

Self-Interacting Dark Matter in Cosmology: accurate numerical implementation and observational constraints

Rafael Yunis^{a,b} Carlos R. Argüelles^{a,c,d} Claudia G. Scóccola^{c,d} Diana López Nacir^{e,f} Gastón Giordano^{g,h}

^aICRANet, Piazza della Repubblica 10, I-65122 Pescara, Italy

^bPhysics Department, La Sapienza University of Rome, P.le Aldo Moro 5, I-00185 Rome, Italy

^cFacultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata, Paseo del Bosque, B1900FWA LaPlata, Argentina

^dCONICET (Consejo Nacional de Investigaciones Científicas y Técnicas), Argentina

^eDepartamento de Física Juan José Giambiagi, FCEyN UBA

^fIFIBA CONICET-UBA, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón I, 1428 Buenos Aires, Argentina

^gInstituto de Física La Plata IFLP

^hDepartamento de Física FCE-UNLP C.C. 67, 1900 La Plata, Argentina

E-mail: yunis121@gmail.com

Abstract. This paper presents a systematic and accurate treatment of the evolution of cosmological perturbations in self-interacting dark matter models, for particles which decoupled from the primordial plasma while relativistic. We provide a numerical implementation of the Boltzmann hierarchies developed in a previous paper [JCAP, 09 (2020) 041] in a publicly available Boltzmann code and show how it can be applied to realistic DM candidates such as sterile neutrinos either under resonant or non-resonant production mechanisms, and for different field mediators. At difference with traditional fluid approximations - also known as a $c_{\text{eff}} - c_{\text{vis}}$ parametrizations- our approach follows the evolution of phase-space perturbations under elastic DM interactions for a wide range of interaction models, including the effects of late kinetic decoupling. Finally, we analyze the imprints left by different self interacting models on linear structure formation, which can be constrained using Lyman- α forest and satellite counts. We find new lower bounds on the particle mass that are less restrictive than previous constraints.

Contents

1	Introduction	2
1.1	WDM Production in the Early Universe	4
2	Boltzmann Formalism for Self Interactions in WDM	6
2.1	Formal Boltzmann Framework	7
2.2	Relaxation Time Approximation	10
2.3	Evolution of f_0 and Self Interaction Decoupling	12
2.4	Limiting forms for the Relaxation Time	16
3	Numerical Implementation in CLASS	18
3.1	The Tight-Coupling Approximation	19
3.2	Non-Relativistic Decoupling	21
3.3	Effects on the Matter Power Spectrum	23
4	Comparison with Observables	25
4.1	Milky Way Satellite Counts	27
4.2	Lyman- α Constraints	27
4.3	Parameter Space Constraints	28
4.3.1	Exploring the SI-WDM parameter space	28
4.3.2	The ν MSM Parameter Space in the Presence of Self Interactions	29
5	Conclusions	30
A	Power Spectrum Oscillations and Tight Coupling	39
B	Sterile Neutrino WDM Model Subsets	43
C	Fitting Formulae for the Transfer Function Coefficients	44
D	Power Spectra and Parameter Space Constraints for Other Mediator and Background Models	45

1 Introduction

The nature, mass and dynamics of the dark matter (DM) particles are still a mystery that have sparked several extensions to the standard model of particle physics. Most of the evidence of its existence comes from astrophysical and cosmological observations, from which we can search for clues about its behaviour [1, 2]. So far, the most likely explanation for the dark matter phenomenology has some sort of microscopic origin [3–5]. In the standard model of structure formation, DM haloes form on a “bottom-up” hierarchy: small scales collapse first to form virialized objects, later merging and accreting into larger ones [1, 6, 7]. This paradigm is in remarkable agreement with observations at large and small scales, such as the distribution of large-scale structure [8, 9], CMB anisotropies [1] and the internal structure of DM haloes [2]. However, recent observations at small scales have become challenging to be described within the paradigm [10]. Such observations include a number of satellite galaxies in the Milky Way that is far smaller than predicted (the “missing satellite problem”)[11–13]; the lack of observation of the most massive haloes, which are predicted to be luminous (the “too big to fail” problem)[14]; and inconsistencies with the inner structure of dSph galaxies (the “core-cusp” problem) [10, 15, 16], among others.

While these tensions by no means rule out the concordance model, as they can be alleviated within Λ CDM itself, they have raised an increasing amount of interest in alternative DM models which may reduce these tensions in a more natural way. Some of the earliest approaches to these alternative models imply modifying the DM fluid’s properties themselves. The earliest approaches include a DM mixture of CDM and a “hot” version of DM; low mass particles with a velocity dispersion such as to erase small scale structure [17, 18]. Related to this approach, recently there has been growing interest in the field of interacting dark matter: while some of these models propose interactions between CDM and standard model particles, others assume a dark mediator field with low mass, interacting with CDM and contributing to the energy density budget of the universe. These types of models are widely known as “Self Interacting Dark Matter” (SIDM): such self interactions, apart from modifying the cosmological evolution of these particles, affect significantly the clustering of matter in N-body simulations, as collisions can flatten the inner regions of most galaxies in CDM, reconciling these values with observation and alleviating the core-cusp problem [19–21]. Some particle physics models for SIDM may involve mirror world DM (see [22, 23] and references therein), light or heavy mediators through a broken $U(1)$ symmetry [24–26], non Abelian interactions [27] or “dark atoms” [28], to name a few.

Other promising approaches include “Warm Dark Matter”(WDM): these models slightly reduce the preferred mass ranges of DM enough to include a moderate amount of initial velocity dispersion and free streaming, sufficient to erase some of the smallest scale structure [29–31]. If this free streaming length today is smaller than the size of galaxy clusters, it can alleviate the missing satellites and too-big-to-fail problems [32–34]. Traditionally, these particles were also expected to generate DM cores in haloes and solve the core-cusp tension, however in recent years it has been shown that the WDM particle mass required for this to happen in N-body simulations is inconsistent with phase-space constraints [35, 36]. Nevertheless, different approaches to DM halo structure may indeed relax these tensions, in particular the ones considering halo formation from a maximum entropy production principle, which can account for the expected DM distribution and halo-sizes in dSphs, while being in agreement with phase-space bounds [37].

A promising extension to the standard model of particle physics that naturally leads to these types of WDM models can be found in ν MSM (neutrino minimal standard model), which includes intermediate mass sterile neutrinos [31, 38]. For such model, the preferred production mechanism involves active-sterile oscillations in the neutrino fluid, around the time of the quark-hadron transition [39–41]. It is possible to strongly constrain these models using a set of cosmological and astrophysical data. Indeed, a combination of limits from DM production; constraints from the predicted X-Ray radiative decay of sterile neutrinos; and structure formation bounds such as MW subhalo counts and Lyman- α forest severely reduced the allowed region in the parameter space of the models [34, 38, 42–45]. In fact, this last set of observations have, in recent years, almost completely ruled out this particular realization of WDM [34, 44, 45].

These tight constraints may be alleviated by performing simple extensions to this basic sterile neutrino WDM paradigm. In this work, we aim at a particular subset of these extensions, that combines both WDM and SIDM approaches, which we label as “Self Interacting Warm Dark Matter” (SI-WDM). Previous studies have suggested that including self interactions in WDM sterile neutrino may contribute to its production in the early universe [31, 46–48], others have studied the effects of WDM in self interacting N-body simulations [49], while others have explored the ν MSM parameter space with the inclusion of self interactions [48]. Moreover, it has been recently suggested that self interactions in WDM may play a significant role in the evolution of cosmological perturbations. That is, many studies have shown that strongly self interacting light DM (with particle masses ranging from $\mathcal{O}(1 - 100)$ keV) will remain in kinetic equilibrium until late times: this significantly modifies the linear evolution of cosmological perturbations, with respect to non-interacting models [50–53]¹.

Following the footsteps of a previous work [54], where a full theoretical framework was given in order to treat the linear evolution of SI-WDM perturbations, we seek to improve these earlier numerical studies on several aspects and provide a more accurate framework for studying the evolution of SI-WDM, at the same time providing a comprehensive theoretical framework and developing a straightforward numerical application. To this end, we develop an extension of the public Boltzmann solver CLASS v2.7.2 [55], and provide a free distribution link at github.com/yunis121/siwdm-class. We begin with an accurate approach to the Boltzmann collision term in cosmological perturbation theory as in [54], and consider several SI-WDM candidates, interactions and production mechanisms. Then, we properly introduce the necessary approximations for a feasible numerical application, paying special attention to the relaxation time and tight coupling approximations. Regarding the evolution of cosmological perturbations, many authors choose to compute the evolution of self interacting warm species using a fluid-like approximation (also known as a $c_{\text{eff}} - c_{\text{vis}}$ parametrization) either during its entire evolution as in [51, 52], or up to self-interaction decoupling [53]. However, it has been argued in [56, 57] and shown in [58], that those fluid-like approximations, typical of CDM scenarios, provide inaccurate results for the computation of the power spectrum for self-interacting neutrinos and WDM respectively, and several doubts have been raised [28, 56, 59] about its physical meaningfulness.

¹While finalizing this manuscript, the work [53] was presented on arXiv which studies WDM together with self interactions including the effects of kinetic coupling until the non-relativistic transition. Though they implemented approximations such as instantaneous self-interaction decoupling and a partial-fluid description which are superseded by our more self-consistent approach, both works show relaxed bounds when including the self interaction effects.

Thus, we choose here to fully follow the evolution of the perturbations to the background DM distribution function, accounting for the elastic DM interactions self-consistently throughout the evolution of the species and up to today. This implies a more rigorous approach to the evolution of SI-WDM perturbations with respect to previous works on the subject (see above), which either use fluid parametrizations or focus solely on the background evolution. Throughout this work, we present these self-interacting models as minimal extensions to existing WDM candidates (such as the sterile neutrino ν MSM) instead of effective toy models (such as the scalar field DM models [60]). This has the advantage of presenting numerous constraints from various sources to the parameter space, and including our models in a more realistic framework that can account for several other processes such as DM production, baryogenesis and neutrino mass. We leave for a future work the possible effects self interactions may have on these predictions. The evolution of DM perturbations typically leave their strongest imprint in observables such as the CMB, Milky Way satellite counts and the Lyman- α forest [10, 31, 61]. We consider these last two and compare our results with observations, obtaining the strongest constraints on the mass of the DM particles from the Lyman- α forest.

The paper is organized as follows. In the remaining of this introduction, we will explore and discuss several production mechanisms for WDM, as the distribution function resulting from these processes will have significant consequences on the evolution of perturbations and, consequently, on structure formation observables. Afterwards, following from [54], we will present in section 2 the full formalism for SI-WDM in linear theory. We will recall some previous results on the Boltzmann collision terms, explore the relaxation time approximation, discuss about different regimes of DM self decoupling (introducing the non-relativistic self decoupling regime) and provide simple forms for the relevant quantities ideal for numerical implementation. This will be fully explored in section 3, where we give details on the development of a SI-WDM linear cosmology solver based on CLASS, together with relevant approximations. We also include in that section a numerical application of non-relativistic self decoupling and present some initial results on the power spectra obtained. With that in mind, in section 4 we compare the results from this solver with observations. We will focus particularly on the predicted number of MW satellites and the Lyman- α forest observations. We explore the results of these analysis for various initial conditions, and compare the self interacting alternative to the standard ν MSM. Finally, in section 5 we draw out conclusions.

1.1 WDM Production in the Early Universe

Before introducing the assumptions and governing equations of SI-WDM as a whole, it is important to discuss briefly about its production in the early universe. This discussion serves a two-fold purpose. Firstly, a relevant model should allow DM to be sufficiently produced in the early universe as to match its abundance today. Secondly, the production mechanism gives the initial condition for the whole structure formation process, which results in additional constraints on the model [31, 41]. An interesting example that is constrained by the latter but not the former is Hot Dark Matter (HDM) [62], which can be efficiently produced in equilibrium with the hot plasma but their resulting high velocity dispersion does not reproduce the observed DM halo assembly history [6]. Thus, it is important to consider the possible production mechanisms for these models as a first step in their analysis.

There have been recent advancements in exploring some models of WDM production in the presence of self interactions, with varying degrees of success in describing the DM abundance as observed today [46, 47]. We choose to avoid these questions for now, as they exceed the scope of this work, mostly aimed at discerning these effects at the level of linear

theory of perturbations. Thus, along this work we will simply explore a few relevant WDM production mechanisms, assuming that they are not significantly altered by the presence of self interactions (other than in a few obvious aspects such as thermalization of the initial distribution function (DF), see section 3) and observe the consequences at the level of structure formation. We remind the reader that we do not explore whether or not dark matter production is significantly modified by the presence of self interactions, and we aim to address these questions and provide a more in-depth treatment in future works.

Having clarified this, we will consider a few production mechanisms for these models: namely thermal production and production through active-sterile neutrino mixing (both on the resonant and non-resonant scenarios). The first scenario is the most straightforward, DM particles are produced in equilibrium in the hot plasma through interactions with other species and, as their interaction rate falls, they decouple from the hot plasma and evolve independently from the other species. This scenario is very similar to WIMP decoupling, however they decouple while relativistic with a relatively high particle mass, generally seen as incompatible with them being produced at the same time as WIMPS or neutrino HDM, when weak force interactions decouple (see e.g. [61]). Indeed, the abundance today for a particle that decouples while relativistic is approximately [63]

$$\Omega_{i,0}h^2 \approx 7.64 \times 10^{-2} \left[\frac{g_{i,\text{eff}}}{g_{*,s}(T_f)} \right] \left(\frac{m_i}{\text{eV}} \right), \quad (1.1)$$

where $g_{i,\text{eff}} = g_i$ for bosons, $g_{i,\text{eff}} = (3/4)g_i$ for fermions, g_i the spin degeneracy of the species and T_f its freeze-out temperature. This indeed results in overproduction for typical WDM masses of $m \sim \mathcal{O}(\text{keV})$, unless the particles are produced much earlier, when the number of relativistic degrees of freedom g_* is much higher than $\sim 10^3$ [61]. Thus, thermal models are assumed to be produced at extremely high energies on an equilibrium background distribution, with an overall temperature of $(T/T_\nu)^3 = \Omega_{\text{DM}}h^2(94 \text{ eV}/m)$ if we assume these to take up the whole DM budget in the universe [61].

While these thermal models are easily extended to many DM candidates, the other models we will consider are particular for the case of sterile neutrinos. Typical sterile neutrinos are heavier, mostly right handed companions of the active neutrinos, with the lightest of these particles acting as WDM. There exists mixing in these models between active and sterile species, measured by the angle θ , which in the models considered here contributes to both their production and decay. When discussing these sterile neutrino models, we will mostly focus on a particular realization of these, which involves a minimal extension to the standard model of particle physics: νMSM (see e.g. [31, 38] for comprehensive reviews).

Indeed, DM production in these models results from the quantum oscillations between active and sterile neutrinos in the hot plasma. In the cases relevant to sterile neutrino DM, the Boltzmann equation for the quantum-damped, collisionally driven sterile neutrino production is [41]

$$\begin{aligned} \frac{\partial f_{\nu_s}}{\partial t} - H p \frac{\partial f_{\nu_s}}{\partial p} = & \sum_{\nu_\mu + a + \dots \rightarrow i + \dots} \int \frac{d^3 p_a}{(2\pi)^3 2E_a} \dots \frac{d^3 p_i}{(2\pi)^3 2E_i} (2\pi)^4 \delta_D^4(p + p_a + \dots - p_i - \dots) \\ & \times \frac{1}{2} \left[\langle P_m(\nu_\mu \rightarrow \nu_s) \rangle (1 - f_{\nu_s}) \sum |\mathcal{M}|_{i+\dots \rightarrow a+\nu_\mu+\dots}^2 f_i \dots (1 \pm f_a)(1 - f_{\nu_\mu}) \dots \right. \\ & \left. - \langle P_m(\nu_s \rightarrow \nu_\mu) \rangle f_{\nu_s} (1 - f_{\nu_\mu}) \sum |\mathcal{M}|_{\mu_\nu+a+\dots \rightarrow i+\dots}^2 f_a \dots (1 \pm f_i) \dots \right]. \end{aligned} \quad (1.2)$$

The r.h.s. of this equation sums up all interactions that produce or consume, in this case, a muon neutrino, with the respective Bose enhancement and Pauli blocking terms $(1 \pm f)$. The P_m are the active-sterile oscillation probabilities in matter modified by interactions with the medium, which are parametrized by neutrino self energy and quantum damping. The oscillation probability P_m then reads

$$\langle P_m(\nu_\mu \leftrightarrow \nu_s; p, t) \rangle = \frac{1}{2} \Delta^2(p) \sin^2 2\theta \left\{ \Delta^2(p) \sin^2 2\theta + D^2(p) + \left[\Delta(p) \cos 2\theta - V^L - V^{\text{th}}(p) \right]^2 \right\} \quad (1.3)$$

where $\Delta(p)$ is the vacuum oscillation rate $\Delta(p) \equiv \delta m_{\nu_\mu, \nu_s} / 2p$, $\delta m_{\nu_\mu, \nu_s}$ is the mass difference between active and sterile neutrino, the quantity $D(p)$ is the quantum damping rate (half the interaction rate of active neutrinos) and the neutrino self energy is split into the lepton asymmetry potential V^L and the thermal potential V^{th} . More information on these potentials, as well as detailed calculations can be found in [41].

Two distinct regimes appear in this case: first, under no initial lepton asymmetry the potential V^L becomes zero. This situation, as originally proposed by [64], is known as non-resonant production. In this case, the resulting distribution function of sterile neutrinos roughly follows the one for active ones, but normalized depending on the mixing angle θ . Thus, in this last production scenario, only a small window of mixing angles can provide the correct DM abundance today.

However, in the presence of a nonzero lepton asymmetry the potential V^L enhances the production rate, allowing for smaller mixing angles and, at the level of the distribution function, resulting in a nonequilibrium function which is generally colder [39, 40]. This mechanism, in the presence of a significant lepton asymmetry, is the scenario known as resonant production. The resulting background distribution functions depend on the mixing angle, particle mass and lepton asymmetry values; and their shapes are typically computed numerically (see [41] for a typical application, together with a public numerical tool).

It is important to remark here that the form of the Boltzmann equation (1.2), particularly the interaction terms, are the ones responsible for DM production and most relevant in the early radiation dominated era. Later in this work, we will use a different form of the Boltzmann equation to calculate the evolution of perturbations (equations (2.2) and (2.16)), in which we neglect this production term and only consider elastic DM collisions. This is due to the fact that the perturbations typically evolve later in cosmological history. However, the details of DM production still affect the evolution of perturbations as they act as initial conditions for the background distribution function.

2 Boltzmann Formalism for Self Interactions in WDM

In a previous paper [54], we introduced a full framework to study the effect of self interactions in WDM in Cosmological Perturbation Theory and its impact on the linear power spectrum. In particular, we considered an extension to [57] where the full collision integrals are calculated for a scalar mediator and effectively massless particles, and extended it for massive particles, maintaining a certain degree of generality in the mediator model.

While the formalism we develop is fairly general, a key task of the present work is to calculate the effects on cosmological observables under a certain scenario, described by the following *hypotheses*:

- We consider a typical (fermionic) WDM particle of mass $\mathcal{O}(1 - 50)$ keV, that is produced and also decouples from the plasma while being relativistic, and becomes non-relativistic in the radiation dominated era (see, for example, [31]). In any case, the formalism itself makes no assumptions on particle mass, so limiting forms (either relativistic or non-relativistic) can provide accurate treatment for light and heavy relics, provided the other assumptions here listed are met.
- We assume that, after production/decoupling from the plasma, the only form of interaction is an elastic self interaction. We further assume that these effects can be modeled by a tree-level scattering process following Feynmann rules of e.g. [65, 66] (Note: the amplitudes presented here apply only for *Majorana* fermions)².
- We neglect the cosmological population of mediator particles, and avoid calculating the evolution of the mediator fields themselves. This is a justified assumption on massive mediator scenarios: the initial population of mediators is assumed to have already decayed at $T_\gamma < m_{\text{med}}$ (the mediator mass), and further production is kinematically suppressed. The case for massless mediators has been briefly discussed in [54] and thoroughly discussed in [56] for light DM.
- In order to model these self interactions, we work under a specific ansatz for the scattering amplitude:

$$|\mathcal{M}|^2 = A_t(s)t^2 + B_t(s)t + C_t(s) = A_u(t)u^2 + B_u(t)u + C_u(t) \quad (2.1)$$

where A_t , B_t and C_t are simple functions of s and A_u , B_u and C_u are simple functions of t , where s, t, u are the Mandelstam variables, and the two expressions are related by $s + t + u = 4m^2$. This ansatz includes various massive mediator cases, as well as a few others, and we explicitly presented results in [54] for three different mediator models: Constant Amplitude, Massive Scalar and Massive Vector Field.

In this section, we will provide a brief summary of the results in [54] and provide some new concepts. First, we will summarize the full framework for studying Boltzmann Hierarchies in the scenarios we introduce in that previous work. Then, we present the Relaxation Time Approximation used to construct Boltzmann Hierarchies and specialize the results in [54] to the specific models we consider in section 1. Finally, we will study the evolution of the background distribution function f_0 under these models and provide here a self-consistent study on the effects of self interaction decoupling, superseding (most) of the previous results in the literature.

