\rm env}$. By definition, it commutes with the unperturbed Hamiltonian: $[\hat{\rho}(t = 0), \hat{H}_0] = 0$.

Our starting point is the von Neumann equation in the interaction picture: $i\dot{\hat{\rho}}_I(t) = [\hat{H}_I(t), \hat{\rho}_I(t)]$, where the density matrix and interaction Hamiltonian in the interaction picture are given by $\hat{\rho}_I = e^{i\hat{H}_0 t}\hat{\rho}e^{-i\hat{H}_0 t}$ and $\hat{H}_I = e^{i\hat{H}_0 t}\hat{H}_{\rm int}e^{-i\hat{H}_0 t}$ respectively. The time evolution of the particle number density is obtained from the expectation value of $\hat{n}_{\mathbf{k}_\chi} \equiv \sum_s \hat{a}_{\mathbf{k}_\chi}^{s\dagger}\hat{a}_{\mathbf{k}_\chi}^s$ as follows:

$$
\begin{aligned}
\dot{n}_\chi &= \frac{1}{{\rm vol}(\mathbb{R}^3)} \int \frac{{\rm d}^3 k_\chi}{(2\pi)^3} \, {\rm Tr}\left[\dot{\hat{\rho}}_I \hat{n}_{\mathbf{k}_\chi}\right] \\
&\simeq -\frac{1}{{\rm vol}(\mathbb{R}^3)} \int \frac{{\rm d}^3 k_\chi}{(2\pi)^3} \lim_{t\to\infty} \int_0^t {\rm d}t' \, {\rm Tr}\left\{\left[\left[\hat{n}_{\mathbf{k}_\chi}, \hat{H}_I(t)\right], \hat{H}_I(t')\right]\hat{\rho}(t = 0)\right\}.
\end{aligned} \qquad (2.4)
$$

In the second line, we perform the time dependent perturbation with respect H_I and take the leading order.[2] We also take $t \to \infty$ by assuming that a typical time scale of our interest is much slow compared to that in the thermal plasma. The time evolution of the anti-particle and bound-state number density can be expressed in the same way. Notice that the expectation value in the right-hand side of Eq. (2.4) should be taken by the initial factorized density matrix.

To derive the change of the particle number density under dipole transitions, we have to evaluate the double commutator in Eq. (2.4) for the dipole interaction Hamiltonian in Eq. (2.2). While the dipole interaction Hamiltonian contains all possible types of two-body conversion processes, only the conversion of a scattering state into a bound state and the reverse process can change the particle number density n_χ. Thus, it is sufficient to consider only the mixed contributions containing the product $c^\dagger ab$ and its hermitian conjugate part in Eq. (2.2). These two contribution lead to bound-state formation and the reverse process called dissociation. The computational details of the double commutator are discussed in Appendix A.

Here we highlight the most important part in the computation in Appendix A. Since the dipole Hamiltonian contains the electric field, the computation of the double commutator in Eq. (2.4) involves photon two-point functions. As already emphasized, this commutator should be evaluated by means of the factorized initial density matrix, i.e., $\hat{\rho}(t = 0) =$

---

[2] Here only the leading order term in the interaction Hamiltonian $\hat{H}_I$ is kept. If one is interested in processes involving photons more than one, such as double photon emission, one has to consider high order terms in $\hat{H}_I$.

$\hat{\rho}_\mathcal{S} \otimes \hat{\rho}_\mathcal{B} \otimes \hat{\rho}_{\text{env}}$ with $\hat{\rho}_{\text{env}} \propto e^{-\hat{H}_{\text{env}}/T}$. Therefore the photon two-point functions appearing in the computation are nothing but the thermal propagators: $D_{\mu\nu}^{-+}(x-y) \equiv \langle A_\mu(x) A_\nu(y) \rangle$ for BSF and $D_{\mu\nu}^{+-}(x-y) \equiv \langle A_\nu(y) A_\mu(x) \rangle$ for dissociation. As well known, thermal propagators fulfill the Kubo-Martin-Schwinger relation [64, 65] (see also chapter "Real-time formalism prerequisites" in [66]):

$$D_{\mu\nu}^{-+}(P) = [1 + f_\gamma^{\text{eq}}(P^0)] D_{\mu\nu}^\rho(P), \quad D_{\mu\nu}^{+-}(P) = f_\gamma^{\text{eq}}(P^0) D_{\mu\nu}^\rho(P), \tag{2.5}$$

with f_γ^{eq} being the equilibrium phase-space distribution obeying Bose-Einstein statistics. The KMS condition relates both two-point functions to the spectral function, defined in coordinate space by $D_{\mu\nu}^\rho(x-y) \equiv \langle [A_\mu(x), A_\nu(y)] \rangle$. This photon spectral function encodes all interactions of the primordial plasma environment. We will outline how to estimate the effects from the thermal plasma perturbatively in Eq. (2.12).

With these relations, the change of the particle number density in Eq. (2.4) due to the dipole interactions in Eq. (2.2) is found to be given by:

$$\dot{n}_\chi = -g_\chi g_{\bar{\chi}} \sum_\mathcal{B} \int \frac{\mathrm{d}^3 k_\chi}{(2\pi)^3} \frac{\mathrm{d}^3 k_{\bar{\chi}}}{(2\pi)^3} \frac{\mathrm{d}^3 p}{(2\pi)^3} D_{\mu\nu}^\rho(\Delta E, \mathbf{p}) \sum_{\text{Spin}} \mathcal{T}_{\mathbf{k},\mathcal{B}}^\mu(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\nu\star}(\Delta E, \mathbf{p})$$
$$\times \left\{ f_\chi(\mathbf{k}_\chi) f_{\bar{\chi}}(\mathbf{k}_{\bar{\chi}}) [1 + f_\gamma^{\text{eq}}(\Delta E)] - f_\mathcal{B}(\mathbf{K} - \mathbf{p}) f_\gamma^{\text{eq}}(\Delta E) \right\}, \tag{2.6}$$

with $\Delta E = \mathcal{E}_\mathbf{k} - \mathcal{E}_\mathcal{B}$, $\mathbf{k} = (\mathbf{k}_\chi - \mathbf{k}_{\bar{\chi}})/2$, $\mathbf{K} = \mathbf{k}_\chi + \mathbf{k}_{\bar{\chi}}$, and $\mathbf{p}$ being the three-momentum of the mediator with fixed $P^0 = \Delta E$. The transition matrix elements $\mathcal{T}$ of scattering and bound states are proportional to the dipole overlap integrals:

$$\mathcal{T}_{\mathbf{k},\mathcal{B}}^0(P^0, \mathbf{p}) \equiv g \frac{i \delta^{ss'} \delta^{rr'}}{\sqrt{g_\chi g_{\bar{\chi}}}} \mathbf{p} \int \mathrm{d}^3 r \psi_\mathcal{B}^\star(\mathbf{r}) \mathbf{r} \psi_\mathbf{k}(\mathbf{r}), \tag{2.7}$$

$$\mathcal{T}_{\mathbf{k},\mathcal{B}}^i(P^0, \mathbf{p}) \equiv g \frac{i \delta^{ss'} \delta^{rr'}}{\sqrt{g_\chi g_{\bar{\chi}}}} P^0 \int \mathrm{d}^3 r \psi_\mathcal{B}^\star(\mathbf{r}) \mathbf{r}^i \psi_\mathbf{k}(\mathbf{r}). \tag{2.8}$$

The transition element fulfills current conservation $P_\mu \mathcal{T}_{\mathbf{k},\mathcal{B}}^\mu(P) = 0$, as a consequence the global symmetry of the model in Eq. (2.1).

The χ fields and the bound states are assumed to be in kinetic equilibrium and dilute. Their phase-space densities take the classical Maxwell-Boltzmann statistics and can be written as: $f_X = f_X^{\text{eq}} n_X / n_X^{\text{eq}}(T)$. By noticing the standard non-relativistic thermal average in Eq. (2.6), a bound-state formation cross section can be defined as:

$$\langle \sigma_\mathcal{B}^{\text{bsf}} v_{\text{rel}} \rangle \equiv \frac{g_\chi g_{\bar{\chi}}}{n_\chi^{\text{eq}} n_{\bar{\chi}}^{\text{eq}}} \int \frac{\mathrm{d}^3 k_\chi}{(2\pi)^3} \frac{\mathrm{d}^3 k_{\bar{\chi}}}{(2\pi)^3} (\sigma_\mathcal{B}^{\text{bsf}} v_{\text{rel}}) f_\chi^{\text{eq}}(\mathbf{k}_\chi) f_{\bar{\chi}}^{\text{eq}}(\mathbf{k}_{\bar{\chi}}). \tag{2.9}$$

Replacing in the l.h.s. of Eq. (2.6), the time with cosmic time derivative and adding the usual DM annihilation part in the r.h.s., the particle number density equation can be brought into the standard form:

$$\dot{n}_\chi + 3H n_\chi = -\sum_\mathcal{B} \langle \sigma_\mathcal{B}^{\text{bsf}} v_{\text{rel}} \rangle \left[n_\chi n_{\bar{\chi}} - n_\mathcal{B} \frac{n_\chi^{\text{eq}} n_{\bar{\chi}}^{\text{eq}}}{n_\mathcal{B}^{\text{eq}}} \right] - \langle \sigma^{\text{an}} v_{\text{rel}} \rangle \left[n_\chi n_{\bar{\chi}} - n_\chi^{\text{eq}} n_{\bar{\chi}}^{\text{eq}} \right]. \tag{2.10}$$

Finally, we can identify the *generalized bound-state formation cross section* as
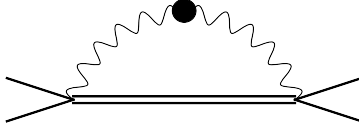
Figure 2. Illustration of the generalized cross section Eq. (2.11) in terms of a self-energy diagram, containing the interacting mediator two-point function. Parallel lines represent the bound state, while open lines correspond to the two-particle scattering state.

$$\sigma_{\mathcal{B}}^{\mathrm{bsf}} v_{\mathrm{rel}} \equiv \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \left[1 + f_\gamma^{\mathrm{eq}}(\Delta E)\right] D_{\mu\nu}^\rho(\Delta E, \mathbf{p}) \sum_{\mathrm{Spin}} \mathcal{T}_{\mathbf{k},\mathcal{B}}^\mu(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\nu\star}(\Delta E, \mathbf{p}). \qquad (2.11)$$

This cross section is one of the central results of this work and can be diagrammatically expressed in terms of the self-energy diagram depicted in fig. 2. The vertices represent the leading order dipole transition contained in $\mathcal{T}_{\mathbf{k},\mathcal{B}}$. The black blob indicates the *interacting spectral correlation function D^ρ* of the mediator, which encodes all interactions with the primordial plasma environment and is the key difference compared to previous literature. It can be seen as a probability density of having a mediator excitation at a certain $\mathbf{p}$ for fixed ΔE. The excitation can be on-shell but also virtually induced through the thermal environment via, e.g., bath-particle scattering. It is convenient to compute the spectral function from the retarded correlator via $D_{\mu\nu}^\rho = 2\,\mathrm{Im}\left[iD_{\mu\nu}^R\right]$, since the latter obeys also in thermal field theory the Dyson-Schwinger equation, given in momentum space by

$$D_{\mu\nu}^R = D_{\mu\nu}^{R,0} + D_{\mu\alpha}^{R,0} \Pi_R^{\alpha\beta} D_{\beta\nu}^{R,0} + \dots . \qquad (2.12)$$

In the next section 3, we demonstrate that BSF via on-shell mediator emission is reproduced from the free retarded propagator. In section 4, the first interaction term containing the retarded self-energy Π_R is analyzed.

Although the cross section was derived for a particular model, we expect the factorization into the interacting mediator spectral function and the transition elements at the Born level to be a rather model-independent feature. In ref. [59], a direct relation between the transition matrix elements $\mathcal{T}_{\mathbf{k},\mathcal{B}}$ and relativistic amplitudes has been given. The advantage of this relation is that results of the previous literature, computing the on-shell emission only, can be used to directly determine $\mathcal{T}_{\mathbf{k},\mathcal{B}}$. For example, ref. [39] provides already expressions for amplitudes for non-abelian mediators at the leading order dipole approximation, ready to be used with Eq. (2.11). In case of a mediating gluon, this would allow to study BSF via gluon scattering (triple vertex), see the fourth diagram in fig. 1, which is expected to occur at the first order of interactions in the gluon spectral function. For Yukawa interactions involving scalar mediators, one can drop the Lorentz indices in Eq. (2.11). The corresponding transition elements, as well as many other cases, can be found in ref. [67, 68].
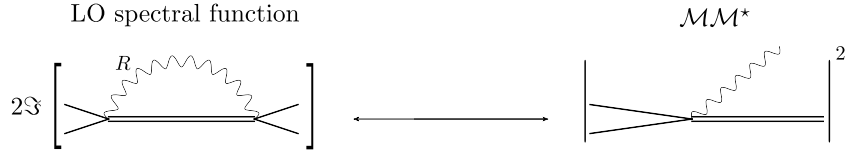
Figure 3. Correspondence between the leading order self-energy contribution and the amplitude squared shown. R indicates the retarded mediator propagator.

3 Recovering on-shell emission at the leading order

Consider the dominant direct capture into the ground state $nlm = 100$ for the QED-like model in Eq. (2.1). The overlap integral in Eq. (2.8) has been analytically computed in, e.g., ref [67]. From this result one can easily get an expression for our transition matrix elements as:

$$\sum_{\text{spins}} \mathcal{T}_{\mathbf{k},100}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},100}^{\nu\star}(\Delta E, \mathbf{p}) = \frac{4\pi h(\zeta)}{\mu^3 \Delta E^2} \left(\frac{\mathbf{p} \cdot \hat{\mathbf{k}}}{\Delta E \, \hat{\mathbf{k}}} \right)^{\mu} \left(\frac{\mathbf{p} \cdot \hat{\mathbf{k}}}{\Delta E \, \hat{\mathbf{k}}} \right)^{\nu}, \tag{3.1}$$

$$h(\zeta) \equiv 2^6 \pi \left(\frac{2\pi\zeta}{1 - e^{-2\pi\zeta}} \right) \frac{\zeta^6}{(1 + \zeta^2)^3} e^{-4\zeta \mathrm{acot}\zeta}, \tag{3.2}$$

with $\zeta \equiv \alpha/v_{\text{rel}}$ and $\Delta E = m_\chi \alpha^2 (1 + \zeta^{-2})/4$ for the ground-state capture. The nice feature now is that Eq. (3.1) can be reused in Eq. (2.11) at any order in the two-point correlation function of the photon field to determine leading and higher-order bound-state formation cross sections inside the primordial plasma environment.

Focusing in this section on the leading order, the free retarded propagator is given by $D_{\mu\nu}^R(P)\big|_{\text{free}} = -ig_{\mu\nu}[P^2 + i\mathrm{sgn}(P^0)\epsilon]^{-1}$. The imaginary part of this expression determines the free photon spectral function, which is nothing but the on-shell contribution:

$$D_{\mu\nu}^{\rho}(P)\big|_{\text{LO}} = -g_{\mu\nu}\mathrm{sgn}(P^0)(2\pi)\delta(P^2). \tag{3.3}$$

The cross section for capture into the ground state via the emission of an on-shell vector mediator is recovered (cf., e.g., [67]) by inserting the free spectral function in Eq. (3.3) together with the transition matrix elements in Eq. (3.1) into the main formula Eq. (2.11):

$$\sigma_{100}^{\text{LO}} v_{\text{rel}} \equiv \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \left[1 + f_\gamma^{\text{eq}}(\Delta E) \right] D_{\mu\nu}^{\rho}(\Delta E, \mathbf{p})\big|_{\text{LO}} \sum_{\text{spin}} \mathcal{T}_{\mathbf{k},100}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},100}^{\nu\star}(\Delta E, \mathbf{p})$$

$$= \frac{4h(\zeta)\Delta E}{3\mu^3} \left[1 + f_\gamma^{\text{eq}}(\Delta E) \right]. \tag{3.4}$$

At leading order of the mediator spectral function, the generalized BSF cross section reduces to the result one would obtain in the Boltzmann formalism based on vacuum amplitudes. In this example, the result is the same as for SM neutral hydrogen recombination, with the emission of a photon. The correspondence between the leading order self-energy diagram and the amplitudes in the vacuum field theory is shown in fig. 3. In the regime where $T \ll \Delta E$, $\zeta \gg 1$, and using $\lim_{\zeta \to \infty} \zeta \mathrm{acot}\zeta = 1$, one obtains that the leading order BSF cross section Eq. (3.4) is by approximately a factor three larger compared to the s-wave Sommerfeld enhanced annihilation cross section into two photons, see ref. [67] for details.
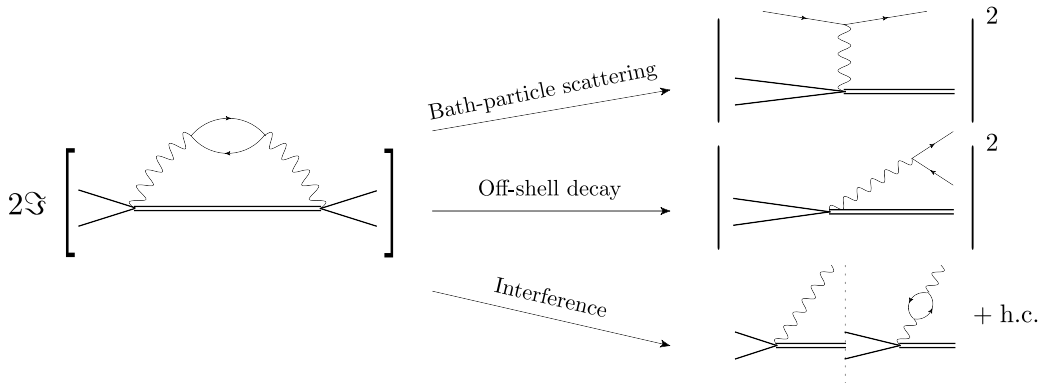
4 Next-to-leading order

Figure 4. Graphic shows the next-to-leading order contributions to the photon spectral function.

