Sterile Neutrinos with Secret Interactions — Cosmological Discord?

Xiaoyong Chu, Basudeb Dasgupta, Mona Dentler, Joachim Kopp, and Ninetta Saviano

I. INTRODUCTION

A quote famously attributed to Isaac Asimov is “the most exciting phrase to hear in science, the one that heralds new discoveries, is not ‘Eureka’ but ‘That’s funny . . . .’”. Neutrino physics is arguably a field of research where this phrase can be heard rather frequently. Currently, it applies for instance to several independent anomalies observed in short baseline neutrino oscillation experiments. Most recently, interest in these anomalies has been renewed when new data from the MiniBooNE experiment at Fermilab corroborated its earlier results. The anomalies have been interpreted as possible hints for the existence of a fourth (“sterile”) neutrino flavor, even though global fits indicate that it is not possible to interpret all experimental results in such a scenario. This conclusion remains true in scenarios with more than one sterile neutrino. However, the possibility remains that some anomalies are heralding new physics while others have mundane explanations. Even more interesting would be the possibility that the new physics is richer than just a sterile neutrino (see for instance). In any case, a very severe trial that sterile neutrino models must face is that of cosmology. More specifically, observations of the cosmic microwave background (CMB), of light element abundances from Big Bang Nucleosynthesis (BBN), and of large scale structures (LSS) in the Universe constrain the total energy in relativistic species, usually expressed in terms of the effective number of neutrino species, . In addition, LSS and the CMB constrain the sum of neutrino masses, , or, more precisely, the sum of the masses of collisionless neutrino species.

However, cosmology can only probe particle species that are abundant in the early Universe. It is therefore interesting to explore scenarios where sterile neutrinos, in spite of having , are not produced in sufficient abundance to have observable consequences. One proposed mechanism to achieve this is the “secret interactions” scenario, in which sterile neutrinos, while being singlets under the SM gauge group, are coupled to a new gauge boson (or to a new pseudoscalar) with mass . Through this new interaction, sterile neutrinos feel a new temperature-dependent matter potential, which dynamically suppresses their mixing with active neutrinos at high temperatures, while being negligible today. To avoid constraints on , it is in particular required that active–sterile neutrino mixing is strongly suppressed at temperatures , the temperature where active neutrinos decouple from the photon bath. Note that introducing new interactions in the sterile neutrino sector may also be one way of reconciling the LSND and MiniBooNE anomalies with other neutrino oscillation data.

While the secret interactions scenario has motivated a number of model building and phenomenology papers, it has also been argued that most of the available parameter space is ruled out. Constraints come mainly from two directions. First, sterile neutrinos will eventually recouple with active neutrinos and are then efficiently produced collisionally via the Dodelson–Widrow mechanism. The temperature at which this recoupling happens depends on the interplay of the effective potential that suppresses flavor-changing collisions and the relevant scattering rates, which can be very large (see...
below) \cite{11}. Even if recoupling happens at $T \ll \text{MeV}$, it will still lead to equilibration between active and sterile neutrinos. This may lead to tension with limits on $\sum m_s$ \cite{39}. Second, mixing of active and sterile neutrinos leads to reduced free-streaming of active neutrinos. A certain amount of active neutrino free-streaming is, however, required by CMB observations \cite{31,42}. Note that this second constraint spoils attempts to keep sterile neutrinos safe from the limit on $\sum m_s$ by postulating that they interact so strongly that they cannot free-stream enough to affect large scale structure \cite{40}.