2.1 Formal Boltzmann Framework

Following the derivations in [54], the linear Boltzmann hierarchy for SI-WDM is given by:

²A notable example where this is not the case is in light mediator models: as the DM particle becomes non-relativistic, scattering increasingly relies on the effects of point-charge potentials. In terms of Feynmann diagrams, this means that events where many mediator particles are exchanged in each process become increasingly relevant [27, 67].

$$\begin{aligned}
\dot{F}_0(k, E_q, \tau) &= -\frac{qk}{E_q} F_1(k, E_q, \tau) + \frac{\dot{h}}{6} \frac{\partial f_0}{\partial \ln q} \\
&\quad - G_0 a F_0(k, E_q, \tau) \Gamma(E_q, \tau) + G_0 a \int dE_l F_0(k, E_l, \tau) \mathcal{K}_0^{(1)}(E_q, E_l, \tau) \\
\dot{F}_1(k, E_q, \tau) &= \frac{qk}{3E_q} F_0(k, E_q, \tau) - \frac{2qk}{3E_q} F_2(k, E_q, \tau) \\
&\quad - G_0 a F_1(k, E_q, \tau) \Gamma(E_q, \tau) + G_0 a \int dE_l F_1(k, E_l, \tau) \mathcal{K}_1^{(1)}(E_q, E_l, \tau) \\
\dot{F}_2(k, E_q, \tau) &= \frac{qk}{5E_q} \left[2F_1(k, E_q, \tau) - 3F_3(k, E_q, \tau) \right] - \frac{\partial f_0}{\partial \ln q} \left[\frac{1}{15} \dot{h} + \frac{2}{5} \dot{\eta} \right] \\
&\quad - G_0 a F_2(k, E_q, \tau) \Gamma(E_q, \tau) + G_0 a \int dE_l F_2(k, E_l, \tau) \mathcal{K}_2^{(1)}(E_q, E_l, \tau) \\
\dot{F}_l(k, E_q, \tau) &= \frac{qk}{(2l+1)E_q} \left[lF_{(l-1)}(k, E_q, \tau) - (l+1)F_{(l+1)}(k, E_q, \tau) \right] \\
&\quad - G_0 a F_l(k, E_q, \tau) \Gamma(E_q, \tau) + G_0 a \int dE_l F_l(k, E_l, \tau) \mathcal{K}_l^{(1)}(E_q, E_l, \tau) \quad , \quad l \geq 3
\end{aligned} \tag{2.2}$$

in the synchronous gauge, defined by the line element $ds^2 = a^2(\tau) \{-d\tau^2 + (\delta_{ij} + h_{ij})dx^i dx^j\}$, where the scalar mode of the perturbation h_{ij} can be described in terms of two fields $h(\vec{k}, \tau)$ and $\eta(\vec{k}, \tau)$, the trace and traceless parts of the h_{ij} perturbation in Fourier space, respectively. An overdot denotes derivation with respect to comoving time τ , and moments $\vec{q} = a\vec{p}$ are comoving proper momenta, with \vec{p} the proper momentum measured by an observer in a fixed spatial coordinate, and $E_q = \sqrt{m^2 + q^2/a^2}$, with m the DM mass.

The various kernel moments are defined as:

$$\Gamma(E_q, \tau) = \frac{1}{E_q q} \int dE_l ds f_0(E_l) \chi(s) , \tag{2.3}$$

$$\mathcal{K}_l^{(1)}(E_q, E_l, \tau) = -\chi_l(E_q, E_l) f_0(E_q) + 2 \frac{1}{E_q q} K_l(E_q, E_l, \tau) , \tag{2.4}$$

with $G_0 = 1/[4(2\pi)^3]$, s referring to the Mandelstam variable $s \equiv (\mathbf{q} + \mathbf{l})^2$ and, for the purposes of the integration kernels Γ, \mathcal{K} the momenta q is *not* comoving: $q \equiv \sqrt{E_q^2 - m^2}$. Integration over these terms (and in all definitions) runs over the whole range of the variable when integration limits are not specified. The l -th moment of the perturbed DF, F_l , is defined as in

$$\begin{aligned}
F(|k|, |q|, \cos \epsilon) &= \sum_{l=0}^{\infty} (-i)^l (2l+1) F_l(|k|, |q|) P_l(\cos \epsilon) , \\
F_l(|k|, |q|) &= \frac{i^l}{2} \int_{-1}^1 d \cos \epsilon F(|k|, |q|, \cos \epsilon) P_l(\cos \epsilon) ,
\end{aligned} \tag{2.5}$$

and K_l, χ_l the Legendre transforms of the K, χ kernel functions are defined as in

$$\mathcal{K}_l(E_q, E_l, \tau) \equiv \int ds \mathcal{K}(E_q, E_l, s, \tau) P_l(\cos \theta(s)) . \tag{2.6}$$

The K, χ functions, the proper kernel functions, encode the behaviour of the particular mediator model and are defined in terms of the coefficients (2.1) as:

$$\chi(s) = \frac{1}{3}A_t(s - 4m^2)^2 + \frac{1}{2}B_t(s - 4m^2) + C_t, \quad (2.7)$$

$$\begin{aligned} K(E_q, E_{q'}, t, \tau) = & \left\{ \frac{A_u}{8|\vec{q} - \vec{q}'|^5} \left\{ \langle f_0 \rangle_2 \left[4t(3(E_q + E_{q'})^2 t - ((E_q - E_{q'})^2 - t)(-4m^2 + t)) \right] \right. \right. \\ & + \langle f_0 \rangle_1 \left[4t(4(E_q - E_{q'})^2(E_q + 3E_{q'})m^2 \right. \\ & \quad \left. \left. - 4(E_q(E_q - E_{q'})(E_q + 2E_{q'}) + (E_q + 3E_{q'})m^2)t + (E_q + 3E_{q'})t^2 \right] \right. \\ & + \langle f_0 \rangle_0 \left[(48(E_q - E_{q'})^4 m^4 - 16(E_q - E_{q'})^2 m^2(2E_q^2 - 3E_q E_{q'} + 6m^2)t \right. \\ & \quad \left. + 8(E_q^2(E_q - E_{q'})^2 + (7E_q^2 - 12E_q E_{q'} + 3E_{q'}^2)m^2 + 6m^4)t^2 \right. \\ & \quad \left. \left. - 4(2E_q^2 - 3E_q E_{q'} + 6m^2)t^3 + 3t^4 \right] \right\} \\ & + \frac{B_u}{2|\vec{q} - \vec{q}'|^3} \left\{ \langle f_0 \rangle_1 [t(E_q + E_{q'})] + \langle f_0 \rangle_0 [2(E_q - E_{q'})^2 m^2 + 2E_q(-E_q + E_{q'})t - 4m^2 t + t^2] \right\} \\ & + \frac{C_u}{|\vec{q} - \vec{q}'|} \langle f_0 \rangle_0 \left. \right\} \end{aligned} \quad (2.8)$$

with the background distribution moments $\langle f_0 \rangle_i$ defined as:

$$\langle f_0 \rangle_n(E_q, E_{q'}, t, \tau) = \int_{R_2}^{\infty} dE_{l'} f_0(E_{l'}, \tau) E_{l'}^n, \quad (2.9)$$

which are functions of $(E_q, E_{q'}, t)$ only through R_2 , defined as

$$R_2 = \frac{1}{2} \left\{ E_q - E_{q'} - |\vec{q} - \vec{q}'| \sqrt{1 - \frac{4m^2}{t}} \right\}.$$

The Boltzmann Hierarchy (2.2) provides an accurate, relatively model-independent approach to the collisional Boltzmann hierarchies in the DM cosmological scenario above described. Once the interaction model amplitude is specified in equation (2.1), these Boltzmann hierarchies can be solved coupled to the Einstein equations in order to provide an accurate description of SI-WDM. While the expressions are lengthy, most of these model-dependent kernels have forms relatively straightforward to evaluate and all these kernels can be pre-computed, once the evolution of f_0 is known. The evolution of the background distribution f_0 due to elastic self interactions is also given in [54], and we refer the reader to section 3.2 there for a full expression of the collision terms.

Still, in order to provide a better understanding of these collision terms and their broad effects on the cosmological evolution, it is always useful to provide some more straightforward approximation to the collision terms in these hierarchies. In the rest of this section we will review how to bring this exact operator into a simpler form, provide approximate Boltzmann hierarchies (yet portraying accurately the general behaviour of elastic collisions) and use this formalism to better comprehend the cosmological evolution of the SI-WDM component.

This approach of providing approximate forms for the collision terms in the Boltzmann equation comes into contrast with previous works on the evolution of SI-WDM, which choose to avoid calculating perturbations to the distribution function entirely by relying on a fluid approximation of the energy-momentum tensor [51, 52, 68]. However, it has been shown in [56, 57] to not accurately portray the evolution of species that decouple while relativistic and methods based on the full Boltzmann hierarchies were favored.

2.2 Relaxation Time Approximation

The evolution of the collisional DM fluid can be approximated in a simpler way by noting the general behavior of the collisions in the phase space density $f(\vec{k}, \vec{q}, \tau)$ as a whole. Statistical systems evolving through the Boltzmann equation with a significant collision term tend to relax to their local equilibrium distributions due to entropy maximization: these equilibrium forms carry the particularity of nullifying the Liouville operator, the collision term and it can be shown to satisfy the detailed balance condition $\ln[f_{eq}(\mathbf{q})] + \ln[f_{eq}(\mathbf{l})] = \ln[f_{eq}(\mathbf{q}')] + \ln[f_{eq}(\mathbf{l}')] + \ln[f_{eq}(\mathbf{l}')] + \ln[f_{eq}(\mathbf{l})]$ with the 4-momenta $\mathbf{q}, \mathbf{l}, \mathbf{q}', \mathbf{l}'$ satisfying conservation of total 4-momentum, where the primed momenta correspond to the outgoing particles in an interaction.

In general, it can be shown [69] that in a FLRW expanding universe there is no general equilibrium distribution function, i.e., there is no function f_{eq} that satisfies $\mathcal{L}(f_{eq}) = 0$ at all times under this metric. However, solutions can be found in two limiting cases:

$$f_{eq}(p, t) = \begin{cases} f_{rel}(p, t) \propto e^{-\beta pa(t)}, & \text{if } p \gg m \\ f_{nrel}(p, t) \propto e^{-\beta \frac{p^2}{2m} a^2(t)}, & \text{if } p \ll m \end{cases} \quad (2.10)$$

with β a constant, usually interpreted as the inverse of the temperature today and p the physical momentum, and we have suppressed quantum effects in f_{eq} for simplicity, see [70] for a full expression. We will refer to f_{rel} as the *relativistic* equilibrium DF and to f_{nrel} as the *non-relativistic* equilibrium DF.

Taking into consideration this behavior, we can construct a much simpler operator for the collision term that recreates the expected behavior for small departures from equilibrium:

$$\left(\frac{\partial f}{\partial t}\right)_{col}^{(1)} = -\frac{f(\vec{x}, \vec{p}, \tau) - f_{eq}(\vec{p}, \tau)}{\tau_{rel}(q)} \quad (2.11)$$

where τ_{rel} is a quantity known as *relaxation time* and f_{eq} is a local equilibrium distribution function - in the context of cosmology, we will specify this local equilibrium function to be the background DF, f_0 , that describes the isotropic evolution.

This approximate form is known as the *Relaxation Time Approximation* (RTA). In general, this operator will qualitatively “erase” all perturbations to f_0 bringing the system back into equilibrium (even erasing collisional invariants such as number and energy densities). The quantity τ_{rel} governs the time scale in which these perturbations relax back into equilibrium and, in general, the rate of change of the distribution function as a whole. This approximate form can be obtained from the full collision term by assuming that the perturbation to the DF is highly localized in momenta or, equivalently, forcing the evolution of the perturbations to be local in momentum space.

It has been shown in a previous paper [54] that this quantity can be expressed in terms of the collision kernel χ as:

$$\tau_{rel}(q)^{-1} = -\frac{\mathcal{D}_2[f_{eq}](q, \tau)}{f_{eq}(q, \tau)} = \frac{g_i^3}{32(2\pi)^3 E_{qQ}} \int dE_l ds f_{eq}(E_l, \tau) \chi(s) \quad (2.12)$$

where f_{eq} is now an equilibrium DF, g_i is the number of spin degrees of freedom of the DM particle and, in the last equality, momenta and energies are defined as in (2.2). In general this expression is momentum dependent, but it is possible to further simplify this approach by constructing a momentum independent thermal average of the relaxation time $\langle \tau_{rel}(q) \rangle_{th} \equiv \tau_{rel} = \int d^3q \tau_{rel}(q) f_{eq}(q, \tau) / \int d^3q f_{eq}(q, \tau)$. Thus, the thermal averaged relaxation time (hereby, just relaxation time) can be expressed in terms of collision kernels as:

$$\tau_{rel}^{-1} = \frac{g_i^3}{32(2\pi)^3} \frac{\int dE_q dE_l ds f_{eq}(E_q, \tau) f_{eq}(E_l, \tau) \chi(s)}{\int dE_q q E_q f_{eq}(E_q, \tau)}, \quad (2.13)$$

assuming the azimuthal angle ψ between \vec{q} and \vec{l} has already been averaged out (see [56]).

At this stage, one can draw a parallelism between this kernel function χ and the invariant rate defined in [71], as $R \propto \chi$. Also, through the invariant flux $F = \frac{1}{2} \sqrt{s(s-4m^2)}$ ³, we find that $\sigma F = 4E_q E_l \sigma v_{mol} = \frac{g_i^2}{4} \chi$ with σ the total cross section of the interaction and v_{mol} the Moller velocity. This allows us to put this relaxation time in terms of more well-known expressions as:

$$\Gamma \equiv \tau_{rel}^{-1} = n \langle \sigma v_{mol} \rangle = n \frac{\int d^3q d^3l \sigma v_{mol} f_{eq}(q, \tau) f_{eq}(l, \tau)}{\int d^3q d^3l f_{eq}(q, \tau) f_{eq}(l, \tau)} \quad (2.14)$$

where Γ is the interaction rate, n is the DM number density and, to keep the same notation/normalization as in [72, 73], we use $\sigma v_{mol} = \frac{g_i^2/4}{4E_q E_l} \chi$. This form allows us to compare with their results and to condense the overall normalization of the distribution function in n^4 .

With this simpler form for the relaxation time, constructing a Boltzmann Hierarchy becomes straightforward: in terms of the perturbations to the cosmological background DF, the approximate form (2.11) becomes

$$\left(\frac{\partial f}{\partial t} \right)_{col,l}^{(1)} \approx - \frac{F_l(k, q, \tau)}{\tau_{rel}} \quad (2.15)$$

where we have used the momentum averaged relaxation time τ_{rel} and, by making use of this, the full Boltzmann Legendre expansion is relatively simple:

$$\begin{aligned} \dot{F}_0(k, E_q, \tau) &\simeq - \frac{qk}{E_q} F_1(k, E_q, \tau) + \frac{\dot{h}}{6} \frac{\partial f_0}{\partial \ln q} \\ \dot{F}_1(k, E_q, \tau) &\simeq \frac{qk}{3E_q} F_0(k, E_q, \tau) - \frac{2qk}{3E_q} F_2(k, E_q, \tau) \\ \dot{F}_2(k, E_q, \tau) &\simeq \frac{qk}{5E_q} \left[2F_1(k, E_q, \tau) - 3F_3(k, E_q, \tau) \right] - \frac{\partial f_0}{\partial \ln q} \left[\frac{1}{15} \dot{h} + \frac{2}{5} \dot{\eta} \right] - a \frac{F_2(k, E_q, \tau)}{\tau_{rel}} \\ \dot{F}_l(k, E_q, \tau) &\simeq \frac{qk}{(2l+1)E_q} \left[lF_{(l-1)}(k, E_q, \tau) - (l+1)F_{(l+1)}(k, E_q, \tau) \right] - a \frac{F_l(k, E_q, \tau)}{\tau_{rel}}, \quad l \geq 3. \end{aligned} \quad (2.16)$$

³In general, as seen in [71], the general expression for the invariant flux is $F = n_1 n_2 \frac{p_1 \cdot p_2}{E_1 E_2} V_r$ with $n_{1,2}$ the number density and V_r the special relativity relative velocity. In this case, the expression reduces to this form assuming that one particle states are normalized to $2E$ and that incoming and outgoing particles are of the same species.

⁴It is important to differentiate this definition for σv_{mol} from the one used in [54]. In that case, the normalization and definition for the cross section matches [74], and differs from this one on a numerical factor.

In this hierarchy, while the $l = 0, 1$ moments of the collision term have the same form as all the others, we choose to set them to zero in order to restore particle number, velocity and energy conservation to the elastic collision term (an approach sometimes referred to as the *Lorentz model*, see [75]).

2.3 Evolution of f_0 and Self Interaction Decoupling

The only missing piece of information regarding the calculation of the relaxation time and its application in a Boltzmann Hierarchy, is the equilibrium distribution function f_{eq} . We will now discuss its behavior, how it evolves in time and what are the mechanisms that may alter it. In the general case, the background distribution function is governed by the zero order Boltzmann equation

$$\frac{\partial f_0}{\partial t} - H(t)p \frac{\partial f_0}{\partial p} = \left(\frac{\partial f_0}{\partial t} \right)_{col}^{(0)} \quad (2.17)$$

where H is the Hubble parameter, \dot{a}/a , and for clarity, we have expressed this equation in terms of time t and physical (not comoving) momentum $p \propto a^{-1}$. In general, this zero order collision term includes all possible interactions that the DM particle suffers: production, annihilation, scattering, etc.

The process by which a particle loses equilibrium with the hot plasma (“decouples”) has been well studied in the past years, and several reviews are dedicated to this problem [63, 70]. In equation (2.17) the two timescales involved are the gravitational (Hubble) timescale and the timescales of all involved interactions $\{\Gamma_{\text{prod}}, \Gamma_{\text{annh}}, \Gamma_{\text{scatt}}, \dots\}$ for production, annihilation, scattering, etc. Whenever the expansion rate of the universe (measured by the Hubble expansion H) overcomes the total interaction rate $\Gamma_{\text{tot}} = \sum_i \Gamma_i$, such a species gets out of equilibrium with the hot plasma and effectively decouples from the evolution of the universe, its evolution governed only by the cosmological redshift of momenta.

While a given species is in equilibrium with the hot plasma, its distribution function is governed by a Fermi-Dirac or Bose-Einstein function:

$$f_{eq}(\vec{p}, \tau) = \frac{1}{\exp \left[\frac{E(p) - \mu(\tau)}{T(\tau)} \right] \pm 1} , \quad (2.18)$$

where E is the particle energy, T is the temperature and μ is the chemical potential (equal to zero for Majorana particles). At the time of decoupling, the temperature of the species is approximately equal to the photon temperature, and after this time the particles move in geodesics, with their physical momenta redshifting with the expansion of the universe. Thus, their distribution functions evolve as:

$$f(\vec{p}, \tau) \equiv f\left(\vec{p} \frac{a}{a_f}, t_f\right) , \quad (2.19)$$

i.e., the distribution function is “frozen-in” at the value it had at decoupling. In the case that the decoupling happens while the particle is still ultra-relativistic ($T \gg m$), this distribution evolves as:

$$f_{rel}(\vec{p}, \tau) = \frac{1}{\exp \left[\frac{p}{T(t)} \right] \pm 1} , \quad (2.20)$$

where we have neglected the impact of the chemical potential μ , and we interpret the temperature evolving as $T = T_f a_f / a \propto a^{-1}$. In the case that it decouples *from the hot plasma* while non-relativistic, the types of interactions that are involved must be carefully considered. For such a species, if it is being kept in equilibrium by non-elastic interactions, its distribution function becomes kinematically suppressed, and it can be expressed as

$$f_{nrel}(\vec{p}, \tau) = \exp\left[-\frac{m}{T_f}\right] \exp\left[-\frac{p^2}{2mT(\tau)}\right], \quad (2.21)$$

where now T evolves as $T = T_f a_f^2 / a^2 \propto a^{-2}$. While these distribution functions can be used to calculate abundances, it is often required to fully solve the dynamics of freeze-out in equation (2.17) in order to obtain accurate results. Studying properly the freeze-out mechanism is necessary to calculate the expected thermal abundance of a given DM species, and is often a major component of DM production in the early universe.

In this work, however, we will focus on a slightly different type of freeze-out mechanism. Throughout this paper, we will assume that number non-conserving interactions have frozen out deep into the relativistic dominated era and well before the epoch relevant to the evolution of perturbations. We will indeed assume that the only relevant interaction in this era is the self interaction of WDM, and here we study this special scenario where the species has decoupled from the plasma long ago and inelastic interactions are no longer relevant. It is important to clarify that these (elastic) self interactions are considered here as a separate process from any interactions the DM component may have with the plasma. All of these interactions between DM and standard model particles (both elastic and inelastic) will be considered to have decoupled at an early time, while the self interactions may decouple at a time relevant to the evolution of perturbations.

Qualitatively, the main effect of this assumption is that, as self interactions conserve number density, total momentum and energy density (referred as *collisional invariants*), these quantities are effectively “locked” since the species gets out of equilibrium with the plasma (or is produced, see [38]). While elastic self interactions might change the form of the background distribution function and the evolution of perturbations, they *cannot* alter the collisional invariant quantities. Furthermore, we will assume that the original production mechanism of the SI-WDM has indeed produced the correct abundance of the species as to be the sole DM component, and leave the discussion of how this might be achieved for future discussions. Regarding the decoupling of these self interactions, we assume that at the early times these maintain the SI-WDM component in kinetic equilibrium and, at some point, they would decouple and equilibrium would be lost until today.

In order to clarify these concepts, we will introduce the following notation: we will label the decoupling temperature *from the hot plasma*, i.e., the typical “freeze-out” temperature, with the standard notation T_f . The other relevant scale we will refer to is the self interaction decoupling temperature T_{SID} , the scale at which $\Gamma_{SI}(T_{SID}) = H$, with Γ_{SI} referring to the interaction rate of self interactions only. According to our assumptions, T_f would have happened at some very high temperature (relative to the scales of Cosmological Perturbation Theory, CPT) and after this moment the only relevant form of interactions are self interactions, which are assumed to decouple at some point during the computation of perturbations: $T_f \gg T_{\text{early}} \gg T_{SID} \gg T_0$, where T_{early} is the earliest computed time in CPT, and T_0 is the temperature of the CMB photons today. We will distinguish between two cases: when $T_{SID} \gg m$ (relativistic self decoupling) and when $T_{SID} \ll m$ (non-relativistic self decoupling).