We turn now to BSF processes arising at the first order of interactions with the primordial plasma environment. The imaginary part of the first interaction term in the Dyson Eq. (2.12) of the retarded mediator correlation function defines the next-to-leading order contribution to the spectral function and encodes the first interactions with the environment:

$$D_{\mu\nu}^{\rho}(P)\big|_{\text{NLO}} = 2\,\text{Im}\left[iD_{\mu\alpha}^{R}(P)\Pi_{R}^{\alpha\beta}(P)D_{\beta\nu}^{R}(P)\right].\tag{4.1}$$

Hereby, the self-energy $\Pi_{R}^{\alpha\beta}(P)$ depends on the specific model. Under the assumption of a SM photon as mediator, its self-energy is a sum over various contributions with W-bosons, quarks and charged leptons running in the thermal loop. This is equivalent with considering all processes interacting with the plasma at next-to-leading order, for instance BSF via bath-particle scattering and off-shell mediator decay into a bath-particle pair is intrinsically taken into account in the thermal self-energy.

In the following, we concentrate on the interactions with ultra-relativistic Dirac fermions ψ in the primordial plasma such that we identify $\mathcal{L}_{\text{env}} = -g\bar{\psi}\gamma^{\mu}\psi A_{\mu}$, which resembles the interaction with SM quarks and charged leptons. Hence, the retarded self-energy contains a fermion loop as illustrated by the left diagram in fig. 4. We derive the retarded self-energy directly from the Wightman functions for massless fermions as described Appendix B. By inserting the obtained expression for the retarded photon self energy

$$\Pi_{\mu\nu}^{R}(P) = g^2 \int \frac{\mathrm{d}^3k_1}{(2\pi)^3 2|\mathbf{k}_1|} \int \frac{\mathrm{d}^3k_2}{(2\pi)^3 2|\mathbf{k}_2|} \,\text{Tr}\big[\gamma_\mu \slashed{K}_1 \gamma_\nu \slashed{K}_2\big](2\pi)^3 \times$$

$$\left\{ \left[1 - f_\psi^{\text{eq}}(|\mathbf{k}_1|) - f_\psi^{\text{eq}}(|\mathbf{k}_2|)\right]\left[\frac{i\delta^3(\mathbf{p}+\mathbf{k}_1+\mathbf{k}_2)}{P^0+|\mathbf{k}_1|+|\mathbf{k}_2|+i\epsilon} - \frac{i\delta^3(\mathbf{p}-\mathbf{k}_1-\mathbf{k}_2)}{P^0-|\mathbf{k}_1|-|\mathbf{k}_2|+i\epsilon}\right]\right.$$

$$\left.+ \left[f_\psi^{\text{eq}}(|\mathbf{k}_2|) - f_\psi^{\text{eq}}(|\mathbf{k}_1|)\right]\left[\frac{i\delta^3(\mathbf{p}+\mathbf{k}_1-\mathbf{k}_2)}{P^0+|\mathbf{k}_1|-|\mathbf{k}_2|+i\epsilon} - \frac{i\delta^3(\mathbf{p}-\mathbf{k}_1+\mathbf{k}_2)}{P^0-|\mathbf{k}_1|+|\mathbf{k}_2|+i\epsilon}\right]\right\}\tag{4.2}$$

into the spectral function Eq. (4.1), and the latter into the general expression for the BSF cross section Eq. (2.11), we obtain the following expression:

$$\sigma_{\mathcal{B}}^{\mathrm{NLO}} v_{\mathrm{rel}} \equiv \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \left[1 + f_{\gamma}^{\mathrm{eq}}(\Delta E)\right] D_{\mu\nu}^{\rho}(\Delta E, \mathbf{p})\big|_{\mathrm{NLO}} \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\nu\star}(\Delta E, \mathbf{p}) \tag{4.3}$$

$$= \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \left[1 + f_{\gamma}^{\mathrm{eq}}(\Delta E)\right] \sum_{\mathrm{spins}} \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},\mathcal{B}}^{\nu\star}(\Delta E, \mathbf{p}) \times$$

$$g^2 2 \operatorname{Im} \left[\left(\frac{-i}{(\Delta E + i\epsilon)^2 - |\mathbf{p}|^2} \right)^2 \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3 2|\mathbf{k}_1|} \int \frac{\mathrm{d}^3 k_2}{(2\pi)^3 2|\mathbf{k}_2|} \operatorname{Tr}\left[\gamma_\mu \slashed{K}_1 \gamma_\nu \slashed{K}_2\right] \times \right.$$

$$\left\{ \frac{\slashed{\delta}^3(\mathbf{p} + \mathbf{k}_1 - \mathbf{k}_2)}{\Delta E + |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon} \left[f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|) \right] + \right.$$

$$\frac{\slashed{\delta}^3(\mathbf{p} - \mathbf{k}_1 + \mathbf{k}_2)}{\Delta E - |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon} \left[f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) \right] +$$

$$\frac{\slashed{\delta}^3(\mathbf{p} - \mathbf{k}_1 - \mathbf{k}_2)}{\Delta E - |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon} \left[1 - f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|) \right] +$$

$$\left. \left. \frac{\slashed{\delta}^3(\mathbf{p} + \mathbf{k}_1 + \mathbf{k}_2)}{\Delta E + |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon} \left[-1 + f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) + f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|) \right] \right\} \right].$$

When integrating the above equation, the imaginary part of the integral contains a product of possible double and single poles. The former originate from the squared photon propagator D^2 and the latter from the self-energy Π_R in Eq. (4.1).

Taking the imaginary part of single poles corresponds to putting the fermions in the loop as on-shell, while the photons remain virtual. From top to bottom, the four single poles belong to: BSF via particle[3] and anti-particle scattering, as well as BSF via off-shell decay of the photon into a pair of bath particles and the reverse process. The reverse process of off-shell decay is an exception and does not contribute to the imaginary part for kinematical reason (Note that ΔE is always positive). All others are graphically represented by the top and middle diagram on the right-hand side of fig. 4.

The contributions arising from the imaginary part of the double pole correspond to the on-shell emission at next-to-leading order with a temperature dependent fermion loop. We call these contributions in the following interference terms, collectively represented by the bottom right diagram in fig. 4.

The total BSF cross section amounts to a sum over all imaginary parts of the aforementioned single and double poles. The bath-particle scattering and off-shell decay, which arise from the imaginary part of single poles, diverge in the case of a massless photon for zero opening angle of the bath particles ($\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2 \to 1$). We will demonstrate that this divergence is exactly canceled by the collinear divergence appearing in the interference terms. In the *conventional* Boltzmann approach (i.e. considering only bath-particle scattering without

---

[3]To identify the bath-particle scattering case, the identity $\left[1 + f_\gamma^{\mathrm{eq}}(\Delta E)\right] \left[f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_1| + \Delta E)\right] = f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) \left[1 - f_\psi^{\mathrm{eq}}(\Delta E + |\mathbf{k}_1|)\right]$ is helpful. Similarly, all three other possibilities can be identified. We checked that the resulting cross sections from the imaginary part of the single poles only are identical to the cross sections obtained in the conventional Boltzmann formalism, and that they are collinear divergent.

on-shell emission at NLO and finite temperature fermion loop), there is no other term regulating this divergence, while in our approach the cancellation will be shown to happen automatically. The loop contains also UV divergent vacuum parts, i.e. f_ψ independent terms in Eq. (4.3), which can be renormalized with the standard counter terms.

In the following, we show in section 4.1 the main steps for computing BSF at next-to-leading order for the special case of capture into the ground state. In section 4.2, we provide a general proof for the cancellation of the appearing collinear divergences in our formalism. The result for the next-to-leading order BSF cross section is compared to the leading order contribution (on-shell mediator emission) in section 4.3. We set our results into context to the massive mediator case in section 4.4.

4.1 Computational results for direct capture into the ground state

The discussion of the cancellation of the forward scattering divergences and the renormalization of the UV divergences can be separated by noticing that the self-energy can be written as a sum over finite temperature and zero temperature parts. According to this separation, it is useful to define cross sections respectively:

$$\frac{1}{4\pi} \int d\Omega_{\mathbf{k}} (\sigma_{100}^{\text{NLO}} v_{\text{rel}}) = (\sigma_{100}^{\text{NLO}} v_{\text{rel}})_{T=0} + (\sigma_{100}^{\text{NLO}} v_{\text{rel}})_{T\neq0}, \tag{4.4}$$

where the angular averaged parts are defined as

$$(\sigma_{100}^{\text{NLO}} v_{\text{rel}})_{T=0} \equiv \frac{1}{4\pi} \int d\Omega_{\mathbf{k}} \int \frac{d^3p}{(2\pi)^3} \left[1 + f_\gamma^{\text{eq}}(\Delta E)\right] \mathcal{T}_{\mathbf{k},100}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},100}^{\nu\star}(\Delta E, \mathbf{p}) \times \tag{4.5}$$
$$2 \, \text{Im} \left[iD_{\mu\alpha}^R(\Delta E, \mathbf{p}) \Pi_R^{\alpha\beta}(\Delta E, \mathbf{p}) D_{\beta\nu}^R(\Delta E, \mathbf{p}) \right]_{T=0},$$

$$(\sigma_{100}^{\text{NLO}} v_{\text{rel}})_{T\neq0} \equiv \frac{1}{4\pi} \int d\Omega_{\mathbf{k}} \int \frac{d^3p}{(2\pi)^3} \left[1 + f_\gamma^{\text{eq}}(\Delta E)\right] \mathcal{T}_{\mathbf{k},100}^{\mu}(\Delta E, \mathbf{p}) \mathcal{T}_{\mathbf{k},100}^{\nu\star}(\Delta E, \mathbf{p}) \times \tag{4.6}$$
$$2 \, \text{Im} \left[iD_{\mu\alpha}^R(\Delta E, \mathbf{p}) \Pi_R^{\alpha\beta}(\Delta E, \mathbf{p}) D_{\beta\nu}^R(\Delta E, \mathbf{p}) \right]_{T\neq0}.$$

These two cross sections are computed separately in the following. The transition elements $\mathcal{T}$ for the direct capture into the ground state are given in Eq. (3.1). Out of computational reasons it is already reasonable to perform here the angular average over the orientation of the DM relative momentum and it naturally arises in the thermal average in Eq. (2.9). The following identities for the direct capture into the ground state will be helpful to simplify

appearing expressions:

$$\frac{1}{4\pi} \int d\Omega_{\mathbf{k}} \sum_{\text{spins}} \mathcal{T}^{\mu}_{\mathbf{k},100}(\Delta E, \mathbf{p}) \mathcal{T}^{\nu\star}_{\mathbf{k},100}(\Delta E, \mathbf{p})(g^{\mu\nu}P^2 - P^{\mu}P^{\nu})$$

$$= \frac{4h(\zeta)\Delta E}{3\mu^3} \frac{3\pi}{\Delta E^3}(\Delta E^2 - \mathbf{p}^2)(\mathbf{p}^2/3 - \Delta E^2), \tag{4.7}$$

$$\frac{1}{4\pi} \int d\Omega_{\mathbf{k}} \sum_{\text{spins}} \mathcal{T}^{\mu}_{\mathbf{k},100}(\Delta E, \mathbf{p}) \mathcal{T}^{\nu\star}_{\mathbf{k},100}(\Delta E, \mathbf{p}) \operatorname{Tr}\left[\gamma_{\mu} \slashed{K}_1 \gamma_{\nu} \slashed{K}_2\right] =$$

$$\frac{16\pi h(\zeta)}{3\mu^3 \Delta E^2} \bigg\{ 2 \left(|\mathbf{k}_1||\mathbf{k}_2|\mathbf{p}^2 - \Delta E|\mathbf{k}_1|\mathbf{p}\cdot\mathbf{k}_2 - \Delta E|\mathbf{k}_2|\mathbf{p}\cdot\mathbf{k}_1 + \Delta E^2 \mathbf{k}_1\cdot\mathbf{k}_2\right)$$

$$- \left(\mathbf{p}^2 - 3\Delta E^2\right)\left(|\mathbf{k}_1||\mathbf{k}_2| - \mathbf{k}_1\cdot\mathbf{k}_2\right)\bigg\}. \tag{4.8}$$

Vacuum contribution The renormalized retarded self-energy at zero temperature can be obtained from the conventional Euclidean time-ordered self-energy through analytic continuation. In the $\overline{\text{MS}}$-scheme, the retarded photon self-energy for vacuum polarization with approximately massless fermions running in the loop is given by:

$$i\Pi^{R}_{\mu\nu}(P)\big|_{T=0} = -\left(g_{\mu\nu}P^2 - P_{\mu}P_{\nu}\right)\frac{g^2}{12\pi^2}\left[\ln\left(\frac{P^2 + i\operatorname{sign}(P^0)\epsilon}{-\mu_0^2}\right) - \frac{5}{3}\right]. \tag{4.9}$$

The logarithm has a complex contribution for time-like P^2, which originates from the off-shell decay into a massless bath-particle pair. Computing the integral over the imaginary parts in Eq. (4.5), we realize that the NLO cross section factorizes

$$(\sigma^{\text{NLO}}_{100} v_{\text{rel}})_{T=0} = (\sigma^{\text{LO}}_{100} v_{\text{rel}})\left[\frac{\alpha}{\pi} \lim_{\epsilon \searrow 0} R_{\epsilon}\right], \tag{4.10}$$

into the LO contribution as given in Eq. (3.4) and the NLO contribution with

$$R_{\epsilon} = \frac{1}{\pi} \int_0^{\infty} d|\mathbf{p}| \operatorname{Im}\left\{\frac{\mathbf{p}^2(\Delta E^2 - \mathbf{p}^2)\left(\mathbf{p}^2 - 3\Delta E^2\right)}{3\Delta E^3 \left[(\Delta E + i\epsilon)^2 - \mathbf{p}^2\right]^2}\left[\ln\left(\frac{(\Delta E + i\epsilon)^2 - \mathbf{p}^2}{-\mu_0^2}\right) - \frac{5}{3}\right]\right\}. \tag{4.11}$$

Performing the integral and taking $\epsilon \to 0$, we obtain a finite result (see appendix C.1 for the details on the contour integration):

$$\lim_{\epsilon \searrow 0} R_{\epsilon} = \frac{1}{3}\left[\ln\left(\frac{\Delta E^2}{\mu_0^2/4}\right) - \frac{10}{3}\right]. \tag{4.12}$$

These vacuum corrections are shown in fig. 12, shared and discussed in more detail in appendix C.2. From now on, the renormalization scale is fixed to the ground state binding energy $\mu_0 = E_{100} = \mu\alpha^2/2$.

As naively expected at $T = 0$ from the KLN theorem, the collinear divergences in real and virtual corrections cancel each other. Hence, for a finite, physical cross section, both contributions, off-shell decays as well as the virtual correction to the on-shell emission have to be taken into account. This similarly suggests that when considering off-shell decays in

a thermal plasma that we have to consider similarly on-shell emission at next-to-leading order at finite temperatures.

