SISSA

Sterile neutrinos with secret interactions—lasting friendship with cosmology

To cite this article: Xiaoyong Chu *et al* JCAP10(2015)011

View the article online for updates and enhancements.

You may also like

Sterile neutrinos with secret interactions — lasting friendship with cosmology

Xiaoyong Chu,[a] **Basudeb Dasgupta**[b] **and Joachim Kopp**[c]

[a]International Center for Theoretical Physics,
 Strada Costiera 11, Trieste, 34014 Italy
[b]Tata Institute of Fundamental Research,
 Homi Bhabha Road, Mumbai, 400005 India
[c]PRISMA Cluster of Excellence and Mainz Institute for Theoretical Physics,
 Johannes Gutenberg University, Staudingerweg 7, Mainz, 55128 Germany

 E-mail: xchu@ictp.it, bdasgupta@theory.tifr.res.in, jkopp@uni-mainz.de

Abstract. Sterile neutrinos with mass $\simeq 1\,\mathrm{eV}$ and order 10% mixing with active neutrinos have been proposed as a solution to anomalies in neutrino oscillation data, but are tightly constrained by cosmological limits. It was recently shown that these constraints are avoided if sterile neutrinos couple to a new MeV-scale gauge boson A'. However, even this scenario is restricted by structure formation constraints when A'-mediated collisional processes lead to efficient active-to-sterile neutrino conversion after neutrinos have decoupled. In view of this, we reevaluate in this paper the viability of sterile neutrinos with such "secret" interactions. We carefully dissect their evolution in the early Universe, including the various production channels and the expected modifications to large scale structure formation. We argue that there are two regions in parameter space — one at very small A' coupling, one at relatively large A' coupling — where all constraints from big bang nucleosynthesis (BBN), cosmic microwave background (CMB), and large scale structure (LSS) data are satisfied. Interestingly, the large A' coupling region is precisely the region that was previously shown to have potentially important consequences for the small scale structure of dark matter halos if the A' boson couples also to the dark matter in the Universe.

Contents

1 Introduction

The possible existence of extra, "sterile", neutrino species with masses at the eV scale and $\mathcal{O}(10\%)$ mixing with the Standard Model (SM) neutrinos is one of the most debated topics in neutrino physics today. Several anomalies in neutrino oscillation experiments [1–6] seem to point towards the existence of such particles, but null results from other experiments that did not observe a signal cast doubt on this hypothesis [7–11]. A multi-faceted experimental program is under way to clarify the issue and either detect, or conclusively rule out, eV-scale sterile neutrinos with large mixing angle.

If the SM is indeed augmented with one or several such sterile neutrinos, but nothing else, some of the tightest constraints come from cosmological observations. In particular, measurements of the effective number of relativistic particle species in the primordial plasma, $N_{\rm eff}$ [12, 13], disfavor the existence of an abundance of light or massless particles beyond the SM neutrinos and the photon in the early Universe. If sterile neutrinos are at the eV scale or above, they are also constrained by the distribution of large scale structure (LSS) in the Universe [14] which would be washed out due to efficient energy transport over large distances by free-streaming neutrinos.

Cosmology, however, only constrains particle species that are abundantly produced in the early Universe. In two recent letters [15, 16], it was demonstrated that the production of sterile neutrinos can be suppressed until relatively late times if they are charged under a new interaction. This idea has elicited interest in detailed model building and has interesting phenomenological consequences [17–25]. However, Mirizzi et al. [22] have recently pointed out that collisions mediated by the new interaction can result in significant late time production of sterile neutrinos and lead to tensions with CMB and LSS data on structure formation. There are, however, several important caveats to this statement. In particular, the bounds from ref. [22] can be evaded if the sterile neutrinos either never recouple with active neutrinos or remain collisional until matter-radiation equality. (We communicated on these caveats with the authors of ref. [22], who were aware of the first possibility but did not mention it as

they found it less interesting in the context of previous literature. They mention the second possibility in the final version of their paper.)

Our aim in the present paper is to understand in detail the role and impact of sterile neutrino collisions, and reevaluate if secretly interacting sterile neutrinos remain cosmologically viable. Section 2 is a review of the main features of self-interacting sterile neutrino scenarios. In section 3, we calculate the additional contribution to $N_{\rm eff}$ at the BBN and CMB epochs, largely confirming previous calculations, but with some new results. In section 4 we consider the impact on the large scale structure in the Universe, focusing in particular on the sensitivity of Sloan Digital Sky Survey (SDSS) and Lyman-α data. We find that there are two regions of parameter space where sterile neutrinos with secret interactions are only weakly constrained. In section 5, we discuss our conclusions and summarize the results.

2 Secret interactions and sterile neutrino production

We assume that the Standard Model is augmented by a sterile neutrino ν_s with mass m_s,[1] and with order 10% mixing with the SM neutrinos. We moreover assume the existence of a new secret $U(1)_s$ gauge interaction, mediated by a vector boson A' of mass M at the MeV scale and coupling to sterile neutrinos through an interaction of the form

$$\mathcal{L}_{\rm int} = e_s \bar{\nu}_s \gamma^\mu P_{\rm L} \nu_s A'_\mu \,. \tag{2.1}$$

Here, e_s is the $U(1)_s$ coupling constant and $P_{\rm L} = (1 - \gamma^5)/2$ is the left-handed chirality projection operator. We define the secret fine structure constant as $\alpha_s \equiv e_s^2/(4\pi)$. The vector boson A' can obtain its mass either via the Stückelberg mechanism [26] or through spontaneous breadking of $U(1)_s$. In any case, a small breaking of $U(1)_s$ is required to allow for mixing between the active and sterile neutrinos, and this will typically require adding extra degrees of freedom to the sterile sector. Since we wish to remain agnostic regarding the precise mechanism that breaks $U(1)_s$, we will in the following analysis neglect these degrees of freedom. Nevertheless, we should keep in mind that the precise results for the sterile sector temperature and for the parameter $N_{\rm eff}$ (see section 3) will depend on them if they do not decouple sufficiently early.

This new interaction generates a large temperature-dependent potential [16]

$$V_{\rm eff} \simeq \begin{cases} -\frac{7\pi^2 e_s^2 E T_s^4}{45 M^4} & \text{for } T_s \ll M \\ +\frac{e_s^2 T_s^2}{8E} & \text{for } T_s \gg M \end{cases} \tag{2.2}$$

for sterile neutrinos of energy E and sterile sector temperature T_s. This potential leads to an in-medium mixing angle θ_m between active neutrinos $\nu_{\rm a}$ and sterile neutrinos ν_s, given by

$$\sin^2 2\theta_m = \frac{\sin^2 2\theta_0}{\left(\cos 2\theta_0 + \frac{2E}{\Delta m^2} V_{\rm eff}\right)^2 + \sin^2 2\theta_0} \,. \tag{2.3}$$

In the following, we will use a vacuum mixing angle $\theta_0 \simeq 0.1$ and an active-sterile mass squared difference $\Delta m^2 \simeq 1\,{\rm eV}^2$. As shown in [15, 16], the secret interactions can suppress θ_m, and thus active to sterile neutrino oscillations, until after neutrino decoupling as long as $|V_{\rm eff}| \gg |\Delta m^2/(2E)|$.

---

[1]Since ν_s is not a mass eigenstate, m_s actually means the mass of the fourth, mostly sterile, mass eigenstate.

The new interaction also leads to collisions of sterile neutrinos. The collision rate for $\nu_s \nu_s \leftrightarrow \nu_s \nu_s$ scattering is given by

$$\Gamma_{\rm coll} = n_{\nu_s}\sigma \sim \begin{cases} n_{\nu_s}e_s^4 \frac{E^2}{M^4} & \text{for } T_s \ll M \\ n_{\nu_s}e_s^4 \frac{1}{E^2} & \text{for } T_s \gg M \end{cases}, \tag{2.4}$$

where n_{ν_s} is the sterile neutrino density. The sterile neutrino production rate Γ_s and the final density depend on this collision rate. Two qualitatively different scenarios must be distinguished:

- *Collisionless production:* if the collision rate $\Gamma_{\rm coll}$ is smaller than the Hubble rate H at all times, the active and sterile neutrinos can be taken to be oscillating without scattering [27].[2] If $\Delta m^2/(2T_{\nu_{\rm a}}) \gg H$, ν_s are then produced only through oscillations, so that the final sterile neutrino number density is $n_{\nu_s} \simeq \frac{1}{2}\sin^2 2\theta_m\, n_{\nu_{\rm a}}$, where $n_{\nu_{\rm a}} = 3\zeta(3)/(4\pi^2)g_{\nu_{\rm a}}T_{\nu_{\rm a}}^3$ is the density of one of active neutrino flavors and $T_{\nu_{\rm a}}$ is the active neutrino temperatur. The final population of sterile neutrinos thus remains small, at most $\mathcal{O}(10^{-2})$ of the active neutrino density, because of the small mixing angle.

- *Collisional production:* if $\Gamma_{\rm coll}$ exceeds the Hubble rate H, then sterile neutrinos cannot be treated as non-collisional [28]. In each collision, the sterile component of a $\nu_{\rm a}$–ν_s superposition changes its momentum, separates from the $\nu_{\rm a}$ component, and continues to evolve independently. Subsequently, the active component again generates a sterile component, which again gets scattered. This process continues for all neutrinos until eventually the phase space distributions of $\nu_{\rm a}$ and ν_s have become identical. Thus, the fraction of $\nu_{\rm a}$ converted to sterile neutrinos is not limited by the mixing angle, and all neutrino flavors end up with equal number densities.