Relativistic Self Decoupling

In the case that the self decoupling happens while the particle is still relativistic, i.e. $T_{\text{SID}} \gg m$, the scenario is rather similar to what happens with no self interactions. In this case, after the decoupling from the hot plasma at T_f takes place, the DM gets out of equilibrium from the rest of the species in the universe. However, as it maintains kinetic equilibrium with itself, it keeps its distribution function in equilibrium according to the DF of a relativistic species (2.18):

$$f_0(p, T_f > T(t) > T_{\text{SID}}) \equiv f_{0,\text{R}} = \mathcal{C}_{\text{R}} \frac{1}{\exp\left[\frac{pa(t)}{T_{0,\text{R}}}\right] \pm 1} = \mathcal{C}_{\text{R}} \frac{1}{\exp\left[\frac{q}{T_{0,\text{R}}}\right] \pm 1}, \quad (2.22)$$

where p is the local momentum, q is the comoving momentum $q = ap$, $T_{0,\text{R}}$ is the temperature the species would have today if it maintains a relativistic DF, and we have introduced the normalization factor \mathcal{C}_{R} in order to enforce the proper abundance today. Here, the temperature T is to be interpreted to evolve as $T \propto a^{-1}$, and the whole distribution function can be cast in terms of the temperature today $T_{0,\text{R}}$ by using comoving quantities. It is important to note that as this species is evolving accordingly to its own temperature, it is not affected by the entropy release in the hot plasma as other species decouple.

When the self interactions finally decouple at T_{SID} , the specie's temperature is frozen out at this temperature and the evolution of the DF becomes governed by the redshift of momenta according to (2.20). This, incidentally, gives the same form for the background DF as in (2.22), yet maintaining it frozen in this ultra relativistic form even when $T < m$.

Non-Relativistic Self Decoupling

The situation becomes drastically different in the case $T_{\text{SID}} \ll m$ and self interactions remain coupled through the non-relativistic transition, $T = m$. In this case, the situation is the same as in the relativistic self decoupling as long as $m < T \ll T_f$: the species maintains its own temperature $T_{0,\text{R}}$ but remains effectively decoupled from the rest of the hot plasma. As soon as the temperature approaches the particle's mass, self interactions cannot maintain equilibrium in the distribution function anymore, as the equilibrium distribution itself becomes ill-defined.

Once the (pseudo, see [76]) temperature of the species falls well below the particle's mass $m \gg T \gg T_{\text{SID}}$, self interactions are able to restore equilibrium into the DM species, but now fall into a *non-relativistic* distribution function, as in (2.10),

$$f_0(p, m > T(t) > T_{\text{SID}}) \equiv f_{0,\text{NR}} = \mathcal{C}_{\text{NR}} \exp\left[-\frac{p^2 a^2(t)}{2mT_{0,\text{NR}}}\right] = \mathcal{C} \exp\left[-\frac{q^2}{2mT_{0,\text{NR}}}\right], \quad (2.23)$$

with \mathcal{C}_{NR} some appropriate normalization and $T_{0,\text{NR}}$ its temperature today (generally, different from the temperature $T_{0,\text{R}}$ for the relativistic decoupling case). Given the properties of the self interactions, this distribution function would be considerably different from (2.21). Since collisional invariance protects the number density, velocity and energy density, this new function should preserve these quantities through the non-relativistic transition, fixing the values for temperature and normalization once the DF at T_f is specified.

After self interaction decoupling, the evolution of this DF should follow from the redshift of momenta according to (2.19), with the particles moving along a geodesic. This ultimately results in this distribution function being frozen out in this same form even after T_{SID} , making the form (2.23) valid even for $T < T_{\text{SID}}$ in the case the species is coupled through $T = m$.

Transition Approximations and Contour Conditions

Now, we turn to the question of how to properly specify the conservation conditions for these collisional invariants. We will not consider overall momentum: while it is conserved, we assume translational invariance for the background distribution function. Instead, we will focus ourselves in other conserved quantities in order to specify the non-relativistic distribution function.

In the case of number density, it is safe to assume that it is conserved throughout the non-relativistic transition. As it evolves due to the cosmic expansion as $n \propto a^{-3}$ regardless of the decoupling scenario, it is enough to impose that the total number density for this new distribution function today (2.23) to be the same as the relativistic version (2.22), valid only in $T_f \gg T \gg m$ but extrapolated until today. Thus, we have our first contour condition

$$n_{0,R} = \mathcal{C}_R \frac{3}{4} \frac{\zeta(3)}{\pi^2} g m T_{0,R}^3 = \mathcal{C}_{NR} g m \left(\frac{T_{0,NR} m}{2\pi} \right)^{3/2} = n_{0,NR} . \quad (2.24)$$

For the next condition, we would have to rely on the conservation of the energy density ρ . Ideally, the evolution of the energy density should be identical in this case to the case where there are no self interactions at all, as the species only interacts with itself through elastic collisions. Indeed if we demand the condition that the number densities follow the same evolution for the two distribution functions (relativistic and non relativistic), it can be readily shown that this condition is automatically fulfilled. In order to determine both temperature and normalization, we would have to make a few assumptions about how particular thermodynamical quantities evolve during the non-relativistic transition itself. Here, we will present three sets of assumptions, which all lead to variations of the same result.

For the first set of contour conditions, we will take a look at the evolution of the energy density. We make the (rough) approximation of extrapolating the expression for energy density in the ultra relativistic regime (with a relativistic DF, temperature $T_R(t)$ and normalization \mathcal{C}_R) and the energy density in the non-relativistic regime (with a non-relativistic DF, temperature $T_{NR}(t)$ and normalization \mathcal{C}_{NR}) into the exact moment of the non-relativistic transition $T = T_R = T_{NR} = m$, and impose continuity in the energy density. So, we impose the condition $\rho_R(T_R = m) = \rho_{NR}(T_{NR} = m)$ at the transition temperature, with ρ_R the relativistic limit of the energy density and ρ_{NR} the non-relativistic limit. This condition, together with the condition $n_{0,R} = n_{0,NR}$ will be our first set of contour conditions, which we will name the *microcanonical* contour conditions.

For the next two sets of contour conditions, we will perform similarly, but considering different thermodynamic variables. First, if we impose continuity in the temperature between the two different distribution functions, we end up with the condition $T_{0,NR} = T_{0,R}^2/m$. Together with the condition on number density today, we name this set as the *canonical* contour conditions. In the case of the final set, we take a look at the entropy density, calculated as $s = (\rho + p)/T$. As before, this case constitutes the *isentropic* contour condition. We summarize the results for all the different contour conditions in table 1.

These contour conditions are set in order to avoid the complicated non-equilibrium dynamics involved in the transition to the non-relativistic coupled regime. A more accurate treatment can be obtained by fully solving the problem and including the results of a full zero-order Boltzmann solution to the DF (2.17) considering a full expression for the collision term. Even when we choose to avoid this problem altogether by making use of these conditions, the question of what sort of background f_0 to use in the regime $T \sim m$ still holds. We

$f_0(T < m) = \mathcal{C}_{\text{NR}} \exp \left[-\alpha q^2 / T_{0,\text{R}}^2 \right]$	$\mathcal{C}_{\text{NR}} / \mathcal{C}_{\text{R}}$	α
<i>Micro-Canonical</i> $n_{0,\text{R}} = n_{0,\text{NR}} , \rho_{\text{R}}(T = m) = \rho_{\text{NR}}(T = m)$	$\frac{7}{8} \frac{\pi^2}{30} (2\pi)^{3/2} \sim 4.534$	$\frac{1}{2} \left(\frac{180}{7} \frac{\zeta(3)}{\pi^4} \right)^{3/2} \sim 1.075$
<i>Canonical</i> $n_{0,\text{R}} = n_{0,\text{NR}} , T_{\text{R}}(T = m) = T_{\text{NR}}(T = m)$	$\frac{3}{4} \frac{\zeta(3)}{\pi^2} (2\pi)^{3/2} \sim 1.439$	1
<i>Isoentropic</i> $n_{0,\text{R}} = n_{0,\text{NR}} , s_{\text{R}}(T = m) = s_{\text{NR}}(T = m)$	$\frac{7}{360} \pi^2 (2\pi)^{3/2} \sim 3.022$	$\frac{\pi^{8/3}}{18} \left(\frac{10}{7} \zeta(3) \right)^{3/2} \sim 0.8201$

Table 1: Summary of results for the distribution functions at $T < m$ for the different contour conditions mentioned in section 2.3. These distributions are characterized by the factors α and $\mathcal{C}_{\text{NR}} / \mathcal{C}_{\text{R}}$

have studied a phenomenological approach to this, by interpolating between the two known solutions, and will come back to this question in section 3.2.

2.4 Limiting forms for the Relaxation Time

Finally, let us obtain limiting expressions for the relaxation time τ_{rel} both in the deep relativistic and non-relativistic regimes. In the relativistic case, we can make use of the results in [72], where a compact expression for the thermal average $\langle \sigma v_{mol} \rangle$ (2.14) is given assuming a background DF $f_0 \propto \exp[-E/T]$, where T is the temperature of the species. In this case, this quantity reduces to the following form, in terms of the kernel $\chi(s)$:

$$\langle \sigma v_{mol} \rangle = \frac{1}{4m^4 T K_2^2(m/T)} \int_{4m^2}^{\infty} ds \chi(s) \frac{g_i^2}{4} \sqrt{s - 4m^2} K_1(\sqrt{s}/T) , \quad (2.25)$$

where K_i are the modified Bessel functions, and we have corrected an error in the prefactors in [54] due to differences in the normalization of $\chi(s)$. While in [72] this expression is always valid, in our case it only applied to the relativistic case. Indeed, in the non-relativistic limit (in both the relativistic and non-relativistic self decoupling scenarios) the background DF is *not* well approximated by $\exp[-E/T]$, which carries the kinematic suppression factor in (2.21). For this case, we take the limit of equation (2.14) in both the relativistic and non-relativistic self decoupling scenarios.

First, we start by reducing the form of the kernels χ in the non-relativistic limit. In this case, we take $s \rightarrow m^2(4 + v_r^2)$ with $v_r^2 = (\vec{v}_q - \vec{v}_l)^2$ the relative velocity of incoming particles. In this limit, the kernel χ takes the form:

$$\begin{aligned} \chi \left[s \rightarrow m^2(4 + v_r^2) \right] &= \frac{C_t^0 v_r}{2} + \frac{1}{16} v_r^3 \left(4B_t^0 m^2 - C_t^0 + 8C_t^1 \right) \\ &+ \frac{1}{768} v_r^5 \left(128A_t^0 m^4 - 24B_t^0 m^2 + 192B_t^1 m^2 + 9C_t^0 - 48C_t^1 + 384C_t^2 \right) + \mathcal{O}(v_r^7) , \end{aligned} \quad (2.26)$$

where we have separated $C_t = \sum_i C_t^i v_r^{2i}$ in powers of v_r^2 , and the same for B_t and A_t . Then, we can take the non-relativistic limit in equation (2.14) [73], where v_{mol} is now replaced by v_r which reads, in terms of $\chi(s)$:

$$\langle \sigma v_{mol} \rangle_{\text{NR}} = \frac{\int d^3v_1 d^3v_2 \sigma v_r f_{eq}(v_1) f_{eq}(v_2)}{\int d^3v_1 d^3v_2 f_{eq}(v_1) f_{eq}(v_2)} \equiv \frac{\int d^3v_1 d^3v_2 \chi [s \rightarrow m^2(4 + v_r^2)] g_i^2/4 f_{eq}(v_1) f_{eq}(v_2)}{4m^2 \int d^3v_1 d^3v_2 f_{eq}(v_1) f_{eq}(v_2)}. \quad (2.27)$$

In the case of the relativistic self decoupling scenario, this would be a distribution matching the one of a relativistic, decoupled particle species. If we assume a Maxwell-Boltzmann form (ignoring quantum effects in f_0 , in this case), the integral (2.27) can be carried out directly

$$\begin{aligned} \langle \sigma v_{mol} \rangle_{\text{NR}}^{\text{R-SID}} &= \frac{35}{16m^2} \left(\frac{C_t^0}{2} \left(\frac{T}{m} \right) + \frac{9}{4} (4B_t^0 m^2 - C_t^0 + 8C_t^1) \left(\frac{T}{m} \right)^3 \right. \\ &\quad \left. + \frac{99}{32} (128A_t^0 m^4 - 24B_t^0 m^2 + 192B_t^1 m^2 + 9C_t^0 - 48C_t^1 + 384C_t^2) \left(\frac{T}{m} \right)^5 \right) \\ &\quad + \mathcal{O}(T/m)^7 \end{aligned} \quad (2.28)$$

In the case of a non-relativistic distribution function, the integral can also be carried out directly: following the results in [71], we reach the following expression

$$\begin{aligned} \langle \sigma v_{mol} \rangle_{\text{NR}}^{\text{NR-SID}} &= \frac{1}{\sqrt{\pi} m^2} \left(\frac{C_t^0}{2} \left(\frac{T}{m} \right)^{1/2} + \frac{1}{2} (4B_t^0 m^2 - C_t^0 + 8C_t^1) \left(\frac{T}{m} \right)^{3/2} \right. \\ &\quad \left. + \frac{1}{8} (128A_t^0 m^4 - 24B_t^0 m^2 + 192B_t^1 m^2 + 9C_t^0 - 48C_t^1 + 384C_t^2) \left(\frac{T}{m} \right)^{5/2} \right) \\ &\quad + \mathcal{O}(T/m)^{7/2} \end{aligned} \quad (2.29)$$

We can see the evolution of the relaxation time for, for example, a vector field case in fig. 1. There we compare the self interaction rate Γ with the Hubble rate H to give a more appropriate idea of the decoupling process: whenever the (self) interaction rate falls below the Hubble rate, the species can be considered as (self) decoupled, thus having lost kinetic equilibrium (with itself). For this case, we can see from both the figure and the limiting expression (2.29) a well known behaviour (e.g. [77]) of the thermal averaged cross section, behaving as $\langle \sigma v_{mol} \rangle \propto T^{1/2}$, but remarkably only when the background distribution function is non relativistic, which is not always the case for a self interacting species. In any case, in both scenarios the interaction rate behaves as $\Gamma \propto a^{-4}$ with the scale factor.

Both in the deep relativistic as well as in the non-relativistic regimes, the expressions (2.29), (2.28) and (2.25) are used along this paper to calculate the relaxation time and the interaction rates for all mediator models. However, in the intermediate regime, where none of these approximations are valid, we are left without any straightforward calculation method for those quantities. Thus, we choose to calculate the interaction rate in the intermediate regime as an interpolation between the relativistic and non-relativistic regimes with the form:

$$\Gamma^i \simeq \frac{1}{2} \left(\text{P}(\Gamma_{\text{NR}}^i) + \text{P}(\Gamma_{\text{R}}^i) \right) \quad (2.30)$$

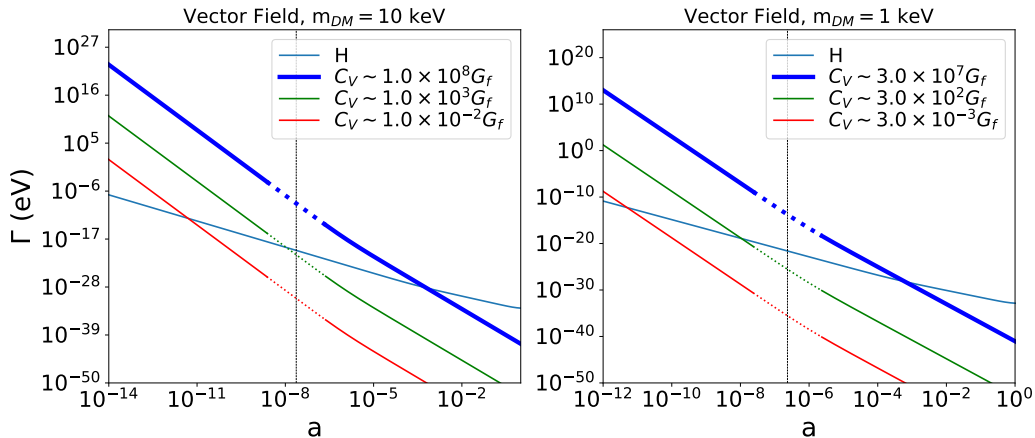


Figure 1: Self Interaction rate for a vector field SI-WDM model for two values of the DM particle mass: 1 keV (*right panel*) and 10 keV (*left panel*), compared to the Hubble expansion rate H as a function of scale factor, considering the effects of non-relativistic self decoupling. Dotted lines show the intermediate regime $0.1m_{\text{DM}} < T < 10m_{\text{DM}}$, where none of the approximate forms described in section 2.4 are used and instead the interpolating form (2.30) is used. The interaction constant $C_V = g_V^4 m_V^{-4} \cos^{-4} \theta'_W$ (with g_V the Lagrangian coupling constant, m_V the mediator mass and θ'_W the dark sector Weinberg angle) corresponding to the Bullet Cluster upper limit constraints (see [78]) are shown in thick lines, and vertical lines label $T_\gamma = m$.

with P indicating a power-law fit and $\Gamma_{\text{R}}^i, \Gamma_{\text{NR}}^i$ indicating the relativistic and non-relativistic approximations of the interaction rates for model i , respectively (valid for both non-relativistic and relativistic self decoupling modes). This approximation provides good results in the case of massive scalar and vector field mediators and, while it is a rough approximation to the true relaxation time, it is a subdominant component of the overall error in our calculations introduced by other approximations.

Now, with all of the ingredients ready (the relaxation time approximation of the Boltzmann Hierarchies, a study on the evolution of the background f_0 and a way of computing the relaxation time τ_{rel}), we are set to implement the SI-WDM hierarchies in a Boltzmann solver. In the next section, we will see the details regarding the implementation of this species in CLASS, and present the main results for this work.

3 Numerical Implementation in CLASS

We implemented the Relaxation Time Approximation (2.16) in the public Boltzmann solver CLASS (version 2.7.2), as a modification to the existing module handling WDM and non cold relics with the objective of assessing the impact of SI-WDM on the power spectrum and on CMB anisotropies⁵. To this end, we have modified the code appropriately in order to handle the thermal history outlined in section 2 for this new species and developed or adapted a few approximations and procedures to provide fast and accurate results, which we describe in this section.

⁵We provide a free distribution link to this modified version of CLASS in github.com/yunis121/siwdm-class, requiring a pre-existing installation of CLASS v2.7.2 (which we also provide in the repository). This particular version does not precompute the relaxation time within CLASS, but is instead supplied with a relaxation time table as an external file.

The CLASS Boltzmann solver introduced a complex method of optimized momentum sampling in order to deal with non cold relics, such that it maximizes the precision in the integrals of the perturbations of the form $q^n F_l(k, E_q, \tau)$ with $n = 2, 3, 4$ and $l = 0, 1, 2$ (see [58], sec. 5 for more details). As the approximated Boltzmann hierarchy (2.16) does not couple different momentum bins, this optimization method remains unchanged in this version of SI-WDM implementation, except for a few minor modifications in the non-relativistic self decoupling scenario which we describe in section 3.2.

Once the relaxation times are already pre-computed, the thermal history for this species is calculated according to what was outlined in section 2. If this thermal history is met (the species is self coupled at the earliest computed time, and decouples at some point afterwards), a few procedures are put in order to obtain self consistence. As the program can be initialized under any arbitrary background distribution function, a few steps are taken in order to i) thermalize the distribution: i.e. if a non-equilibrium distribution is given, construct an equilibrium distribution that matches its number and energy densities, according to

$$n_{eq} = \mathcal{C}_R \left(\frac{T_0^{eq}}{a} \right)^3 F_2(0) = 4\pi \left(\frac{T_0^{ncdm}}{a} \right)^3 \int_0^\infty dq q^2 f_0(q) = n_{init} \quad (3.1)$$

$$\rho_{eq} = \mathcal{C}_R \left(\frac{T_0^{eq}}{a} \right)^4 F_3(0) = 4\pi \left(\frac{T_0^{ncdm}}{a} \right)^4 \int_0^\infty dq q^2 f_0(q) \epsilon(q, m) = \rho_{init} \quad (3.2)$$

where T_0^{eq} , T_0^{ncdm} are the equilibrium and initial temperatures of the species today, and ii) re-normalize the distribution function in order to match the desired abundance today (a function already implemented in standard CLASS for non cold relics). These steps are taken in order to fulfill the assumptions we have taken in section 2 and reflect the expected effects of self interactions at early times.

Once these self consistency steps are taken, the program further checks what decoupling scenario it should implement, whether that may be relativistic or non-relativistic self decoupling. As the distribution (in comoving variables) is not modified in the first scenario, the evolution of f_0 continues the same way as for standard WDM, but for the non-relativistic decoupling scenario a few extra steps are taken, described in section 3.2 and according to what is discussed in section 2.3.

Once the evolution of f_0 and the decoupling scenario is specified, the evolution of perturbations F_i in the RTA (2.16) can be implemented, based on the existing implementation of non cold relics in CLASS [58]. The hierarchy (2.16) is implemented by adding a momentum independent collision term to the existing WDM hierarchies, with a time dependent relaxation time calculated as in section 2.4. Afterwards, the steps regarding the calculation of the angular power spectrum of CMB anisotropies, C_l 's, as well as obtaining the linear power spectrum today follow directly from the WDM implementation. The timescale τ_{rel} of this new interaction term can overcome the other phenomenological timescales of the system $1/\mathcal{H} = a/\dot{a}$ and $\tau_k = \epsilon/qk$, suppressing high- l multipoles and warranting a Tight-Coupling approximation (TCA), as was described in [57] and extended here to SI-WDM in section 3.1.