While a consistent treatment of BSF at NLO is non-trivial in the conventional approach (e.g. with respect to double counting of real intermediate states), our advocated formalism will take care of all crucial subtleties such as finite temperature effects, double counting of real intermediate states, and the "automatic" cancellation of all appearing infrared divergences, as we will discuss in the following.

Finite temperature contributions Similar to Eq. (4.10), the finite temperature part of the NLO cross section can be written as

$$(\sigma_{100}^{\mathrm{NLO}} v_{\mathrm{rel}})_{T \neq 0} = (\sigma_{100}^{\mathrm{LO}} v_{\mathrm{rel}}) \left[\frac{\alpha}{\pi} \lim_{\epsilon \searrow 0} \sum_{\sigma_1, \sigma_2} R_\epsilon^{\sigma_1 \sigma_2} \right], \tag{4.13}$$

where the remaining part to compute is the dimensionless function:[4]

$$R_\epsilon^{\sigma_1 \sigma_2} \equiv \frac{1}{\pi} \int_0^\infty \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \int_0^\infty \mathrm{d}|\mathbf{p}| \operatorname{Im} \left[G_\epsilon^{\sigma_1 \sigma_2}(|\mathbf{p}|, \tau, |\mathbf{k}|) \right]. \tag{4.14}$$

In this compact notation, the summation over $\sigma_i \in \{+, -\}$ in Eq. (4.13) takes into account all the finite temperature contributions contained in the self-energy in Eq. (4.3). To arrive here, integration over $\mathbf{k}_2$ in Eq. (4.3) was performed over the momentum conserving delta function and $\mathbf{k}_1$ was relabeled by $\mathbf{k}$. The angular integration variable is $\tau \equiv \hat{\mathbf{p}} \cdot \hat{\mathbf{k}}$. The function $G_\epsilon^{\sigma_1 \sigma_2}$ contains a product of the single and double poles, as given by

$$G_\epsilon^{\sigma_1 \sigma_2}(|\mathbf{p}|, \tau, |\mathbf{k}|) \equiv \frac{F^{\sigma_1 \sigma_2}(|\mathbf{p}|, \tau, |\mathbf{k}|)}{\left([\Delta E + i\epsilon]^2 - \mathbf{p}^2\right)^2 \left(\Delta E + \sigma_1 |\mathbf{k}| - \sigma_2 |\mathbf{p} + \sigma_1 \mathbf{k}| + i\epsilon\right)}, \tag{4.15}$$

where for the numerator we get

$$F^{\sigma_1 \sigma_2}(|\mathbf{p}|, \tau, |\mathbf{k}|) \equiv \left[\sigma_1 f_\psi^{\mathrm{eq}}(|\mathbf{k}|) - \sigma_2 f_\psi^{\mathrm{eq}}(|\mathbf{p} + \sigma_1 \mathbf{k}|)\right] \frac{\mathbf{k}^2 \mathbf{p}^2}{\Delta E^3 |\mathbf{p} + \sigma_1 \mathbf{k}|} \times \tag{4.16}$$

$$\left\{ \left[\mathbf{p}^2 - 3\Delta E^2\right] \left[|\mathbf{p} + \sigma_1 \mathbf{k}| - \sigma_2 (\sigma_1 |\mathbf{k}| + |\mathbf{p}|\tau)\right] - \right.$$

$$\left. 2 \left[\mathbf{p}^2 |\mathbf{p} + \sigma_1 \mathbf{k}| - \Delta E \sigma_2 |\mathbf{p}|(|\mathbf{p}| + \sigma_1 |\mathbf{k}|\tau) - \Delta E |\mathbf{p} + \sigma_1 \mathbf{k}||\mathbf{p}|\tau + \Delta E^2 \sigma_2 (\sigma_1 |\mathbf{k}| + |\mathbf{p}|\tau)\right] \right\}.$$

---

[4]In principle, it does not matter which integral is chosen to perform the integration over the poles. However, we would like to share some insights why we find this particular order especially suited. Here, we have chosen the integral over $\mathbf{k}_2$ in Eq. (4.3) to first perform the simple integration over the momentum conserving delta functions, and then $\mathbf{p}$ to perform integration over the poles. In this order, the collinear divergence occurs at $\tau \equiv \hat{\mathbf{p}} \cdot \hat{\mathbf{k}}_1 \to 1$, which is physically not anymore the zero opening angle between the bath particles but related to that divergence. In a different order where $\mathbf{p}$ integration over the delta functions is performed first, the collinear divergences occur when the opening angle of the bath particles approaches zero, i.e., $\hat{\mathbf{k}}_1 \cdot \hat{\mathbf{k}}_2 \to 1$. This order introduces coordinate singularities in addition, which one can get rid off by transforming into certain elliptical coordinates. The final result in these coordinates would be the same as in the order of integration chosen in this section, but more difficult to physically interpret. Finally, we would like to remark that one is not allowed to choose different integration orders for the single and double poles. This is because only the simultaneous integral over both contributions is in general finite, while the individual terms can diverge. When separating the divergent integrals over single and double poles, we find that one can still show that the collinear divergences cancel at the end, but one misses important boundary terms and the final result would be different.

For the integration over $|\mathbf{p}|$ in Eq. (4.14), we consider the analytic continuation $|\mathbf{p}| \to z$ and perform the integration over the single and double poles of the function in Eq. (4.15). In appendix C.3, we show that due to the imaginary part in Eq. (4.14), $F^{\sigma_1\sigma_2}$ holomorphic on the real line, and integration range from 0 to infinity, only the real and positive poles contribute. All the real and positive poles are listed in table 1, together with their existence criteria. We denote the double pole as z_0 and the possible two single poles as $z_{p/m}$ for each $\sigma_1\sigma_2$ configuration in Eq. (4.15). The table can be used to simply write down the $R^{\sigma_1\sigma_2}$ functions expressed in terms of the residues as:

$$R_0^{++} = \int_0^\infty \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \left[\mathrm{Res}(G_0^{++}, z_0) + \mathrm{Res}(G_0^{++}, z_p) \right], \tag{4.17}$$

$$R_0^{+-} = \int_0^\infty \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \left[\mathrm{Res}(G_0^{+-}, z_0) \right], \tag{4.18}$$

$$R_0^{-+} = \int_0^{\Delta E/2} \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \left[\mathrm{Res}(G_0^{-+}, z_0) + \mathrm{Res}(G_0^{-+}, z_p) \right] \tag{4.19}$$

$$+ \int_{\Delta E/2}^{\Delta E} \mathrm{d}|\mathbf{k}| \left\{ \int_{\sqrt{\frac{2\Delta E|\mathbf{k}|-\Delta E^2}{|\mathbf{k}|^2}}}^1 \mathrm{d}\tau \left[\mathrm{Res}(G_0^{-+}, z_0) + \mathrm{Res}(G_0^{-+}, z_p) - \mathrm{Res}(G_0^{-+}, z_m) \right] \right.$$

$$\left. + \int_{-1}^{\sqrt{\frac{2\Delta E|\mathbf{k}|-\Delta E^2}{|\mathbf{k}|^2}}} \mathrm{d}\tau \left[\mathrm{Res}(G_0^{-+}, z_0) \right] \right\}$$

$$+ \int_{\Delta E}^\infty \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \left[\mathrm{Res}(G_0^{-+}, z_0) \right],$$

$$R_0^{--} = R_0^{++}. \tag{4.20}$$

The different signs in front of the residues originate from the $i\epsilon$ in Eq. (4.15), where the z_m pole was constructed to lie always in the lower complex half plane for finite epsilon and the other ones in the upper half plane, see appendix C.3 for details. The last equality follows from symmetry considerations, which one can directly see from Eq. (4.3) by interchanging the labels $\mathbf{k}_1$ and $\mathbf{k}_2$ for the corresponding term [5].

The list of all finite temperature contributions, from Eq. (4.17) to Eq. (4.20), is subject to further discussion. BSF via bath-particle scattering is contained in the single pole contribution $\mathrm{Res}(G_0^{++}, z_p)$ and the finite temperature part of the off-shell decay of the vector mediator into a bath-particle pair is contained in the z_p residues in R^{-+}. The individual residue would diverge in the limit $\tau \to 1$, reflecting the collinear divergence. However, the double pole contribution with z_0 diverges in the limit as well, but with opposite sign, resulting in the fact that the sum over double and single pole contributions remains finite in collinear direction. A rather general proof for the cancellation of collinear divergences is presented in the next section. Therein, it is also explained why the terms where no single poles exist are finite. The remaining $\mathbf{k}$ and τ integrals are all finite and we show their numerically obtained values in fig. 14, shared in the appendix C.4. Worthwhile to note is that the most dominant contributions are R^{++} and R^{--}, which contain BSF via particle and anti-particle, respectively.

---

[5]Physically, this symmetry states that BSF via particle or anti-particle scattering is the same.

$\sigma_1\sigma_2$	z_0	z_p	z_m
$++$	ΔE	$-\lvert\mathbf{k}\rvert\tau + \sqrt{\mathbf{k}^2\tau^2 + \Delta E^2 + 2\Delta E\lvert\mathbf{k}\rvert}$	$-$
$+-$	ΔE	$-$	$-$
$-+$	ΔE	$\lvert\mathbf{k}\rvert\tau + \sqrt{\mathbf{k}^2\tau^2 + \Delta E^2 - 2\Delta E\lvert\mathbf{k}\rvert}$, $\text{if } \left\{ \Delta E/2 \le \lvert\mathbf{k}\rvert \le \Delta E \right.$ $\left. \wedge\tau \ge \sqrt{\frac{2\Delta E\lvert\mathbf{k}\rvert - \Delta E^2}{\lvert\mathbf{k}\rvert^2}} \right\}$ $\text{or } 0 \le \lvert\mathbf{k}\rvert \le \Delta E/2.$	$\lvert\mathbf{k}\rvert\tau - \sqrt{\mathbf{k}^2\tau^2 + \Delta E^2 - 2\Delta E\lvert\mathbf{k}\rvert}$, $\text{if } \Delta E/2 \le \lvert\mathbf{k}\rvert \le \Delta E$ $\wedge\tau \ge \sqrt{\frac{2\Delta E\lvert\mathbf{k}\rvert - \Delta E^2}{\lvert\mathbf{k}\rvert^2}}.$

Table 1. Summary of the real and positive poles of Eq. (4.15), as well as their existence criteria.

4.2 Proof for the cancellation of collinear divergences

From the mathematical point of view, collinear divergences occur since in the collinear limit $\tau \to 1$ the single pole z_p approaches the double pole z_0 as

$$z_p - z_0 = (1 - \tau)\frac{\sigma_1\lvert\mathbf{k}\rvert\Delta E}{\Delta E + \sigma_1\lvert\mathbf{k}\rvert} + \mathcal{O}((1 - \tau)^2). \tag{4.21}$$

To make that clear, let us take a closer look on the residues of a function with a double pole at z_0 and a single pole at z_p, approaching each other in the collinear limit as, e.g., in Eq. (4.21). Every holomorphic function having such kind of pole structure can be written as

$$G(z) = \frac{H(z)}{(z - z_0)^2(z - z_p)}, \tag{4.22}$$

where H is holomorphic at $z = z_0$ and $z = z_p$. Using this general form, the residues of G are given by:

$$\text{Res}(G, z_0) = -\frac{H(z_0)}{(z_p - z_0)^2} - \frac{H'(z_0)}{z_p - z_0}, \tag{4.23}$$

$$\text{Res}(G, z_p) = \frac{H(z_p)}{(z_p - z_0)^2}$$

$$= \frac{H(z_0)}{(z_p - z_0)^2} + \frac{H'(z_0)}{z_p - z_0} + \frac{1}{2}H''(z_0) + \mathcal{O}(z_p - z_0), \tag{4.24}$$

In the last line we expanded $H(z_p)$ around $z_p = z_0$. Clearly, each individual residue is divergent in the collinear limit, when $z_p \to z_0$. However, the colliner divergent terms from Eq. (4.23) and Eq. (4.24) occur with opposite sign and hence cancel each other in their sum. Let us comment on the remaining term $\frac{1}{2}H''(z_0)$ whose appearance one can intuitive understand. Since the double pole merges with the single pole in the collinear limit, the function $G\vert_{\tau=1} = \frac{H(z)}{(z-z_0)^3}$ has a triple pole. And indeed, its residue gives the result expected

from Eq. (4.23) and Eq. (4.24) as

$$\begin{aligned}
\text{Res}(G|_{\tau=1}, z_0) &= \frac{1}{2} \lim_{z \to z_0} \frac{\mathrm{d}^2}{\mathrm{d}z^2} \left[(z - z_0)^3 \frac{H(z)}{(z - z_0)^3} \right] \\
&= \frac{1}{2} H''(z_0) \\
&= \lim_{\tau \to 1} \left[\text{Res}(G, z_p) + \text{Res}(G, z_0) \right],
\end{aligned} \tag{4.25}$$

without any collinear divergence. This proves the collinear finiteness for *any* H holomorphic at $z = z_0$ and $z = z_p$ and is in particular fulfilled for any $F^{\sigma_1 \sigma_2}$ in Eq. (4.16) for capture into the ground state.

From the discussion above, it follows that R_0^{++} as well as R_0^{--} in Eq. (4.17) and Eq. (4.20) are finite. For $\sigma_1 = +$, $\sigma_2 = -$, the single poles do not exists as $\Delta E + |\mathbf{k}_1| + |\mathbf{k}_2|$ is always nonzero. Hence the singularity at $z = z_p$ is removable and Eq. (4.23) as well as the singular part of Eq. (4.24) vanish. Therefore Eq. (4.18) is collinear finite. It remains to discuss Eq. (4.19), where $\sigma_1 = -$ and $\sigma_2 = +$. As one can see, the collinear divergences are canceled in the first and the second line. In the third line τ is smaller than 1. The last line of Eq. (4.19) is collinear finite since there are no single poles as $\Delta E - |\mathbf{k}_1| - |\mathbf{k}_2|$ is always negative for $|\mathbf{k}_1| > \Delta E$. Analog to the case $\sigma_1 = +$, $\sigma_2 = -$ the singularities vanish.

While the residue of the single pole $\text{Res}(G, z_p)$ and the residue of the double pole $\text{Res}(G, z_0)$ diverge individually in the collinear limit, we have shown in summary that the sum of both terms remains finite. In the Boltzmann formalism only the single pole $\text{Res}(G, z_p)$ would occur, and hence the collision term would be ill-defined. This makes the thermal field theory approach necessary for studying BSF at higher order, at least if massless gauge bosons or light mediators (for how light see section 4.4) are involved.

It remains to discuss the residue at z_m, only present in R_0^{-+}. Similarly, $\text{Res}(G_0, z_p)$ and $\text{Res}(G_0, z_m)$ have poles if $z_p = z_m$. This situation occurs if $|\tau| = \sqrt{\frac{2\Delta E|\mathbf{k}| - \Delta E^2}{|\mathbf{k}|^2}} =: \tau_*$ and therefore it is only present in the second line of Eq. (4.19). Because of the relative signs these singularities do not cancel. However, due to the choice of the order of the integration (see discussion in Sec. 4.1, paragraph **Finite temperature contributions**) the singularities only grow like $(\tau - \tau_*)^{-\frac{1}{2}}$. For that reason the integrating over τ does not cause any problems. Finally, the last exception occurs if $z_0 = z_p = z_m$. This is the case for $\tau = 1$, $|\mathbf{k}| = \Delta E$. In this case the function F^{-+} given in Eq. (4.16) goes to zero and the integral remains finite.