The $\nu_{\rm a} \to \nu_s$ production rate in this case is $\Gamma_s \simeq \frac{1}{2}\sin^2 2\theta_m \cdot \Gamma_{\rm coll}$ [29], where we can interpret the first factor as the average probability that an initially active neutrino is in its sterile state at the time of collision. The second factor gives the scattering rate that keeps it in the sterile state. We note that the production rate Γ_s is proportional to n_{ν_s} and thus rapidly approaches its final value,

$$\Gamma_s \simeq \frac{1}{2}\sin^2 2\theta_m \times \frac{3}{4}n_{\nu_{\rm a}}^{\rm SM} \cdot \begin{cases} e_s^4 \frac{E^2}{M^4} & \text{for } T_s \ll M \\ e_s^4 \frac{1}{E^2} & \text{for } T_s \gg M \end{cases}. \tag{2.5}$$

Note that, when $\Gamma_{\rm coll}$ is much larger than the oscillation frequency, using the average oscillation probability $\frac{1}{2}\sin^2 2\theta_m$ is inappropriate, and in fact the production rate Γ_s goes to zero in this case. Such a situation is, however, not realized for the parameter values explored in this work.

In the following, we will look at both collisionless and collisional production of sterile neutrinos in more detail,[3] with a special focus on the latter where more sterile neutrinos may be produced.

---

[2]We ignore the SM matter potential and scattering experienced by active neutrinos because we will be interested in the regime where the secret interaction dominates over the SM interaction.

[3]There is also the possibility that Mikheyev-Smirnov-Wolfenstein (MSW) type resonant effects, e.g., because of the sign-flip of the secret potential $V_{\rm eff}$ around $T_s \simeq M$, modify the ν_s production probability. In this work we treat all MSW transitions to be completely non-adiabatic and thus ignore them. A careful momentum-dependent treatment, which we defer to future work, is needed to accurately describe resonant conversion.

3 Constraints on N_{eff}

Cosmology is sensitive to the presence of relativistic degrees of freedom through their contribution to the overall energy density. At early times the sterile sector was presumably in equilibrium with the SM plasma, so that ν_s and A' were thermally populated. We assume that the sterile sector decouples from the SM sector well above the QCD scale and that oscillations remain suppressed until active neutrinos also decouple. Since the temperature T_γ of the SM sector drops more slowly than the sterile sector temperature T_s when extra entropy is produced during the QCD phase transition, T_s at BBN is significantly smaller than T_γ.

It is useful to track the ratio

$$\xi \equiv \frac{T_s}{T_\nu^{\text{SM}}} \tag{3.1}$$

of the sterile sector temperature T_s and the the temperature T_ν^{SM} of a standard neutrino. Before e^+e^- annihilation, $T_\nu^{\text{SM}} = T_\gamma$, while afterwards $T_\nu^{\text{SM}} = (4/11)^{1/3}T_\gamma$. Assuming comoving entropy is conserved, the ratio ξ at BBN is

$$
\xi_{\text{BBN}} = \begin{cases} \left(\frac{10.75}{106.7}\right)^{\frac{1}{3}} \left(\frac{2\cdot 7/8+3}{2\cdot 7/8}\right)^{\frac{1}{3}} & \text{for } M \gg 0.5\,\text{MeV} \\ \left(\frac{10.75}{106.7}\right)^{\frac{1}{3}} & \text{for } M \ll 0.5\,\text{MeV} \end{cases}
$$

$$
= \begin{cases} 0.649 & \text{for } M \gg 0.5\,\text{MeV} \quad (\text{case A}) \\ 0.465 & \text{for } M \ll 0.5\,\text{MeV} \quad (\text{case B}) \end{cases} . \tag{3.2}
$$

Here, the factor $(10.75/106.7)^{1/3}$ gives the ratio of the sterile sector temperature to the active sector temperature before A' decay, assuming that the two sectors have decoupled above the electroweak scale. It is based on counting the SM degrees of freedom that freeze out between the electroweak and BBN epochs.

A' is present in the Universe at the BBN epoch if $M \ll 3T_s|_{\text{BBN}} \simeq 0.5\,\text{MeV}$, and has decayed away if heavier. The factor $(2\cdot 7/8+3)/(2\cdot 7/8)$ in the first row of eq. (3.2) corresponds to the ratio of sterile sector degrees of freedom[4] before and after the decay of A' at $T_s \simeq M/3$.

The extra radiation in the Universe is parameterized as $N_{\text{eff}} \equiv (\rho_{\nu_{\text{a}}} + \rho_{\nu_s,A'})/\rho_\nu^{\text{SM}}$, i.e., the energy density of all non-photon relativistic species, measured in units of the energy density of a SM neutrino species. The primordial population of ν_s and A' leads to N_{eff} marginally larger than 3. For $M \gg 0.5$ MeV,

$$N_{\text{eff,BBN(A)}} = N_{\nu_{\text{a}}} + \xi^4_{\text{BBN(A)}} \simeq 3.22\,, \tag{3.3a}$$

at the BBN epoch. The first term, $N_{\nu_{\text{a}}}$, on the right hand side accounts for the active neutrinos and is equal to 3.045. The second term includes the relativistic sterile sector particles, i.e., only ν_s if $M \gg 0.5$ MeV. If the A' bosons are lighter, i.e., $M \ll 0.5$ MeV, they are present during BBN and contribute $g_{A'} = 3$ degrees of freedom in the sterile sector, in addition to the $g_{\nu_s} = 2 \times 7/8$ degrees of freedom of a sterile neutrino. Using the fact that also each active neutrino species has $g_{\nu_{\text{a}}} = 2 \times 7/8$ degrees of freedom, we find

$$N_{\text{eff,BBN(B)}} = N_{\nu_{\text{a}}} + \frac{g_{\nu_s} + g_{A'}}{g_{\nu_{\text{a}}}} \xi^4_{\text{BBN(B)}} \simeq 3.17\,, \tag{3.3b}$$
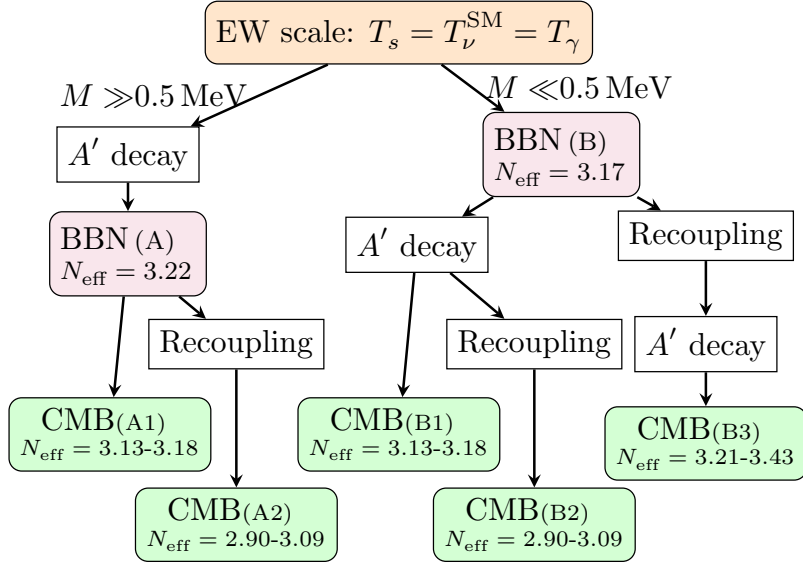
Figure 1. Possible cosmological histories of the active neutrinos ν_a, the sterile neutrinos ν_s, and the sterile sector gauge bosons A' below the electroweak (EW) scale. Various possibilities, labeled as A1, A2 and B1, B2, B3, are determined by the values of the A' mass M and the $U(1)_s$ fine structure constant α_s and lead to testable predictions for $N_{\rm eff}$ at both the BBN and CMB epochs. See text for details.

In figure 1, these two cases are summarized as BBN (A) and BBN (B), respectively. In either case, $N_{\rm eff,BBN(A/B)}$ remains consistent with the current BBN bound on extra radiation, $\Delta N_{\rm eff,BBN} = 0.66^{+0.47}_{-0.45}$ (68% C.L.) [12].

After BBN, the next important event is a possible secret *recoupling* of ν_a and ν_s. If the sterile neutrino production rate $\Gamma_s > H$, a new hotter population of ν_s can be collisionally produced from ν_a, and they achieve kinetic equilibrium with the primordially produced colder population of ν_s. Also, the A' can decay and heat up the sterile neutrinos. Both processes change the number and energy density of neutrinos, and $N_{\rm eff}$ at CMB depends on the order in which they occur.

In figure 2, we show the collisional $\nu_a \rightarrow \nu_s$ production rate Γ_s, normalized to the Hubble expansion rate H, as a function of the photon temperature T_γ: Γ_s/H is suppressed at high temperatures (say, above GeV), where $\sin^2 2\theta_m$ is small due to the large $V_{\rm eff}$. Since $\Gamma_s \propto T_\gamma^{-3}$ in this regime (see eqs. (2.2), (2.3) and (2.5)) and $H \propto T_\gamma^2$, Γ_s/H increases with T_γ^{-5} as the temperature decreases. We define the recoupling temperature $T_{\rm re}$ as the temperature where $\Gamma_s/H > 1$ for the first time since the primordial decoupling of the active and sterile sectors above the QCD phase transition. When $T_s \sim M$, the energy and temperature dependence of Γ_s changes (see eq. (2.5)), and when also $V_{\rm eff}$ drops below $\Delta m^2/(2T_s)$ at $T_s < M$, Γ_s begins to drop again. The asymptotic behavior is $\Gamma_s/H \propto T_\gamma^3$ at $T_s \ll M$ and $\theta_m \simeq \theta_0$. There are then three possible sequences of events:

1. *No recoupling*: for a sufficiently small interaction strength α_s, the scattering rate Γ_s always stays below the Hubble rate and there is no recoupling (solid black curve in figure 2). It roughly corresponds to $10^{-16} \leq \alpha_s \leq 10^{-11}$ and $M \leq O(1)\,{\rm GeV}$.