3.1 The Tight-Coupling Approximation

The Tight Coupling Approximation, described in [57], holds whenever the timescale of self interactions, in this case τ_{rel} dominates over all of the other timescales of the system. For the Boltzmann hierarchies (2.16), these would be the conformal Hubble rate $1/\mathcal{H} = a/\dot{a}$ related to

the metric perturbations and the combination $\tau_k = \epsilon/qk$ related to the streaming of particles to higher l moments.

In the case TCA holds, the interaction terms have the effect of strongly suppressing the perturbations $l > 1$ as the term $C_l[F] = -a\Gamma(\tau)F_l(k, q, \tau)$ dominates. Thus, the species becomes essentially a perfect fluid with zero shear stress, and the only non trivial equations for the system become the ones corresponding to the $l = 0, 1$ moments. The proper implementation is especially important in the case of numerical integration, as the strong suppression of high l -moments that takes place in this regime can result in a numerically “stiff” system.

While the validity of TCA is in this case with $m \neq 0$ and dependent on the momentum bin considered, for the sake of simplicity we take the approximation as valid as long as it holds for $q \sim 1$. We can see in fig. 2 the regions of validity of the TCA approximation in the k, z plane. As was pointed out in [57], given a fixed interaction coupling constant G_{eff} the TCA always fails at large wavenumbers, then starts failing later in time for progressively smaller k until it formally fails for all k at the same time.

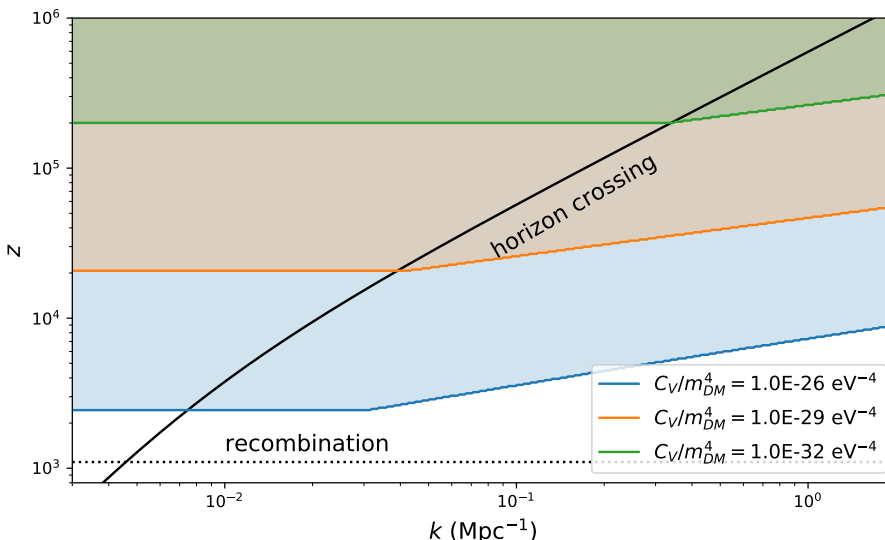


Figure 2: Regions of validity of the Tight-Coupling approximation for different values of the coupling constant, in the case of a vector field mediator for $m_{DM} = 10 \text{ keV}$. The black line represents the moment when a mode k undergoes horizon crossing ($k = aH$). For each C_V , the TCA is valid within their respective colored regions. The colored lines represent the moment when TCA stops being valid, i.e. $\tau_{\text{rel}}^{-1} \equiv a\Gamma = \max(aH, (qk/\epsilon)|_{q=1})$. We have used, for simplicity, $\Gamma \sim C_V T_{\text{ncdm}}^5$ [57].

High l -moments in TCA

A further refinement on the tight coupling approximation can be achieved by making use of the suppressed moments on $l \geq 2$. If we assume the rate of variation of the interaction term is much higher than the variation of other source terms, the equations for these high l -moments in the TCA regime can be approximated as:

$$\begin{aligned}\frac{\Psi}{\tau_{rel}} &\sim \frac{qk}{5\epsilon}(2\Psi_1 - 3\Psi_3) - \left(\frac{1}{15}\dot{h} + \frac{2}{5}\dot{\eta}\right) \frac{d\ln f_0}{d\ln q}, \\ \frac{\Psi_{l\geq 3}}{\tau_{rel}} &\sim \frac{qk}{(2l+1)\epsilon}[l\Psi_{l-1} - (l+1)\Psi_{l+1}].\end{aligned}\tag{3.3}$$

where $\Psi_l(k, q, \tau) = F_l(k, q, \tau)/f_0(q, \tau)$. By making use of the closing assumption for $\Psi_{l_{max}+1} \approx \Psi_{l_{max}}(2l_{max}+1)\epsilon/(qk\tau) - \Psi_{l_{max}-1}$ (see [68]) and only using the value of the first moment in the hierarchy (obtained through the Boltzmann solver), this system can be inexpensively solved by tridiagonal matrix solving algorithms [79]. Thus, a value for F_2 (suppressed by a factor of τ_{rel}) can be included back on the evolution equation for F_1 . This has the advantage of increasing the precision and providing extra numerical stability as the TCA approaches the limit of its validity range, allowing us to be less reliant on the precise region of validity of TCA at a marginal computational cost.

3.2 Non-Relativistic Decoupling

Once the background evolution for f_0 (described at the beginning of this section) has been calculated, we are in the position of discerning between the two scenarios described in section 2.3. In the case that non-relativistic self decoupling is detected, a few extra steps are taken in order to implement this scenario in the evolution of the perturbations and background quantities.

First, we tackle the problem of specifying the new background distribution itself. Once the contour conditions are specified (from the ones we calculated in section 2.3), we have both limiting cases for the background distribution function: at $T \gg m$ we have $f_{0,R} \sim (\exp[q/T] \pm 1)^{-1}$ (2.22) and when $T \ll m$ we have $f_{0,NR} \sim [-q^2/2mT]$ (2.23). However, in the intermediate regime, the distribution function still has to be specified: as a first approximation we choose to interpolate between these two regimes using the following function:

$$f_{Sw}(q, T/m) = Sw(T/m)f_{0,NR}(q) + (1 - Sw(T/m))f_{0,R}(q),\tag{3.4}$$

where the switching function $Sw(x)$ is defined as:

$$Sw(x) = \frac{1}{2}\text{erfc}[\log(x)],\tag{3.5}$$

and $\text{erfc}(x)$ is the complementary error function.

When the non-relativistic decoupling scenario is specified, all function calls to the background distribution function can be replaced by this new function. This includes both calls to f_0 itself and the logarithmic derivative $\partial \ln f_0 / \partial \ln q$ that appears in the Boltzmann hierarchy (2.2). In practice, we also precalculate values for $f_0(q_i)$ and $[\partial \ln f_0 / \partial \ln q](q_i)$, where q_i are the momentum samples, in the deep relativistic and non-relativistic regimes and replace the function evaluations for these precomputed values whenever appropriate.

An interesting question is how to evolve the perturbations through the relativistic to non-relativistic transition itself. As it was the case for the background distribution function, the problem of transitioning between the relativistic and non relativistic regimes appears in the perturbations too. Ideally, a full approach to the Boltzmann hierarchies, complete with moment dependent collision terms such as the ones implemented in (2.2) would naturally include this transition, and no extra care would be needed. However, when using an approximate framework such as the RTA, it is important to consider this transition separately.

One simple approach to this would be to simply assume that the RTA hierarchies (2.16) are valid during the whole evolution, particularly during this transition. In that case, the perturbation moments $F_l(q, k, \tau)$ would evolve without changes through the transition. However, as the background distribution gets suppressed severely during the transition for high q , the ratio $F/f_0|_{q \gg 1}$ would greatly increase and the perturbations would leave the linear regime. This is not representative of the behavior of these relaxation processes, as we expect the interactions that transfer phase space density to lower momenta at this stage to both affect background and perturbation. This phase space density transfer from high to low momenta cannot be described through the RTA, as it is local in momentum space, making this simple approach a poor choice. We instead opt for another way of accounting for these effects that does not imply solving the full collision integrals: we set the background-scaled perturbation $\Psi_l = F_l/f_0$ to be constant through the transition, as a straightforward approximation to this complex process.

Special care needs to be taken in the case of computing the integrals resulting in the energy-momentum perturbations. These quantities are obtained via integration rules of the form

$$\mathcal{I} = \int_0^\infty dq f_0(q) g(q) \approx \sum_{i=1}^n W_i g(q_i) \quad (3.6)$$

where the integrated function can be either $g(q) = \hat{g}(q)$ describing a background quantity or $g(q) = \hat{g}(q)\Psi_i(q, \tau)$ describing an energy-momentum perturbation. The weights W_i are optimized via a complex method (see [58]) in order to obtain maximum precision in the calculation of these integrals for $\hat{g}(q) = \{q^2, q^3, q^4\}$ (describing energy, momentum, pressure and shear integrals). In order to take advantage of the methods already implemented in CLASS, but extend it to this time dependent background DF we choose to re-define the integration rule (3.6) as:

$$\sum_{i=1}^n W_i g(q_i) \rightarrow \sum_{i=1}^{n'} \left(\frac{f_{\text{Sw}}(q_i, T/m)}{f_{0,\text{R}}(q_i)} \right) W'_i g(q_i) , \quad (3.7)$$

where the weights W'_i are now optimized in order to obtain the desired precision in the function

$$t(q) = a_2 q^2 + a_3 q^3 + a_4 q^4 + b_2 \left(\frac{f_{0,\text{NR}}}{f_{0,\text{R}}} \right) q^2 + b_3 \left(\frac{f_{0,\text{NR}}}{f_{0,\text{R}}} \right) q^3 + b_4 \left(\frac{f_{0,\text{NR}}}{f_{0,\text{R}}} \right) q^4 \quad (3.8)$$

and the coefficients are chosen such as

$$a_n \int dq \frac{q^n}{e^q + 1} = 1 \quad , \quad b_n \int dq q^n e^{-q^2} = 1 .$$

As before, we precompute the values for $(f_{\text{Sw}}(q_i, T/m)/f_{0,\text{R}}(q_i))$ in the deep non-relativistic limit and use them appropriately.

With this machinery in place, it is possible to compute the evolution for SI-WDM under the RTA approximation (2.16) and obtain desired observables such as CMB perturbations and power spectrum, among others. We will further explore the results of this extension to CLASS in section 4, in the light of the particle models we describe in section 1.

3.3 Effects on the Matter Power Spectrum

Having developed this extension to CLASS in order to implement the evolution of SI-WDM, we take a first look in this section to the cosmological results in the linear regime. We show some of the main features that these alternative DM models introduce in the power spectrum, in order to discuss observable effects in section 4. One of the most noticeable effects is a suppression on the matter power spectrum at high k . We compare here the suppression produced by SI-WDM to the one produced by other alternative models, such as WDM or SIDM models. Typically, both the tightest constraints and the alleviation of the small scale tensions mentioned in section 1 have been related to this suppression in the power spectrum for the case of WDM [10, 80], therefore in this work we will focus on it and leave the analysis of other quantities such as the C_l 's to future works.

We show a few results of the matter power spectrum $P(k)$ for different masses and interaction constants in fig. 3 and fig. 4, for the case of a massive Vector Field self interaction. In particular, fig. 4 shows the full effect of the non-relativistic self interaction decoupling mentioned in section 3.2, while in fig. 3 this phenomenon is ignored altogether. These very same plots, this time for the case of a Massive Scalar mediator, can be found in figures D.2 and D.1 in appendix D.

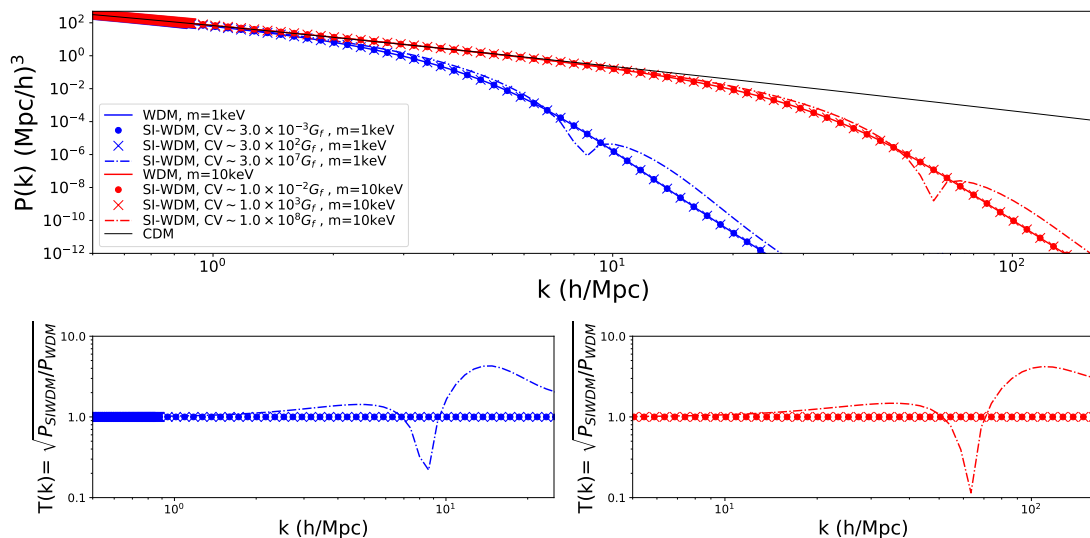


Figure 3: Power Spectrum (*top panel*) and Transfer Functions with respect to standard WDM (*bottom panels*) for a vector field SI-WDM model, simulated using the modification to CLASS described in section 3. In this case, the effects of non-relativistic self decoupling are ignored and we assume a relativistic DF (2.18) holds until today. We assume the relaxation time approximation (2.16), and consider two values of the DM particle mass: 1 and 10 keV. Also plotted are the power spectra of a CDM model and of WDM models with DM mass of 1 and 10 keV. All WDM and SI-WDM models consider a nonresonant production scenario (Dodelson-Widrow mechanism, [64]) with $T \sim (4/11)^{1/3} T_\gamma$.

In these figures, we compare the power spectrum with their most similar counterpart, standard WDM, as well as a reference CDM model. We have chosen two different particle masses of 1 and 10 keV, the first one to provide a parallelism with [77], and the second one to show a model with predictions that agree better with current observations [42, 44]. We have assumed a simple background model: a thermal background with a temperature $T_0 = (4/11)^{1/3} T_\gamma$ identical to SM neutrinos, renormalized to give the correct DM abundance

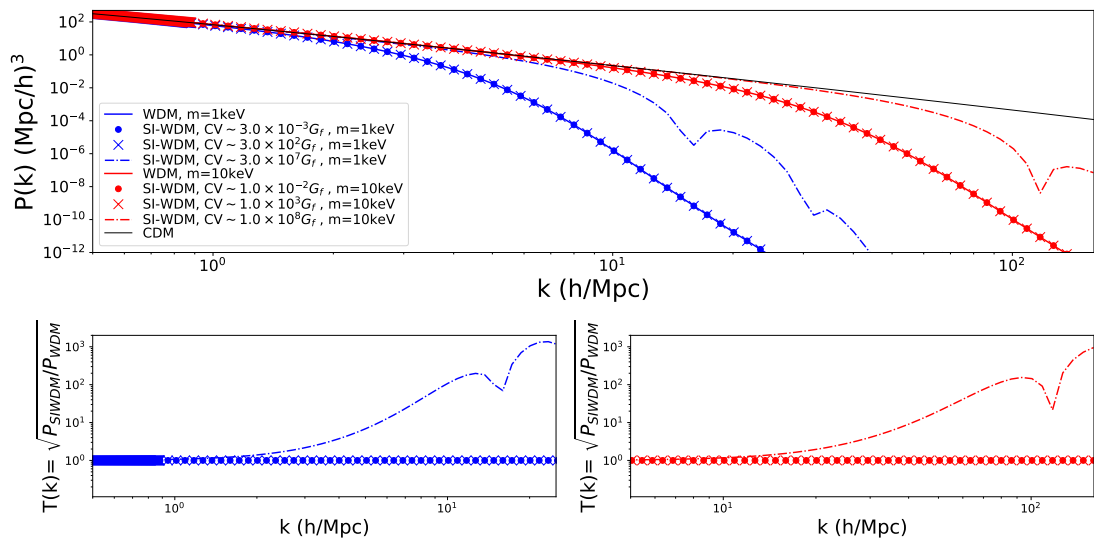


Figure 4: Simulated Power Spectrum and Transfer Functions with respect to standard WDM for several SI-WDM models under the same assumptions as in fig. 3, but considering the effects of non-relativistic self decoupling when appropriate, as described in section 2.3 and section 3.2.

(non-resonant sterile neutrino production [64]). Over these, we show the results of several SI-WDM cosmologies, for the mediator models described in [54] and in section 1. We show three examples for each, ranging from a very weak interaction to one matching the upper limits for the scattering amplitude of DM self interactions coming from Bullet Cluster constraints ($\sigma/m_{\text{DM}} \leq 0.1 \text{cm}^2/g$, see [78]). We show also the transfer functions for the self interacting models with respect to their WDM counterparts, defined as $T_i(k) = (P_i(k)/P_{\text{WDM}}(k))^{1/2}$.

We parametrize these mediator models via their effective interaction constants, which are defined according to the expressions in section 3.3 in [54]. In particular, we considered in these figures a Massive Vector Field interaction, where $\bar{C}_V = (g_V m_V^{-1} \cos^{-1} \theta'_W)^4$, where g_V is the coupling constant appearing in the Lagrangian interaction term, m_V the mediator mass and θ'_W is the (dark sector) Weinberg angle, in units of the Fermi constant G_F . The other interaction model we consider, a Massive Scalar mediator interaction, is also defined according to [54] as $C_m = (g_m/m_\Phi)^4$, with g_m the coupling constant of the interaction and m_Φ the mediator mass.

The first obvious conclusion we can draw from both figures is that, for low enough interaction constants the results are, predictably, indistinguishable from the ones of WDM. This is evident for a large range of values of the interaction constants, and is indeed the case if interactions are completely irrelevant during the matter dominated era when metric perturbations start to grow. Even for constants when this is not the case, the “base” WDM model can give us an idea of the general behaviour of SI-WDM: a suppression in the power spectrum, typical of WDM models, is also seen in SI-WDM models. This suppression is identical for low enough interaction constant and, as can be seen from these figures, it would remain almost identical even for higher constants were it not for the effect of non-relativistic Self Decoupling.

This brings us to a second feature of these power spectra: the effects of non-relativistic Self Decoupling. For the models shown in both fig. 3 and fig. 4, the highest interaction

constants would correspond to a situation of non-relativistic Self Decoupling as explained in section 2.3 and section 3.2. We can see that both of these high-constant spectra are the ones that show significant departures from its corresponding WDM equivalent but only a small oscillation at high k is present if we don't consider these effects. If we consider the effects described in section 3.2, then the suppression scale moves to a significantly higher k and a completely different pattern of oscillation appears. This is remarkable, as a high interaction constant might make a SI-WDM model “colder” than its WDM counterpart, i.e. as if it corresponds to a much higher particle mass: we will explore this further in section 4. Also, regarding the various contour conditions that were considered in section 2.3, we have found that all of them produce virtually indistinguishable results in the power spectrum.

Finally, we can discuss about the final evident feature in these spectra, the oscillations themselves. Indeed, these oscillations are not necessarily exclusive to SI-WDM models, and we can see them in several other self interaction models [21] as well as a few approximation schemes to WDM itself [58]. In fact, this is common to most fluid or pseudo-fluid approximations to the full Boltzmann hierarchy and we will explore quantitatively how, for the case of SI-WDM, these oscillations come as a consequence of the tight-coupling regime in appendix A.

4 Comparison with Observables

Having discussed about the underlying particle models for SI-WDM in section 1, developed theoretical knowledge about their cosmological evolution in section 2, and incorporated this theory into a numerical tool in section 3, it is time to use this framework to put bounds in the parameter space of these models. As we mentioned before, in this work we have focused on obtaining and constraining the simulated matter power spectrum, which is likely the quantity that is more sensitive to the evolution of small scale structure. Thus, in this section we will use some observables of the small-scale structure in order to identify the simulated spectra that are allowed by cosmology and thus restrict the available parameter space for SI-WDM.

However, comparing the outcome of simulated *linear* cosmology with the structure observed today may not be a straightforward task to begin with. In fact, the smallest gravitationally bound structures today have typically collapsed long ago, a highly nonlinear phenomenon that is not well captured by linear cosmology and is traditionally modeled via N-body numerical simulations [16, 35, 49]. However, certain predictions about the collapsed nonlinear structure are possible by exploiting the statistical properties of the cosmic density field. In particular, the extended Press-Schechter (PS) formalism allows us to obtain predictions about, for example, the estimated mass functions of haloes and subhaloes [34]. These predictions are directly related to the *linear* power spectrum extrapolated until today, thus we can obtain constraints on it by using these theoretical predictions and comparing them against direct observation of structure such as, in this case, the number of subhaloes in the Milky Way.

In [81] it is shown that the linear power spectrum can be parametrized for various particle physics models via a phenomenological transfer function defined as

$$T(k) = [1 + (\alpha k)^\beta]^\gamma, \quad (4.1)$$

where $\{\alpha, \beta, \gamma\}$ are fitting coefficients to the transfer function $T(k)$. Such a *transfer function based* approach was used in that work to derive predictions about both Milky Way satellite galaxy counts and Lyman- α forest data.

In this work, however, we use exclusively the simulated data from our modified version of CLASS in order to put constraints on these observables. Still, we observe that our numerical results can be parametrized using this phenomenological function and, moreover, that the coefficients $\{\alpha, \beta, \gamma\}$ can be put in terms of the physical parameters $\{m, C_i\}$. We remind the reader that, while this phenomenological fitting function does not capture some richer behaviour, such as the oscillatory patterns, it describes well the power spectrum damping and is accurate when used to calculate the observables relevant for this section, such as subhalo counts and Lyman- α forest measurements.