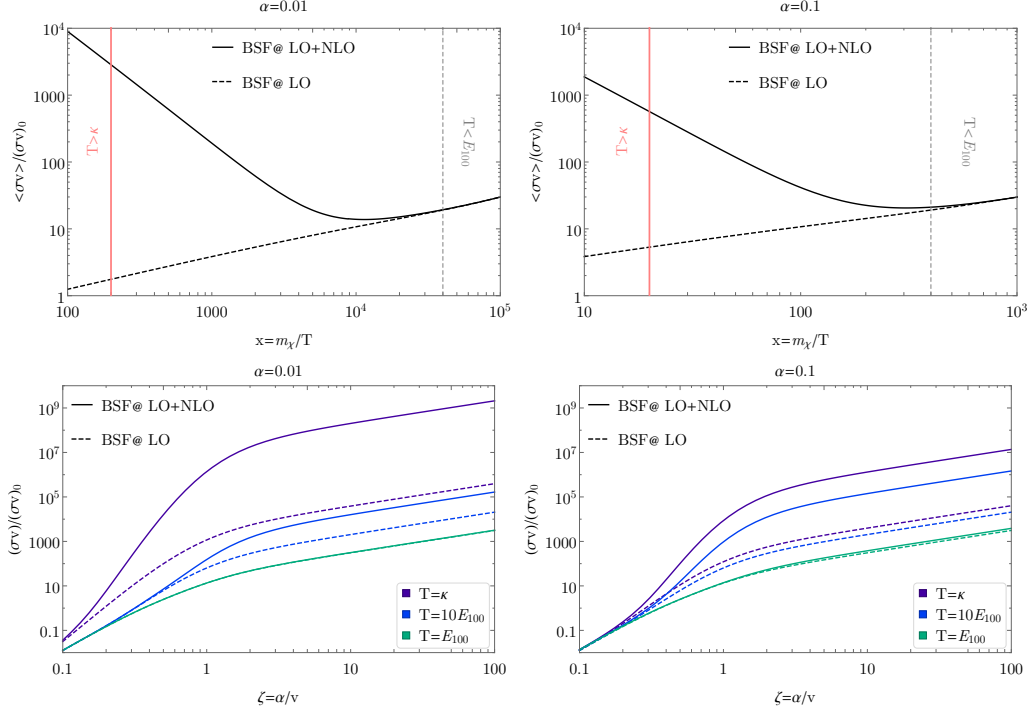## 4.3 Comparison of ground state capture at leading and next-to-leading order

Figure 5. Lower row shows the bound-state formation cross sections. Upper row compares the thermally averaged quantities. Validity of dipole approximation breaks down for $T \gtrsim \kappa = \mu\alpha$. The cross sections are normalized to $(\sigma v)_0 \equiv 2\pi\alpha^2/m_\chi^2$.

The leading order cross section for capture into the ground state ($\sigma_{100}^{\mathrm{LO}} v_{\mathrm{rel}}$) in Eq. (3.4) and our result of the next-to-leading order cross section ($\sigma_{100}^{\mathrm{NLO}} v_{\mathrm{rel}}$) in Eq. (4.4) are compared to each other. The thermally averaged quantities are shown in the top row of fig. 5. Overall one can recognize that *the NLO starts to dominate over the LO cross section for temperatures larger than the absolute value of the ground state binding energy*. While the thermally averaged BSF cross section via the on-shell mediator emission increases for decreasing temperature, the NLO contribution becomes larger for higher temperature.

The reason for this behavior can be understood from the lower panel of fig. 5, where the non-averaged quantities as a function of inverse relative-velocity are shown. In the high temperature regime around the Bohr-momentum κ, the total cross section differs in shape and amplitude compared to the LO. The overall increase in the amplitude towards higher temperatures is mainly caused by the number density of relativistic bath particles, scaling as T^3. Additionally, the NLO cross section has a stronger dependence on the relative-velocity for $v_{\mathrm{rel}} \gtrsim \alpha$, which turns into the milder scaling v_{rel}^{-1} for $v_{\mathrm{rel}} \ll \alpha$. The temperature and velocity enhancement leads to the fact that the overall BSF probability distribution in the thermal average is sharply peaked at the DM relative momentum $\sim \kappa$ and contributes less at the conventional one $\sim (m_\chi T)^{1/2}$. These features are less pronounced in the on-shell emission case, qualitatively explaining the differences of the thermally averaged quantities.

4.4 When is a thermal field theory approach required?

In the Boltzmann formalism, the bound-state formation cross section for bath-particle scattering is always finite for massive mediators. However in the limit mass to zero, the cross section has a logarithmic divergence, which originates from the forward scattering divergence of the bath-particles, see also ref. [59]. In this section, we would like to answer the following question for practical purposes: what is the critical mediator mass, above which the conventional Boltzmann formalism is a sufficient description and below which a more sophisticated thermal field theory analysis as presented in our work is required?
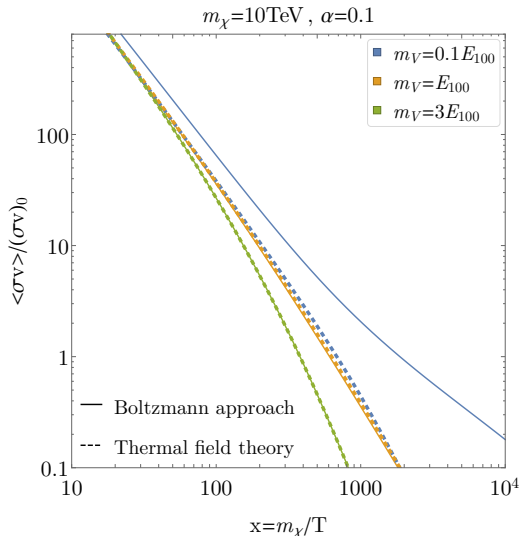

Figure 6. Graphic compares the thermally averaged bound-state formation cross section for bath-particle scattering in the conventional Boltzmann and the thermal field theory approach.

To estimate the critical mediator mass, only the dominant contributions of R^{++} and R^{--} are considered, which contain the BSF via particle and anti-particle scattering, respectively. Additionally, these functions also contain interference terms originating from the double poles of the photon propagator. For a vector mediator with a mass m_V, one can simply replace the photon propagator by a massive one. This changes only the double pole into:

$$G_\epsilon^{++}(|\mathbf{p}|, \tau, |\mathbf{k}|) = \frac{F^{++}(|\mathbf{p}|, \tau, |\mathbf{k}|)}{\left([\Delta E + i\epsilon]^2 - \mathbf{p}^2 - m_V^2\right)^2 (\Delta E + |\mathbf{k}| - |\mathbf{p} + \mathbf{k}| + i\epsilon)}, \qquad (4.26)$$

where the single pole and F^{++} remain unaffected. One can immediately recognize that the double pole at $|\mathbf{p}| = \tilde{z}_0 = (\Delta E^2 - m_V^2)^{1/2}$ only exists for $\Delta E > m_V$. This is always true for mediator masses smaller than the absolute value of the binding energy. In general, for massive mediators one can replace R^{++} by:

$$\lim_{\epsilon \searrow 0} R_\epsilon^{++} = \int_0^\infty \mathrm{d}|\mathbf{k}| \int_{-1}^1 \mathrm{d}\tau \left[\theta(\tilde{z}_0^2)\mathrm{Res}(G_0^{++}, \tilde{z}_0) + \mathrm{Res}(G_0^{++}, z_p)\right]. \qquad (4.27)$$

As a reminder, the term $\mathrm{Res}(G_0^{++}, z_p)$ corresponds to BSF via bath-particle scattering and reproduces ref. [59]. From this equation one can see that the double pole contribution, which causes the difference to the Boltzmann formalism, can be neglected for *mediator masses much larger than the binding energy* due to $\theta(\tilde{z}_0^2)$. This statement is confirmed numerically as shown in fig. 6, where we compute the BSF cross section based on the full Eq. (4.27) and without the double pole contribution. The Boltzmann formalism overestimates the effect of the bath-particle scattering for $m_V \ll E_{100}$. For $m_V \gg E_{100}$, the Boltzmann formalism is reliable.

5 Impact on thermal relic abundance

The results of the previous sections allow us to explore the impact of higher-order BSF processes on the evolution of the thermal relic abundance for the following simplified model with N bath particle species:

$$\mathcal{L} \supset -g\bar{\chi}\gamma^\mu\chi A_\mu - \sum_{i=1}^{N} g\bar{\psi}_i\gamma^\mu\psi_i A_\mu. \tag{5.1}$$

An s-wave χ-pair can annihilate into 2γ and N effectively massless ψ pairs. The total s-wave Sommerfeld enhanced annihilation cross section, averaged over the initial spin and summed over final degrees of freedom, can be written as

$$\sigma^{\mathrm{an}}v_{\mathrm{rel}} = (1+N)\,\frac{\pi\alpha^2}{m_\chi^2}|\psi_{\mathbf{k},l=0}(r=0)|^2, \tag{5.2}$$

where the first two factors account for the tree-level part and $|\psi_{\mathbf{k},l=0}(0)|^2 = 2\pi\zeta(1-e^{-2\pi\zeta})^{-1}$ is the Sommerfeld enhancement factor [9, 10]. For the lowest s-wave χ bound states, we consider the spin singlet (S) decay into 2γ [69, 70], and the spin triplet decay into 3γ [71] and N ψ-pairs [72]. The decay widths can be written as

$$\Gamma_S^{\mathrm{dec}} = 4 \times \frac{\pi\alpha^2}{m_\chi^2}|\psi_{100}(r=0)|^2, \tag{5.3}$$

$$\Gamma_T^{\mathrm{dec}} = \left(c_{3\gamma} + \frac{N}{3}\right)\Gamma_S^{\mathrm{dec}}, \tag{5.4}$$

where $|\psi_{100}(0)|^2 = (\mu\alpha)^3/\pi$, and $c_{3\gamma} = 4(\pi^2-9)\alpha/(9\pi)$ [71]. We compute the BSF cross section in Eq. (2.11) for the capture into the ground state up to NLO in the dark photon spectral function:

$$\langle\sigma_{100}^{\mathrm{bsf}}v_{\mathrm{rel}}\rangle \simeq \langle\sigma_{100}^{\mathrm{LO}}v_{\mathrm{rel}}\rangle + N\langle\sigma_{100}^{\mathrm{NLO}}v_{\mathrm{rel}}\rangle. \tag{5.5}$$

For the LO cross section we can directly use the expression in Eq. (3.4), while for the NLO contribution we take N times our results in Eq. (4.4). The number density evolution of the lowest bound states can be included into the Boltzmann Eq. (2.10) for the scattering states approximately as [34, 36]:[6]

$$\dot{n}_\chi + 3Hn_\chi = -\left[\langle\sigma^{\mathrm{an}}v_{\mathrm{rel}}\rangle + W(T)\right]\left[n_\chi n_{\bar{\chi}} - n_\chi^{\mathrm{eq}}n_{\bar{\chi}}^{\mathrm{eq}}\right], \tag{5.6}$$

$$W(T) \equiv \frac{\langle\sigma_{100}^{\mathrm{bsf}}v_{\mathrm{rel}}\rangle}{4}\left[\frac{\Gamma_S^{\mathrm{dec}}}{\Gamma_S^{\mathrm{dec}}+\Gamma_S^{\mathrm{dis}}} + 3\times\frac{\Gamma_T^{\mathrm{dec}}}{\Gamma_T^{\mathrm{dec}}+\Gamma_T^{\mathrm{dis}}}\right]. \tag{5.7}$$

The effective cross section $W(T)$ comprises all information of the bound states. It consists out of a sum over the individual singlet and triplet BSF cross sections (here spin-independent), weighted by branching ratios containing the individual decay and dissociation rates. The latter corresponds to the inverse process of the bound-state formation and

---

[6]For the numerical solution of this equaion, we use a private mathematica version of the DarkSUSY [73] relic density routine. For the case $N=0$, we reproduced the results of ref. [34] as a check.
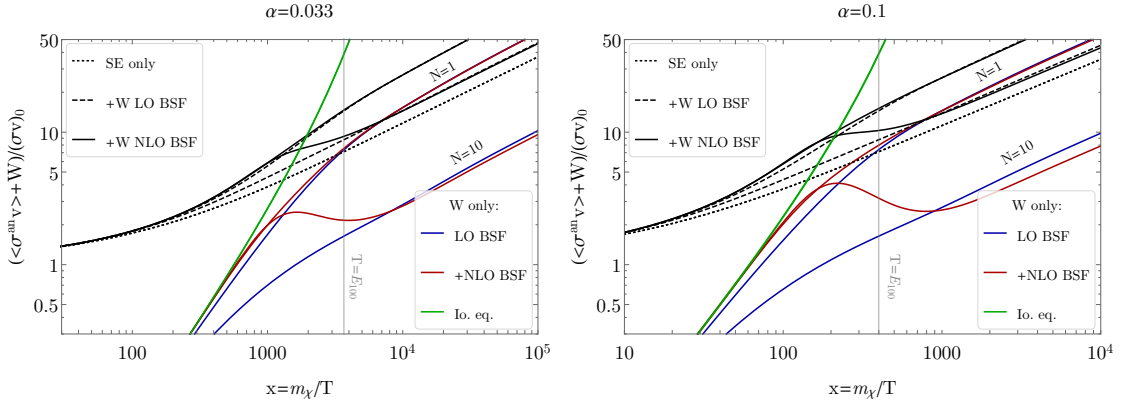
Figure 7. Graphic shows the increasing importance of NLO BSF processes for larger N. Upper (lower) solid and dashed black line, as well as upper (lower) red and blue line correspond to the $N = 1$ ($N = 10$) case. All quantities are normalized to $(\sigma v)_0 \equiv (1+N)\pi\alpha^2/m_\chi^2$, such that the curve for SE annihilation and for W in ionization equilibrium are independent of N.

is therefore related via detailed balance as

$$\Gamma_i^{\rm dis} = \frac{\langle\sigma_{100}^{\rm bsf}v_{\rm rel}\rangle}{4}\frac{n_\chi^{\rm eq}n_{\bar\chi}^{\rm eq}}{n_{100}^{\rm eq}}\,, \tag{5.8}$$

where the equilibrium number density of the bound state contains one spin d.o.f.

The effective cross section W features two asymptotic regimes:

$$W(T) \simeq \begin{cases} (\Gamma_S^{\rm dec} + 3\Gamma_T^{\rm dec})n_{100}^{\rm eq}/(n_\chi^{\rm eq}n_{\bar\chi}^{\rm eq}) & \text{for } \Gamma_i^{\rm dec} \ll \Gamma_i^{\rm dis} \text{ (ionization equilibrium),}\\ \langle\sigma_{100}^{\rm bsf}v_{\rm rel}\rangle & \text{for } \Gamma_i^{\rm dec} \gg \Gamma_i^{\rm dis} \text{ (out-off io. eq.).} \end{cases} \tag{5.9}$$

In the first asymptotic regime, the bound and scattering states are in *ionization equilibrium* (io. eq.). Here, W is i) *independent* of the BSF cross section and dissociation rate [66], and ii) *maximum* for a given bound-state decay rate and temperature. In fig. 7, the maximum value of W is indicated by the green line for electroweak (left panel) and strong couplings (right panel). In the second asymptotic regime at later times, the number density depletion depends only on the total BSF cross section, since the bound states immediately decay without being dissociated back into the scattering states (BSF without "back reaction"). The transition between these two regimes occurs when the dissociation rate is comparable to the decay rates.

We estimate the maximum relative size of W with respect to the SE annihilation cross section [7], assuming io. eq.:

$$\frac{W(T)}{\langle\sigma^{\rm an}v_{\rm rel}\rangle} \approx 2 \times \frac{|E_{100}|}{T}e^{|E_{100}|/T}. \tag{5.10}$$

In this model, the ratio is independent of the number N of bath-particle species. In io. eq., W first grows rapidly for decreasing temperature with a power law $(m_\chi/T)^{3/2}$, becomes

---

[7]In the instantaneous approximation of the SE, one can estimate $\langle\sigma^{\rm an}v_{\rm rel}\rangle \approx (1+N)(\sigma v_{\rm rel})_0(x/x_0)^{1/2}$ for $x \gg x_0$ and $x_0 = 3/(4\pi^2\alpha^2)$, see, e.g., appendix in ref. [57] for details.
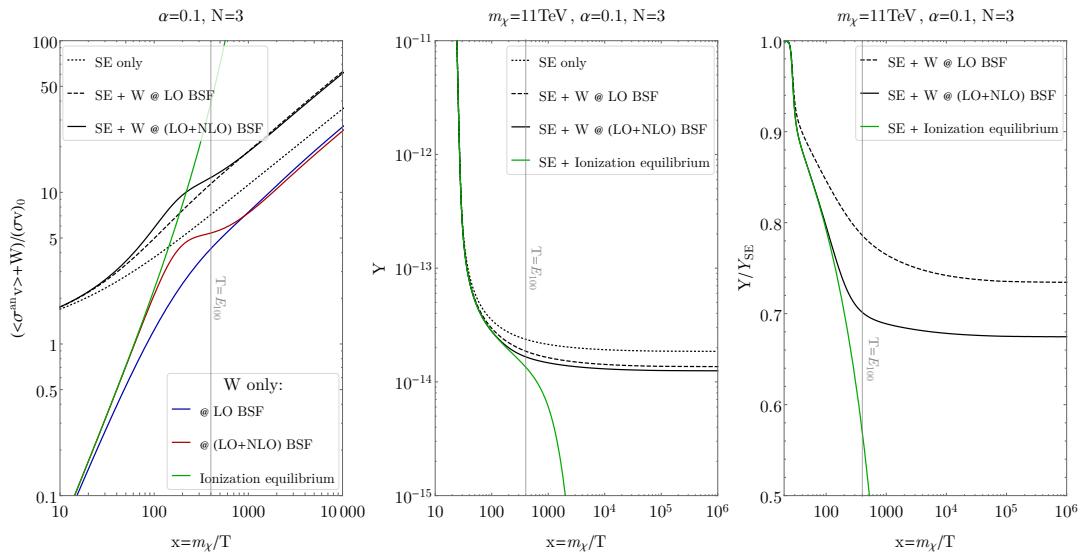
Figure 8. Graphics show the impact of bound-state formation at NLO on the evolution of the thermal relic abundance. Here, $Y \equiv n_\chi/s$, where s is the SM entropy density, and Y_{SE} assumes Sommerfeld enhanced annihilation only.

comparable to the SE for temperature around twice the binding energy, and has an exponentially growing factor for lower temperature.