---

[4]Note that in complete models, for instance in scenarios including a dark Higgs sector to break the $U(1)_s$ symmetry, more degrees of freedom may need to be taken into account in the above equations.
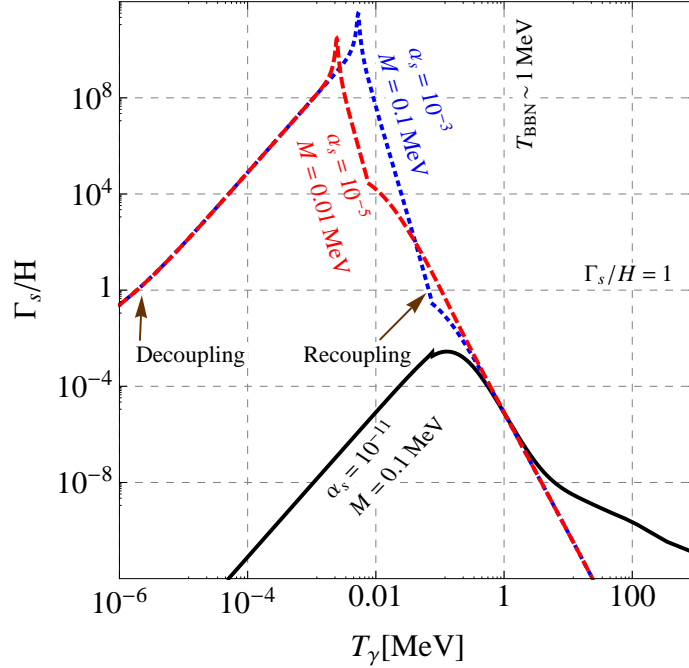
Figure 2. Evolution of the collisional $\nu_a \to \nu_s$ production rate Γ_s, normalized by the Hubble rate H, versus the photon temperature T_γ, for different representative choices of the secret gauge boson mass M and the secret fine structure constant α_s. When $\Gamma_s/H > 1$, collisional production of ν_s from the thermal bath of ν_a is effective. The solid black curve shows a case where this never happens. The shoulder around $T_\gamma \sim M$ is where the A' decay away. The dotted blue and dashed red curves correspond to recoupling after and before A' decay, respectively.

For even weaker interactions A' practically decouples, so the *recoupling* between ν_a and ν_s happens before BBN, which is excluded by N_{eff} measurements. On the other hand, if the interaction is stronger, a recoupling of ν_a and ν_s only happens after BBN, either *after* or *before* A' decay:

2. *Recoupling after A' decay*: if $M > \text{few} \times 10^{-2}\,\text{MeV}$, the recoupling happens after A' have decayed (dotted blue curve in figure 2).

3. *Recoupling before A' decay*: if $M < \text{few} \times 10^{-2}\,\text{MeV}$, the recoupling happens before A' have decayed (dashed red curve in figure 2).

In the second and third cases, there is also a secret *decoupling* when Γ_s/H again drops below one. If $e_s^2/M^2 \leq \mathcal{O}(10\,\text{MeV}^{-2})$, this decoupling happens while ν_s are still relativistic, i.e. $T_s \gtrsim m_s/3$.[5]

In the following, we discuss the three aforementioned cases in detail. Refs. [17, 18, 22, 23, 25] have previously obtained some of these results.

---

[5]In this paper, we will always assume this to be the case since we will find that the parameter region with $e_s^2/M^2 \geq \mathcal{O}(10\,\text{MeV}^{-2})$ is already disfavored by the requirement that active neutrinos should free stream sufficiently early [30] (see sections 4 and 5). If ν_s and ν_a are still coupled when the ν_s become non-relativistic, the mostly sterile mass eigenstate ν_4 will undergo a non-relativistic freeze-out and partly annihilate to pairs of mostly active neutrinos. Similarly, there is the possibility that the A' decay after the decoupling, but this does not happen for the range of parameters we will discuss here.

3.1 No recoupling

In the *no recoupling* cases, labeled as A1 and B1 in figure 1, the cosmological evolution after BBN is very straightforward. Vacuum oscillations convert a small fraction $\simeq \frac{1}{2}\sin^2 2\theta_0 \simeq 0.01$ of active neutrinos into sterile neutrinos (and vice versa), but this has negligible impact on the cosmological observables. Therefore, the temperature ratio ξ at CMB can be derived from the separate conservation of entropy in the active neutrino sector and in the sterile neutrino sector. It is independent of when the A' decay (provided that it happens before the CMB epoch and approximately in chemical equilibrium.) That is, $\xi_{\text{CMBA/B}} \simeq \xi_{\text{BBN(A)}} = 0.649$.

$N_{\text{eff,CMB}}$ can be estimated in analogy to eq. (3.3a). For the assumed sterile neutrino mass $\simeq 1\,\text{eV}$, the ν_s contribution to the relativistic energy density has to be weighted by an extra factor because they are already semi-relativistic at the CMB epoch, where the photon temperature is $T_\gamma \simeq 0.30\,\text{eV}$ and the kinetic temperature of the sterile sector is $T_s = \xi_{\text{CMB}} \cdot (4/11)^{1/3} T_\gamma \simeq 0.14\,\text{eV}$. As in [22], we assume that the extra weight factor is characterized by the pressure P. (See appendix A for the definition and calculations of the kinetic temperature and the pressure P used here.) We thus obtain

$$N_{\text{eff,CMB}} = N_{\nu_a} + \left.\frac{P_{m_s=1\,\text{eV}}}{P_{m_s=0}}\right|_{\text{CMB}} \cdot \xi_{\text{CMB}}^4 \simeq 3.13\,. \tag{3.4}$$

It is worth noting that the CMB temperature spectrum does not exactly measure the value of $N_{\text{eff,CMB}}$. Instead, the observed spectrum depends on the evolution of the energy density in relativistic degrees of freedom between the epoch of matter-radiation equality ($T_{\gamma,\text{eq}} \sim 0.7\,\text{eV}$) and recombination ($T_{\gamma,\text{CMB}} \simeq 0.30\,\text{eV}$) [13]. Therefore, the value of $N_{\text{eff,CMB}}$ measured from the CMB temperature power spectrum lies between the values of N_{eff} at $T_{\gamma,\text{CMB}}$ and $T_{\gamma,\text{eq}}$. The latter value, which we will denote by $N_{\text{eff,eq}}$, is given by

$$N_{\text{eff,eq}} = N_{\nu_a} + \left.\frac{P_{m_s=1\,\text{eV}}}{P_{m_s=0}}\right|_{\text{eq}} \cdot \xi_{\text{eq}}^4 \simeq 3.18\,. \tag{3.5}$$

Both of these values agree with the bound from the 2015 Planck data release, $N_{\text{eff}} = 3.15 \pm 0.23$ (68% C.L.) [13].

3.2 Recoupling after A' decay

The cases of *recoupling after A' decay* are labeled as A2 and B2 in figure 1. In both cases, entropy conservation in the sterile sector before recoupling leads to a temperature ratio just after A' decay of $\xi_{\text{M}} \simeq \xi_{\text{BBN(A)}} = 0.649$, which in turn implies

$$N_{\text{eff},M} = 3.045 + \xi_{\text{M}}^4 = 3.22\,. \tag{3.6}$$

Here, we have assumed that during A' decay chemical equilibrium holds in the sterile sector.

After recoupling, efficient neutrino oscillations and collisions lead to equilibration of the number densities and energy densities of all active and sterile neutrino species. Here, the flavor equilibration by oscillations is an irreversible process and does not conserve entropy. Nevertheless, since number-changing interactions are strongly suppressed at $T_s \ll M$, this recoupling cannot change the total (active + sterile) neutrino number density beyond what is necessitated by cosmological expansion. Moreover, relativistic neutrino energy densities

merely redshift as $1/a^4$ (a is the cosmic scale factor), guaranteed by general relativity, independently of the recoupling. This differs from the treatment in some previous papers [22, 25]. The kinetic temperature $T_{\nu,>\text{re}}$ shared by all neutrinos *after* recoupling is then given by

$$T_{\nu,>\text{re}} \simeq \frac{3.045 \cdot (T_{\nu,<\text{re}}^{\text{SM}})^4 + 1 \cdot T_{s,<\text{re}}^4}{3.045 \cdot (T_{\nu,<\text{re}}^{\text{SM}})^3 + 1 \cdot T_{s,<\text{re}}^3} \simeq 0.97 \, T_{\nu,<\text{re}}^{\text{SM}} \,, \tag{3.7}$$

where $T_{\nu,<\text{re}}^{\text{SM}}$ is the active neutrino temperature just *prior* to recoupling (which is at its SM value) and $T_{s,<\text{re}}$ denotes the sterile sector temperature just prior to recoupling.