In this paper, we take two more steps in the ongoing exploration of secretly interacting sterile neutrinos. First, we update our earlier results from ref. \cite{40}, confirming in particular the findings by Cherry et al. \cite{41}. A detailed account of cosmological constraints on secret interactions has also been given recently in \cite{43}. In comparison to that paper, we focus less on a complete fit to cosmological data based on simulations, but instead derive constraints from physical arguments and estimates that can be much more easily generalized to other models. Second, and perhaps more importantly, we show that although the vanilla secret interactions model is indeed disfavored by cosmological data, the general idea underlying it remains viable and interesting. We give explicit examples of models that show this. The core assumption of the secret interactions scenario is that the sterile neutrino is hidden in cosmology because it gets a large temperature-dependent mass at high $T$ due to its interactions. We show that, if the vector boson employed in the original works \cite{26,27} is replaced by a scalar mediator with suitable symmetries and potential, the above-mentioned constraints from BBN, CMB and LSS appear to be avoidable. The mechanism we propose generates a large mass for $\nu_s$ in the high-temperature phase of the scalar potential, precluding efficient $\nu_s$ production. Only after a late phase transition in the scalar sector is the sterile neutrino mass reduced to the value observed today. We also outline more mundane ways to reconcile the vanilla scenario with data, by simply adding more free-streaming particles, or by allowing neutrinos to decay.

We begin with a review of the basic features of the secret interactions scenarios (section \[II\]), followed by the derivation of detailed cosmological constraints (section \[III\]). We then discuss several possible modifications to the original secret interactions models that could render the scenario phenomenologically viable again (section \[IV\]). We summarize and conclude in section \[V\].

II. THE SECRET INTERACTIONS SCENARIO

We augment the Standard Model (SM) with an extra, sterile, neutrino flavor $\nu_s$. We assume $\nu_s$ has appreciable, $\mathcal{O}(10\%)$, mixing with the three active neutrino flavors, collectively denoted by $\nu_\mu$. For the neutrino mass eigenstates, we use the notation $\nu_j$, with $j = 1 \ldots 4$, where $\nu_1$, $\nu_2$, $\nu_3$ have masses $\ll 1\text{ eV}$ and are mostly composed of $\nu_4$. For the mostly-sterile mass eigenstate $\nu_4$, we assume a mass around 1 eV, as motivated by the short baseline oscillation anomalies \cite{13,14,16}. We finally introduce the secret interaction by charging the sterile flavor eigenstate $\nu_s$ under a new $U(1)$, gauge group, with a gauge boson $A'$ at the MeV scale or somewhat below. The relevant interaction term reads

$$\mathcal{L}_{\text{int}} = e_s \bar{\nu}_s \gamma^\mu P_L \nu_s A'_\mu,$$

(1)

where $e_s$ is the $U(1)_s$ coupling constant, and $P_L = \frac{1}{2}(1 - \gamma^5)$ is the projection operator onto left-chiral fermion states. In the following, we will be agnostic about the mechanism that breaks $U(1)_s$, and endows the $A'$ boson with a mass. In particular, we will neglect the possible additional degrees of freedom—for instance sterile sector Higgs bosons—that may be introduced to achieve this breaking. If $A'$ gets its mass $M$ via the St"{u}ckelberg mechanism, this approximation becomes exact. When a sterile neutrino with energy $E$ propagates through a thermalized background of sterile neutrinos and $A'$ bosons at temperature $T_s$, it experiences a potential \cite{27}

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

(2)

Like a conventional Mikheyev–Smirnov–Wolfenstein (MSW) potential \cite{45,47}, $V_{\text{eff}}$ changes the neutrino mixing angle. In the $1 + 1$ flavor approximation, the $\nu_s - \nu_\mu$ mixing angle $\theta_m$ in a thermal $\nu_s$ background is given by

$$\sin^2 2\theta_m = \frac{\sin^2 2\theta_0}{(\cos 2\theta_0 + \frac{2E}{\Delta m^2} V_{\text{eff}})^2 + \sin^2 2\theta_0},$$

(3)

where $\theta_0$ is the mixing angle in vacuum, and $\Delta m^2 \equiv m_4^2 - m_1^2$ is the mass squared difference between the mostly sterile and mostly active neutrino mass eigenstates. Our qualitative results will remain unchanged even when more than one active neutrino flavor is considered. Equations (2) and (3) show that, at the high temperatures prevalent in the early Universe, mixing is strongly suppressed because $|V_{\text{eff}}| \gg \Delta m^2/(2E)$. Experiments today, on the other hand, will observe $\theta_m = \theta_0$ to a very good approximation.