The relevance of NLO BSF contributions depends, for instance, on for how long the LO BSF can keep the system in ionization equilibrium. In fig. 7, we show that for only one bath-particle species in the plasma, the on-shell emission of a vector mediator is still effective enough to keep W close to its maximum value for sufficiently long time. In this case, the effect of the NLO contributions is only marginally relevant. For larger N, however, the relevance of the NLO contributions is more important. This is because the SE annihilation, NLO BSF cross section, and triplet decay rate are proportional to N, while the LO BSF is independent. BSF via bath-particle scattering as the dominant NLO process, can keep the system in io. eq. until temperatures close to the binding energy even for large N.

Moreover, the relevance of NLO BSF contributions depends on the relative size of W compared to the SE annihilation around the decoupling time from ionization equilibrium. Close to the binding energy, the depletion of the DM number density is exponentially sensitive to the ionization decoupling temperature according to Eq. (5.10). While in the conventional Boltzmann framework one can artificially push the decoupling temperature below the binding energy by lowering the vector mediator mass, our thermal field theory result suggests even for a massless vector mediator a decoupling before but close to the strong exponentially enhanced regime. Based on this result, we may expect a significant impact on the predicted relic abundance in the model under consideration mainly for strong couplings $\alpha \sim 0.1$.

To demonstrate this more clearly, we choose a strong coupling value and more conservatively $N = 3$ ("one quark") in fig. 8. Here, the effect of NLO corrections can significantly change the relic abundance due to a delayed decoupling from ionization equilibrium. The
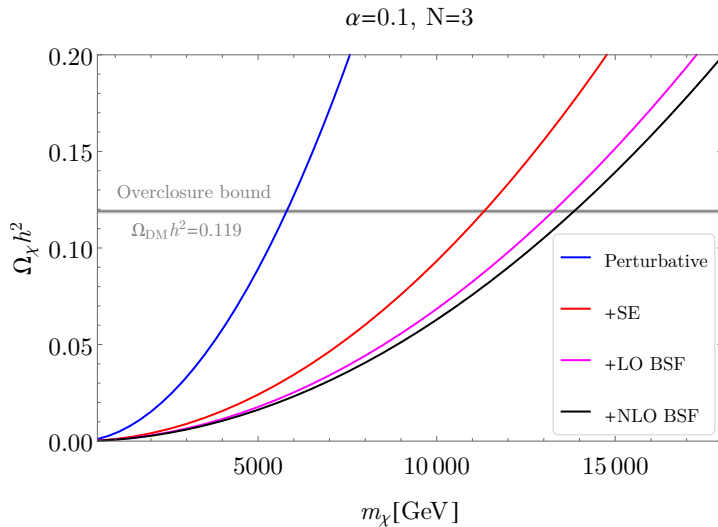
Figure 9. Under the assumption of a strong coupling $\alpha = 0.1$ and one quark in the plasma, we show the predicted relic abundance when considering a tree-level cross section only (blue), including the Sommerfeld effect (red), considering BSF via LO on-shell emission (pink), as well as BSF at NLO (black).

resulting corrections to the upper limit of the DM mass, above which the Universe would contain too much DM (overclosure bound), are shown in fig. 9. For choices of larger N, the differences between LO and NLO would further increase.

6 Discussion

From the results of previous sections, we learned that DM bound-state formation inside a relativistic thermal bath can be dominated by higher order processes for a certain temperature range for massless mediators. It was demonstrated that next-to-leading order processes contained in the mediator spectral function have the potential to change the thermal relic abundance significantly. Although we analyzed a simplified model, this message should be recognized in the broader context of coannihilation scenarios, where the number of SM bath particles can be much larger than considered here in this work.

The formalism of this work also allows to study the impact of the ambient primordial plasma on excited DM state transitions. One can simply change the initial scattering state into a bound state, which transforms the BSF cross section Eq. (2.11) into a level-transition rate. The factorization between the spectral function and the level-transition matrix element holds in this case as well. Contributions from higher DM states were however dropped in the previous section and remain a major uncertainty in the prediction of the final relic abundance.

The bound-state formation and the reverse process have also been investigated in the context of heavy annihilating quarkonia inside the quark-gluon plasma produced, e.g., at the Large-Hadron-Collider, which has many similarities compared with heavy annihilating dark matter in the primordial plasma. For fairness, we compare our approach to the exist-
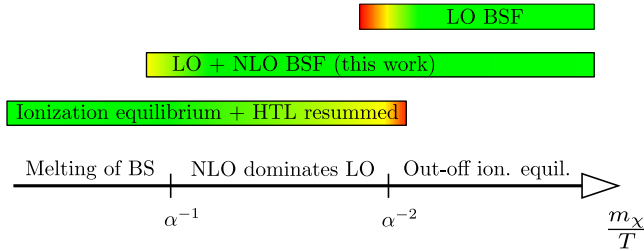
Figure 10. Graphic illustrates goodness of different descriptions, based on the assumption that the mediator is coupled to the primordial plasma background. LO and LO+NLO stand for the bound-state formation description via the emission of an on-shell mediator and higher order couplings of the mediator to the plasma, respectively. The bottom bar line represents HTL resummed corrections to the SE annihilation and bound-state decay rate under the assumption that ionization equilibrium is maintained.

ing methods in the quarkonia literature in the following. A similar formalism is developed in ref. [74], where the LO cross section for quarkonium formation and dissociation is derived from the Lindblad equation. In ref. [75], a particular part of cancellation of collinear divergences is suggested while an ad hoc insertion of a finite temperature self-energy into vacuum amplitudes was performed. Our work can be regarded as a first complete analysis at NLO in the mediator spectral function. Starting from first principles, we have derived Eq. (2.11) and demonstrated that the mediator spectral function evaluated at NLO automatically leads to a collinear safe cross section and includes all the possible processes simultaneously. For example, we have shown that even the finite part of the interference terms is important for predicting the relic abundance precisely (see Fig. 14), which cannot be addressed without the full NLO treatment.

The limitation of the formalism presented in this work is set by the assumptions of potential non-relativistic effective field theory, including the leading order dipole approximation. For temperatures larger than the Bohr-momentum κ, the plasma can entirely probe the typical size of the bound state, and the validity of our generalized BSF cross section Eq. (2.11) breaks down. At the critical value $T \sim \kappa$, the bound and scattering states become non-separable because of the rapid transition between them, which calls for another formalism capable of coping with this situation. At the same time, it was shown that BSF and the back reaction is such efficient, that the system is robustly in ionization equilibrium. We emphasize that the collision term is independent of the BSF cross section in such a state (see Eq. (5.9)).

By using another formalism valid at high temperature but limited to ionization equilibrium, environmental corrections to the upper formula of (5.9) have been studied in the literature [76–81]. The method is based on an effective in-medium potential, which shows next to the expected Debye screening mass, a temperature dependent energy shift (Salpeter correction), as well as an imaginary thermal width [82, 83]. A conservative estimate shows that the thermal width in the effective in-medium potential can lead to an entire *melting* of all bound states for $T \gtrsim \kappa$ [84, 85]. Such melting phenomena have been experimentally observed, e.g., in the decay spectra of bottomonium systems in a quark-gluon plasma, see

fig.1 in ref. [86]. In ref. [66], it has been emphasized that the formalism for computing the relic abundance in the effective in-medium potential description is limited to the assumption of ionization equilibrium (see also ref. [87]), illustrated by the bottom bar line in fig. 10.

One of the main formal achievements of this work was to develop a complementary method, which overlaps with the validity region of the previous effective in-medium potential approach. Combined, we can now describe the evolution of the DM system for a broader range of temperatures, ranging from the melting of bound states in the high temperature regime down to arbitrarily below the decoupling from ionization equilibrium. In particular, we have now addressed the temperature regime below the Bohr-momentum, where the thermal width can be treated as a perturbation and bound states can be characterized by their usual quantum numbers. Note again that the expansion given in Eq. (2.3) implicitly assumes a clear separation between the scattering and bound states, which is a crucial assumption when the continuum starts to overlap with the discrete energy spectrum in the high temperature regime. For lower temperatures around the binding energy, these corrections are expected to be negligible compared to our improved description of the decoupling from ionization equilibrium. In our model, we found that the depletion of the dark matter density is *exponentially sensitive* to the precise decoupling temperature from ionization equilibrium. The presented formalism in this work now allows to accurately resolve the number density evolution especially during this regime.

7   Summary and conclusion

In the conventional Boltzmann formalism, the amplitude for higher-order bound-state formation processes can become collinear divergent in the case of massless mediators. A large number of possible bound-state formation channels through SM particle scattering, via the virtual exchange of, e.g., photons or gluons in coannihilation scenarios, can not be investigated in this context. Based on non-equilibrium quantum field theory techniques, we derived a more general cross section, Eq. (2.11), which addressed this issue without the need of an ad hoc screening mass regulator. We presented for the first time a full computation and analysis of the thermal one-loop correction for the dominant capture into the ground state in the case of ultra-relativistic fermions in the primordial plasma environment, resembling the interactions with light SM leptons and quarks. For temperatures larger than the absolute value of the binding energy, we found that bound-state formation via bath-particle scattering dominates also for massless vector mediators over the so far considered on-shell mediator emission. Our results are complementary to the one in ref. [59], where massive mediators where investigated.

The key quantity in the generalized bound-state formation cross section is the spectral two-point correlation function of the interacting mediator. It was demonstrated that a perturbative expansion in the coupling parameter of this spectral function successively generates on-shell and virtual mediator contributions in a proper thermal field theoretical framework. The known result for the capture into the ground state via the emission of an on-shell mediator was reproduced at the lowest order in the perturbative expansion. The higher-order BSF processes via bath-particle scattering and off-shell mediator decay were

identified at the first interaction level. These processes are collinear divergent for massless mediators in the conventional Boltzmann approach. It was shown that other terms automatically occur in the spectral function in addition, canceling the collinear divergences and resulting in a finite collision term. A rather general analytic proof for the collinear finiteness, applying in particular to our full first interaction term, was presented in section 4.2.

Based on our extended analysis in section 4.4, we conclude that a thermal field theory approach is required for models where the mediator mass is smaller than the absolute value of the binding energy. This implies that our approach can become important also for massive SM gauge bosons or the Higgs field during the electroweak cross over, featuring a large number of possible BSF channels via SM particle scattering as well.

In the case of our simplified model, we demonstrated in section 5 that the impact of the new higher-order bound-state formation effects on the thermal relic abundance can be especially significant if the on-shell emission is not enough to keep the system in ionization equilibrium until temperatures close to the binding energy. Regarding more realistic coannihilation scenarios, the size of the corrections to the upper bound on the DM mass from higher-order bound-state formation processes still remains an open question from the perspective of this work, since the details of the underlying model might play a crucial role. This is mainly due to the fact that the impact on the relic abundance is exponentially sensitive to the precise decoupling time from ionization equilibrium if it is maintained until temperatures around the binding energy. Model-dependent order one variations in the ionization equilibrium decoupling temperature are crucial for the impact of the higher-order effects. With an accurate formalism by hand, together with the insights from simplified DM models, makes it nevertheless worthwhile to analyze more realistic scenarios in future work.

Acknowledgments

A   Computation of double commutator

In this appendix, we give a detailed derivation of Eq. (2.6) starting from Eq. (2.4). Throughout this paper, we are interested in the BSF and dissociation rate. For this purpose, one may focus on a particular part of the whole interaction Hamiltonian, which is nothing but the dipole interactions given in Eq. (2.2). Inserting the mode expansion given in Eq. (2.3), we obtain the following form for the operator responsible for the transition between scattering

and bound states:

$$\hat{H}_{\text{dip}} \supset - \int\limits_{\mathbf{K},\mathbf{k},P} \sum_{\mathcal{B},\text{Spin}} e^{i(\Delta E - P_0)t} \, E^i(P) L^{i\star}_{\mathbf{k},\mathcal{B}} \, \hat{a}^{s\dagger}_{\mathbf{K}/2+\mathbf{k}} \hat{b}^{r\dagger}_{\mathbf{K}/2-\mathbf{k}} \hat{c}^{sr}_{\mathcal{B},\mathbf{K}} + \text{H.c.,} \qquad (A.1)$$

where the overlap integral is given by

$$L^i_{\mathbf{k},\mathcal{B}} = g \int \mathrm{d}^3 r \, \psi^\star_{\mathcal{B}}(\mathbf{r}) r^i \psi_{\mathbf{k}}(\mathbf{r}). \qquad (A.2)$$

Here we use a shorthand notation for integrals: $\int_{\mathbf{k}} \equiv \int \frac{\mathrm{d}^3 k}{(2\pi)^3}$ and $\int_P \equiv \int \frac{\mathrm{d}^4 P}{(2\pi)^4}$. The positive quantity $\Delta E \equiv \mathcal{E}_{\mathbf{k}} - \mathcal{E}_{\mathcal{B}}$ is the total energy emitted in the inelastic conversion, i.e. relative kinetic energy plus the absolute value of the negative binding energy.

All one needs to do is to evaluate the double commutator in the right-hand side of Eq. (2.4) by using Eq. (A.1) in order to obtain the collision term for the BSF and dissociation rates. Below we summarize commutators and expectation values relevant for our computations:

$$\left[\hat{n}_{\mathbf{k}_\chi}, \hat{a}^s_{\mathbf{k}'_\chi}\right] = -(2\pi)^3 \delta^3(\mathbf{k}_\chi - \mathbf{k}'_\chi)\hat{a}^s_{\mathbf{k}'_\chi}, \qquad \left[\hat{n}_{\mathbf{k}_\chi}, \hat{a}^{s\dagger}_{\mathbf{k}'_\chi}\right] = (2\pi)^3 \delta^3(\mathbf{k}_\chi - \mathbf{k}'_\chi)\hat{a}^{s\dagger}_{\mathbf{k}'_\chi}, \qquad (A.3)$$

for commutators, and

$$\left\langle a^{s\dagger}_{\mathbf{k}_\chi} a^{s'}_{\mathbf{k}'_\chi} \right\rangle = f_\chi(\mathbf{k}_\chi)(2\pi)^3 \delta^3(\mathbf{k}_\chi - \mathbf{k}'_\chi)\delta_{ss'}, \qquad \left\langle b^{s\dagger}_{\mathbf{k}_{\bar{\chi}}} b^{s'}_{\mathbf{k}'_{\bar{\chi}}} \right\rangle = f_{\bar{\chi}}(\mathbf{k}_{\bar{\chi}})(2\pi)^3 \delta^3(\mathbf{k}_{\bar{\chi}} - \mathbf{k}'_{\bar{\chi}})\delta_{ss'},$$

$$\left\langle c^{sr\dagger}_{\mathcal{B},\mathbf{K}} c^{s'r'}_{\mathcal{B}',\mathbf{K}'} \right\rangle = f_{\mathcal{B}}(\mathbf{K})(2\pi)^3 \delta^3(\mathbf{K} - \mathbf{K}')\delta_{\mathcal{B}\mathcal{B}'}\delta_{ss'}\delta_{rr'}, \qquad (A.4)$$

for expectation values of creation and annihilation operators. Here note that these expectation values are taken by the initial factorized density matrix: $\langle \hat{\bullet} \rangle = \text{Tr}[\hat{\bullet}\hat{\rho}(t=0)]$. We also need two-point functions of the electric field defined by

$$E^{-+}_{ij}(P) \equiv \int \mathrm{d}^4(x-y)e^{-iP\cdot(x-y)} \left\langle E^i(x)E^j(y) \right\rangle,$$

$$E^{+-}_{ij}(P) \equiv \int \mathrm{d}^4(x-y)e^{-iP\cdot(x-y)} \left\langle E^j(y)E^i(x) \right\rangle. \qquad (A.5)$$

For later convenience, we rewrite this electric correlator in terms of its gauge field as:

$$E^{-+/-+}_{ij}(P) = (ip^0 g^{i\mu} - ip^i g^{0\mu})(-ip^0 g^{j\mu} + ip^j g^{0\nu})D^{-+/+-}_{\mu\nu}(P), \qquad (A.6)$$

where in coordinate space $D^{-+}_{\mu\nu}(x-y) \equiv \langle A_\mu(x)A_\nu(y) \rangle$ and $D^{+-}_{\mu\nu}(x-y) \equiv \langle A_\nu(y)A_\mu(x) \rangle$. Since the expectation value is taken by the factorized initial density matrix and the environment is assumed to be in thermal equilibrium, these correlators coincide with the thermal correlator, fulfilling the KMS relation:

$$D^{-+}_{\mu\nu}(P) = [1 + f^{\text{eq}}_\gamma(P^0)]D^\rho_{\mu\nu}(P), \quad D^{+-}_{\mu\nu}(P) = f^{\text{eq}}_\gamma(P^0)D^\rho_{\mu\nu}(P), \qquad (A.7)$$

where the photon spectral function is given by $D^\rho_{\mu\nu}(x-y) \equiv \langle [A_\mu(x), A_\nu(y)] \rangle$. f^{eq}_γ represents the Bose-Einstein distribution for the gauge boson.