Eventually, the mostly sterile eV-scale mass eigenstate decouples from the light mass eigenstates and becomes semi-relativistic at the CMB time. Its kinetic temperature at this epoch is $T_{s,\text{CMB}} \simeq 0.13 \, \text{eV}$. The effective number of relativistic species at the CMB epoch is given by

$$N_{\text{eff,CMB}} = N_{\text{eff},M} \left(\frac{3}{4} + \frac{1}{4} \frac{P_{m_s=1\,\text{eV}}}{P_{m_s=0}} \Big|_{\text{CMB}} \right)$$
$$\simeq 2.51 + 0.39 \simeq 2.90 \,. \tag{3.8}$$

Note that this is smaller than the SM value 3.045. This happens because part of the energy of active neutrinos has been transferred to the mostly sterile mass eigenstate ν_4, whose kinetic energy gets redshifted away more efficiently after it becomes non-relativistic. Refs. [22, 25] also found $N_{\text{eff}} < 3$ for this scenario. Similarly, we obtain for the time of matter-radiation equality:

$$N_{\text{eff,eq}} = 3.09 \,. \tag{3.9}$$

Both values are in agreement with the Planck bound [13].

3.3 Recoupling before A' decay

The last possibility, labeled as case B3 in figure 1, is that recoupling happens before A' decay. In this case, all neutrinos, together with A', reach a common chemical equilibrium, which lasts until most of the A' particles have decayed. During the formation of chemical equilibrium, again the total energy is conserved while entropy increases. Energy conservation allows us to calculate the temperature $T_{\nu,>\text{re}}$ of the active + sterile neutrino sector immediately after recoupling:

$$(3g_{\nu_a} + g_{\nu_s} + g_{A'}) \, T_{\nu,>\text{re}}^4 = \left[3g_{\nu_a} + (g_{\nu_s} + g_{A'}) \, \xi_{\text{BBN(B)}}^4 \right] (T_{\nu,<\text{re}}^{\text{SM}})^4 \,, \tag{3.10}$$

where $T_{\nu,<\text{re}}^{\text{SM}}$ is again the active neutrino temperature just prior to recoupling. Plugging in numbers for the effective numbers of degrees of freedom g_{ν_a}, g_{ν_s}, $g_{A'}$ and using $\xi_{\text{BBN(B)}} = 0.465$, we obtain

$$T_{\nu,>\text{re}} = 0.861 \, T_{\nu,<\text{re}}^{\text{SM}} \,. \tag{3.11}$$

Later, the A' decay and the thermal bath of neutrinos gets reheated by a factor $[(3g_{\nu_a} + g_{\nu_s} + g_{A'})/(3g_{\nu_a} + g_{\nu_s})]^{1/3} \simeq 1.125$. The effective number of relativistic species after A' decay is then

$$N_{\text{eff},M} \simeq \frac{3g_{\nu_a} + g_{\nu_s}}{g_{\nu_a}} (1.125 \cdot 0.861)^4 \simeq 3.568. \tag{3.12}$$

The next steps are the decoupling of sterile neutrinos and active neutrinos, and then the freeze-out of sterile neutrino self-interactions. Since the number densities and energy densities of the different species do not change during these decouplings, the total effective number of relativistic degrees of freedom at the CMB epoch is given by

$$N_{\mathrm{eff,CMB}} = N_{\mathrm{eff},M}\left(\frac{3}{4} + \frac{1}{4}\frac{P_{m_s = 1\,\mathrm{eV}}}{P_{m_s = 0}}\bigg|_{\mathrm{CMB}}\right) \simeq 3.21\,, \tag{3.13}$$

where again the pressure characterizes the contribution of the semi-relativistic ν_4 to the radiation density in the Universe. Similarly, at matter-radiation equality we have

$$N_{\mathrm{eff,eq}} \simeq 3.43\,. \tag{3.14}$$

This number is still within the 2σ error of the Planck bound [13].

4 Structure formation

Besides the constraints on extra radiation measured by N_{eff}, CMB data also prefers that most of the active (massless) neutrinos start to free-stream before redshift $z \sim 10^5$ [30, 31]. On the other hand, matter power spectrum observations forbid these free-streaming degrees of freedom from carrying so much energy as to suppress small scale structures [32]. Therefore measurements of the matter power spectrum put the most stringent upper bound on the mass of all fully thermalized neutrino species: $\sum m_\nu \lesssim 0.2$–$0.7\,\mathrm{eV}$ (95% C.L.) [13]. This concern [22] excludes a large proportion of the parameter region for self-interacting sterile neutrinos considered in [16]. However, like the constraint on N_{eff} discussed in section 3, it is avoided if Γ_s never exceeds H after the epoch when V_{eff} drops below the oscillation frequency (cases A1 and B1 in figure 1).

Interestingly, structure formation constraints are *also* significantly relaxed when the $\mathrm{U}(1)_s$ gauge coupling e_s is large and/or the gauge boson mass M is small. In this case, sterile neutrinos, although produced abundantly through collisional production (see section 2), cannot free-stream until late times, long after matter-radiation equality. Thus, their influence on structure formation is significantly reduced. Quantitatively, we will require that the sterile neutrinos remain collisional until *after* the CMB epoch at $T_\gamma \simeq 0.3\,\mathrm{eV}$. We will now discuss this observation in more detail.

After the active and sterile neutrinos have equilibrated through A'-mediated collisions, they should be treated as an incoherent mixture of the four mass eigenstates ν_i ($i = 1\ldots 4$). The reason is that for $m_4 \sim 1\,\mathrm{eV}$, their oscillation time scales are much smaller than both the Hubble time and the time interval between scatterings. For simplicity, assume that only the mostly sterile mass eigenstate ν_4 is massive with mass $m_4 \simeq 1\,\mathrm{eV}$, and that it only mixes appreciably with one of the mostly active mass eigenstates, say ν_1:

$$\nu_s \simeq \sin\theta_0\,\nu_1 + \cos\theta_0\,\nu_4\,. \tag{4.1}$$

We take the vacuum mixing angle to be $\theta_0 \simeq 0.1$ and we take into account that matter effects are negligible at temperatures relevant for structure formation (after matter-radiation equality). Since it is the flavor eigenstate ν_s that is charged under $\mathrm{U}(1)_s$, the mass eigenstates ν_1 and ν_4 interact with relative rates $\sin^2\theta_0$ and $\cos^2\theta_0$, respectively, while ν_2 and ν_3 essentially free-stream.

To study the influence of the secret interaction on structure formation, we estimate the mean comoving distance λ_s that each ν_4 can travel in the early Universe. Since neutrinos can transport energy efficiently over scales smaller than λ_s, the matter power spectrum will be suppressed on these scales. As long as neutrinos are collisional, they do not free stream, but diffuse over scales of order [33]

$$(\lambda_s^{\mathrm{coll}})^2 \simeq \int_0^{t_s^{\mathrm{dec}}} dt \, \frac{\langle v_s \rangle^2}{a^2(t)} \, \frac{1}{n_s \, \langle \sigma v \rangle_s} \,, \tag{4.2}$$

where $a(t)$ is the scale factor of the Universe, t_s^{dec} is the time at which sterile neutrino self-interactions decouple,

$$\langle \sigma v \rangle_s \sim \langle v_s \rangle \frac{e_s^4 \cos^2 \theta_0}{(M^2 + T_s^2)^2} \, (T_s + m_s)^2 \tag{4.3}$$

is the thermally averaged interaction cross section of the mostly sterile mass eigenstate ν_4, estimated here by naïve dimensional analysis, and n_s is the number density of sterile neutrinos. We have checked that the results do not deviate appreciably from the more accurate tree-level cross section. For simplicity, we take the kinetic temperature T_s of the sterile sector equal to the active neutrino temperature in this section, i.e. $T_s = T_\nu^{\mathrm{SM}} = (4/11)^{1/3} T_\gamma \propto a^{-1}(t)$, as long as ν_4 are relativistic. If sterile neutrinos become non-relativistic ($T_s < m_s$) while they are still strongly self-coupled, the kinetic temperature of the sterile sector scales as $T_s \propto a^{-2}(t)$ until T_s drops below $T_{s,\mathrm{dec}}$. After that, the sterile neutrino momenta are simply redshifted proportional to $a^{-1}(t)$. This implies in particular that, at $T_s \gg m_s$, we have $n_s \simeq T_s^3$, while after ν_4 become non-relativistic, but are still strongly coupled, this changes to $n_s \propto T_s^{3/2}$. The computation of the average velocity $\langle v_s \rangle$ of ν_4 entering eq. (4.3) is discussed in appendix A.

The decoupling temperature $T_{s,\mathrm{dec}}$ and the corresponding time t_s^{dec} are defined by the condition that the sterile neutrino interaction rate is just equal to the Hubble rate:

$$n_s \, \langle \sigma v \rangle_s \big|_{t=t^{\mathrm{dec}}} = H(t^{\mathrm{dec}}) \,. \tag{4.4}$$

After t^{dec}, sterile neutrinos start to free stream. The total comoving distance that a ν_4 travels between the time t^{dec} and the present epoch t^0 is [33]

$$\lambda_s^{\mathrm{fs}} = \int_{t^{\mathrm{dec}}}^{t_0} dt \, \frac{\langle v_s(t) \rangle}{a(t)} \,. \tag{4.5}$$

The overall damping scale is then given by

$$\lambda_s^2 = (\lambda_s^{\mathrm{coll}})^2 + (\lambda_s^{\mathrm{fs}})^2 \tag{4.6}$$

At scales larger than λ_s, structure formation is unaffected by the existence of sterile neutrinos, while at smaller scales, structures are washed out.