Note that $V_{\text{eff}}$ has opposite sign at $T_s \ll M$ compared to $T_s \gg M$. This implies that there should be a temperature range around $T_s \sim M$ where the potential passes through zero and has a small magnitude. In other words, sterile neutrinos could be produced in this interval. However, as shown explicitly in ref. \cite{27}, this temperature interval is very short, therefore it is unclear what its impact on the final $\nu_s$ abundance is. Answering this question is one of the goals of this paper.

A typical cosmological history in the secret interactions model begins at high temperature with a negligible
abundance of \( \nu_s \) and \( A' \). As soon as a small number of \( \nu_s \) and \( A' \) are produced through oscillations or through some high-scale interactions, a large effective potential \( V_{\text{eff}} \) arises, suppressing mixing and preventing further \( \nu_s \) production. When the temperature drops so low that \( |V_{\text{eff}}| \lesssim \Delta m^2/(2E) \), sterile neutrinos recouple and thermalize with \( \nu_a \) through unsuppressed oscillations and \( A' \)-mediated scattering processes. At \( |V_{\text{eff}}| \approx \Delta m^2/(2E) \), oscillations can even be resonantly enhanced if the recoupling temperature is \( \lesssim M \) so that \( V_{\text{eff}} \) is negative. If this recoupling between \( \nu_a \) and \( \nu_s \) happens after the \( \nu_a \) have decoupled from the thermal bath at temperatures \( \sim \) MeV, \( \nu_s \) production does not change \( N_{\text{eff}} \). The predicted value of \( N_{\text{eff}} \) is then similar to that in the SM, \( N_{\text{eff}} \approx 3.046 \) [48, 49], and the model is consistent with the observed value \( N_{\text{eff}} \approx 3.15\pm0.23 \) (68\% CL) [21]. (The value of \( N_{\text{eff}} \) measured at recombination can still be reduced compared to the SM prediction because \( \nu_s \) become non-relativistic earlier than \( \nu_a \).) Nevertheless, the conversion of \( \nu_s \) into \( \nu_s \) increases the prediction for \( \sum m_\nu \), and for eV-scale \( \nu_s \), this puts the model in tension with the constraint \( \sum m_\nu < 0.23 \) eV [21]. Note that, to be precise, these bounds should be slightly modified in the secret interactions scenario as \( \nu_s-\nu_s \) recoupling at sub-MeV temperatures lowers the temperature of the neutrino sector compared to the standard \( \Lambda \)CDM model [39]. Computing this correction would require modified simulations of structure formation, which is beyond the scope of this work.

In ref. [40], we had proposed two possible ways out:

(i) Recoupling between \( \nu_a \) and \( \nu_s \) never happens because the gauge coupling \( e_a \) is so small that the sterile neutrino scattering rate \( \Gamma_s \) drops below the Hubble rate before \( |V_{\text{eff}}| \) drops below \( \Delta m^2/(2E) \). Of course, \( e_a \) still needs to be large enough to make sure that \( |V_{\text{eff}}| \gg \Delta m^2/(2E) \) until active neutrino scattering decouples. However, using a refined calculation, that we will confirm below in section III, the authors of ref. [41] have argued that these two contrary requirements cannot be fulfilled simultaneously.

(ii) Recoupling between \( \nu_a \) and \( \nu_s \) happens, but the gauge coupling \( e_a \) is so large that \( \nu_s \) cannot free-stream until very late times, after matter–radiation equality. In this case, bounds on \( \sum m_\nu \), which are effectively bounds on free-streaming species, do not apply. A possible problem with this option is that the free-streaming of \( \nu_a \) will be delayed as well through the \( \nu_s-\nu_a \) mixing. The rough estimates given in [40], suggested that the scenario might be marginally consistent with the data and only a dedicated analysis of CMB data would allow us to draw definitive conclusions. Forastieri et al. have recently carried out such an analysis and have shown that the scenario appears to be in tension with data [42].