Thus, we fit this phenomenological transfer function (4.1) to the simulated data and we find formulas for the fitting coefficients $\{\alpha, \beta, \gamma\}$ in terms of the model parameters that fit well our results in appendix C. We remark that these fitting functions, while providing relatively accurate approximations for the transfer function, can incur in significant errors in the absolute values of the parameters $\{\alpha, \beta, \gamma\}$, along with the parametrization of the transfer function itself ignoring more complex effects such as power spectrum oscillations as described in appendix A.

We show a few examples of simulated power spectra in fig. 5. For these examples, we also show the best fit spectra using the parametrization 4.1 for the transfer function in terms of $\{\alpha, \beta, \gamma\}$. We also compare these with the resulting parameters from the phenomenological fitting formulae outlined in appendix C, in terms of $\{m, C_i\}$ and the relevant DM interaction model and thermal history scenario.

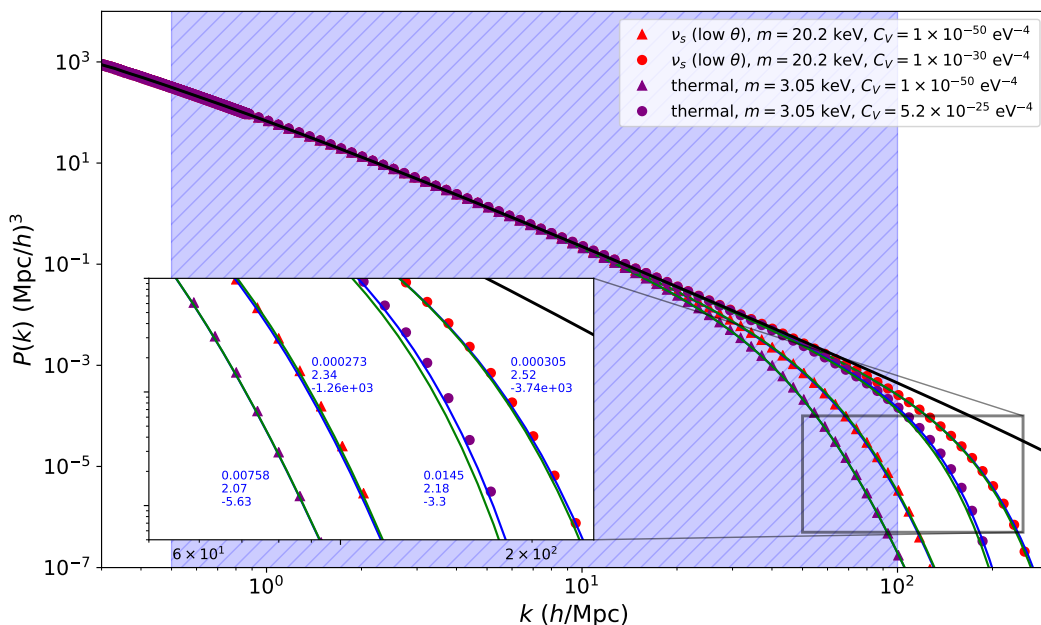


Figure 5: Power spectra obtained through fitting functions, compared to the simulated data. We considered four different simulations, for the cases of thermal production (shown in red) and low θ sterile neutrino WDM (shown in purple, see fig. B.1) and the cases of relativistic (triangles) and nonrelativistic (circles) self interaction decoupling (note that θ is the value of the sterile neutrino mixing angle with the SM neutrinos). On top of these, we plot the power spectra obtained from the best fit of the transfer function eq. (4.1) and its corresponding values $\{\alpha, \beta, \gamma\}$, in blue. We show as well the power spectra corresponding to the values of $\{\alpha, \beta, \gamma\}$ obtained from the fitting formulae shown in appendix C, in green. As a reference we also show a CDM power spectrum (black), and the shaded region corresponds to the relevant values of k for the Lyman- α analysis outlined in section 4.2.

4.1 Milky Way Satellite Counts

We provide here short summaries of the methods used in [81] to compare the linear power spectrum with both the observations of Milky Way satellite counts and the Lyman- α forest data. In general, the Λ CDM model predicts a number of satellite galaxies of the Milky Way that is too large, as many as ~ 1000 , compared to the observed ~ 50 [10]. In order to get an estimate number of observed satellites of the Milky Way, we add the number of classical MW satellites to the number of ultra-faint ones observed by SDSS, accounting for the limited sky coverage of the survey and a 10% sampling variance, arriving to a conservative number of observed MW satellites of $N_{sat} = 57$.

We compare that with the estimated number of subhaloes predicted by the PS formalism. According to [34, 81], the subhalo mass function, for a given host halo of mass M_{halo} , is given by:

$$\frac{dN}{dM_{sub}} = \frac{1}{44.5} \frac{1}{6\pi^2} \frac{M_{halo}}{M_{sub}^2} \frac{P(1/R_{sub})}{R_{sub}^3 \sqrt{2\pi(S_{sub} - S_{halo})}}, \quad (4.2)$$

with M_{sub} , S_{sub} the mass and variance of a given subhalo and M_{halo} , S_{halo} the ones for the main host halo, defined as:

$$S_i = \frac{1}{2\pi^2} \int_0^{1/R_i} dk k^2 P(k) \quad , \quad M_i = \frac{4\pi}{3} \Omega_m \rho_c (cR_i)^3 \quad , \quad c = 2.5 \quad (4.3)$$

with Ω_m the matter density parameter, ρ_c the critical density of the universe (today) and $P(k)$ the linear power spectrum for a given cosmological model, extrapolated to $z = 0$. We can obtain a prediction for the total number of subhaloes for the MW using an assumption for its mass of $M_{MW} \sim 1.2 \times 10^{12} M_\odot/h$ [82] and integrating expression (4.2) from a minimum subhalo mass of $10^8 M_\odot/h$ up to M_{MW} . We obtain a constraint by requiring that the predicted number of subhaloes must be equal or greater than the conservative bound for the observed MW satellites, and report the results for various SI-WDM cosmologies in section 4.3.1.

4.2 Lyman- α Constraints

Another relevant probe of the small scale matter distribution is the so called Lyman- α forest: at a quick glance, the Lyman- α absorption spectrum produced by the intergalactic medium on the spectra of distant quasars. This absorption spectra is produced by the inhomogeneous distribution of neutral hydrogen and has been widely used as an accurate probe of the matter power distribution on small and intermediate scales, namely $0.5 h/\text{Mpc} \leq k \leq 100 h/\text{Mpc}$ [80, 81, 83, 84]. While computing accurate bounds would imply a full statistical analysis, we here instead follow the methods of [34, 81] in order to investigate the plausibility of small deviations to already well-explored non-cold DM cosmologies, namely the thermal 3.5 keV constraints obtained though a full analysis of Lyman- α forest data [84].

We use the criterion set in [81] to characterize the deviation of a model from a Λ CDM cosmology. At a given scale k , we quantify this by the ratio of the 1-dimensional power spectra

$$r(k) = \frac{P_{1D}(k)}{P_{1D}^{\Lambda\text{CDM}}(k)} \quad (4.4)$$

which is itself defined as the following integral of the 3-dimensional power spectrum

$$P_{1D}(k) = \frac{1}{2\pi} \int_k^\infty dk' k' P(k') \quad (4.5)$$

This quantity allows us to obtain a rough estimate of how suppressed is a given power spectrum with respect to Λ CDM. We thus can adopt the following criterion in order to determine whether or not a particular cosmology is allowed by this analysis: if a given power spectrum is “more suppressed” than the reference 3.5 keV thermal WDM model, then this model would be excluded. We can quantify this suppression at all scales k by introducing the following estimator

$$\delta A = \frac{A_{\Lambda\text{CDM}} - A}{A_{\Lambda\text{CDM}}} \quad (4.6)$$

where A is the integral of the ratio $r(k)$ in the scales probed by the Lyman- α forest observations: $0.5\text{Mpc}/h \leq \lambda \leq 100\text{Mpc}/h$ for the MIKE/HIRES+XQ-100 combined dataset used in [81, 84].

Thus, by using the 3.5 keV thermal WDM power spectrum, the authors in [81] obtain an estimated suppression value of $\delta A = 0.38$. This serves as the most extreme model that can be allowed according to this analysis: any model that shows a higher value of this suppression estimator δA is considered to be excluded by the Lyman- α forest data at at least 95% C.L. These authors also include a tighter constraint of $\delta A = 0.21$ for a fixed thermal history, which we will include in our analysis for completeness.

It is important, however, to discuss about the validity of this method. A full analysis of the Lyman- α forest data requires a careful statistical analysis and a complex set of hydrodynamical simulation in order to compute the most convenient statistics for this set of observations, namely the flux power spectrum $P_F(k)$ [34, 80, 85]. However, in [34, 81] it is argued that as the relation between flux and linear 3-dimensional matter power spectra can be accurately modeled as $P_F(k) = b^2(k)P_{3D}(k)$, with $b^2(k)$ a bias factor. This bias factor differs little between Λ CDM and modified DM models such as SI-WDM, which are “fairly close” to the reference case, and thus equation (4.4) can be reasonably applied to the flux power spectra as well. We remind the reader that this is an approximate method and a proper, robust bound should be obtained via a full hydrodynamical simulation as in, for example, [80].

4.3 Parameter Space Constraints

4.3.1 Exploring the SI-WDM parameter space

In this section, we present the results of the Milky Way satellite count and Lyman- α analysis for a suite of SI-WDM cosmological parameters. We analyze the resulting power spectra for different mediator models and initial conditions for the background distribution function, each of them with approximately ~ 400 simulated spectra. For each combination of mediator model and initial conditions, we covered a uniform grid in the SI-WDM parameters (particle mass and interaction constant), and performed the analysis we outline in section 4.1, section 4.2 for each of them. We plot the results for a vector field mediator under a high asymmetry, resonantly produced sterile neutrino model in fig. 6, and we show the results for the other mediator-initial condition combinations in appendix D.

For the specific case of resonant sterile neutrino production, we considered two different scenarios depending on the values of the sterile neutrino mixing angle with the SM neutrinos (θ^2). As in this scenario sterile neutrinos are produced both by the non-resonant [64] and resonant [39, 41] mechanisms, and the latter depends on the amount of initial lepton asymmetry in the universe, one can arrive to the proper DM abundance for a multitude of SM mixing angles by varying this asymmetry value. Therefore, in this case we have considered two cases: one with high θ (low asymmetry) and one with low θ (high asymmetry), sampled from the

slices in the sterile neutrino parameter space for ν MSM we show in fig. B.1. For each mass value sampled along these lines, we have obtained the initial distribution functions at $T = 10$ MeV by using the `sterile-dm` routine in [41].

We can see many features in fig. 6, and most of them are shared across all production mechanisms and interaction models, albeit with small differences in their constraints. First of all, we can see that for small enough interaction constants (depending on the interaction model itself), the results for each initial conditions are comparable to the ones for standard WDM. We find similar results to the analysis on [34, 81] in the lower end of interaction constants, and indeed during the whole parameter space corresponding to the relativistic self decoupling scenario. Following the same arguments as in the discussion following fig. 4, we can attribute this to the fact that these perturbations do not grow during the radiation dominated era, thus the effects of self interactions deep into the particle’s relativistic regime (which is itself deep into the radiation dominated era) is indeed negligible, and the results are comparable to WDM cosmology.

Consistently with the findings of [34, 81, 84], we find that the results for Lyman- α forest analysis are indeed much tighter than the ones from MW satellite counts for all models, and in particular looking at fig. 6 is possible to exclude almost all of the available parameter space for resonantly produced ν MSM sterile neutrinos using these observations, at least in the non interacting case. We see, however, that the situation changes drastically if we instead include strong enough self interactions.

A clear transition in both N_{sub} and δA can be seen whenever the models transition to the non-relativistic self decoupling regime. A sharp change takes place for this regime, as the spectra become immediately much colder (as we have discussed after fig. 4). This basically results in spectra which are much less suppressed than in the relativistic decoupling regime when compared to Λ CDM, with a much higher suppression scale in k space. This can be thought of as a “shift to the left” in mass space: models that originally correspond to a mass that is highly suppressed now result in similar predictions as models with a much higher mass through the introduction of self interactions. Thus, spectra corresponding to WDM become “colder” through the effect of self interactions.

For the models themselves, these colder spectra have two main consequences: first it re-allows models previously excluded by observations. As these models now become much more similar to a Λ CDM cosmology, the predictions arising from the linear power spectrum become increasingly more consistent with observation. At the same time WDM models that are allowed by observations become, through the inclusion of self interactions in this regime, not able to alleviate small scale structure tensions to Λ CDM through the linear power spectrum alone. Still, we remind the reader that some of these tensions can be alleviated via the inclusion of self interactions in the later, non-linear stages of structure formation [10, 49].

4.3.2 The ν MSM Parameter Space in the Presence of Self Interactions

From the previous section we have concluded that, even in a resonant production scenario relevant to ν MSM, sufficiently strong self interactions may significantly relax the constraints on MW satellites and Lyman- α . Indeed, we have seen that, for a given particle mass and mixing angle θ , the differences between low interaction constant under relativistic self decoupling and high interaction with non-relativistic self decoupling are particularly marked, and a number of models that are excluded by the analysis in the first scenario are admitted in the second.

Thus, we can go a step further and analyze the full parameter space of ν MSM, discerning exactly how much of the $\{\theta, m\}$ parameter space can be readmitted under the effects of

self interactions. This new model would be comprised of three parameters now: particle mass m , mixing angle θ and the interaction coupling constant C_i . For this analysis, we are interested in which section of the $\{\theta, m\}$ parameter space can be readmitted by using *any* interaction model (quantified by C_i). As we have seen from the past section, for fixed $\{\theta, m\}$ a higher interaction constant always results in equal or colder spectra. Thus if for these fixed parameters the model with the highest allowed interaction constant is allowed by a given observation, we can conclude that there is indeed a finite range of interacting models for which that particular value of $\{\theta, m\}$ is allowed.

So, we calculated these constraints in the ν MSM parameter space and show these results in figure 7. In this analysis, for each point in parameter space, we consider that particular model to be admitted by each observable (MW satellites and Lyman- α) if they are allowed for the highest value of the interaction constant consistent with Bullet Cluster constraints (our current upper limit on the interaction strength).

On top of these bounds, we can see the other sources of constraints for this particular model, namely production bounds (based on upper limits to the amount of lepton asymmetry consistent with Big Bang Nucleosynthesis) and X-Ray constraints from observation of various galactic and extragalactic objects. For comparison, we show also the Lyman- α constraint for standard ν MSM as was calculated in [34]. We see a striking difference with the non interacting case, with the latter expanding considerably the available parameter space with respect to the former, which nearly excludes the whole available space for these particular WDM models.

5 Conclusions

Following on the initial theory of self interactions in WDM linear theory set in [54], we aimed in this paper to provide a relatively complete practical application of this framework. This included the relevant approximations that can be used to simplify the treatment, a numerical application in CLASS and comparison of the available parameter space of this self interacting models with its WDM counterparts.

Thus, in section 2 we started introducing the Boltzmann formalism required to implement these SI-WDM models. After stating the particular scenario of thermal history considered, we recalled the results obtained in [54] and expanded upon them. In that section, we provided an in-depth explanation of the relaxation time approximation and the resulting Boltzmann hierarchies, introduced the situation of non-relativistic self decoupling and considered an approximate way of treating the relativistic to non-relativistic transition for both background and perturbations. We considered as well limiting forms for the relaxation time in all of these scenarios.

Then, having discussed thoroughly about these approximations to the full linear evolution of perturbations in SI-WDM, we implemented them in the public Boltzmann solver CLASS. We made this modification publicly available at github.com/yunis121/siwdm-class, and discussed its implementation in section 3. We first considered a few further approximations in various regimes that are necessary to the numerical implementation, such as the tight-coupling regime, as well as several improvements to the built-in integrator and the background DF modules for the case of non-relativistic self decoupling. Afterwards, we took a first look at the resulting power spectra and discussed both the effects of self interactions and non-relativistic self decoupling in the resulting distributions.

Having obtained these spectra, in section 4 we compared these results with observation. From comparing these spectra with observables from structure formation such as the number of Milky Way subhaloes and the Lyman- α forest measurements, we can obtain a very interesting resulting parameter space for several underlying (self interacting extensions of) WDM models, for various mediator models.

In this work, we obtained several interesting advancements in the area of SI-WDM, together with conclusions relevant to the field of WDM. Not only we show a readily applicable version of the results in [54], a proper discussion of the evolution of these species in the case of the Relaxation Time Approximation and an introduction to the scenario of non-relativistic self decoupling; but also we provide a full numerical application, which implies a very relevant leap forward in the development of SI-WDM models. This can pose these models as a very interesting alternative and extension to typical WDM models and providing a careful analysis of linear perturbations can both help encase past works which included such WDM extensions in a more complete cosmological framework, as well as motivate further study in the topic. Particularly in the case of self interacting ν MSM, we have seen in section 4 that such an approach can severely relax the structure formation constraints, and possibly readmit a large portion of its parameter space, as was suggested in [48].

It is important to know, however, that what is presented here is an approximate view of a general class of SI-WDM models, and proper consideration of any particular realization of such models would indeed require further analysis, from their particle production in the early universe, to studies on their clustering as well as observational constraints. Also, the approximate view of SI-WDM linear theory presented here can be improved in many ways, some of the most obvious being extending this approach to more general models (such as light mediators or long-range interactions). Other improvements involve the implementation of the full collision terms shown in [54], or calculating precisely the relativistic to classical transition.

In general, we believe this work can serve as an important stepping stone in the study of self interacting extensions to WDM. We have indeed shown here that the inclusion of self interactions to these general class of models can alter the resulting power spectra, thermal history and distribution today on a significant way, with important consequences for a wide range of DM models.

Acknowledgments

CRA has been supported by CONICET, Secretary of Science and Technology of FCAG and UNLP (grants G140 and G175), National Agency for the Promotion of Science and Technology (ANPCyT) of Argentina (grant PICT-2018-03743) and ICRANet. CGS is supported by ANPCyT grant PICT-2016-0081; and grants G140, G157 and G175 from UNLP. DLN has been supported by CONICET, ANPCyT and UBA.

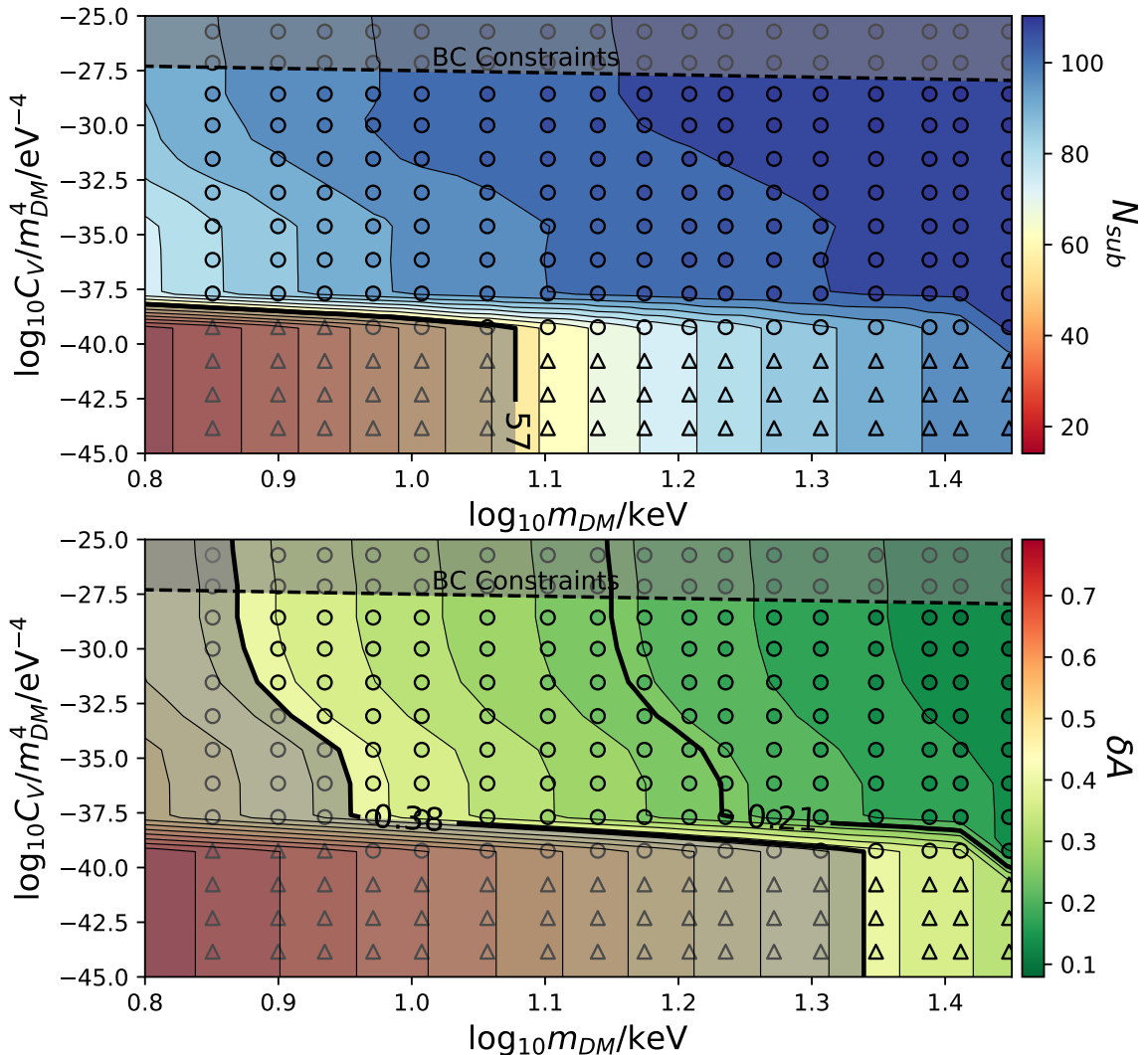


Figure 6: Parameter space constraints for SI-WDM cosmologies, according to the analysis outlined in section 4. We show here the case of a vector field mediator for self interactions as outlined in [54] and resonant production (Shi-Fuller mechanism [39, 41]) in the high lepton asymmetry case. We have also considered the case of a heavy scalar field for the mediator model and three other different production mechanisms for the DM particles: non-resonant production (Dodelson-Widrow mechanism [64]), thermal WDM production in the early universe (as outlined in [61, 63]) and resonant production with low lepton asymmetry. The bounds for all combinations of background DM production and mediator models are shown in appendix D. For this model, we show the results of the predicted total number of subhaloes for the MW according to the analysis in section 4.1 in the top panel, and the results of relative suppression with respect to Λ CDM regarding Lyman- α observations in the bottom panel, according to the analysis in section 4.2. For each plot, we show in markers the individual simulated cosmologies corresponding to each pair of parameters: DM mass and effective coupling constant. Empty triangles correspond to a case of relativistic (standard) decoupling of self interactions while circles denote a case of non-relativistic self decoupling (as outlined in section 2.3 and section 3.2). We show in colors the results for N_{sub} and δA , and we plot as well the level curves for an interpolating function over all data points. In grey we mark the regions of parameter space ruled out by these analyses: for the upper panels the lower shaded regions mark $N_{sub} \leq 57$ and for the lower panels they mark the conservative constraint $\delta A \geq 0.38$. We plot also in the lower panels the tighter constraint of $\delta A \geq 0.21$, according to [81]. The upper grey shaded regions in all panels mark regions excluded by Bullet Cluster constraints [78]: we take the approximate bound $\sigma/m \leq 0.1 \text{cm}^2/g$, with σ the non-relativistic self interaction cross section today, taken as $\sigma \sim 0.144 C_v^2/m^2$ for the vector field mediator and $\sigma \sim C_m^2/m^2$ for the massive scalar mediator.