As a numerical example, for $M = 0.1\,\mathrm{MeV}$, $e_s = 0.1$, we find $\lambda_s^{\mathrm{coll}} \simeq 29\ \mathrm{Mpc}/h$, $\lambda_s^{\mathrm{fs}} \simeq 68\ \mathrm{Mpc}/h$ and thus

$$\lambda_s \simeq 74\ \mathrm{Mpc}/h \,, \tag{4.7}$$

corresponding to a wave number of

$$k_s \equiv 2\pi/\lambda_s \simeq 0.085\ h/\mathrm{Mpc} \,. \tag{4.8}$$

This should be compared to the free streaming scale of a decoupled sterile neutrino with a mass $\lesssim 1\,\text{eV}$,

$$k_s^{\text{no self-int.}} \simeq 0.018 \sqrt{\frac{m}{\text{eV}}} \, h/\text{Mpc} \,. \tag{4.9}$$

This factor of ~ 5 decrease in the free streaming scale compared to a conventional sterile neutrinos without self-interactions implies that data on large scale structure (LSS) and baryon acoustic oscillations (BAO) will be in much better agreement with our model than with sterile neutrino models that do not feature self-interactions. The strongest constraints will come from data probing very small scales, in particular Lyman-α forests.

Even at scales $k > k_s$, the suppression of the matter power spectrum $P_{\text{M}}(|\vec{k}|)$ does not set in abruptly, but increases gradually. For non-interacting sterile neutrinos, numerical simulations show that the suppression saturates at $k \simeq 50\,k_s$. At even smaller scales (even larger k), the deviation from the prediction of standard cosmology is [32, 37]

$$\frac{\delta P_{\text{M}}(|\vec{k}|)}{P_{\text{M}}(|\vec{k}|)} \simeq -8 f_\nu \tag{4.10}$$

in the linear structure formation regime. Here, $f_\nu = 3 m_s \zeta(3)/(2\pi^2)\,T_s^3(t_0) \times 8\pi G/(3H^2(t_0))/\Omega_m \simeq 0.07$ is the ratio of the sterile neutrino mass density Ω_s to the total mass density $\Omega_m \simeq 0.3$ today. In the regime of non-linear structure formation, $\delta P_{\text{M}}(|\vec{k}|)/P_{\text{M}}(|\vec{k}|)$ is somewhat larger [32, 37], but N-body simulations show that it decreases again at scales $k \gtrsim$ few h/Mpc [35, 38].

It is, however, difficult to directly measure $P_{\text{M}}(|\vec{k}|)$ at these nonlinear scales. The most sensitive data sets are Lyman-α forests, from which the 1-dimensional *flux* power spectrum $P_{\text{F}}(k)$ of Lyman-α photons can be extracted. Translating $P_{\text{F}}(k)$ into a measurement of $P_{\text{M}}(|\vec{k}|)$ requires a determination of the bias $b(k)$, which is obtained from numerical simulations of structure formation that include the dynamics of the gas clouds in which Lyman-α photons from distant quasars are absorbed. For SM neutrinos, such simulations have been performed for instance in [35], and we can estimate from figure 13 of that paper that the maximal suppression of $P_{\text{F}}(k)$ is of order

$$\frac{\delta P_{\text{F}}(k)}{P_{\text{F}}(k)} \sim -0.1 \times \left(\frac{\sum m_\nu}{1\ \text{eV}} \right) , \tag{4.11}$$

where $\sum m_\nu$ is the sum of all neutrino masses. This estimate is crude but conservative, and ignores the fact that the maximal suppression is actually smaller at lower redshifts. The suppression of $P_{\text{F}}(k)$ described by eq. (4.11) is *smaller* than the suppression of $P_{\text{M}}(|\vec{k}|)$ from eq. (4.10) because of the nonlinear k-dependent relation between the matter power spectrum and the flux power spectrum (see for instance [39], especially figure 16 in that paper). Since no dedicated simulations are available for our self-interacting sterile neutrino model, we will assume in the following that $\delta P_{\text{F}}(k)/P_{\text{F}}(k)$ saturates at the value given by eq. (4.11) even when $\sum m_\nu$ is dominated by the sterile neutrino mass m_s. This amounts to assuming that the impact of secretly interacting sterile neutrinos on these small scales is *qualitatively* similar to that of active neutrinos. A more detailed treatment requires a dedicated simulation including these secretly interacting sterile neutrinos. Note that neutrino free-streaming after CMB decoupling may lead to less suppression than described in eqs. (4.10) and (4.11) because perturbation modes well within the horizon have already grown significantly by that time. We will not include this effect in the following discussion to remain conservative.

We show the qualitative impact of self-interacting sterile neutrinos on large scale structure in figure 3. Panel (a) compares theoretical predictions in models with and without sterile neutrinos to data on the three-dimensional matter power spectrum $P_M(|\vec{k}|)$ from the Sloan Digital Sky Survey (SDSS) Luminous Red Galaxy (LRG) catalog [34]. Panel (b) compares to one-dimensional flux power spectra $\Delta^2(k) \equiv k\,P_F(k)/\pi$ from Lyman-α forest data [36]. SDSS-LRG data corresponds to a mean redshift of $z \simeq 0.35$, while Lyman-α data is split up according to redshift and reaches up to $z \simeq 5.4$. Note that the data in [36] is presented as a function of the wave number k_v in velocity space, measured in units of sec/km. The conversion to the wave number k in coordinate space, measured in units of h/Mpc, is done according to the formula $k = k_v\,H(z)/(1+z)$, where $H(z)$ is the Hubble rate at redshift z. The theoretical predictions for the SM with vanishing neutrino mass (solid green curves in figure 3) are taken from [34] and [36], respectively. Our (qualitative) predictions for sterile neutrino models with and without self-interactions are obtained in the following way: we start from the numerical prediction for the neutrino-induced suppression of the matter power spectrum from ref. [32]. In particular, we use the curve corresponding to $f_\nu = 0.07$ from figure 7 in that paper. We then shift this curve such that the onset of the suppression coincides with our calculated damping scale k_s (for the self-interacting scenario) or $k_s^{\text{no self-int.}}$ (for the conventional scenario without self-interactions). Moreover, we rescale the curves such that the maximal suppression is $-8f_\nu$ in figure 3a (linear regime) and 10% in figure 3b (nonlinear regime), see eqs. (4.10) and (4.11). We then multiply with the SM prediction to obtain the dotted green curves for sterile neutrinos without self-interactions and the red dashed curves for sterile neutrinos with self-interactions in figure 3. We use $e_s = 0.1$, $M = 0.1\,\text{MeV}$ and $m_s = 1\,\text{eV}$. Since we neglect a possible upturn of the power spectrum at $k \gtrsim 1\,h$/Mpc [38], our estimates are very conservative.

From figure 3a, we observe that the suppression of the matter power spectrum at scales $\lesssim 0.2\,h$/Mpc due to self-interacting sterile neutrinos is completely negligible, while a fully thermalized non-interacting sterile neutrino with the same mass leads to a clear suppression already at these scales. This implies that self-interacting sterile neutrinos with the parameters chosen here are not constrained by data on linear structure formation. Going to smaller scales or larger k (figure 3b), where non-linear effects become relevant, we see that both the sterile neutrino model with self-interactions and the one without lead to suppression, but the amount of suppression is reduced in the self-interacting case. It was shown in ref. [36] that the data disfavors suppression larger than 10% at $k = 10\,h$/Mpc. Self-interacting sterile neutrinos at the benchmark point shown in figure 3 appear to be marginally consistent with these constraints. It should be kept in mind, however, that our predictions are only qualitative. Therefore, only a detailed fit using simulations of non-linear structure formation that include sterile neutrino self-interactions could provide a conclusive assessment of the viability of such a scenario.

Let us finally discuss how the cosmological effects of the three active neutrinos are modified in the self-interacting sterile neutrino scenario. The dynamics of the mass eigenstates ν_2 and ν_3, which we assume not to mix with ν_4, is the same as in standard cosmology: they start to free stream at redshift $z \gg 10^5$. ν_1, however, starts to free stream later than a non-interacting neutrino, but earlier than ν_4. The free-streaming condition for ν_1 is, in analogy to eq. (4.4),

$$(T_\nu^{\text{SM}})^3 \cdot (T_\nu^{\text{SM}})^2 \left(\frac{e_s^2}{M^2} \right)^2 \sin^2 \theta_0 \lesssim H\,. \tag{4.12}$$
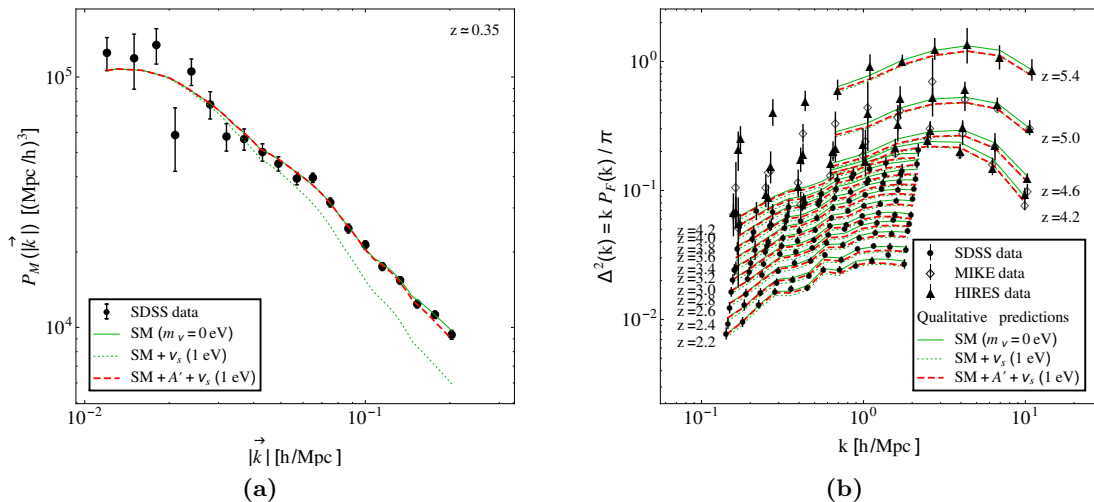
Figure 3. (a) Three-dimensional matter power spectrum $P_M(|\vec{k}|)$ derived from the SDSS Luminous Red Galaxy (LRG) Sample [34] and (b) one-dimensional flux power spectrum $\Delta^2(k) = k\,P_F(k)/\pi$ of Lyman-α photons at various redshifts, compared to the qualitative predictions of sterile neutrino models with (red dashed curves) and without (green dotted curves) self-interactions. Note that the relation between $P_M(|\vec{k}|)$ and $P_F(k)$ is non-linear, see e.g. [35]. The assumed self-interaction parameters are $e_s = 0.1$, $M = 0.1\,\text{MeV}$, and the assumed sterile neutrino mass is $1\,\text{eV}$. The data points and the SM prediction (solid green curves) are taken from [34] and and from [36], respectively. The predictions including sterile neutrinos are obtained by multiplying the SM predictions by the k-dependent suppression profile from figure 7 of [32], shifted such that the onset of the suppression is at our calculated damping scale k_s (eq. (4.8)) or $k_s^{\text{no self-int.}}$ (eq. (4.9)); see text for details), and scaled such that the maximum suppression is given by eq. (4.10) for panel (a) and by (4.11) for panel (b). Note that the error bars shown here are statistical only, and large systematic uncertainties, especially at small scales (large k) should be kept in mind.