III. CONSTRAINTS ON STERILE NEUTRINOS WITH SECRET INTERACTIONS

In the following, we will derive updated constraints on the secret interactions model introduced in section II. We employ two complementary approaches: our first approach, outlined in section IIIA, is a computation of the recoupling temperature, \( T_{\text{rec}} \), i.e., the temperature at which the sterile and active neutrinos recouple in the 1+1 scenario. This is similar to our previous calculations in ref. [40], but includes several improvements including those pointed out in ref. [11]. Assuming that sterile and active neutrinos equilibrate instantaneously at \( T_{\text{rec}} \), this computation allows us to estimate which cosmological data sets are sensitive to the resulting abundance of sterile neutrinos. In the second approach, presented in section IIIB, we go one step further and explicitly simulate the flavor evolution of the neutrino sector after recoupling. In doing so, we also go beyond the 1+1 flavor approximation and use instead a 2+1 flavor approximation, i.e., two active flavors and one sterile flavor. There are several motivations to do this, as we will explain later.

A. Recoupling Temperature Computation

In our first approach, we work in the mass basis and compute the production rate \( \Gamma_s \) of the mostly sterile mass eigenstate. The following reactions contribute to \( \nu_s \approx \nu_4 \) production:

1. \( W \) and \( Z \)-mediated processes

(i) \( e^- + e^+ \rightarrow \nu_1 + \nu_4 \) via s-channel \( Z \) exchange or t-channel \( W \) exchange;

(ii) \( e^- + \nu_1 \rightarrow e^- + \nu_4 \) via t-channel \( Z \) exchange or s-channel \( W \) exchange;

(iii) \( e^+ + \nu_1 \rightarrow e^+ + \nu_4 \) via t-channel \( Z \) exchange or s-channel \( W \) exchange;

(iv) \( \nu_1 + \nu_1 \rightarrow \nu_1 + \nu_4 \) via \( Z \) exchange in the t- or u-channel;

(v) \( \nu_1 + \nu_1 \rightarrow \nu_1 + \nu_4 \) via \( Z \) exchange in the s- or t-channel;

2. \( A' \)-mediated processes

(vi) \( \nu_4 + \nu_1 \rightarrow \nu_4 + \nu_4 \) via \( A' \) exchange in the s- or t-channel;

(vii) \( \nu_4 + \nu_1 \rightarrow \nu_4 + \nu_4 \) via \( A' \) exchange in the t- or u-channel;

Of course, the corresponding \( CP \)-conjugate processes contribute equally. Analytical expressions for the cross sections of these reactions are given in the appendix. Of the processes listed here, the first five are SM reactions involving electrons and/or electron neutrinos that produce sterile neutrinos through the mixing of the light neutrinos with the heavy mass eigenstate \( \nu_4 \). The remaining two...
processes are mediated by $A'$ and produce sterile neutrinos from electron neutrinos through the overlap of $\nu_e$ and $\bar{\nu}_e$. Note that process (vi), $\bar{\nu}_e + \nu_e \rightarrow \bar{\nu}_e + \nu_e$, involves $s$-channel $A'$ exchange and is thus resonantly enhanced in a specific part of the neutrino spectrum. Note also that $A'$-mediated $t$-channel scattering is enhanced in the forward direction if the $A'$ mass is much smaller than the neutrino temperature.

We first compute the temperature $T_{\text{rec}}$ at which sterile and active neutrinos recouple via scattering. We define $T_{\text{rec}}$ as the temperature at which $\Gamma_s$ becomes equal to the Hubble rate, i.e., $\Gamma_s = H$. In terms of the scattering cross sections given in the appendix A, $\Gamma_s$ is given by

$$
\Gamma_s = c_{QZ} \left[ \{\sigma v\}_{ee \rightarrow 14}^2 n_e^2 + \{\sigma v\}_{e1 \rightarrow e4} n_e \right. \\
\left. + \{\sigma v\}_{11 \rightarrow 14} n_\nu + \{\sigma v\}_{14 \rightarrow 44} n_s \right],
$$

(4)

Here, the notation $\langle \cdot \rangle$ refers to averaging over the momentum distributions of the involved particles. We assume these distributions to have a Fermi-Dirac form at all times.