Using these equations, one finally obtain the collision term associated with the BSF and dissociation:

$$\mathcal{C}_{\text{BSF+diss}} = -g_\chi g_{\bar{\chi}} \sum_{\mathcal{B}} \int \frac{\mathrm{d}^3 k_\chi}{(2\pi)^3} \frac{\mathrm{d}^3 k_{\bar{\chi}}}{(2\pi)^3} \frac{\mathrm{d}^3 p}{(2\pi)^3} D^\rho_{\mu\nu}(\Delta E, \mathbf{p}) \sum_{\text{Spin}} \mathcal{T}^\mu_{\mathbf{k},\mathcal{B}}(\Delta E, \mathbf{p}) \mathcal{T}^{\nu\star}_{\mathbf{k},\mathcal{B}}(\Delta E, \mathbf{p})$$

$$\times \left\{ f_\chi(\mathbf{k}_\chi) f_{\bar{\chi}}(\mathbf{k}_{\bar{\chi}})[1 + f^{\text{eq}}_\gamma(\Delta E)] - f_\mathcal{B}(\mathbf{K} - \mathbf{p}) f^{\text{eq}}_\gamma(\Delta E) \right\}, \quad \text{(A.8)}$$

with $\Delta E = \mathbf{k}^2/(2\mu) + |\mathcal{E}_\mathcal{B}|$, $\mathbf{k} = (\mathbf{k}_\chi - \mathbf{k}_{\bar{\chi}})/2$, and $\mathbf{K} = \mathbf{k}_\chi + \mathbf{k}_{\bar{\chi}}$. Here we have dropped the Bose enhancement and Fermi suppression factors for the scattering and bound states because they are dilute. To obtain this compact form, we have used the following relation between the overlap integral and tensors defined in Eqs. (2.7) and (2.8):

$$\frac{1}{g_\chi g_{\bar{\chi}}} E^\rho_{ij}(P) \sum_{\text{Spin}} L^i_{\mathbf{k},\mathcal{B}} \delta^{ss'} \delta^{rr'} L^{j\star}_{\mathbf{k},\mathcal{B}} \delta^{ss'} \delta^{rr'} = D^\rho_{\mu\nu}(P) \sum_{\text{Spin}} \mathcal{T}^\mu_{\mathbf{k},\mathcal{B}}(P) \mathcal{T}^{\nu\star}_{\mathbf{k},\mathcal{B}}(P), \quad \text{(A.9)}$$

where g_χ and $g_{\bar{\chi}}$ represent the spin degrees of freedom for particle and anti-particle respectively. The collision term can be expressed as the following simple form

$$\mathcal{C}_{\text{BSF+diss}} = -\sum_{\mathcal{B}} \langle \sigma^{\text{bsf}}_\mathcal{B} v_{\text{rel}} \rangle \left[n_\chi n_{\bar{\chi}} - n_\mathcal{B} \frac{n^{\text{eq}}_\chi n^{\text{eq}}_{\bar{\chi}}}{n^{\text{eq}}_\mathcal{B}} \right], \quad \text{(A.10)}$$

where

$$\sigma^{\text{bsf}}_\mathcal{B} v_{\text{rel}} \equiv \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \left[1 + f^{\text{eq}}_\gamma(\Delta E) \right] D^\rho_{\mu\nu}(\Delta E, \mathbf{p}) \sum_{\text{Spin}} \mathcal{T}^\mu_{\mathbf{k},\mathcal{B}}(\Delta E, \mathbf{p}) \mathcal{T}^{\nu\star}_{\mathbf{k},\mathcal{B}}(\Delta E, \mathbf{p}). \quad \text{(A.11)}$$

One might wonder why this cross section contains $1 + f^{\text{eq}}_\gamma(\Delta E)$ even for the inverse process. To avoid this confusion, we provide an explicit proof for the inverse process starting from Eq. (A.8):

$$\text{reverse} = g_\chi g_{\bar{\chi}} \sum_{\mathcal{B}} \int_{\mathbf{k}_\chi, \mathbf{k}_{\bar{\chi}}} \sigma_\mathcal{B} v_{\text{rel}} \frac{f^{\text{eq}}_\gamma(\Delta E)}{1 + f^{\text{eq}}_\gamma(\Delta E)} f_\mathcal{B}(\mathbf{K} - \mathbf{p})$$

$$= g_\chi g_{\bar{\chi}} \sum_{\mathcal{B}} \int_{\mathbf{k}_\chi, \mathbf{k}_{\bar{\chi}}} \sigma_\mathcal{B} v_{\text{rel}} e^{-\frac{\Delta E}{T}} f^{\text{eq}}_\mathcal{B}(\mathcal{E}_\mathcal{B} + \mathbf{K}^2/8\mu) \times \frac{n_\mathcal{B}}{n^{\text{eq}}_\mathcal{B}}$$

$$= \sum_{\mathcal{B}} \left(\frac{g_\chi g_{\bar{\chi}}}{n^{\text{eq}}_\chi n^{\text{eq}}_{\bar{\chi}}} \int_{\mathbf{k}_\chi, \mathbf{k}_{\bar{\chi}}} e^{-\mathbf{k}^2_\chi/4\mu} e^{-\mathbf{k}^2_{\bar{\chi}}/4\mu} \sigma_\mathcal{B} v_{\text{rel}} \right) \times \frac{n^{\text{eq}}_\chi n^{\text{eq}}_{\bar{\chi}}}{n^{\text{eq}}_\mathcal{B}} n_\mathcal{B}$$

$$= \sum_{\mathcal{B}} \langle \sigma_\mathcal{B} v_{\text{rel}} \rangle \times \frac{n^{\text{eq}}_\chi n^{\text{eq}}_{\bar{\chi}}}{n^{\text{eq}}_\mathcal{B}} n_\mathcal{B}. \quad \text{(A.12)}$$

In the first line we insert the definition of the generalized cross section. In the second line we use the relation $f^{\text{eq}}_\gamma(\Delta E)/[1 + f^{\text{eq}}_\gamma(\Delta E)] = e^{-\Delta E/T}$ and the approximation of kinetic equilibrium $f_\mathcal{B} = f^{\text{eq}}_\mathcal{B} n_\mathcal{B}/n^{\text{eq}}_\mathcal{B}$. In the third line we use the diluteness of the bound state $f^{\text{eq}}_\mathcal{B} = e^{-(\mathcal{E}_\mathcal{B} + \mathbf{K}^2/8\mu)/T}$ and the energy conservation $e^{-\Delta E/T} e^{-(\mathcal{E}_\mathcal{B} + \mathbf{K}^2/8\mu)/T} = e^{-\mathbf{k}^2_\chi/4\mu} e^{-\mathbf{k}^2_{\bar{\chi}}/4\mu}$. This completes the proof of the second term in Eq. (A.10).

B Retarded self-energy for massless fermions

The retarded photon self-energy is defined in terms of greater and lesser self-energies as

$$\Pi_{\mu\nu}^{R}(x-y) = \theta(x^0 - y^0)\left[\Pi_{\mu\nu}^{-+}(x-y) - \Pi_{\mu\nu}^{+-}(x-y)\right], \tag{B.1}$$

$$\Pi_{\mu\nu}^{+-}(x-y) = g^2\,\mathrm{Tr}\left[\gamma_\mu S^{+-}(x-y)\gamma_\nu S^{-+}(y-x)\right], \tag{B.2}$$

$$\Pi_{\mu\nu}^{-+}(x-y) = g^2\,\mathrm{Tr}\left[\gamma_\mu S^{-+}(x-y)\gamma_\nu S^{+-}(y-x)\right]. \tag{B.3}$$

where the two-point functions of the fermionic bath-particles are defined as

$$S_{ij}^{-+}(x-y) \equiv \langle \psi_i(x)\bar{\psi}_j(y)\rangle, \tag{B.4}$$

$$S_{ij}^{+-}(x-y) \equiv -\langle \bar{\psi}_j(y)\psi_i(x)\rangle. \tag{B.5}$$

In the free limit and in thermal equilibrium, their Fourier transform is given by

$$S^{+-}(K) = -\slashed{K}(2\pi)\delta(K^2)\left[-\theta(-K^0) + f_\psi^{\mathrm{eq}}(|K^0|)\right], \tag{B.6}$$

$$S^{-+}(K) = -\slashed{K}(2\pi)\delta(K^2)\left[-\theta(+K^0) + f_\psi^{\mathrm{eq}}(|K^0|)\right]. \tag{B.7}$$

f_ψ^{eq} is the Fermi-Dirac equilibrium phase-space distribution. In the free and equilibrium limit, the terms needed for the self-energy can be expressed in Fourier space as:

$$
\begin{aligned}
S^{+-}(x-y) &= \int \frac{\mathrm{d}^4 k}{(2\pi)^4} e^{-ik(x-y)} S^{+-}(k) \\
&= \int \frac{\mathrm{d}^3 k}{(2\pi)^3 2|\mathbf{k}|}\slashed{K}\left[-f_\psi^{\mathrm{eq}}(|\mathbf{k}|)e^{-ik(x-y)} - (1 - f_\psi^{\mathrm{eq}}(|\mathbf{k}|))e^{ik(x-y)}\right] \\
&\equiv \int \frac{\mathrm{d}^3 k}{(2\pi)^3 2|\mathbf{k}|}\slashed{K}\left[-\alpha_{\mathbf{k}}e^{-ik(x-y)} - \beta_{\mathbf{k}}e^{ik(x-y)}\right], \tag{B.8}
\end{aligned}
$$

$$
\begin{aligned}
S^{-+}(x-y) &= \int \frac{\mathrm{d}^4 k}{(2\pi)^4} e^{-ik(x-y)} S^{-+}(k) \\
&= \int \frac{\mathrm{d}^3 k}{(2\pi)^3 2|\mathbf{k}|}\slashed{K}\left[(1 - f_\psi^{\mathrm{eq}}(|\mathbf{k}|))e^{-ik(x-y)} + f_\psi^{\mathrm{eq}}(|\mathbf{k}|)e^{ik(x-y)}\right] \\
&\equiv \int \frac{\mathrm{d}^3 k}{(2\pi)^3 2|\mathbf{k}|}\slashed{K}\left[\beta_{\mathbf{k}}e^{-ik(x-y)} + \alpha_{\mathbf{k}}e^{ik(x-y)}\right], \tag{B.9}
\end{aligned}
$$

$$\theta(x^0 - y^0) = -\frac{1}{2\pi i}\int \mathrm{d}k\,\frac{1}{k^0 + i\epsilon}e^{-ik^0(x^0 - y^0)}. \tag{B.10}$$

Where α and β are short notation for the phase space density and the Pauli blocking, respectively.

Using these equations, the retarded self-energy can be written in Fourier space as

$$\Pi_{\mu\nu}^R(P) = g^2 \int \mathrm{d}^4 z e^{iPz} \int \frac{\mathrm{d}k^0}{(2\pi)} e^{-ik^0 z^0} \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3 2|\mathbf{k}_1|} \int \frac{\mathrm{d}^3 k_2}{(2\pi)^3 2|\mathbf{k}_2|} \operatorname{Tr}\left[\gamma_\mu \slashed{K}_1 \gamma_\nu \slashed{K}_2\right] \frac{i}{k^0 + i\epsilon} \times$$

$$\left\{ \left(\beta_{\mathbf{k}_1} e^{-ik_1 z} + \alpha_{\mathbf{k}_1} e^{ik_1 z}\right)\left(-\alpha_{\mathbf{k}_2} e^{ik_2 z} - \beta_{\mathbf{k}_2} e^{-ik_2 z}\right) - \right.$$

$$\left.\left(-\alpha_{\mathbf{k}_1} e^{-ik_1 z} - \beta_{\mathbf{k}_1} e^{ik_1 z}\right)\left(\beta_{\mathbf{k}_2} e^{ik_2 z} + \alpha_{\mathbf{k}_2} e^{-ik_2 z}\right)\right\}$$

$$= g^2 \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3 2|\mathbf{k}_1|} \int \frac{\mathrm{d}^3 k_2}{(2\pi)^3 2|\mathbf{k}_2|} \operatorname{Tr}\left[\gamma_\mu \slashed{K}_1 \gamma_\nu \slashed{K}_2\right] (2\pi)^3 \times$$

$$\left\{ \alpha_{\mathbf{k}_1}\alpha_{\mathbf{k}_2}\left(\frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 - \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 + \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon}\right)\right.$$

$$+ \beta_{\mathbf{k}_1}\beta_{\mathbf{k}_2}\left(\frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 + \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 - \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon}\right)$$

$$+ \alpha_{\mathbf{k}_1}\beta_{\mathbf{k}_2}\left(\frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 + \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 - \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon}\right)$$

$$+ \beta_{\mathbf{k}_1}\alpha_{\mathbf{k}_2}\left(\frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 - \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 + \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon}\right)\right\}. \tag{B.11}$$

After rearrangement, the latter equation can be brought into a convenient form

$$\Pi_{\mu\nu}^R(P) = g^2 \int \frac{\mathrm{d}^3 k_1}{(2\pi)^3 2|\mathbf{k}_1|} \int \frac{\mathrm{d}^3 k_2}{(2\pi)^3 2|\mathbf{k}_2|} \operatorname{Tr}\left[\gamma_\mu \slashed{K}_1 \gamma_\nu \slashed{K}_2\right] (2\pi)^3 \times$$

$$\left\{ \left[1 - f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|)\right]\left[\frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 + \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 - \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon}\right]\right.$$

$$+ \left[f_\psi^{\mathrm{eq}}(|\mathbf{k}_2|) - f_\psi^{\mathrm{eq}}(|\mathbf{k}_1|)\right]\left[\frac{i\delta^3(\mathbf{p} + \mathbf{k}_1 - \mathbf{k}_2)}{P^0 + |\mathbf{k}_1| - |\mathbf{k}_2| + i\epsilon} - \frac{i\delta^3(\mathbf{p} - \mathbf{k}_1 + \mathbf{k}_2)}{P^0 - |\mathbf{k}_1| + |\mathbf{k}_2| + i\epsilon}\right]\right\}, \tag{B.12}$$

which is used in section 4 to analyze the various NLO contributions.

C Next-to-leading order contributions in more detail

In this appendix, we share some details of the next-to-leading order computation of the mediator spectral function, as well as some more detailed discussion about the individual contributions.