As shown in ref. [30], free-streaming of active neutrinos before redshift $z \sim 10^5$ is required to sufficiently suppress the acoustic peaks in the CMB power spectrum. The change from three to two truly free streaming neutrino species in our model will lead to minor modifications of the CMB power spectrum, but the analysis from [30] suggests that these are unlikely to spoil the fit to CMB data, in particular since they may be compensated by changes in the best fit values of other cosmological parameters. For definiteness, we require that ν_1 starts to free-stream before the CMB epoch while $v_{2,3}$ free-stream before $z \sim 10^5$.

Note that also Planck CMB data alone, without including data on large scale structure observations, imposes an upper limit on the mass of sterile neutrinos, which, for a fully thermalized species is $m_s \lesssim 0.5\,\text{eV}$ at 95% C.L. [13]. A much weaker bound is expected if self- interactions among sterile neutrinos are so strong that they remain collisial until after the CMB epoch. In this case, the early Integrated Sachs-Wolfe (ISW) effect induced by ν_s perturbations at low multipole order ($50 \leq l \leq 200$) will be suppressed [32]. Thus, the main effect on the CMB will come from the shift of matter-radiation equality, to which the sensitivity is, however, much weaker.

To sum up, our discussion in this section suggests that there exists a limited parameter region where both *LSS* and *CMB* bounds on the sterile neutrino mass are considerably lifted. It requires the secret interaction to be so large that sterile neutrinos can only free-stream after recombination, but still small enough for ν_1 to free-stream sufficiently early. Numerically the

required interaction strength is

$$0.02 \lesssim \alpha_s \left(\frac{\text{MeV}}{M} \right)^2 \lesssim 0.2 \,. \tag{4.13}$$

If recoupling occurs after A' decay, the temperature of sterile neutrinos is slightly smaller than the value assumed in the derivation of this constraints. This implies less free-streaming, but this is not noticeable in practice.

We emphasize that a global analysis of the secret interaction scenario including a fit to both CMB and Lyman-α data is desirable. Such an analysis, which is well beyond the scope of the present work, would provide more robust constraints and would allow us to draw definitive conclusios regarding the viability of the parameter region indicated by eq. (4.13).

5   Discussion and conclusions

As we have seen in the previous section, there are two main scenarios in which self-interacting sterile neutrinos do not run into conflict with cosmological data:

(i) The ν_s production rate Γ_s drops below the Hubble expansion rate H before the effective potential $|V_{\text{eff}}|$ drops below the oscillation frequency $|\Delta m^2/(2E)|$ and the dynamic suppression of active-sterile mixing due to V_{eff} ends. In this case, sterile neutrinos are not produced in significant numbers in the early Universe and hence cosmology is not sensitive to their existence. An explanation of small scale structure anomalies as advocated in [16] is, however, not possible in this scenario.[6] In particular, even if the new interaction also couples to dark matter as proposed in [16], it is too weak to have phenomenological consequences.

This disadvantage can be avoided if more than one self-interacting sterile neutrino exists. Consider for example, a model with three mostly sterile neutrino mass eigenstates ν_4, ν_5, ν_6. Let ν_4 and ν_5 have a relatively large mixing $\theta_0 \sim 0.1$ with active neutrinos, as motivated for instance by the short baseline oscillation anomalies. Let their coupling to the A' gauge boson be $e_s^{(4,5)} \simeq 10^{-5}$, large enough to dynamically suppress their mixing with the mostly active mass eigenstates until after BBN, but small enough to prevent their equilibration afterwards. On the other hand, let ν_6 have a vanishing mixing with $\nu_{1,2,3}$, but a larger secret gauge coupling $e_s^{(6)} \simeq 0.1$. Due to its small mixing, it is never produced through oscillations. However, its primordial population — the relic density produced before the visible and sterile sectors decoupled in the very early Universe — still acts as a thermal bath to which the dark matter may be strongly coupled, thus potentially solving the missing satellites problem [42–46].

(ii) The self-interactions are so strong that sterile neutrinos remain collisional at least until matter-radiation equality. In this scenario, sterile neutrinos are produced when $|V_{\text{eff}}| \leq |\Delta m^2/(2E)|$. However, as shown in section 3, the effective number of relativistic degrees of freedom in the Universe, N_{eff}, remains close to 3 because equilibration between active and sterile neutrinos happens after neutrinos have decoupled from the photon bath. Moreover, as argued in section 4, the impact of self-interacting sterile neutrinos on structure formation is much smaller in this scenario than the impact of conventional

---

[6]Note that recent simulations of cosmological structure formation suggest that these anomalies may be resolved once baryons are included in the simulations [40, 41].

non-interacting sterile neutrino because they cannot transport energy efficiently over large distances due to their reduced free-streaming. Structure formation constraints could be further relaxed in models containing, besides an eV-scale mass eigenstate ν_4, one or more additional mostly sterile states with much lower masses [25]. It is intriguing that the parameter region corresponding to scenario *(ii)* contains the region where small scale structure anomalies can be explained, as shown in [16].

We summarize these results in figure 4. The yellow cross-hatched region on the right is ruled out because active and sterile neutrinos come into thermal equilibrium before the active neutrino decoupling from the SM plasma. In the lower part of this region, this happens simply because V_{eff} is negligibly small. In the upper part, V_{eff} is large, but also Γ_s is large so that collisional production of sterile neutrinos is efficient in spite of the suppressed in-medium mixing θ_m. This leads to constraints from N_{eff} and from the light element abundances in BBN [21]. In the blue vertically hatched region, sterile neutrinos are produced after ν_{a} decoupling, so that CMB constraints on N_{eff} remain satisfied. However, sterile neutrinos free-stream early on in this region and violate the CMB and structure formation constraints on their mass. This mass constraint can be considerably relaxed if the sterile neutrinos remain collisional until *after* the CMB epoch at $T_\gamma \simeq 0.3\,\text{eV}$. This defines the upper edge of the blue hatched region. In the red shaded region in the upper left corner, the secret interaction is too strong and ν_1 free streams too late. CMB data requires that active neutrinos free stream early enough, and thus strongly disfavors this region. Two white regions remain allowed: scenario *(i)* with weak self-interactions, corresponds to the wedge-shaped white region in the lower part of the plot. Scenario *(ii)*, with strong self-interactions, is realized in the thin white band between the blue vertically hatched region and the red shaded region. As explained above, systematic uncertainties at Lyman-α scales, and interplay of CMB and Lyman-α data, as well the possible existence of additional states with masses $\ll 1\,\text{eV}$, affect this conclusion and a definitive contraint requires further work.

There are several important caveats and limitations to the above analysis. First, we have only worked with thermal averages for the parameters characterizing each particle species, such as energy, velocity, pressure, etc. To obtain more accurate predictions, it would be necessary to solve momentum-dependent quantum kinetic equations. This would be in particular interesting in the temperature regions where V_{eff} changes sign and where $V_{\text{eff}} \sim \Delta m^2/(2T_{\nu_{\text{a}}}) \times \cos 2\theta_0$. We expect that our modeling of flavor conversions in this region as fully non-adiabatic transitions is accurate, but this assumption remains to be checked explicitly. Moreover, the impact of self-interacting sterile neutrinos on non-linear structure formation at the smallest scales probed by Lyman-α data should be calculated more carefully. Improving these issues is left for future work.