Note also that all cross sections depend on the sterile sector temperature $T_s$ through the mixing angle $\theta_m$. The shorthand notation $\langle \sigma v\rangle_{ee \rightarrow 14}$ refers to process (i) above, $\langle \sigma v\rangle_{e1 \rightarrow e4}$ refers to the sum of processes (ii) and (iii), $\langle \sigma v\rangle_{11 \rightarrow 14}$ to the sum of processes (iv) (multiplied by a factor 1/2 to account for the identical particles in the initial state) and (v), and $\langle \sigma v\rangle_{14 \rightarrow 44}$ to the sum of processes (vi) and (vii). The factors $n_e$, $n_\nu$, and $n_s$ are the electron, active neutrino, and sterile neutrino number densities, respectively, not including their anti-particles. They are chosen such that each term in eq. (4) gives the production rate per active neutrino, i.e., the number of sterile neutrinos produced per unit time in a spatial volume element occupied on average by one active neutrino.

The prefactor $c_{QZ}$ accounts for the quantum Zeno effect, i.e., for the suppression of $\nu_e$ production when the scattering rate is faster than the oscillation frequency $|\Sigma(k)|$. In this case, oscillations have no time to develop before they are interrupted by scattering. To account for this effect, we define $c_{QZ}$ as

$$
c_{QZ} = \frac{(L_{\text{scat}}/L_{\text{osc}})^2}{1 + (L_{\text{scat}}/L_{\text{osc}})^2},
$$

(5)

where $L_{\text{scat}}$ is the $\nu_e - \nu_s$ scattering length and $L_{\text{osc}}$ is the oscillation length in medium. With this definition, $c_{QZ}$ is close to one when $L_{\text{scat}} \gg L_{\text{osc}}$ and approaches zero when $L_{\text{scat}} \ll L_{\text{osc}}$.

### B. Multi-flavor evolution

To understand the dynamics of sterile neutrino production in more detail, we have also simulated the evolution of a $2 + 1$ system (two active species and one sterile species) numerically. We do so, (i) to verify that thermalization between active and sterile neutrinos is indeed quasi-instantaneous after recoupling, (ii) to assess the impact of a nearly vanishing $V_{\text{eff}}$ at $T_s \sim M$, (iii) to check that the simplified treatment of the quantum Zeno correction in section III.A is valid, and (iv) to investigate the possible impact of going beyond the two-flavor approximation.

As a complete numerical simulation of the flavor evolution including the exact temperature-dependence of $V_{\text{eff}}$ is numerically highly challenging, we focus on the evolution during the epochs where the effective potential is small compared to the vacuum oscillation frequency, so that sterile neutrino mixing is unsuppressed. We use the exact temperature-dependence of $V_{\text{eff}}$ from ref. [27] to determine the relevant temperature intervals, and then simulate the flavor evolution within these intervals, setting $V_{\text{eff}} = 0$. Our simulation code is based on refs. [35,36,42].

The effective potential for a sterile neutrino with 4-momentum $k$ is given by [27]

$$
V_{\text{eff}} = -\frac{1}{2k^2} \left[ \left[ (k^0)^2 - k^2 \right] \text{tr}(\tilde{\Sigma}(k)) - k^0 \text{tr}(k \Sigma(k)) \right],
$$

(6)

with $\Sigma(k)$ the temperature-dependent sterile neutrino self-energy at one-loop and $u = (1,0,0,0)$ the 4-momentum of the heat bath. We use the ultra-relativistic approximation $k^0 \approx |k| + V_{\text{eff}}$ and expand eq. (6) in $V_{\text{eff}}$. We can then solve numerically for the critical points where the condition $|V_{\text{eff}}| = \Delta m^2/(2E)$ is fulfilled.