C.1 Contour integration for vacuum part

To evaluate the vacuum part of the cross section in Eq. (4.10), one has to compute

$$R_0 = \lim_{R\to\infty} \lim_{\epsilon\searrow 0} \frac{1}{\pi} \int_0^R \mathrm{d}|\mathbf{p}| \operatorname{Im}\left[G_\epsilon(|\mathbf{p}|)\right], \tag{C.1}$$

for

$$G_\epsilon(|\mathbf{p}|) = \frac{\mathbf{p}^2(\Delta E^2 - \mathbf{p}^2)\left(\mathbf{p}^2 - 3\Delta E^2\right)}{3\Delta E^3 \left[(\Delta E + i\epsilon)^2 - \mathbf{p}^2\right]^2}\left[\ln\left(\frac{(\Delta E + i\epsilon)^2 - \mathbf{p}^2}{-\mu_0^2}\right) - \frac{5}{3}\right]. \tag{C.2}$$
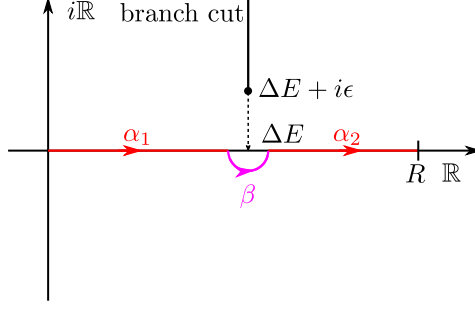
Figure 11. Contour used to compute the vacuum part. Black dot: the pole of the function G_ϵ; α_1, α_2: paths on the real line; β: path in the complex plane (half circle) avoiding the singularities of G_ϵ.

Consider the analytic continuation $|\mathbf{p}| \to z$ and choose the branch cut as shown in fig. 11. In order to avoid integrating thought the singularity at $z = \Delta E + i\epsilon$ we deform the contour and integrate over $\gamma = \alpha_1 + \beta + \alpha_2$ (see fig. 11), where

$$
\begin{align}
\alpha_1(t) &= t, & t &\in [0, \Delta E - r], & \text{(C.3)}\\
\beta(t) &= \Delta E + re^{it}, & t &\in [-\pi, 0], & \text{(C.4)}\\
\alpha_2(t) &= t, & t &\in [\Delta E + r, R]. & \text{(C.5)}
\end{align}
$$

As the value of the integral does not change when integrating along γ and G_0 has no singularities on γ we get

$$
\lim_{\epsilon \searrow 0} \frac{1}{\pi} \int_0^R d|\mathbf{p}| \operatorname{Im}\left[G_\epsilon(|\mathbf{p}|)\right] = \operatorname{Im}\left[\lim_{\epsilon \searrow 0} \frac{1}{\pi} \int_\gamma dz\, G_\epsilon(z)\right] = \operatorname{Im}\left[\frac{1}{\pi} \int_\gamma dz\, G_0(z)\right], \qquad \text{(C.6)}
$$

where

$$
\begin{align}
G_0(z) &= \frac{z^2\left(z^2 - 3\Delta E^2\right)}{3\Delta E^3(\Delta E^2 - z^2)}\left[\ln\left(\frac{z^2 - \Delta E^2}{\mu_0^2}\right) - \frac{5}{3}\right]\\
&= \frac{1}{3\Delta E}\left(2 - \frac{z^2}{\Delta E^2} - \frac{\Delta E}{z + \Delta E} + \frac{\Delta E}{z - \Delta E}\right)\left[\ln\left(\frac{z^2 - \Delta E^2}{\mu_0^2}\right) - \frac{5}{3}\right]. \qquad \text{(C.7)}
\end{align}
$$

Note that $\int_\gamma dz\, G_0(z)$ does not depend on r. Moreover, as G_0 is purely real on $\{x \in \mathbb{R} \mid x > \Delta E\}$ the integration along α_2 in Eq. (C.6) gives no contribution. By considering each summand in Eq. (C.7) separately and taking the limit $r \searrow 0$ we finally obtain

$$
R_0 = \lim_{r \searrow 0} \operatorname{Im}\left[\frac{1}{\pi} \int_{\alpha_1 + \beta} dz\, G_0(z)\right] = \frac{1}{3}\left[\ln\left(\frac{\Delta E^2}{\mu_0^2/4}\right) - \frac{10}{3}\right]. \qquad \text{(C.8)}
$$

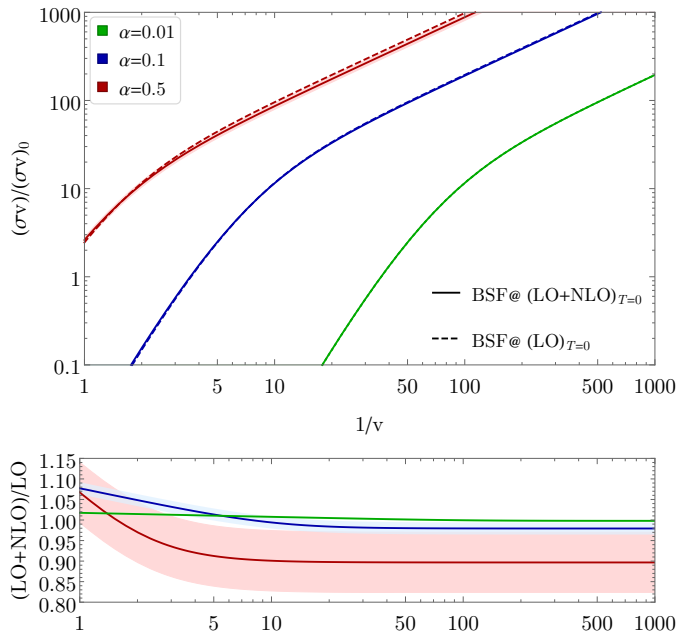## C.2 Scale dependence of the vacuum part

Figure 12. *Upper panel:* BSF cross section normalized to the tree-level annihilation cross section at LO (dashed lines) and NLO (solid lines) in dependence of the inverse relative velocity for different coupling strengths α while fixing the renormalization scale to $\mu_0 = E_{100} = m_\chi \alpha^2/4$ (assuming $N = 1$). Shaded areas (barely visible) indicate the scale uncertainty by varying the renormalization scale between $\mu_0 = 2E_{100}$ and $\mu_0 = E_{100}/2$. *Lower panel:* Relative correction of the NLO contribution with respect to the LO cross section with the renormalization scale fixed at $\mu_0 = E_{100} = m_\chi \alpha^2/4$ (solid lines). The shaded area indicates the renormalization uncertainty as in the upper panel.

As discussed in section 4.1, in particular around Eq. (4.10), the NLO contribution at zero temperature can be factorized in terms of the LO contribution,

$$(\sigma_{100}^{\text{LO}} v_{\text{rel}}) + (\sigma_{100}^{\text{NLO}} v_{\text{rel}})_{T=0} = (\sigma_{100}^{\text{LO}} v_{\text{rel}}) \left\{ 1 + N \times \frac{\alpha}{3\pi} \left[\ln\left(\frac{\Delta E^2}{\mu_0^2/4}\right) - \frac{10}{3} \right] \right\}, \qquad \text{(C.9)}$$

where N specifies the number of particle species in the thermal plasma and thus also the multiplicity of particles that are running in the loop in the NLO contribution to the on-shell emission. As expected from the renormalization procedure of the occurring UV divergence, the final cross section will feature a renormalization scale dependence μ_0 that needs to be fixed by a physical motivated scale. As the characteristic scale of the BSF process lies around the binding energy, we choose accordingly $\mu_0 = E_{100} = m_\chi \alpha^2/4$.

In fig. 12, we show the vacuum contribution at LO and NLO in dependence of the inverse relative velocity v_{rel} for coupling strengths up to the order of the unitarity bound for s-wave SE annihilation, $\alpha = 0.5$ [88]. In order to estimate the uncertainty that arises from this particular choice of scale, we vary the latter between $\mu_0 = 2E_{100}$ and $\mu_0 = E_{100}/2$ (shaded area). The upper panel features the general expected behaviour: With larger velocity the relative cross section decreases as BSF is suppressed due to the small overlap

of the wave functions. Moreover, larger couplings α lead to a relatively larger BSF cross section. From Eq. C.9, we can infer that for larger couplings the NLO correction becomes more relevant and hence also the dependence of the scale uncertainty, as visible in the lower panel. Note that we have taken the conservative choice of $N = 1$. For larger N not only the correction itself, but also the scale uncertainty is expected to be more pronounced.

For the renormalization of the occurring UV divergence, we had chosen the simplest choice for the renormalization scheme, namely the $\overline{\text{MS}}$ scheme. However, a more physical motivated approach would be to use an on-shell scheme for the renormalization of the photon propagator since the photon is on-shell in the corresponding interference terms. However, this unnecessarily complicates the cancellation of the infrared divergences. While it would be academically interesting to study the application of the on-shell scheme and to compare it with the $\overline{\text{MS}}$ scheme in more detail, we leave this investigation for future work.

C.3 Contour integration for the finite temperature part

The function $G_\epsilon^{\sigma_1\sigma_2}(z) = G_\epsilon^{\sigma_1\sigma_2}(z, \tau, |\mathbf{k}|)$ in the finite temperature part of the cross section in Eq. (4.13) has three types of poles:

- the double poles of the squared mediator propagator at $\pm(\Delta E + i\epsilon)$,

- the single poles of the retarded self-energy at $w_{p/m}$, w_p lying in the upper half and w_m lying in the lower half of the complex plane,

- and the poles of $F^{\sigma_1\sigma_2}$, which do not lie on the real line.

When ϵ goes to 0, $w_0 \equiv \Delta E + i\epsilon$ goes to $z_0 = \Delta E$ which lies on the positive real line. w_p and w_m go to

$$z_{p/m} = \lim_{\epsilon \searrow 0} w_{p/m} \equiv -\tau\sigma_1|\mathbf{k}| \pm \text{sign}(\Delta E + \sigma_1|\mathbf{k}|)\sqrt{\tau^2|\mathbf{k}|^2 + \Delta E^2 + 2\sigma_1\Delta E|\mathbf{k}|}, \quad (C.10)$$
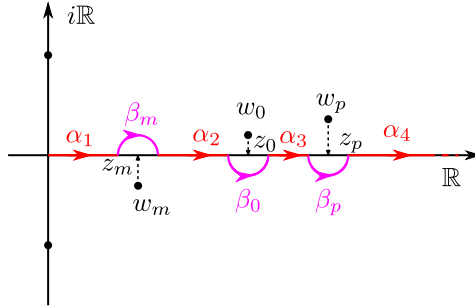
which can lie on the positive real line, as well.

Figure 13. Contour used to compute the finite temperature part. Black dots: poles of the function $G_\epsilon^{\sigma_1\sigma_2}$; z_0, z_p, z_m: location of the poles of $G_0^{\sigma_1\sigma_2}$ which lie on the real line; w_0, w_p, w_m: location of the poles of $G_\epsilon^{\sigma_1\sigma_2}$ which go to z_0, z_p, z_m as $\epsilon \searrow 0$; α_1, α_2, α_3, α_4: paths on the real line; β_0, β_p, β_m: paths in the complex plane (half circles) avoiding the singularities of $G_\epsilon^{\sigma_1\sigma_2}$.

For the moment let us assume that $z_{p/m}$ are real and positive. To avoid integrating through the singularities we deform the path of integration as shown in fig. 13. The deformation of the path does not affect the value of the integral over $|\mathbf{p}|$ as $F^{\sigma_1\sigma_2}$ is holomorphic in some neighborhood of the real line. Since $G_0^{\sigma_1\sigma_2}$ has no poles on the new path γ, we get

$$\lim_{\epsilon\searrow 0}\int_0^\infty \mathrm{d}|\mathbf{p}|\ G_\epsilon^{\sigma_1\sigma_2}(|\mathbf{p}|,\tau,|\mathbf{k}|) = \lim_{\epsilon\searrow 0}\int_\gamma \mathrm{d}z\ G_\epsilon^{\sigma_1\sigma_2}(z,\tau,|\mathbf{k}|) = \int_\gamma \mathrm{d}z\ G_0^{\sigma_1\sigma_2}(z,\tau,|\mathbf{k}|). \quad \text{(C.11)}$$

As $G_0^{\sigma_1\sigma_2}(z)$ is real valued on the real line, the integration along $\alpha_1,\dots,\alpha_4$ only gives a real contribution. Moreover, it follows that the Laurent coefficients of $G_0^{\sigma_1\sigma_2}$ at $z = z_0$, $z = z_p$, and $z = z_m$, respectively, are real. Using Laurent series expansion it is straightforward to verify that

$$\mathrm{Im}\left(\int_{\beta_0} G_0^{\sigma_1\sigma_2}(z)\mathrm{d}z\right) = \pi\,\mathrm{Res}(G_0^{\sigma_1\sigma_2}, z_0), \quad \text{(C.12)}$$

$$\mathrm{Im}\left(\int_{\beta_{p/m}} G_0^{\sigma_1\sigma_2}(z)\mathrm{d}z\right) = \pm\pi\,\mathrm{Res}(G_0^{\sigma_1\sigma_2}, z_{p/m}). \quad \text{(C.13)}$$

Finally we get

$$\lim_{\epsilon\searrow 0}\int_0^\infty \mathrm{d}|\mathbf{p}|\ \mathrm{Im}\left[G_\epsilon^{\sigma_1\sigma_2}(|\mathbf{p}|,\tau,|\mathbf{k}|)\right] = \lim_{\epsilon\searrow 0}\int_0^\infty \mathrm{d}z\ \mathrm{Im}\left[G_\epsilon^{\sigma_1\sigma_2}(z,\tau,|\mathbf{k}|)\right]$$

$$= \pi\cdot\Big[\mathrm{Res}(G_0^{\sigma_1\sigma_2}, z_0) + \mathrm{Res}(G_0^{\sigma_1\sigma_2}, z_p) - \mathrm{Res}(G_0^{\sigma_1\sigma_2}, z_m)\Big], \text{ for } z_{p/m} \in \mathbb{R}^+. \quad \text{(C.14)}$$

If one of the poles $z_{p/m}$ is negative or not real then the corresponding term in Eq. (C.14) vanishes as there is no need to deform the path of integration. In particular, by the same argument the poles on the imaginary axis, originating from the equilibrium phase-space distributions in Eq. (4.3), do no contribute.

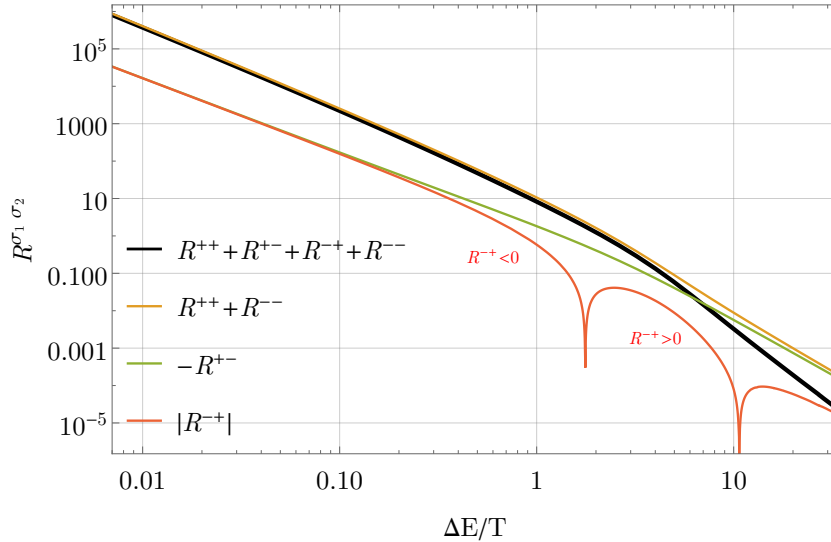## C.4   Individual contributions of the finite temperature part

Figure 14. The plot shows the individual $R^{\sigma_1\sigma_2}$ functions and their dependence on $\Delta E/T$.

The $R^{\sigma_1\sigma_2}$ functions, in Eq. (4.17) to Eq. (4.20), are all dimensionless. Therefore they can only dependent on the ratio $\Delta E/T$ for massless mediators. This fact makes the relic abundance computation efficient, since one can tabulate the sum over all these functions in advance depending on a single variable.

For a selected range we show the individual contributions in fig. 14. R^{++} and R^{--} contain the BSF via bath-particle and anti-particle scattering and are the most dominant contributions. The double pole in R^{+-} contributes negatively, and R^{-+} has some oscillating features. The latter contains the finite temperture part of the off-shell mediator decay into bath particles as well as other single and double pole residues. Depending on which part dominates inside R^{-+}, the function takes positive and negative values.

Most important is the value of the sum over all functions at $T/\Delta E \sim 1$, since there one would roughly enter the exponential depletion phase of the relic abundance if ionization equilibrium is maintained. One can see that for a very precise determination of the relic abundance, the inclusion of all terms are needed at least in this case. Throughout this work we take all of them into account.