In conclusion, we have argued in this paper that self-interacting sterile neutrinos remain a cosmologically viable extension of the Standard Model. As long as the self-interaction dynamically suppresses sterile neutrino production until neutrinos decouple from the photon bath, the abundance produced afterwards is not in conflict with constraints on N_{eff}. Moreover, if the self-interaction is either weak enough for scattering to be negligible after the dynamic mixing suppression is lifted, or strong enough to delay free streaming of sterile neutrinos until sufficiently late times, also structure formation constraints can be avoided or significantly relaxed.
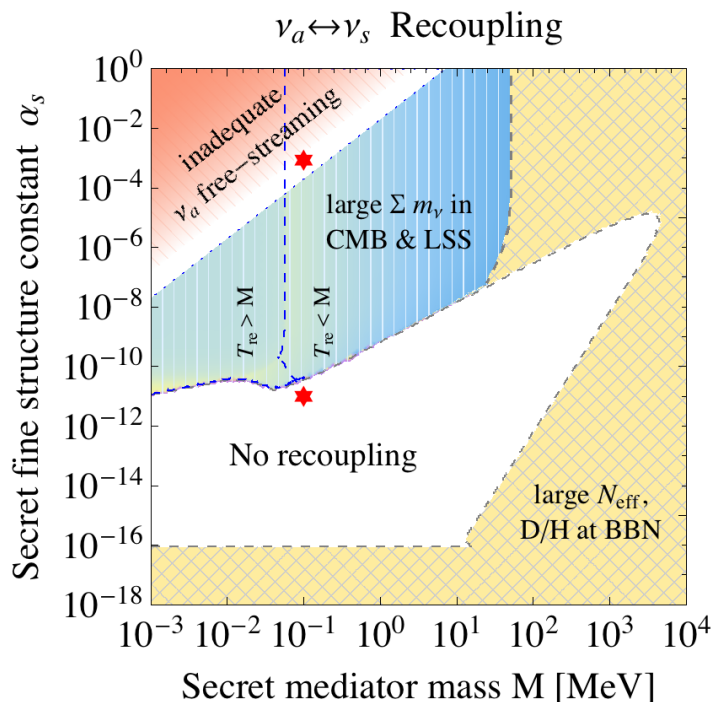
Figure 4. Schematic illustration of the parameter space for eV-scale sterile neutrinos coupled to a new "secret" gauge boson with mass M and a secret fine structure constant α_s. The vacuum mixing angle between active and sterile neutrinos was taken to be $\theta_0 = 0.1$. The white region in the lower half of the plot is allowed by all constraints, while the narrow white band in the upper left part satisfies all constraints except possibly large scale structure (LSS) limits from Lyman-α data at the smallest scales. The red stars show representative models in scenarios *(i)* and *(ii)*. The colored regions are excluded, either by LSS observations (blue vertically hatched), by the requirement that active neutrinos should free stream early enough (red shaded), or by a combination of CMB and BBN data (yellow cross-hatched).

Acknowledgments

A   Kinetic temperature and pressure

In the following, we give more details on the momentum distribution function $f(p,t)$ of sterile neutrinos ν_s after they have decoupled from all other particle species. $f(p,t)$ is essential in the calculation of the pressure P in section 3 and the average velocity $\langle v_s \rangle$ in section 4.

Even when ν_s are decoupled from other particles, they may still couple to themselves via strong self-interactions. If the self-interaction conserves the number of particles, such as $\nu_s\nu_s \leftrightarrow \nu_s\nu_s$, it only maintains kinetic equilibrium, but not chemical equilibrium. Number conservation and entropy maximization force the ν_s momentum distribution function in kinetic equilibrium to take the form

$$f(p,t) = \frac{1}{e^{[E(p)-\mu_s(t)]/T_s(t)} + 1} \, , \tag{A.1}$$

where $T_s(t)$ is defined as the *kinetic temperature*, $\mu_s(t)$ is the chemical potential, and $E(p) = (p^2 + m_s^2)^{1/2}$. Here and in the following, we use the definition $p \equiv |\vec{p}|$. Since we are interested in the evolution at relatively late times, when the sterile neutrino density is low compared to the density of a degenerate fermion gas and thus $\langle f(p,t) \rangle \ll 1$, the classical approximation

$$f(p,t) \simeq e^{-[E(p) - \mu_s(t)]/T_s(t)} \tag{A.2}$$

is adequate.

Our goal is to solve for the functions $T_s(t)$ and $\mu_s(t)$ with the initial condition of a relativistic thermal ensemble of sterile neutrinos. This means that, initially, $T_s = T_i \gg m_s$ and $\mu_s = 0$ at $a = a_i$. Note that the sterile neutrino mass will lead to a non-zero μ_s soon after neutrinos go out of chemical equilibrium. Although it is difficult to analytically solve the corresponding Boltzmann equation, there are two conditions that can be used to numerically obtain T_s and μ_s as functions of the scale factor $a(t)$. One is number conservation. The other is entropy conservation, which holds approximately for kinetic equilibrium in the classical limit [47].

The number density is

$$n_s(t) = \int \frac{d^3p}{(2\pi)^3} \, f(p,t) \tag{A.3}$$

and the classical entropy density is defined as

$$s_s(t) \equiv \int \frac{d^3p}{(2\pi)^3} \, f(p,t) \left[1 - \ln f(p,t)\right]. \tag{A.4}$$

It is straightforward to obtain the asymptotic solutions [47]

$$T_s(t) \propto \begin{cases} a^{-1}(t) & \text{for } T_s \gg m_s \\ a^{-2}(t) & \text{for } T_s \ll m_s \end{cases} \tag{A.5}$$
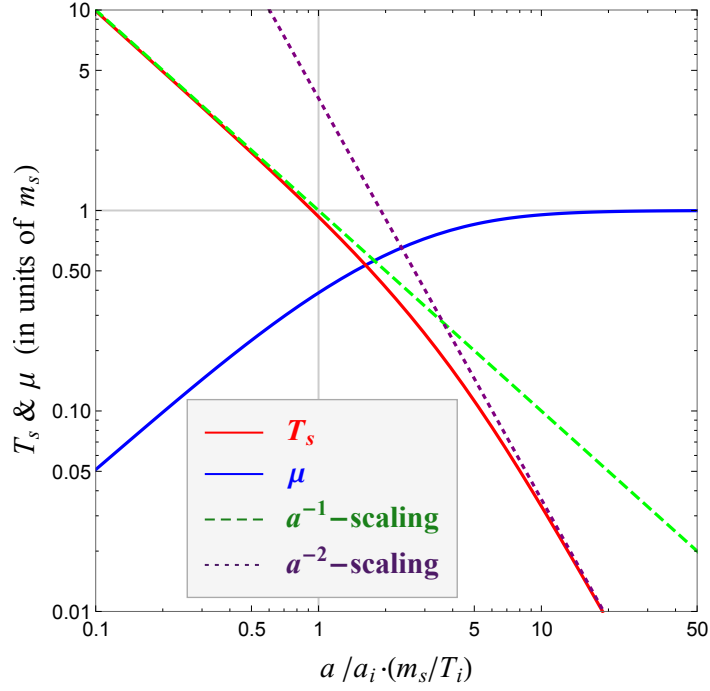
and

$$\mu_s(t) \propto \begin{cases} a(t) & \text{for } T_s \gg m_s \\ \text{const} & \text{for } T_s \ll m_s \end{cases}. \tag{A.6}$$

In the transition region $T_s \sim m_s$, the solution needs to be obtained numerically. The result is plotted in figure 5.

Finally, we comment on the calculation of the pressure P and the average velocity $\langle v_s \rangle$ of sterile neutrinos. The pressure is given by [47]

$$
\begin{aligned}
P &\equiv \int \frac{d^3p}{(2\pi)^3} \frac{p^2}{3E} \, f(p,t) \\
&= -T_s \, e^{\mu_s/T_s} \int \frac{d^3p}{(2\pi)^3} \frac{p}{3} \frac{d}{dp} e^{-E(p)/T_s} \\
&= T_s \cdot n_s \,,
\end{aligned}
\tag{A.7}
$$

Figure 5. Kinetic temperature T_s and chemical potential μ_s of sterile neutrinos, as functions of the scale factor $a(t)$ during the transition from the relativistic regime to the non-relativistic regime, with initial conditions $T_s = T_i \gg m_s$ and $\mu_s = 0$ at $a = a_i$.

and the average velocity is

$$\langle v_s \rangle \simeq \frac{1}{N} \int \frac{d^3 p}{(2\pi)^3} \frac{p}{E(p)} f(p,t) \,. \tag{A.8}$$

Here, $N \equiv \int dp \, 4\pi p^2/(2\pi)^3 \times f(p,t)$ is a normalization factor. Besides the conditions of kinetic equilibrium discussed above, we also need to take into account that sterile neutrino self-interactions freeze out at a time t^{dec} and sterile sector temperature $T_s = T_{s,\mathrm{dec}}$, after which kinetic equilibrium is lost and sterile neutrino momenta are simply redshifted as $a^{-1}(t)$. This implies for the momentum distribution function:

$$f(p,t) = \begin{cases} \dfrac{1}{\exp\left[\frac{1}{T_s(t)}\left(\sqrt{p^2+m_s^2}-\mu_s(t)\right)\right]+1} & \text{for } t < t^{\mathrm{dec}} \\[3mm] \dfrac{1}{\exp\left[\frac{1}{T_{s,\mathrm{dec}}}\left(\sqrt{\frac{p^2 a^2(t)}{a^2(t^{\mathrm{dec}})}+m_s^2}-\mu_s(t^{\mathrm{dec}})\right)\right]+1} & \text{for } t > t^{\mathrm{dec}} \end{cases} \tag{A.9}$$

Here, $\mu_s(t^{\mathrm{dec}})$ is the chemical potential at the time of decoupling. We have checked that the exact decoupling time only slightly changes the evolution of P, so we regard our solution in figure 5 as universal for all parameter values of interest. In section 4, we have for simplicity assumed a sudden decoupling of self-interactions though. For the value $e_s^2/M^2 \simeq 1\,\mathrm{MeV}^{-2}$ chosen there, this leads to $T_{s,\mathrm{dec}} \sim 0.0024\,\mathrm{eV}$, corresponding to a photon temperature of $0.038\,\mathrm{eV}$.