At high enough temperatures, $|V_{\text{eff}}|$ always exceeds $\Delta m^2/(2E)$ as long as the fine structure constant $\alpha_s (= \epsilon_s^2/4\pi)$ is not zero, but as temperatures become smaller two possibilities present themselves. The first possibility is that once $|V_{\text{eff}}|$ falls below $\Delta m^2/(2E)$, it never exceeds it again. An example of this is shown in the left panel of fig. 1. The second possibility is that $V_{\text{eff}}$ crosses through zero but then takes large negative values so that $|V_{\text{eff}}|$ again, as shown in the right panel of fig. 1. We refer to the temperature at which $|V_{\text{eff}}|$ intersects the vacuum term for the last time as the “last crossing” temperature. In the second scenario, $|V_{\text{eff}}|$ intersects the vacuum term around the zero crossing as well (fig. 1 right), and we call the corresponding temperature interval the “zero-crossing” interval.

We describe the neutrino ensembles in terms of a momentum-integrated $3 \times 3$ matrix of densities,

$$
\rho = \begin{pmatrix}
\rho_{ee} & \rho_{e\mu} & \rho_{e\tau} \\
\rho_{\mu e} & \rho_{\mu\mu} & \rho_{\mu\tau} \\
\rho_{\tau e} & \rho_{\tau\mu} & \rho_{\tau\tau}
\end{pmatrix},
$$

(7)

and a similar expression for antineutrinos, denoted by $\tilde{\rho}$. The diagonal entries are the respective number densities, while the off-diagonal ones encode phase information and vanish for zero mixing. In the standard situation, the equilibrium initial condition for the active neutrino
number densities is $\rho_{ee} = \rho_{\mu\mu} = 1$ (and similarly for $\bar{\rho}$), while for the sterile species we have the initial condition $\rho_{ss} = \bar{\rho}_{ss} \simeq 0$. The normalization of $\rho$ and $\bar{\rho}$ is chosen such that a diagonal entry of 1 corresponds to the abundance of a single neutrino (or antineutrino) species in the Standard Model.

The evolution equation for $\rho$ is

$$\frac{d\rho}{dt} = [\Omega, \rho] + C[\rho]. \quad (8)$$

Once again, a similar equation holds for the antineutrino density matrix $\bar{\rho}$. Here, $t$ is the comoving observer’s proper time. The evolution can be easily recast into a function of the photon temperature $T_\gamma$. The first term on the right-hand side of eq. (8) describes flavor oscillations, with the Hamiltonian given by

$$\Omega = U^\dagger \left( \frac{m^2_\nu}{2p} U \right) + \sqrt{2} G_F \left[ - \frac{8}{3} \langle p(\nu) \rangle \left( \frac{E_\ell}{M_W^2} + \frac{E_\nu}{M_Z^2} \right) \right], \quad (9)$$

where $m_\nu = \text{diag}(m_1, m_2, m_3)$ is the neutrino mass matrix in the mass basis, and $U$ is the $3 \times 3$ neutrino mixing matrix. The latter depends on three mixing angles, $\theta_{e\mu}, \theta_{e\tau}$, and $\theta_{\mu\tau}$, using the same parameterization as in ref. [53]. We take $\theta_{e\nu}$ equal to the active neutrino mixing angle $\theta_{13}$ [54], and we fix the active–sterile mixing angles and mass-squared differences at the best-fit values obtained from a global fit to the short-baseline anomalies [16]. The terms proportional to the Fermi constant $G_F$ in eq. (9) encode SM matter effects in neutrino oscillations. In particular, the term containing $E_\ell$ describes charged current interactions of neutrinos with the isotropic background medium, related to the energy density ($\propto T_\gamma^4$) of $e^\pm$ pairs. The term containing $E_\nu$ describes instead interactions of neutrinos with themselves (self-interaction term), related to the energy density ($\propto (\rho + \bar{\rho})T_\gamma^4$) of $\nu$ and $\bar{\nu}$. Note that in both terms, it is necessary to go beyond the low energy effective field theory of SM weak interactions (Fermi theory) and take into account momentum-dependent corrections. These correction terms can compete with the leading term from pure Fermi theory because the latter is proportional to the tiny lepton asymmetry of the Universe. Further details are given in ref. [53]. We remind the reader that we set $V_{\text{eff}} = 0$ below the last crossing temperature and during the zero-crossing interval. Moreover, we also neglect the small neutrino–antineutrino asymmetry $\propto (\rho - \bar{\rho})T_\gamma^3$.