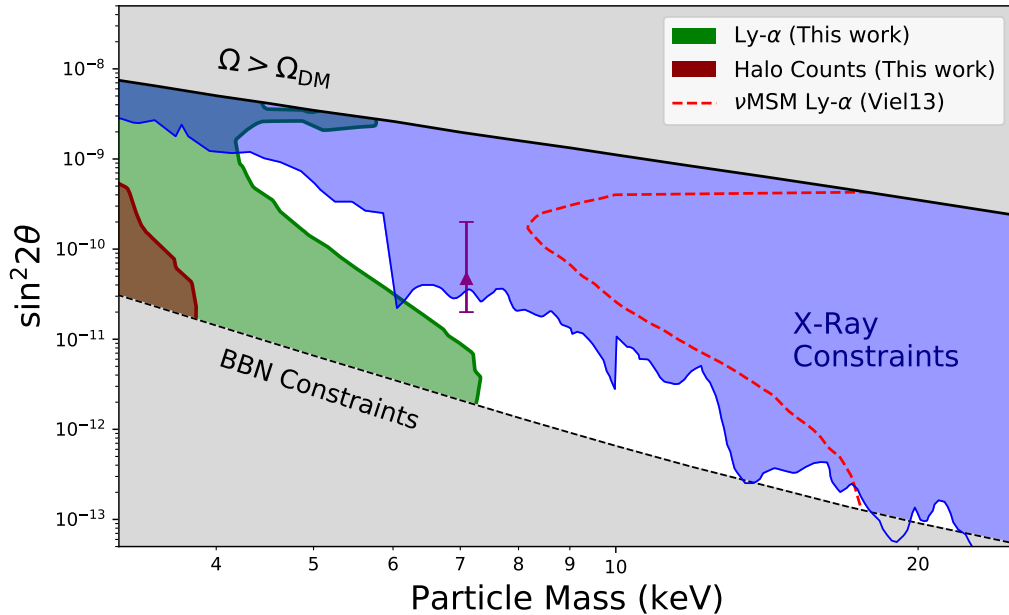


Figure 7: Parameter space constraints for ν MSM, where MW satellite halo counts and Lyman- α forest bounds are analyzed under a self interacting model as outlined above. For each point (θ, m) in the parameter space we consider a self interacting model under a vector field mediator, with its interaction constant given by $\sigma/m \sim 0.144C_v^2/m^3 = 0.1\text{cm}^2/g$, the upper limit given by Bullet Cluster constraints [78]. A particular point in the parameter space is considered to be admitted by any given analysis if it is so for this maximal model. For comparison, we plot the Lyman-alpha bounds for the non interacting case for a comparable analysis, according to the results in [80]. We also plot other bounds to the ν MSM parameter space for informative purposes, namely X-Ray indirect detection bounds [34, 43, 44] (in blue) and sterile neutrino production bounds [31, 41] (in grey), again reminding the reader that these bounds do not include the effects of self interactions in sterile neutrino production. Also for informative purposes, we plot the sterile neutrino model compatible with a tentative 3.5 keV DM signal, subject of debate in recent years, as a purple triangle [86–88].

References

- [1] Planck Collaboration, N. Aghanim, Y. Akrami, M. Ashdown, J. Aumont, C. Baccigalupi et al., *Planck 2018 results. VI. Cosmological parameters*, *arXiv e-prints* **641** (2020) A6 [[1807.06209](#)].
- [2] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu et al., *Introducing the Illustris Project: Simulating the coevolution of dark and visible matter in the Universe*, *Mon. Not. Roy. Astron. Soc.* **444** (2014) 1518 [[1405.2921](#)].
- [3] G. Bertone and T. Tait, *A new era in the search for dark matter*, *Nature* **562** (2018) .
- [4] P. Bull and Others, *Beyond Λ CDM: Problems, solutions, and the road ahead*, *Phys. Dark Univ.* **12** (2016) 56 [[1512.05356](#)].
- [5] M. Milgrom, *MOND vs. dark matter in light of historical parallels*, 2019, [1910.04368](#).
- [6] P. Hut and S. D. M. White, *Can a Neutrino Dominated Universe be Rejected?*, *Nature* **310** (1984) 637.
- [7] N. A. Bahcall, J. P. Ostriker, S. Perlmutter and P. J. Steinhardt, *The Cosmic triangle: Assessing the state of the universe*, *Science* **284** (1999) 1481 [[astro-ph/9906463](#)].
- [8] T. Tröster and Others, *Cosmology from large-scale structure: Constraining Λ CDM with BOSS*, *Astron. Astrophys.* **633** (2020) L10 [[1909.11006](#)].
- [9] G. D’Amico, J. Gleyzes, N. Kokron, K. Markovic, L. Senatore, P. Zhang et al., *The cosmological analysis of the SDSS/BOSS data from the Effective Field Theory of Large-Scale Structure*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) 5 [[1909.05271](#)].
- [10] J. S. Bullock and M. Boylan-Kolchin, *Small-Scale Challenges to the Λ cDM Paradigm*, *Annual Review of Astronomy and Astrophysics* **55** (2017) 343 [[arXiv:1707.04256v2](#)].
- [11] B. F. Griffen, A. P. Ji, G. A. Dooley, F. A. Gómez, M. Vogelsberger, B. W. O’Shea et al., *The Caterpillar Project: A Large Suite of Milky Way Sized Halos*, *The Astrophysical Journal* **818** (2016) 10.
- [12] W. Janesh, K. L. Rhode, J. J. Salzer, S. Janowiecki, E. A. K. Adams, M. P. Haynes et al., *Five Gas-rich Ultrafaint Dwarf Galaxy Candidates Discovered in WIYN Imaging of ALFALFA Sources*, *The Astronomical Journal* **157** (2019) 183.
- [13] The DES Collaboration, A. Drlica-Wagner, K. Bechtol, E. S. Rykoff, E. Luque, A. Queiroz et al., *Eight Ultra-faint Galaxy Candidates Discovered in Year Two of the Dark Energy Survey*, *Astrophys. J.* **813** (2015) 109 [[1508.03622](#)].
- [14] M. Boylan-Kolchin, J. S. Bullock and M. Kaplinghat, *Too big to fail? The puzzling darkness of massive Milky Way subhaloes*, *Monthly Notices of the Royal Astronomical Society: Letters* **415** (2011) 1 [[1103.0007](#)].
- [15] S.-H. Oh, D. A. Hunter, E. Brinks, B. G. Elmegreen, A. Schruba, F. Walter et al., *High-resolution mass models of dwarf galaxies from LITTLE THINGS*, *Astronomical Journal* **149** (2015) [[1502.01281](#)].
- [16] J. F. Navarro, C. S. Frenk and S. D. M. White, *A Universal density profile from hierarchical clustering*, *Astrophys.J.* **490** (1997) 493 [[astro-ph/9611107](#)].
- [17] A. Klypin, J. Holtzman, J. Primack and E. Regos, *Structure formation with cold plus hot dark matter*, *Astrophys. J.* **416** (1993) 1 [[astro-ph/9305011](#)].
- [18] A. D. Dolgov, *Neutrinos in cosmology*, *Phys. Rept.* **370** (2002) 333 [[hep-ph/0202122](#)].
- [19] S. Tulin and H.-B. Yu, *Dark Matter Self-interactions and Small Scale Structure*, *Phys. Rept.* **730** (2018) 1 [[1705.02358](#)].

- [20] J. Zavala, M. Vogelsberger and M. G. Walker, *Constraining self-interacting dark matter with the Milky Way's dwarf spheroidals*, *Monthly Notices of the Royal Astronomical Society: Letters* **431** (2013) L20.
- [21] M. Vogelsberger, J. Zavala, F.-Y. Y. Cyr-Racine, C. Pfrommer, T. Bringmann and K. Sigurdson, *ETHOS - an effective theory of structure formation: dark matter physics as a possible explanation of the small-scale CDM problems*, *Mon. Not. Roy. Astron. Soc.* **460** (2016) 1399 [[1512.05349](#)].
- [22] M. Y. Khlopov, *Fundamental Particle Structure in the Cosmological Dark Matter*, *International Journal of Modern Physics A* **28** (2013) [[1311.2468](#)].
- [23] M. Khlopov, *What comes after the Standard Model?*, *Progress in Particle and Nuclear Physics* **116** (2021) 103824.
- [24] S. Tulin, H.-B. Yu and K. M. Zurek, *Resonant Dark Forces and Small Scale Structure*, *Physical Review Letters* **110** (2012) [[1210.0900](#)].
- [25] L. G. van den Aarssen, T. Bringmann and C. Pfrommer, *Is dark matter with long-range interactions a solution to all small-scale problems of Λ CDM cosmology?*, *Physical Review Letters* **109** (2012) [[1205.5809](#)].
- [26] D. N. Spergel and P. J. Steinhardt, *Observational evidence for self-interacting cold dark matter*, *Physical Review Letters* **84** (1999) 3760 [[9909386](#)].
- [27] M. A. Buen-Abad, G. Marques-Tavares and M. Schmaltz, *Non-Abelian dark matter and dark radiation*, *Physical Review D* **92** (2015) .
- [28] F.-Y. Cyr-Racine and K. Sigurdson, *Limits on Neutrino-Neutrino Scattering in the Early Universe*, *Physical Review D - Particles, Fields, Gravitation and Cosmology* **90** (2013) [[1306.1536](#)].
- [29] P. Bode, J. P. Ostriker and N. Turok, *Halo Formation in Warm Dark Matter Models*, *The Astrophysical Journal* **556** (2001) 93 [[0010389](#)].
- [30] M. R. Lovell, V. Eke, C. S. Frenk, L. Gao, A. Jenkins, T. Theuns et al., *The haloes of bright satellite galaxies in a warm dark matter universe*, *Monthly Notices of the Royal Astronomical Society* **420** (2012) 2318.
- [31] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens and O. Ruchayskiy, *Sterile Neutrino Dark Matter*, *Prog. Part. Nucl. Phys.* **104** (2019) 1 [[1807.07938](#)].
- [32] M. R. Lovell, S. Bose, A. Boyarsky, S. Cole, C. S. Frenk, V. Gonzalez-Perez et al., *Satellite galaxies in semi-analytic models of galaxy formation with sterile neutrino dark matter*, *Monthly Notices of the Royal Astronomical Society* **461** (2016) 60.
- [33] A. Schneider, R. Smith, A. Macciò and B. Moore, *Non-linear evolution of cosmological structures in warm dark matter models*, *MNRAS* **424** (2012) 684 [[1112.0330](#)].
- [34] A. Schneider, *Astrophysical constraints on resonantly produced sterile neutrino dark matter*, *Journal of Cosmology and Astroparticle Physics* **1604** (2016) 59 [[1601.07553](#)].
- [35] A. V. Macciò, S. Paduroiu, D. Anderhalden, A. Schneider and B. Moore, *Cores in warm dark matter haloes: a Catch 22 problem*, *Monthly Notices of the Royal Astronomical Society* **424** (2012) 1105 [[1202.1282](#)].
- [36] S. Shao, L. Gao, T. Theuns and C. S. Frenk, *The phase-space density of fermionic dark matter haloes*, *Monthly Notices of the Royal Astronomical Society* **430** (2013) 2346.
- [37] C. Argüelles, A. Krut, J. A. Rueda and R. Ruffini, *Novel constraints on fermionic dark matter from galactic observables II: galaxy scaling relations*, *Physics of the Dark Universe* **24** (2019) 100278.
- [38] R. Adhikari, M. Agostini, N. Ky, T. Araki, M. Archidiacono, M. Bahr et al., *A White Paper on keV Sterile Neutrino Dark Matter*, *JCAP* **1701** (2017) 25 [[1602.04816](#)].

- [39] X.-D. Shi and G. M. Fuller, *A New dark matter candidate: Nonthermal sterile neutrinos*, *Phys. Rev. Lett.* **82** (1999) 2832 [[astro-ph/9810076](#)].
- [40] M. Laine and M. Shaposhnikov, *Sterile neutrino dark matter as a consequence of ν MSM-induced lepton asymmetry*, *JCAP* **0806** (2008) 31 [[0804.4543](#)].
- [41] T. Venumadhav, F.-Y. Y. Cyr-Racine, K. N. Abazajian and C. M. Hirata, *Sterile neutrino dark matter: Weak interactions in the strong coupling epoch*, *Physical Review D* **94** (2016) 1 [[arXiv:1507.06655v1](#)].
- [42] K. Perez, K. C. Y. Ng, J. F. Beacom, C. Hersh, S. Horiuchi and R. Krivonos, *Almost closing the ν MSM sterile neutrino dark matter window with NuSTAR*, *Phys. Rev.* **D95** (2017) 123002 [[1609.00667](#)].
- [43] J. F. Cherry and S. Horiuchi, *Closing in on Resonantly Produced Sterile Neutrino Dark Matter*, *Physical Review D* **D95** (2017) 83015 [[1701.07874](#)].
- [44] K. C. Y. Ng, B. M. Roach, K. Perez, J. F. Beacom, S. Horiuchi, R. Krivonos et al., *New Constraints on Sterile Neutrino Dark Matter from NuSTAR M31 Observations*, *Physical Review D* **99** (2019) [[1901.01262](#)].
- [45] W. Enzi, R. Murgia, O. Newton, S. Vegetti, C. Frenk, M. Viel et al., *Joint constraints on thermal relic dark matter from a selection of astrophysical probes*, *MNRAS* **000** (2020) 1 [[2010.13802](#)].
- [46] L. Johns and G. M. Fuller, *Self-interacting sterile neutrino dark matter: the heavy-mediator case*, *Phys. Rev.* **D100** (2019) 23533 [[1903.08296](#)].
- [47] A. De Gouvêa, M. Sen, W. Tangarife and Y. Zhang, *Dodelson-Widrow Mechanism in the Presence of Self-Interacting Neutrinos*, *Phys. Rev. Lett.* **124** (2020) 81802 [[1910.04901](#)].
- [48] R. Yunis, C. R. Argüelles, N. E. Mavromatos, A. Moliné, A. Krut, M. Carinci et al., *Galactic Center constraints on self-interacting sterile neutrinos from fermionic dark matter ("ino") models*, *Physics of the Dark Universe* **30** (2020) [[2008.08464](#)].
- [49] A. Fitts, M. Boylan-Kolchin, B. Bozek, J. S. Bullock, A. Graus, V. Robles et al., *Dwarf galaxies in CDM, WDM, and SIDM: disentangling baryons and dark matter physics*, *Monthly Notices of the Royal Astronomical Society* **490** (2019) 962 [[1811.11791](#)].
- [50] R. Huo, *Matter Power Spectrum of Light Freeze-in Dark Matter: With or without Self-Interaction*, *Physics Letters, Section B: Nuclear, Elementary Particle and High-Energy Physics* **802** (2019) [[1907.02454](#)].
- [51] S. Heimersheim, N. Schöneberg, D. C. Hooper and J. Lesgourgues, *Cannibalism hinders growth: Cannibal Dark Matter and the S_8 tension*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) [[2008.08486](#)].
- [52] M. Garny, T. Konstandin, L. Sagunski and S. Tulin, *Lyman- α forest constraints on interacting dark sectors*, *Journal of Cosmology and Astroparticle Physics* **2018** (2018) [[1805.12203](#)].
- [53] D. Egana-Ugrinovic, R. Essig, D. Gift and M. LoVerde, *The Cosmological Evolution of Self-interacting Dark Matter*, *Journal of Cosmology and Astroparticle Physics* **2021** (2021) [[2102.06215](#)].
- [54] R. Yunis, C. R. Argüelles and D. L. Nacir, *Boltzmann hierarchies for self-interacting warm dark matter scenarios*, *Journal of Cosmology and Astroparticle Physics* **2020** (2020) [[2002.05778](#)].
- [55] J. Lesgourgues, *The Cosmic Linear Anisotropy Solving System (CLASS) I: Overview*, 2011.
- [56] I. M. Oldengott, C. Rampf and Y. Y. Y. Wong, *Boltzmann hierarchy for interacting neutrinos I: formalism*, *JCAP* **1504** (2015) 016 [[1409.1577](#)].
- [57] I. M. Oldengott, T. Tram, C. Rampf and Y. Y. Y. Wong, *Interacting neutrinos in cosmology: exact description and constraints*, *JCAP* **1711** (2017) 027 [[1706.02123](#)].

- [58] J. Lesgourgues and T. Tram, *The Cosmic Linear Anisotropy Solving System (CLASS) IV: efficient implementation of non-cold relics*, *Journal of Cosmology and Astro-Particle Physics* **2011** (2011) 032 [[1104.2935](#)].
- [59] E. Sellentin and R. Durrer, *Detecting the cosmological neutrino background in the CMB*, *Physical Review D - Particles, Fields, Gravitation and Cosmology* **92** (2014) [[1412.6427](#)].
- [60] M. A. Buen-Abad, R. Emami and M. Schmaltz, *Cannibal dark matter and large scale structure*, *Phys. Rev. D* **98** (2018) 083517 [[1803.08062](#)].
- [61] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese and A. Riotto, *Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest*, *Phys. Rev.* **71** (2005) 1 [[0501562](#)].
- [62] J. R. Primack and M. A. K. Gross, *Hot dark matter in cosmology*, [astro-ph/0007165](#).
- [63] H. Mo, F. C. van den Bosch and S. White, *Galaxy Formation and Evolution*. 2010.
- [64] S. Dodelson and L. M. Widrow, *Sterile Neutrinos as Dark Matter*, *Physical Review Letters* **72** (1993) 17 [[9303287](#)].
- [65] J. Gluza and M. Zralek, *Feynman rules for Majorana neutrino interactions*, *Phys. Rev. D* **45** (1992) 1693.
- [66] A. Denner, H. Eck, O. Hahn and J. Kublbeck, *Compact Feynman rules for Majorana fermions*, *Phys. Lett. B* **291** (1992) 278.
- [67] J. L. Feng, M. Kaplinghat and H.-B. Yu, *Sommerfeld enhancements for thermal relic dark matter*, *Physical Review D* **82** (2010) .
- [68] C.-P. Ma and E. Bertschinger, *Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges*, *The Astrophysical Journal* **455** (1995) 7 [[9506072](#)].
- [69] J. Bernstein, *KINETIC THEORY IN THE EXPANDING UNIVERSE*, Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge, U.K., 1988, [10.1017/CBO9780511564185](#).
- [70] E. W. Kolb and M. S. Turner, *Book-Review - the Early Universe, Sky and Telescope* **80** (1990) 381.
- [71] M. Cannoni, *Relativistic $\langle\sigma v_{rel}\rangle$ in the calculation of relics abundances: A closer look*, *Physical Review D* **89** (2014) .
- [72] P. Gondolo and G. Gelmini, *Cosmic abundances of stable particles: Improved analysis*, *Nucl. Phys.* **B360** (1991) 145.
- [73] B. W. Lee and S. Weinberg, *Cosmological lower bound on heavy-neutrino masses*, *Phys. Rev. Lett.* **39** (1977) 165.
- [74] S. R. De Groot, W. A. Van Leeuwen and C. G. Van Weert, *Relativistic Kinetic Theory. Principles and Applications*. Amsterdam, Netherlands: North-holland (1980) 417p, 1980.
- [75] P. L. Krapivsky, S. Redner and E. Ben-Naim, *A Kinetic View of Statistical Physics*. Cambridge University Press, 2010, [10.1017/CBO9780511780516](#).
- [76] S. Hofmann, D. J. Schwarz and H. Stöcker, *Damping scales of neutralino cold dark matter*, *Physical Review D* **64** (2001) .
- [77] S. Hannestad and R. J. Scherrer, *Selfinteracting warm dark matter*, *Phys. Rev.* **D62** (2000) 043522 [[astro-ph/0003046](#)].
- [78] C. R. Argüelles, N. E. Mavromatos, J. A. Rueda and R. Ruffini, *The role of self-interacting right-handed neutrinos in galactic structure*, *Journal of Cosmology and Astroparticle Physics* **1604** (2016) 38 [[1502.00136](#)].