References

[1] LSND collaboration, A.A. Aguilar-Arevalo et al., *Evidence for neutrino oscillations from the observation of $\bar{\nu}_e$ appearance in a $\bar{\nu}_\mu$ beam*, *Phys. Rev.* **D 64** (2001) 112007 [hep-ex/0104049] [INSPIRE].

[2] MINIBOONE collaboration, A.A. Aguilar-Arevalo et al., *Improved Search for $\bar{\nu}_\mu \to \bar{\nu}_e$ Oscillations in the MiniBooNE Experiment*, *Phys. Rev. Lett.* **110** (2013) 161801 [arXiv:1207.4809] [INSPIRE].

[3] T.A. Mueller et al., *Improved Predictions of Reactor Antineutrino Spectra*, *Phys. Rev.* **C 83** (2011) 054615 [arXiv:1101.2663] [INSPIRE].

[4] G. Mention et al., *The Reactor Antineutrino Anomaly*, *Phys. Rev.* **D 83** (2011) 073006 [arXiv:1101.2755] [INSPIRE].

[5] A.C. Hayes, J.L. Friar, G.T. Garvey, G. Jungman and G. Jonkmans, *Systematic Uncertainties in the Analysis of the Reactor Neutrino Anomaly*, *Phys. Rev. Lett.* **112** (2014) 202501 [arXiv:1309.4146] [INSPIRE].

[6] M.A. Acero, C. Giunti and M. Laveder, *Limits on ν_e and $\bar{\nu}_e$ disappearance from Gallium and reactor experiments*, *Phys. Rev.* **D 78** (2008) 073009 [arXiv:0711.4222] [INSPIRE].

[7] J. Kopp, M. Maltoni and T. Schwetz, *Are there sterile neutrinos at the eV scale?*, *Phys. Rev. Lett.* **107** (2011) 091801 [arXiv:1103.4570] [INSPIRE].

[8] J. Kopp, P.A.N. Machado, M. Maltoni and T. Schwetz, *Sterile Neutrino Oscillations: The Global Picture*, *JHEP* **05** (2013) 050 [arXiv:1303.3011] [INSPIRE].

[9] J.M. Conrad, C.M. Ignarra, G. Karagiorgi, M.H. Shaevitz and J. Spitz, *Sterile Neutrino Fits to Short Baseline Neutrino Oscillation Measurements*, *Adv. High Energy Phys.* **2013** (2013) 163897 [arXiv:1207.4765] [INSPIRE].

[10] J.R. Kristiansen, Ø. Elgarøy, C. Giunti and M. Laveder, *Cosmology with sterile neutrino masses from oscillation experiments*, arXiv:1303.4654 [INSPIRE].

[11] C. Giunti, M. Laveder, Y.F. Li and H.W. Long, *Pragmatic View of Short-Baseline Neutrino Oscillations*, *Phys. Rev.* **D 88** (2013) 073008 [arXiv:1308.5288] [INSPIRE].

[12] G. Steigman, *Neutrinos And Big Bang Nucleosynthesis*, *Adv. High Energy Phys.* **2012** (2012) 268321 [arXiv:1208.0032] [INSPIRE].

[13] PLANCK collaboration, P.A.R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, arXiv:1502.01589 [INSPIRE].

[14] J. Hamann, S. Hannestad, G.G. Raffelt and Y.Y.Y. Wong, *Sterile neutrinos with eV masses in cosmology: How disfavoured exactly?*, *JCAP* **09** (2011) 034 [arXiv:1108.4136] [INSPIRE].

[15] S. Hannestad, R.S. Hansen and T. Tram, *How Self-Interactions can Reconcile Sterile Neutrinos with Cosmology*, *Phys. Rev. Lett.* **112** (2014) 031802 [arXiv:1310.5926] [INSPIRE].

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

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

[18] P. Ko and Y. Tang, *Self-interacting scalar dark matter with local Z_3 symmetry*, *JCAP* **05** (2014) 047 [arXiv:1402.6449] [INSPIRE].

[19] K.C.Y. Ng and J.F. Beacom, *Cosmic neutrino cascades from secret neutrino interactions*, *Phys. Rev.* **D 90** (2014) 065035 [arXiv:1404.2288] [INSPIRE].

[20] J. Kopp and J. Welter, *The Not-So-Sterile 4th Neutrino: Constraints on New Gauge Interactions from Neutrino Oscillation Experiments*, *JHEP* **12** (2014) 104 [arXiv:1408.0289] [INSPIRE].

[21] N. Saviano, O. Pisanti, G. Mangano and A. Mirizzi, *Unveiling secret interactions among sterile neutrinos with big-bang nucleosynthesis*, *Phys. Rev.* **D 90** (2014) 113009 [arXiv:1409.1680] [INSPIRE].

[22] A. Mirizzi, G. Mangano, O. Pisanti and N. Saviano, *Collisional production of sterile neutrinos via secret interactions and cosmological implications*, *Phys. Rev.* **D 91** (2015) 025019 [arXiv:1410.1385] [INSPIRE].

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

[24] C. Kouvaris, I.M. Shoemaker and K. Tuominen, *Self-Interacting Dark Matter through the Higgs Portal*, *Phys. Rev.* **D 91** (2015) 043519 [arXiv:1411.3730] [INSPIRE].

[25] Y. Tang, *More Is Different: Reconciling eV Sterile Neutrinos with Cosmological Mass Bounds*, *Phys. Lett.* **B 750** (2015) 201 [arXiv:1501.00059] [INSPIRE].

[26] E.C.G. Stueckelberg, *Interaction forces in electrodynamics and in the field theory of nuclear forces*, *Helv. Phys. Acta* **11** (1938) 299 [INSPIRE].

[27] S. Dodelson and L.M. Widrow, *Sterile-neutrinos as dark matter*, *Phys. Rev. Lett.* **72** (1994) 17 [hep-ph/9303287] [INSPIRE].

[28] L. Stodolsky, *On the Treatment of Neutrino Oscillations in a Thermal Environment*, *Phys. Rev.* **D 36** (1987) 2273 [INSPIRE].

[29] T.D. Jacques, L.M. Krauss and C. Lunardini, *Additional Light Sterile Neutrinos and Cosmology*, *Phys. Rev.* **D 87** (2013) 083515 [arXiv:1301.3119] [INSPIRE].

[30] F.-Y. Cyr-Racine and K. Sigurdson, *Limits on Neutrino-Neutrino Scattering in the Early Universe*, *Phys. Rev.* **D 90** (2014) 123533 [arXiv:1306.1536] [INSPIRE].

[31] M. Archidiacono and S. Hannestad, *Updated constraints on non-standard neutrino interactions from Planck*, *JCAP* **07** (2014) 046 [arXiv:1311.3873] [INSPIRE].

[32] J. Lesgourgues and S. Pastor, *Neutrino mass from Cosmology*, *Adv. High Energy Phys.* **2012** (2012) 608515 [arXiv:1212.6154] [INSPIRE].

[33] E.W. Kolb and M.S. Turner, *The Early Universe*, Westview Press (1994).

[34] SDSS collaboration, M. Tegmark et al., *Cosmological Constraints from the SDSS Luminous Red Galaxies*, *Phys. Rev.* **D 74** (2006) 123507 [astro-ph/0608632] [INSPIRE].

[35] G. Rossi et al., *Suite of hydrodynamical simulations for the Lyman-α forest with massive neutrinos*, *Astron. Astrophys.* **567** (2014) A79 [arXiv:1401.6464] [INSPIRE].

[36] M. Viel, G.D. Becker, J.S. Bolton and M.G. Haehnelt, *Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data*, *Phys. Rev.* **D 88** (2013) 043502 [arXiv:1306.2314] [INSPIRE].

[37] J. Lesgourgues and S. Pastor, *Massive neutrinos and cosmology*, *Phys. Rept.* **429** (2006) 307 [astro-ph/0603494] [INSPIRE].

[38] J. Brandbyge, S. Hannestad, T. Haugbolle and B. Thomsen, *The Effect of Thermal Neutrino Motion on the Non-linear Cosmological Matter Power Spectrum*, *JCAP* **08** (2008) 020 [arXiv:0802.3700] [INSPIRE].

[39] R.A.C. Croft et al., *Towards a precise measurement of matter clustering: Lyα forest data at redshifts 2–4*, *Astrophys. J.* **581** (2002) 20 [astro-ph/0012324] [INSPIRE].

[40] M. Vogelsberger et al., *Properties of galaxies reproduced by a hydrodynamic simulation*, *Nature* **509** (2014) 177 [arXiv:1405.1418] [ɪɴSPIRE].

[41] T. Sawala et al., *Local Group galaxies emerge from the dark*, arXiv:1412.2748 [ɪɴSPIRE].

[42] A.A. Klypin, A.V. Kravtsov, O. Valenzuela and F. Prada, *Where are the missing Galactic satellites?*, *Astrophys. J.* **522** (1999) 82 [astro-ph/9901240] [ɪɴSPIRE].

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

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

[45] L.G. van den Aarssen, T. Bringmann and C. Pfrommer, *Is dark matter with long-range interactions a solution to all small-scale problems of ΛCDM cosmology?*, *Phys. Rev. Lett.* **109** (2012) 231301 [arXiv:1205.5809] [ɪɴSPIRE].

[46] I.M. Shoemaker, *Constraints on Dark Matter Protohalos in Effective Theories and Neutrinophilic Dark Matter*, *Phys. Dark Univ.* **2** (2013) 157 [arXiv:1305.1936] [ɪɴSPIRE].

[47] J. Bernstein, *Kinetic Theory in the Expanding Universe*, Cambridge University Press (1988).