References

[1] B. W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino Masses*, *Phys.Rev.Lett.* **39** (1977) 165.

[2] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive and M. Srednicki, *Supersymmetric Relics from the Big Bang*, *Nucl. Phys.* **B238** (1984) 453.

[3] G. Arcadi, M. Dutra, P. Ghosh, M. Lindner, Y. Mambrini, M. Pierre et al., *The waning of the WIMP? A review of models, searches, and constraints*, *Eur. Phys. J.* **C78** (2018) 203 [1703.07364].

[4] PLANCK collaboration, *Planck 2015 results. XIII. Cosmological parameters*, *Astron. Astrophys.* **594** (2016) A13 [1502.01589].

[5] J. Hisano, S. Matsumoto and M. M. Nojiri, *Unitarity and higher order corrections in neutralino dark matter annihilation into two photons*, *Phys.Rev.* **D67** (2003) 075014 [hep-ph/0212022].

[6] J. Hisano, S. Matsumoto and M. M. Nojiri, *Explosive dark matter annihilation*, *Phys.Rev.Lett.* **92** (2004) 031303 [hep-ph/0307216].

[7] J. Hisano, S. Matsumoto, M. M. Nojiri and O. Saito, *Non-perturbative effect on dark matter annihilation and gamma ray signature from galactic center*, *Phys.Rev.* **D71** (2005) 063528 [hep-ph/0412403].

[8] J. Hisano, S. Matsumoto, M. Nagai, O. Saito and M. Senami, *Non-perturbative effect on thermal relic abundance of dark matter*, *Phys.Lett.* **B646** (2007) 34 [hep-ph/0610249].

[9] A. Sommerfeld, *über die beugung und bremsung der elektronen*, *Annalen der Physik* **403** 257.

[10] A. D. Sakharov, *Interaction of an Electron and Positron in Pair Production*, *Zh. Eksp. Teor. Fiz.* **18** (1948) 631.

[11] A. Freitas, *Radiative corrections to co-annihilation processes*, *Phys. Lett.* **B652** (2007) 280 [0705.4027].

[12] M. Cirelli, A. Strumia and M. Tamburini, *Cosmology and Astrophysics of Minimal Dark Matter*, *Nucl.Phys.* **B787** (2007) 152 [0706.4071].

[13] C. F. Berger, L. Covi, S. Kraml and F. Palorini, *The Number density of a charged relic*, *JCAP* **0810** (2008) 005 [0807.0211].

[14] M. Drees, J. M. Kim and K. I. Nagao, *Potentially Large One-loop Corrections to WIMP Annihilation*, *Phys. Rev.* **D81** (2010) 105004 [0911.3795].

[15] A. Hryczuk, R. Iengo and P. Ullio, *Relic densities including Sommerfeld enhancements in the MSSM*, *JHEP* **03** (2011) 069 [1010.2172].

[16] A. Hryczuk, *The Sommerfeld enhancement for scalar particles and application to sfermion co-annihilation regions*, *Phys.Lett.* **B699** (2011) 271 [1102.4295].

[17] A. Hryczuk and R. Iengo, *The one-loop and Sommerfeld electroweak corrections to the Wino dark matter annihilation*, *JHEP* **01** (2012) 163 [1111.2916].

[18] M. Beneke, C. Hellmann and P. Ruiz-Femenia, *Non-relativistic pair annihilation of nearly mass degenerate neutralinos and charginos I. General framework and S-wave annihilation*, *JHEP* **03** (2013) 148 [1210.7928].

[19] K. Harigaya, K. Kaneta and S. Matsumoto, *Gaugino coannihilations*, *Phys. Rev.* **D89** (2014) 115021 [1403.0715].

[20] M. Beneke, C. Hellmann and P. Ruiz-Femenia, *Heavy neutralino relic abundance with Sommerfeld enhancements - a study of pMSSM scenarios*, *JHEP* **03** (2015) 162 [1411.6930].

[21] M. Beneke, A. Bharucha, F. Dighera, C. Hellmann, A. Hryczuk, S. Recksiegel et al., *Relic density of wino-like dark matter in the MSSM*, *JHEP* **03** (2016) 119 [1601.04718].

[22] M. Beneke, A. Bharucha, A. Hryczuk, S. Recksiegel and P. Ruiz-Femenia, *The last refuge of mixed wino-Higgsino dark matter*, *JHEP* **01** (2017) 002 [1611.00804].

[23] S. El Hedri, A. Kaminska and M. de Vries, *A Sommerfeld Toolbox for Colored Dark Sectors*, *Eur. Phys. J.* **C77** (2017) 622 [1612.02825].

[24] T. R. Slatyer, *The Sommerfeld enhancement for dark matter with an excited state*, *JCAP* **1002** (2010) 028 [0910.5713].

[25] M. Beneke, C. Hellmann and P. Ruiz-Femenia, *Non-relativistic pair annihilation of nearly mass degenerate neutralinos and charginos III. Computation of the Sommerfeld enhancements*, *JHEP* **05** (2015) 115 [1411.6924].

[26] K. Blum, R. Sato and T. R. Slatyer, *Self-consistent Calculation of the Sommerfeld Enhancement*, *JCAP* **1606** (2016) 021 [1603.01383].

[27] E. Braaten, E. Johnson and H. Zhang, *Zero-range effective field theory for resonant wino dark matter. Part I. Framework*, *JHEP* **11** (2017) 108 [1706.02253].

[28] E. Braaten, E. Johnson and H. Zhang, *Zero-range effective field theory for resonant wino dark matter. Part II. Coulomb resummation*, *JHEP* **02** (2018) 150 [1708.07155].

[29] E. Braaten, E. Johnson and H. Zhang, *Zero-range effective field theory for resonant wino dark matter. Part III. Annihilation effects*, *JHEP* **05** (2018) 062 [1712.07142].

[30] Y.-L. Tang and G.-L. Zhou, *Calculations of the Sommerfeld Effect in a Unified Wave Function Framework*, *Phys. Rev.* **D99** (2019) 036016 [1806.10124].

[31] M. Beneke, A. Broggio, C. Hasner and M. Vollmann, *Energetic γ-rays from TeV scale dark matter annihilation resummed*, *Phys. Lett.* **B786** (2018) 347 [1805.07367].

[32] M. Beneke, A. Broggio, C. Hasner, K. Urban and M. Vollmann, *Resummed photon spectrum from dark matter annihilation for intermediate and narrow energy resolution*, *JHEP* **08** (2019) 103 [1903.08702].

[33] M. Beneke, C. Hasner, K. Urban and M. Vollmann, *Precise yield of high-energy photons from Higgsino dark matter annihilation*, 1912.02034.

[34] B. von Harling and K. Petraki, *Bound-state formation for thermal relic dark matter and unitarity*, *JCAP* **12** (2014) 033 [1407.7874].

[35] A. Mitridate, M. Redi, J. Smirnov and A. Strumia, *Cosmological Implications of Dark Matter Bound States*, *JCAP* **1705** (2017) 006 [1702.01141].

[36] J. Ellis, J. L. Evans, F. Luo and K. A. Olive, *Scenarios for Gluino Coannihilation*, *JHEP* **02** (2016) 071 [1510.03498].

[37] S. P. Liew and F. Luo, *Effects of QCD bound states on dark matter relic abundance*, *JHEP* **02** (2017) 091 [1611.08133].

[38] H. Fukuda, F. Luo and S. Shirai, *How Heavy can Neutralino Dark Matter be?*, *JHEP* **04** (2019) 107 [1812.02066].

[39] J. Harz and K. Petraki, *Radiative bound-state formation in unbroken perturbative non-Abelian theories and implications for dark matter*, *JHEP* **07** (2018) 096 [1805.01200].

[40] J. Harz and K. Petraki, *Higgs Enhancement for the Dark Matter Relic Density*, *Phys. Rev.* **D97** (2018) 075041 [1711.03552].

[41] J. Harz and K. Petraki, *Higgs-mediated bound states in dark-matter models*, *JHEP* **04** (2019) 130 [1901.10030].

[42] S. Tulin and H.-B. Yu, *Dark Matter Self-interactions and Small Scale Structure*, *Phys. Rept.* **730** (2018) 1 [1705.02358].

[43] M. R. Buckley and P. J. Fox, *Dark Matter Self-Interactions and Light Force Carriers*, *Phys.Rev.* **D81** (2010) 083522 [0911.3898].

[44] 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 Łambda CDM cosmology?*, *Phys.Rev.Lett.* **109** (2012) 231301 [1205.5809].

[45] S. Tulin, H.-B. Yu and K. M. Zurek, *Beyond Collisionless Dark Matter: Particle Physics Dynamics for Dark Matter Halo Structure*, *Phys.Rev.* **D87** (2013) 115007 [1302.3898].

[46] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg and P. Walia, *Strong constraints on self-interacting dark matter with light mediators*, 1612.00845.

[47] H. An, M. B. Wise and Y. Zhang, *Effects of Bound States on Dark Matter Annihilation*, *Phys. Rev.* **D93** (2016) 115020 [1604.01776].

[48] I. Baldes and K. Petraki, *Asymmetric thermal-relic dark matter: Sommerfeld-enhanced freeze-out, annihilation signals and unitarity bounds*, *JCAP* **1709** (2017) 028 [1703.00478].

[49] I. Baldes, M. Cirelli, P. Panci, K. Petraki, F. Sala and M. Taoso, *Asymmetric dark matter: residual annihilations and self-interactions*, *SciPost Phys.* **4** (2018) 041 [1712.07489].

[50] A. Kamada, K. Kaneta, K. Yanagi and H.-B. Yu, *Self-interacting dark matter and muon $g - 2$ in a gauged $U(1)_{L_\mu-L_\tau}$ model*, *JHEP* **06** (2018) 117 [1805.00651].

[51] A. Kamada, M. Yamada and T. T. Yanagida, *Self-interacting dark matter with a vector mediator: kinetic mixing with the $U(1)_{(B-L)_3}$ gauge boson*, *JHEP* **03** (2019) 021 [1811.02567].

[52] S. Matsumoto, Y.-L. S. Tsai and P.-Y. Tseng, *Light Fermionic WIMP Dark Matter with Light Scalar Mediator*, *JHEP* **07** (2019) 050 [1811.03292].

[53] A. Kamada, M. Yamada and T. T. Yanagida, *Unification for the Darkly Charged Dark Matter*, 1908.00207.

[54] P. Ko, T. Matsui and Y.-L. Tang, *Dark Matter Bound State Formation in Fermionic Z_2 DM model with Light Dark Photon and Dark Higgs Boson*, 1910.04311.

[55] A. Kamada, M. Kaplinghat, A. B. Pace and H.-B. Yu, *How the Self-Interacting Dark Matter Model Explains the Diverse Galactic Rotation Curves*, 1611.02716.

[56] M. Kaplinghat, T. Ren and H.-B. Yu, *Dark Matter Cores and Cusps in Spiral Galaxies and their Explanations*, 1911.00544.

[57] T. Binder, M. Gustafsson, A. Kamada, S. M. R. Sandner and M. Wiesner, *Reannihilation of self-interacting dark matter*, *Phys. Rev.* **D97** (2018) 123004 [1712.01246].

[58] T. Bringmann, F. Kahlhoefer, K. Schmidt-Hoberg and P. Walia, *Converting nonrelativistic dark matter to radiation*, *Phys. Rev.* **D98** (2018) 023543 [1803.03644].

[59] T. Binder, K. Mukaida and K. Petraki, *Rapid bound-state formation of Dark Matter in the Early Universe*, 1910.11288.

[60] T. Kinoshita, *Mass singularities of Feynman amplitudes*, *J. Math. Phys.* **3** (1962) 650.

[61] T. D. Lee and M. Nauenberg, *Degenerate Systems and Mass Singularities*, *Phys. Rev.* **133** (1964) B1549.

[62] E. Braaten and R. D. Pisarski, *Soft Amplitudes in Hot Gauge Theories: A General Analysis*, *Nucl. Phys.* **B337** (1990) 569.

[63] A. Pineda and J. Soto, *Effective field theory for ultrasoft momenta in NRQCD and NRQED*, *Nucl. Phys. Proc. Suppl.* **64** (1998) 428 [hep-ph/9707481].

[64] R. Kubo, *Statistical mechanical theory of irreversible processes. 1. General theory and simple applications in magnetic and conduction problems*, *J. Phys. Soc. Jap.* **12** (1957) 570.

[65] P. C. Martin and J. S. Schwinger, *Theory of many particle systems. 1.*, *Phys. Rev.* **115** (1959) 1342.

[66] T. Binder, L. Covi and K. Mukaida, *Dark Matter Sommerfeld-enhanced annihilation and Bound-state decay at finite temperature*, *Phys. Rev.* **D98** (2018) 115023 [1808.06472].

[67] K. Petraki, M. Postma and M. Wiechers, *Dark-matter bound states from Feynman diagrams*, *JHEP* **06** (2015) 128 [1505.00109].

[68] K. Petraki, M. Postma and J. de Vries, *Radiative bound-state-formation cross-sections for dark matter interacting via a Yukawa potential*, *JHEP* **04** (2017) 077 [1611.01394].

[69] J. Pirenne, *The proper field and the interaction of dirac particles. i-iii*, *Arch. Sci. phys. nat* **28** (1946) 233.

[70] J. A. Wheeler, *Polyelectrons*, *Annals of the New York Academy of Sciences* **48** (1946) 219 [https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1946.tb31764.x].

[71] A. Ore and J. L. Powell, *Three-Photon Annihilation of an Electron-Positron Pair*, *Physical Review* **75** (1949) 1696.

[72] S. M. Bilenky, V. H. Nguyen, L. L. Nemenov and F. G. Tkebuchava, *Production and decay of (muon-plus muon-minus)-atoms*, *Yad. Fiz.* **10** (1969) 812.

[73] T. Bringmann, J. Edsjö, P. Gondolo, P. Ullio and L. Bergström, *DarkSUSY 6 : An Advanced Tool to Compute Dark Matter Properties Numerically*, *JCAP* **1807** (2018) 033 [1802.03399].

[74] X. Yao and T. Mehen, *Quarkonium in-medium transport equation derived from first principles*, *Phys. Rev.* **D99** (2019) 096028 [1811.07027].

[75] X. Yao and B. Müller, *Quarkonium inside the quark-gluon plasma: Diffusion, dissociation, recombination, and energy loss*, *Phys. Rev.* **D100** (2019) 014008 [1811.09644].

[76] S. Biondini and M. Laine, *Re-derived overclosure bound for the inert doublet model*, *JHEP* **08** (2017) 047 [1706.01894].

[77] S. Biondini, *Bound-state effects for dark matter with Higgs-like mediators*, *JHEP* **06** (2018) 104 [1805.00353].

[78] S. Biondini and M. Laine, *Thermal dark matter co-annihilating with a strongly interacting scalar*, *JHEP* **04** (2018) 072 [1801.05821].

[79] S. Biondini and S. Vogl, *Coloured coannihilations: Dark matter phenomenology meets non-relativistic EFTs*, *JHEP* **02** (2019) 016 [1811.02581].

[80] S. Kim and M. Laine, *Studies of a thermally averaged p-wave Sommerfeld factor*, 1904.07882.

[81] S. Biondini and S. Vogl, *Scalar dark matter coannihilating with a coloured fermion*, 1907.05766.

[82] M. Laine, O. Philipsen, P. Romatschke and M. Tassler, *Real-time static potential in hot QCD*, *JHEP* **03** (2007) 054 [hep-ph/0611300].

[83] N. Brambilla, J. Ghiglieri, A. Vairo and P. Petreczky, *Static quark-antiquark pairs at finite temperature*, *Phys. Rev.* **D78** (2008) 014017 [0804.0993].

[84] S. Kim and M. Laine, *On thermal corrections to near-threshold annihilation*, *JCAP* **1701** (2017) 013 [1609.00474].

[85] S. Kim and M. Laine, *Rapid thermal co-annihilation through bound states in QCD*, *JHEP* **07** (2016) 143 [1602.08105].

[86] CMS collaboration, *Measurement of nuclear modification factors of $\Upsilon(1S)$, $\Upsilon(2S)$, and $\Upsilon(3S)$ mesons in PbPb collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV*, *Phys. Lett.* **B790** (2019) 270 [1805.09215].

[87] S. Biondini, S. Kim and M. Laine, *Non-relativistic susceptibility and a dark matter application*, 1908.07541.

[88] J. Smirnov and J. F. Beacom, *TeV-Scale Thermal WIMPs: Unitarity and its Consequences*, 1904.11503.