- [79] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, second ed., 2002, [10.1137/1.9780898718027](https://doi.org/10.1137/1.9780898718027), [<https://epubs.siam.org/doi/pdf/10.1137/1.9780898718027>].
- [80] M. Viel, G. Becker, J. Bolton and M. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-alpha forest data*, *Phys. Rev. D* **88** (2013) 43502 [[1306.2314](https://arxiv.org/abs/1306.2314)].
- [81] R. Murgia, A. Merle, M. Viel, M. Totzauer and A. Schneider, *"non-cold" dark matter at small scales: A general approach*, *Journal of Cosmology and Astroparticle Physics* **2017** (2017) [[1704.07838](https://arxiv.org/abs/1704.07838)].
- [82] E. V. Karukes, M. Benito, F. Iocco, R. Trotta and A. Geringer-Sameth, *A robust estimate of the Milky Way mass from rotation curve data*, *Journal of Cosmology and Astroparticle Physics* **2020** (2019) [[1912.04296](https://arxiv.org/abs/1912.04296)].
- [83] J. Baur, N. Palanque-Delabrouille, C. Yèche, C. Magneville and M. Viel, *Lyman-alpha Forests cool Warm Dark Matter*, *Journal of Cosmology and Astroparticle Physics* **1608** (2016) 12 [[1512.01981](https://arxiv.org/abs/1512.01981)].
- [84] V. Iršič, M. Viel, M. G. Haehnelt, J. S. Bolton, S. Cristiani, G. D. Becker et al., *New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman- α forest data*, *Phys. Rev. D* **96** (2017) 23522 [[1702.01764](https://arxiv.org/abs/1702.01764)].
- [85] M. Viel, G. D. Becker, J. S. Bolton, M. G. Haehnelt, M. Rauch and W. L. W. Sargent, *How cold is cold dark matter? Small scales constraints from the flux power spectrum of the high-redshift Lyman-alpha forest*, *Physical Review Letters* **100** (2007) [[0709.0131](https://arxiv.org/abs/0709.0131)].
- [86] E. Bulbul, M. Markevitch, A. Foster, R. K. Smith, M. Loewenstein and S. W. Randall, *Detection of An Unidentified Emission Line in the Stacked X-ray spectrum of Galaxy Clusters*, *Astrophys. J.* **789** (2014) 13 [[1402.2301](https://arxiv.org/abs/1402.2301)].
- [87] T. E. Jeltema and S. Profumo, *Discovery of a 3.5 keV line in the Galactic Centre and a critical look at the origin of the line across astronomical targets*, *Mon. Not. Roy. Astron. Soc.* **450** (2015) 2143 [[1408.1699](https://arxiv.org/abs/1408.1699)].
- [88] N. Cappelluti, E. Bulbul, A. Foster, P. Natarajan, M. C. Urry, M. W. Bautz et al., *Searching for the 3.5 keV Line in the Deep Fields with Chandra : The 10 Ms Observations*, *The Astrophysical Journal* **854** (2018) 179 [[1701.07932](https://arxiv.org/abs/1701.07932)].

A Power Spectrum Oscillations and Tight Coupling

Here in this appendix we take a look at the phenomena of oscillations in the final linear power spectra we have obtained in section 3.3. This oscillatory phenomenon only appears at a reasonably high value of k , and it only appears when the self interaction constant reaches values high enough as to have a significant effect in the matter dominated era. Surprisingly, the frequency of these oscillations seems unchanged with increasing coupling constant and indeed retains similar shape along different interaction models, only ever changing with different particle masses. The interaction constant only seems to increase or decrease the amplitude of such oscillations.

Indeed, phenomena like these appear to be linked to the phenomenon of tight coupling and fluid approximations themselves, as these oscillations in the power spectra appear not only for SI-WDM, as we show in fig. 4, but also in interacting applications of CDM [21] or even fluid approximations to WDM itself [58]. Here, we will try to shed some light on this phenomenon by reviewing initial insights in [58] and expanding upon them in this work, relating them with SI-WDM itself.

We start by recalling some results initially shown in [58], where the authors developed the homogeneous part of the non-interacting Boltzmann hierarchy for WDM (see [68]) into a more convenient form:

$$\dot{\Psi} = \frac{qk}{\epsilon} \mathbf{A} \Psi \equiv \alpha(\tau) \mathbf{A} \Psi, \quad (\text{A.1})$$

where $\Psi = (\Psi_0, \Psi_1, \dots)$ and A is the following matrix:

$$A = \begin{bmatrix} 0 & -1 & & & & \\ 1/3 & 0 & -2/3 & & & \\ & \ddots & \ddots & \ddots & & \\ & & l/(2l+1) & 0 & -(l+1)/(2l+1) & \\ & & & \ddots & \ddots & \ddots \end{bmatrix} \quad (\text{A.2})$$

There, they note that the formal solution to this homogeneous system is indeed the following matrix exponential:

$$\Psi(\tau) = \mathbf{U} e^{\int_{\tau_i}^{\tau} d\tau' \alpha(\tau') \mathbf{D}} \mathbf{U}^{-1} \Psi(\tau_i) \quad (\text{A.3})$$

with \mathbf{D} a diagonal matrix and $\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^{-1}$. There, they show that the highest eigenvalue goes to $\pm i$ for $l_{max} \rightarrow \infty$ and the time dependent phase accompanying this oscillation is:

$$\phi(\tau) = k \int_{\tau_i}^{\tau} d\tau' \frac{q}{\epsilon(\tau')} \quad (\text{A.4})$$

In this appendix, we will start exploring this claim by analyzing the eigenvalue structure of the matrix A . Indeed, we can check in fig. A.1 that the eigenvalues of this matrix are purely complex, with its highest eigenvalues in modulo approaching $\pm i$ as $l_{max} \rightarrow \infty$ (we use $l_{max} = 500$, for higher values the shown structure simply becomes more densely populated). The eigenvectors also show a rich structure, alternating between real and complex and the

l -th component possessing l zeros along the range of eigenvalues. This implies that an initial condition in $l = 0$ (such as the initial condition for CPT we consider) would populate densely all the normal modes of oscillation in the system.

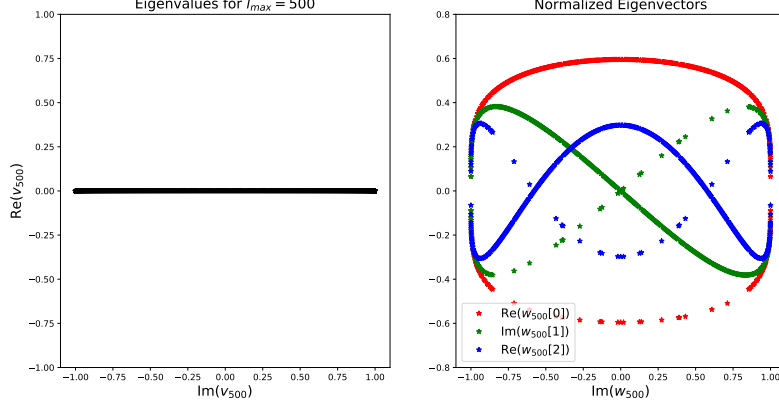


Figure A.1: Eigenvalue and eigenvector structure in the homogeneous Boltzmann Hierarchy without a damping term, for $l_{max} = 500$. The left panel shows the eigenvalue distribution for this case in the complex plane, with the real part of most eigenvalues being proportional to the damping term in the case $\Gamma \gg 1$. The right panel shows the zero, first and second components of the normalized eigenvectors in the case $l_{max} = 500$, as a function of the imaginary part of their eigenvalue. The real/imaginary parts of the eigenvalues not shown in the right hand plot are negligible.

But, what happens in the case of a damping term? In this case, the homogeneous system to resolve would be:

$$\dot{\Psi} = \alpha(\tau) [\mathbf{A}\Psi - \mathbf{\Gamma}] \quad (\text{A.5})$$

with the damping term $\mathbf{\Gamma} = \text{diag}(0, 0, \Gamma, \Gamma, \dots)$. In the case $\Gamma \gg 1$ the eigenvalue structure, shown in fig. A.2, becomes much different. Indeed, the only eigenvalues with a real part that is not equal to $-\Gamma \ll 0$ (thus, with modes that become highly suppressed during evolution) are the ones corresponding to the eigenvalues 0 and $\pm i/\sqrt{3}$. We can also see in the eigenvector structure that only these last two modes are the ones initially populated by a perturbation in the 0-th component (i.e. an initial density perturbation).

Thus, in this regime where a strong damping term is dominant over the homogeneous structure of the system, it will oscillate with a frequency of $\pm i/\sqrt{3}$ and a time dependent phase according to (A.4). This damped system in particular represents the tightly-coupled regime in the case of SI-WDM, but indeed as this term would have the overall effect of suppressing all Boltzmann hierarchy modes with $l \geq 2$ we can think of it as a representation of Boltzmann systems in general where higher modes are suppressed, such as what can happen in the various fluid or quasi fluid approximations for either WDM or other species. Obviously, according to the specific dynamics of the species the 0-th and first equations in the hierarchy may involve different terms but the oscillatory nature is a common characteristic.

This oscillatory nature of the system naturally has some impact in the power spectrum today. If this oscillation would continue until today, the linear power spectrum would be modulated by the oscillation in the density perturbation, itself governed by the oscillations

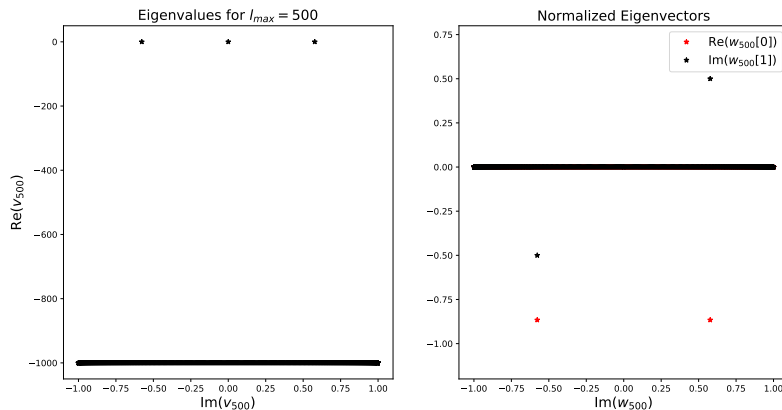


Figure A.2: Eigenvalue and eigenvector structure in the homogeneous Boltzmann Hierarchy with a damping term $\Gamma \sim 1 \times 10^3$, for $l_{max} = 500$. The left panel shows the eigenvalue distribution for this case in the complex plane, with the real part of most eigenvalues being proportional to the damping term in the case $\Gamma \gg 1$. The right panel shows the zero and first of the normalized eigenvalues in this case, as a function of the imaginary part of their eigenvalue. The real/imaginary parts of the eigenvalues not shown in the right hand plot are negligible.

in the 0-th component of the Boltzmann hierarchy. If we focus on only the matter power spectrum, and ignore other effects we obtain the following approximate expression

$$P_{osc}(k) \propto \delta\rho^2(k) = \left[4\pi T_{ncdm,0}^4 \int dq q^2 \epsilon \left(e^{-i\phi/\sqrt{3}} + e^{i\phi/\sqrt{3}} \right) f_0(q) \right]^2 \quad (\text{A.6})$$

where we see the both frequencies selected, $\pm 1/\sqrt{3}$, and this is just assuming that the WDM remains self coupled until today: the evolution after the self-decoupling is not considered here and is expected to modify the spectra. We can, however, compare this approximate power spectrum with the simulation results and we can see them side by side in fig. A.3 for the case without non-relativistic self decoupling and in fig. A.4 for the full scenario, using a background DF as in (2.23).

We can see in these figures that the frequency of the oscillation in the approximate form (A.6) is somewhat similar to the simulated data, even considering the surprising amount of approximations considered in the formula. However, the amplitude of the peaks and valleys are considerably different, especially in the first valley which has disappeared completely, which seems to point to more complex evolution governing the actual amplitude of the oscillations themselves. Nevertheless, we have gained some intuition about the origin of this behavior in the power spectrum. So, in summary:

- While the system is in tight-coupling (or potentially any other fluid approximation with negligible power in $l \geq 2$), the species shows an oscillatory nature that is imprinted in the power spectrum.
- This oscillatory nature is not present in the full hierarchy and, by extension, in non-interacting WDM as the system shows a dense distribution of eigenvalues and eigenvectors that are uniformly populated for $l_{max} \gg 1$.

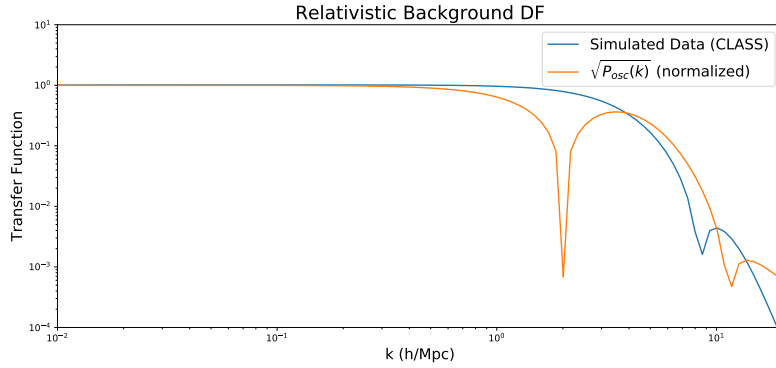


Figure A.3: Transfer function obtained with the approximate form (A.6) compared to simulated data using CLASS. We estimate the approximate transfer function using the value of $\sqrt{P_{osc}(k)}$ normalized to 1 for $k \ll 1$. In this case, we take a self interaction model such as in fig. 3 for the highest interaction constant, with a relativistic background distribution function and assuming $T_{SI-WDM} = T_{\nu_a}$ with T_{ν_a} the temperature of active neutrinos.

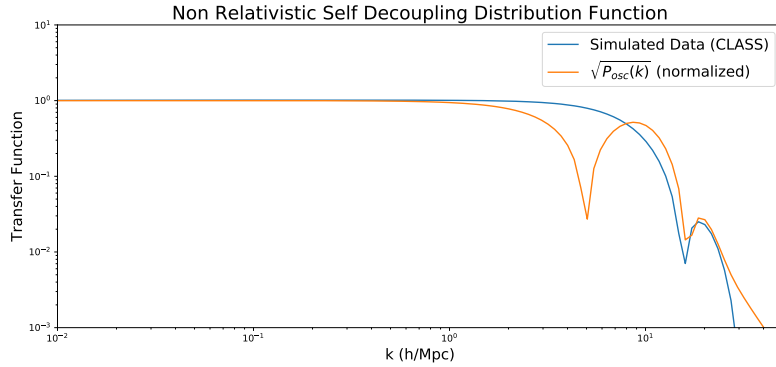


Figure A.4: Transfer function obtained with the approximate form (A.6) compared to simulated data using CLASS. We estimate the approximate transfer function using the value of $\sqrt{P_{osc}(k)}$ normalized to 1 for $k \ll 1$. In this case, we take a self interaction model such as in fig. 4 for the highest interaction constant, with a background distribution function from a non-relativistic self decoupling scenario, according to the derivations in 2.3 and 3.2.

- It is reasonable to assume that, after exiting the tight-coupling phase, free evolution erases some of the oscillations in the power spectrum, and that the longer the system remains tightly coupled the more evident these oscillations are.
- Questions still remain about the mechanisms that govern the amplitude of the oscillations themselves, and the differences in amplitude between different valleys in the power spectrum/transfer function.

B Sterile Neutrino WDM Model Subsets

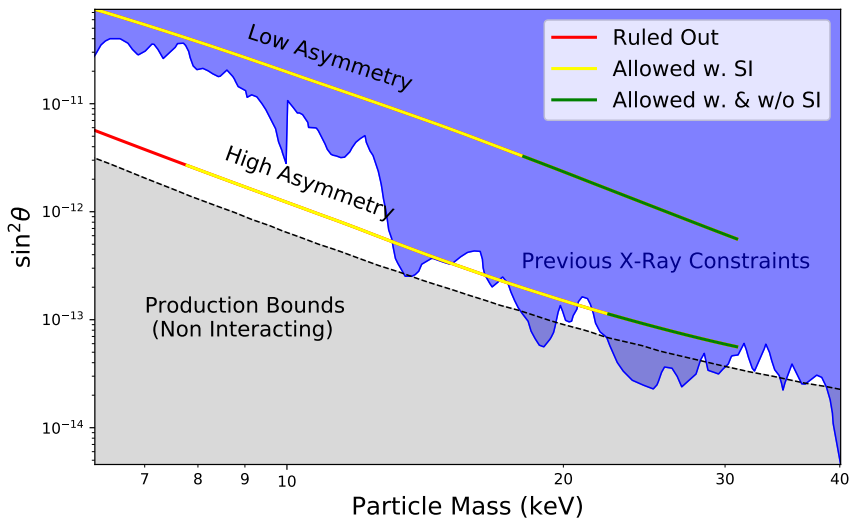


Figure B.1: Model subsets used for the resonant ν MSM production model runs plotted over the ν MSM model parameter space $\{\sin^2\theta, m\}$, where θ is the sterile-active neutrino mixing angle and m is the sterile neutrino mass. The models used here are sampled from the line labeled as “high asymmetry” for the high lepton asymmetry, low mixing angle models and from the line labeled as “low asymmetry” for the low lepton asymmetry, high mixing angle models. Over these lines, we distinguish the qualitative results of the analysis in section 4.1 and section 4.2. In red, we represent models that are excluded by this analysis, for any values of the SI coupling constant and any SI model. In yellow we represent models that are allowed for some SI models but become forbidden as $C_i \rightarrow 0$, i.e. models that were previously forbidden but become allowed by the inclusion of self interactions. Finally, in green we represent models that are allowed for any SI coupling constant and any model in the ranges tested. We also plot other bounds to the ν MSM parameter space for informative purposes, namely X-Ray indirect detection bounds [34, 43, 44] and sterile neutrino production bounds [31, 41], again reminding the reader that these bounds do not include the effects of self interactions in sterile neutrino production.

C Fitting Formulae for the Transfer Function Coefficients

	Production Mechanism	Relativistic Decoupling	non-relativistic Decoupling
Massive Scalar Mediator	Non-Resonant	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 4.20 \times 10^{-4}$, $b_\alpha = -0.894$ $a_\beta = 2.03$, $b_\beta = -0.0181$ $a_\gamma = -3.42$, $b_\gamma = 0.0101$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \cos^2 \left[(C/\text{eV}^{-4})/d_i + c_i \right]$ $a_\alpha = 1.14 \times 10^{-4}$, $b_\alpha = -9.45 \times 10^{-1}$, $c_\alpha = 0.583$, $d_\alpha = 1.002193 \times 10^{-32}$ $a_\beta = 2.17$, $b_\beta = -7.89 \times 10^{-3}$, $c_\beta = -6.17 \times 10^{-2}$, $d_\beta = 9.953530 \times 10^{-33}$ $a_\gamma = -1.79 \times 10^{-4}$, $b_\gamma = 0.134$, $c_\gamma = -14.1$, $d_\gamma = -2.692007 \times 10^{-31}$
	Resonant (High Asymmetry)	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 1.59 \times 10^{-4}$, $b_\alpha = -0.981$ $a_\beta = 2.13$, $b_\beta = 0.00663$ $a_\gamma = 3.02$, $b_\gamma = -0.173$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 8.48 \times 10^{-5}$, $b_\alpha = -0.694$, $c_\alpha = 0.207$, $d_\alpha = 1.871225 \times 10^{-33}$, $e_\alpha = 0.217$ $a_\beta = -0.128$, $b_\beta = -0.00326$, $c_\beta = 0.653$, $d_\beta = 8.591565 \times 10^{-32}$, $e_\beta = -16.5$ $a_\gamma = 18.3$, $b_\gamma = -0.967$, $c_\gamma = 1.262$, $d_\gamma = 1.192073 \times 10^{-32}$, $e_\gamma = -2.01$
	Resonant (Low Asymmetry)	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 9.62 \times 10^{-5}$, $b_\alpha = -1.01$ $a_\beta = 1.84$, $b_\beta = -0.0239$ $a_\gamma = -16.7$, $b_\gamma = 0.189$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 2.43 \times 10^{-5}$, $b_\alpha = -0.945$, $c_\alpha = 0.485$, $d_\alpha = 1.458468 \times 10^{-33}$, $e_\alpha = 0.411$ $a_\beta = -0.0567$, $b_\beta = -0.00130$, $c_\beta = 0.940$, $d_\beta = 1.583597 \times 10^{-31}$, $e_\beta = -37.9$ $a_\gamma = 178$, $b_\gamma = 0.00807$, $c_\gamma = 1.09$, $d_\gamma = 1.918661 \times 10^{-32}$, $e_\gamma = -2.10$
	Thermal	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 1.631 \times 10^{-5}$, $b_\alpha = -1.173$ $a_\beta = 2.023$, $b_\beta = -0.01332$ $a_\gamma = -2.919$, $b_\gamma = -0.01952$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 1.0956 \times 10^{-5}$, $b_\alpha = -1.0093$, $c_\alpha = -1.0257$, $d_\alpha = 6.50224 \times 10^{-32}$, $e_\alpha = 8.260 \times 10^{-2}$ $a_\beta = -1.1856 \times 10^{-1}$, $b_\beta = -4.141 \times 10^{-2}$, $c_\beta = 9.391 \times 10^{-1}$, $d_\beta = 1.244112 \times 10^{-31}$, $e_\beta = -1.65 \times 10^1$ $a_\gamma = 3.711 \times 10^1$, $b_\gamma = -6.82 \times 10^{-1}$, $c_\gamma = 1.077$, $d_\gamma = 4.77270 \times 10^{-33}$, $e_\gamma = -1.948$
	Non-Resonant	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 4.20 \times 10^{-4}$, $b_\alpha = -0.894$ $a_\beta = 2.03$, $b_\beta = -0.0181$ $a_\gamma = -3.42$, $b_\gamma = 0.0101$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \cos^2 \left[(C/\text{eV}^{-4})/d_i + c_i \right]$ $a_\alpha = 3.14 \times 10^{-4}$, $b_\alpha = -9.16 \times 10^{-1}$, $c_\alpha = 1.02$, $d_\alpha = 9.98 \times 10^{-33}$ $a_\beta = 2.27$, $b_\beta = -6.16 \times 10^{-3}$, $c_\beta = -0.201$, $d_\beta = 9.96 \times 10^{-33}$ $a_\gamma = -4.17 \times 10^4$, $b_\gamma = 5.11 \times 10^{-2}$, $c_\gamma = -14.0$, $d_\gamma = -2.79 \times 10^{-32}$
	Resonant (High Asymmetry)	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 1.59 \times 10^{-4}$, $b_\alpha = -0.981$ $a_\beta = 2.13$, $b_\beta = 0.00663$ $a_\gamma = 3.02$, $b_\gamma = -0.173$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 6.45 \times 10^{-5}$, $b_\alpha = -0.766$, $c_\alpha = 0.125$, $d_\alpha = 6.506785 \times 10^{-34}$, $e_\alpha = 0.211$ $a_\beta = -0.126$, $b_\beta = -0.00282$, $c_\beta = 0.454$, $d_\beta = 1.777554 \times 10^{-32}$, $e_\beta = -17.1$ $a_\gamma = 24.2$, $b_\gamma = -0.898$, $c_\gamma = 1.213926 \times 10^{-32}$, $e_\gamma = -2.01$
	Resonant (Low Asymmetry)	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 9.58 \times 10^{-5}$, $b_\alpha = -1.01$ $a_\beta = 1.83$, $b_\beta = -0.0241$ $a_\gamma = -16.8$, $b_\gamma = 0.192$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 3.02 \times 10^{-5}$, $b_\alpha = -0.909$, $c_\alpha = 0.00988$, $d_\alpha = 3.831799 \times 10^{-34}$, $e_\alpha = 0.390$ $a_\beta = -0.0598$, $b_\beta = -1.021 \times 10^{-2}$, $c_\beta = 0.543$, $d_\beta = 5.533459 \times 10^{-32}$, $e_\beta = -36.3$ $a_\gamma = 166$, $b_\gamma = 0.00665$, $c_\gamma = 1.22$, $d_\gamma = 1.442606 \times 10^{-33}$, $e_\gamma = -2.08$
	Thermal	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i}$ $a_\alpha = 1.63 \times 10^{-5}$, $b_\alpha = -1.173$ $a_\beta = 2.022$, $b_\beta = -0.0134$ $a_\gamma = -2.92$, $b_\gamma = -0.0190$	$i \equiv a_i (m_{DM}/\text{MeV})^{b_i} \left\{ \text{erfc} \left[\log_{10}(C/\text{eV}^{-4})/d_i - c_i \right] + e_i \right\}$ $a_\alpha = 5.99 \times 10^{-6}$, $b_\alpha = -1.12$, $c_\alpha = -1.88$, $d_\alpha = 1.098403 \times 10^{-31}$, $e_\alpha = 8.32 \times 10^{-2}$ $a_\beta = -1.19 \times 10^{-1}$, $b_\beta = -4.6 \times 10^{-2}$, $c_\beta = 0.598$, $d_\beta = 6.87690 \times 10^{-32}$, $e_\beta = -1.60 \times 10^1$ $a_\gamma = 5.73$, $b_\gamma = -0.968$, $c_\gamma = 0.772$, $d_\gamma = 7.93625 \times 10^{-33}$, $e_\gamma = -1.95$

Table 2: Approximate fitting formula for the transfer function parameters in SI-WDM models, with respect to CDM cosmology, for all combinations of mediator model and production scenario analyzed in section 4. This table provides fitting functions for the parameters $\{\alpha, \beta, \gamma\}$ used in the parametrization (4.1) of the suppressed transfer function for “non cold” DM models. For each model/production mechanism combination, we provide the functional form that all parameters follow as a function of $\{m_{DM}, C_i\}$ with a number of free parameters, and the appropriate best fit values for each parameters (for example, for a given cell, a fitting formula for α is obtained by replacing the values of $\{a_\alpha, b_\alpha, \dots\}$ in the functional form at the top of the cell).

D Power Spectra and Parameter Space Constraints for Other Mediator and Background Models

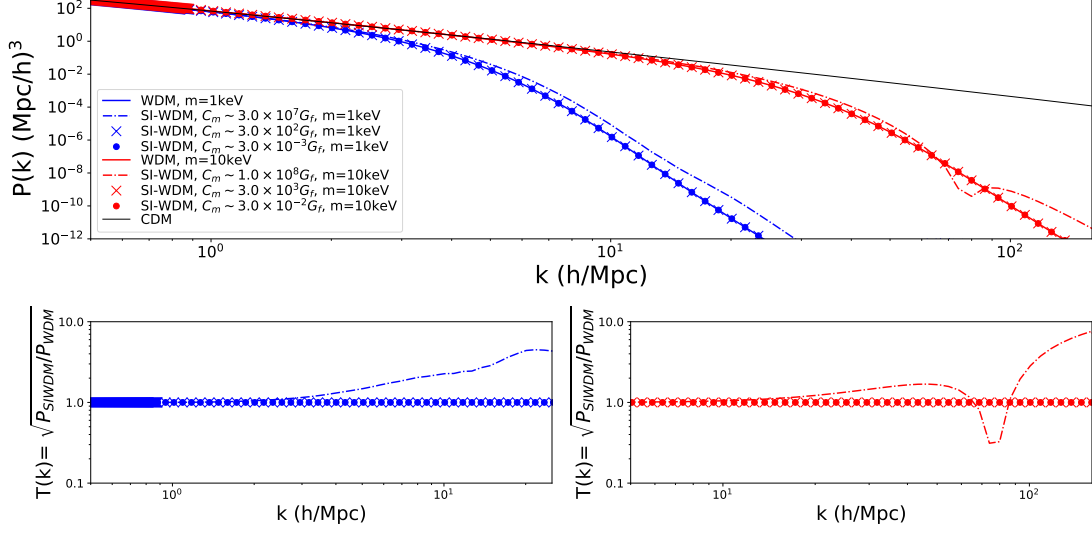


Figure D.1: Power Spectrum and Transfer Functions with respect to standard WDM for several SI-WDM models, ignoring the effects of non-relativistic self decoupling, for the case of a Massive Scalar Mediator. See the caption in fig. 3 for more information.

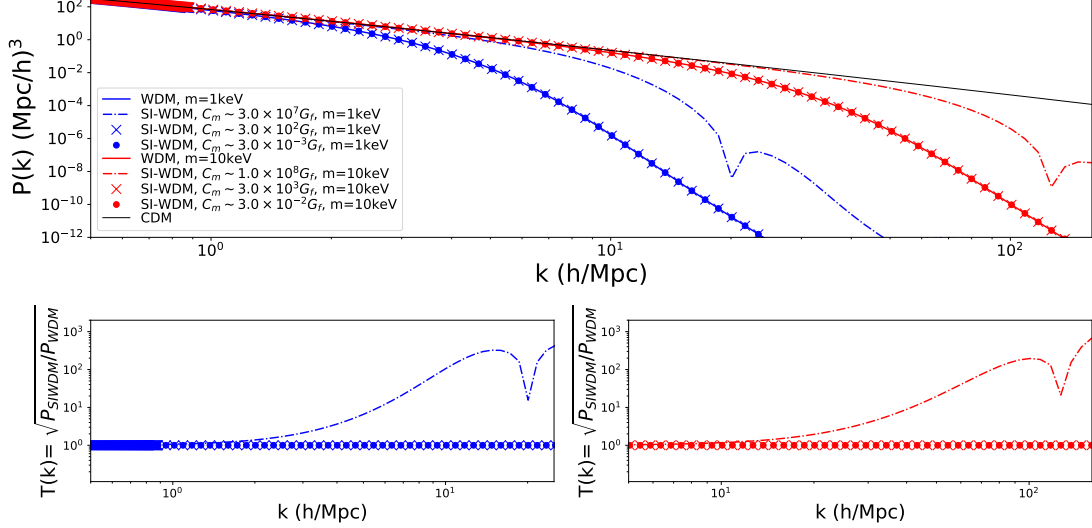
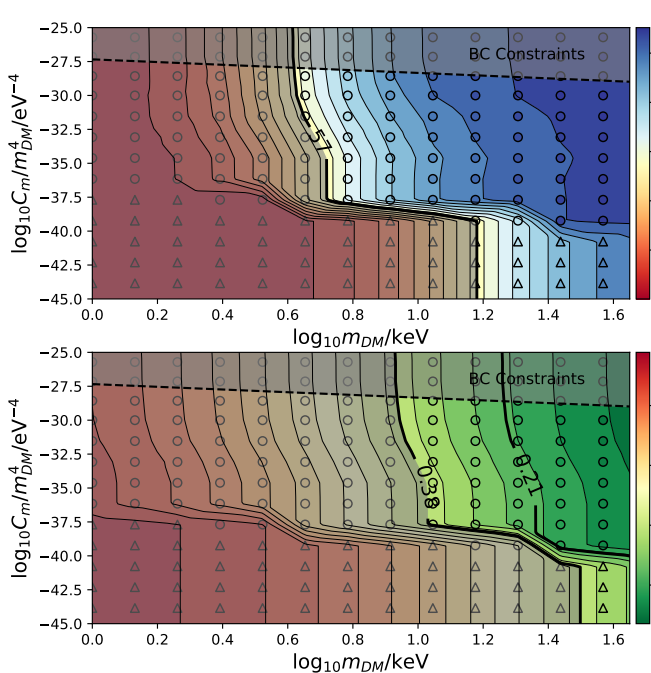
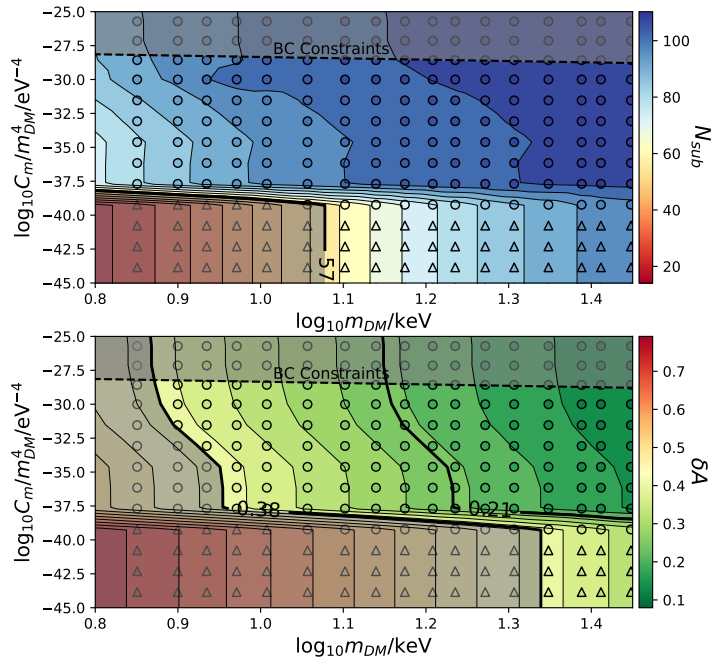


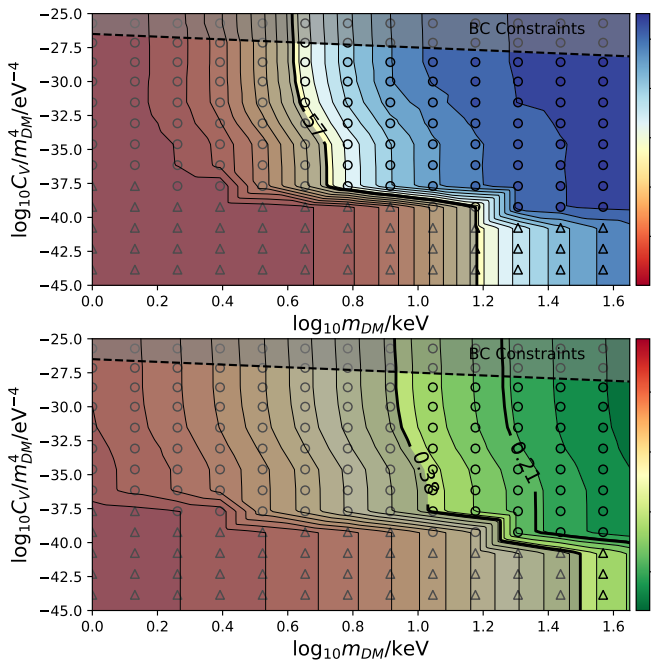
Figure D.2: Simulated Power Spectrum and Transfer Functions with respect to standard WDM for several SI-WDM models, considering the effects of non-relativistic self decoupling when appropriate, for the case of a Massive Scalar Mediator. See the caption in fig. 3 for more information.



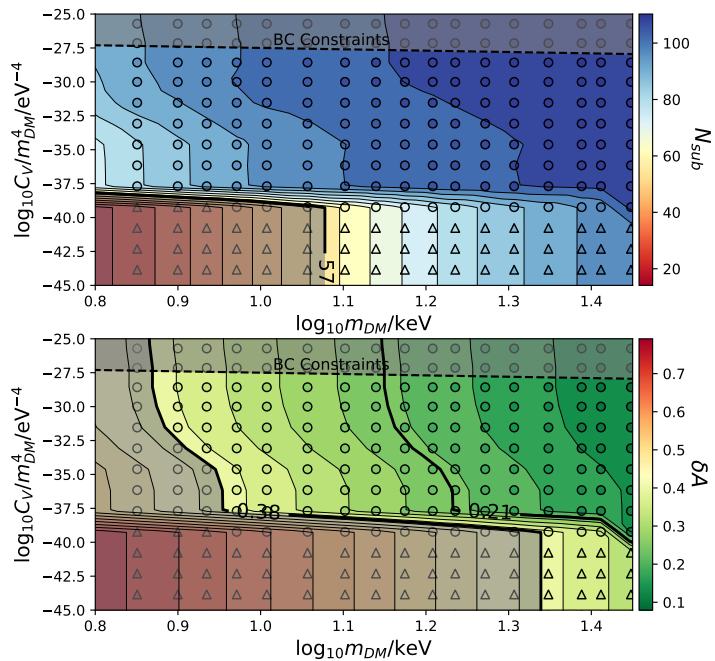
(a) Massive scalar mediator, non-resonant production



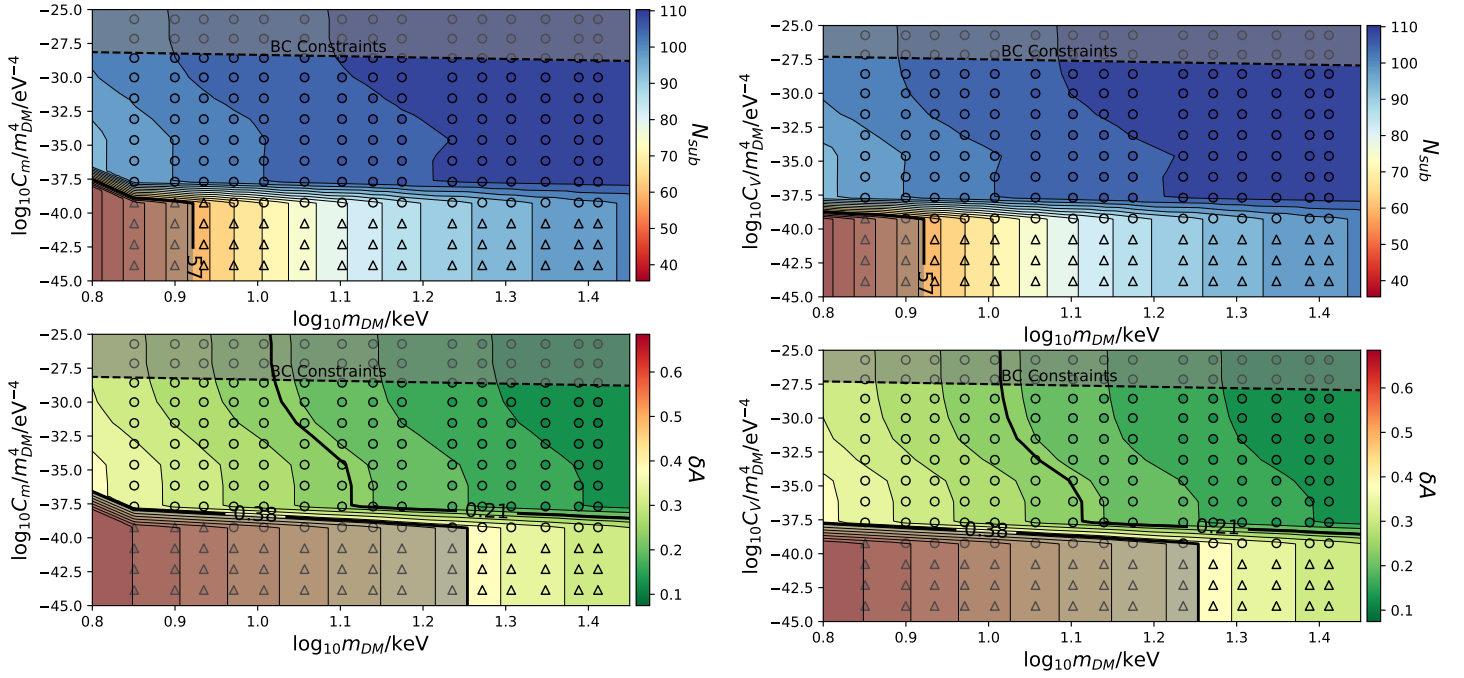
(b) Massive scalar mediator, resonant production (high asymmetry)



(c) Vector field mediator, non-resonant production

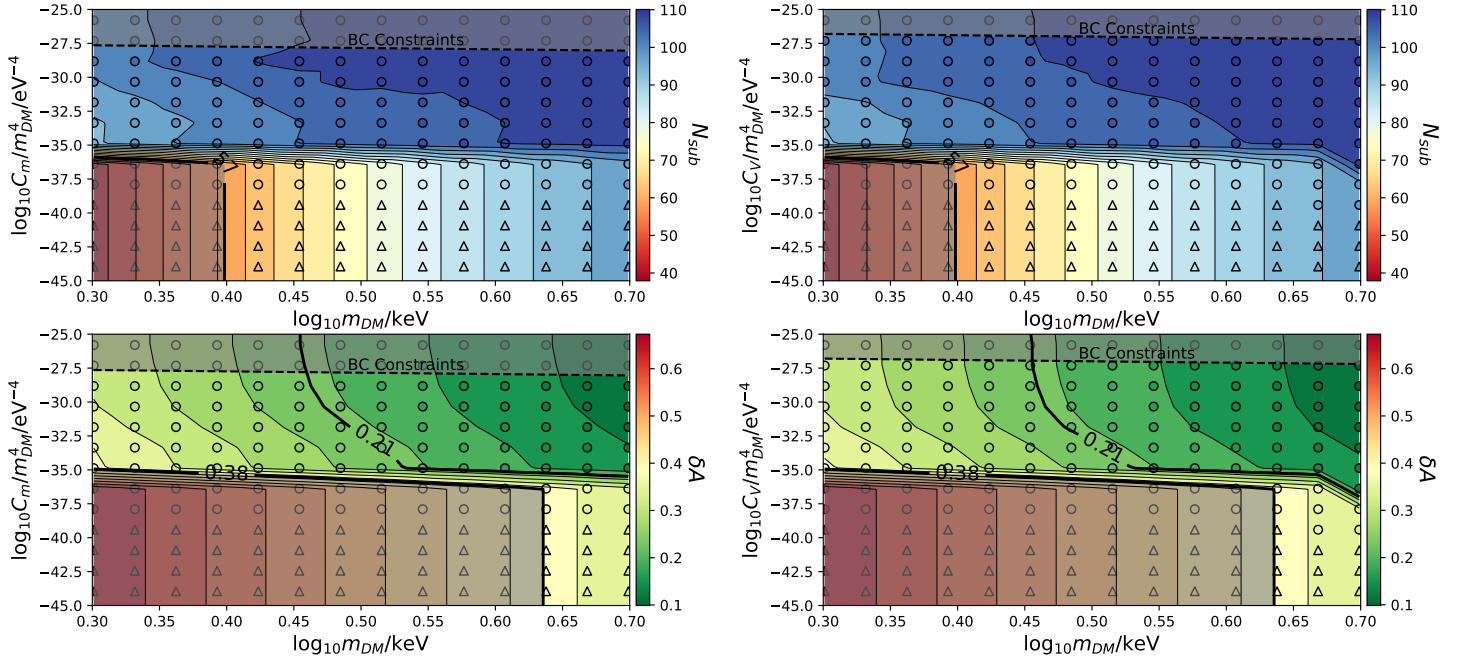


(d) Vector field mediator, resonant production (high asymmetry)



(e) Massive scalar mediator, resonant production (low asymmetry)

(f) Vector field mediator, resonant production (low asymmetry)



(k) Massive scalar mediator, thermal production

(l) Vector field mediator, thermal production

Figure D.3: Parameter space constraints for SI-WDM cosmologies, according to the analysis outlined in section 4, for the combinations of Mediator models and background/production models not considered in fig. 6. See the caption in fig. 6 for more information.