The second term on the right-hand side of eq. (8) is the collisional term. It receives contributions from both SM and secret interactions:

$$C[\rho] = C_{\text{SM}}[\rho] + C_A[\rho]. \quad (10)$$

Following [53], we write the SM collision term as

$$C_{\text{SM}}[\rho] = -\frac{i}{2} G_F \left[ (S^2, \rho - 1) - 25 (\rho - 1) S \right] + \left[ (A^2, \rho - 1) + 2A(\rho - 1)A \right], \quad (11)$$
where the matrices $S$ and $A$ contain the numerical coefficients for the scattering and annihilation of the different flavors. In flavor space, they are given by $S = \text{diag}(g_e^s, g_\mu^s, 0)$ and $A = \text{diag}(g_e^\mu, g_\mu^e, 0)$. Numerically one finds $\frac{(g_e^s)^2}{(g_e^\mu)^2} = 3.06$, $(g_\mu^s)^2 = 2.22$, $(g_\mu^e)^2 = 0.5$, $(g_\mu^e)^2 = 0.28$.

The collision term corresponding to secret interactions in the sterile sector can be written schematically as

$$C_{A'}[\rho] = -\frac{1}{2}(\Gamma_{\text{no-res}} + \Gamma_{\text{res}}) \times \left[(S_{A'}^2, \rho - 1) - 2S_{A'}(\rho - 1)S_{A'}\right],$$

with the coefficient matrix $S_{A'} = \text{diag}(0, 0, 1)$ [35]. Note that here, we write the scattering processes in the flavor basis, whereas in section III A we had worked in the mass basis. Thus, the processes contributing to $C_{A'}$ are $\nu_e\nu_s \rightarrow \nu_e\nu_s$, $\nu_\mu\bar{\nu}_s \rightarrow \nu_\mu\bar{\nu}_s$, and $\nu_\mu\bar{\nu}_s \rightarrow A' A'$. The coupling to $\nu_e$ is generated by the oscillation terms in the equations of motion, but not explicitly present in $C_{A'}$. After all, the new interaction couples only to sterile neutrinos.

To highlight the qualitative differences between different contributions to the collision term, we have artificially split $C_{A'}$ into a piece containing non-resonant scattering processes (including $t$-channel processes) and a piece containing the contribution from resonantly enhanced $\nu_e\bar{\nu}_s \rightarrow \nu_e\bar{\nu}_s$ scattering through $s$-channel $A'$ exchange. The former piece contains the scattering rate

$$\Gamma_{\text{no-res}} \simeq \frac{16\pi^2 \alpha_s^2 \alpha_\mu^5}{T^2 M^2 + M^2},$$

while the latter one contains

$$\Gamma_{\text{res}} \simeq \frac{M}{T^2} \cdot n_s^\text{res} \cdot \sigma_{\text{CM}} \cdot v \simeq 6 \times 10^{-2} \alpha_s \frac{M^2}{T^2}. \quad (14)$$

Note that, at $T \ll M$, $\Gamma_{\text{res}}$ would receive an extra Boltzmann suppression factor. In these expressions, we omit $O(1)$ numerical prefactors for simplicity. We use the notation $n_s^\text{res} = 0.06 T^2 M \Gamma_{A'}$ for the number density of neutrinos participating in the resonant $s$-channel process, i.e. particles whose energies fall within the resonance window of width $\Gamma_{A'} = \alpha_s M/3\pi$ (in the center of mass frame); $\sigma_{\text{CM}} = \pi/M^2$ is the $s$-channel cross section in the center of mass frame, and $v = M/T$ is the relative velocity of the two neutrinos forming a resonant pair.

In fig. 2 we show the results of the numerical flavor evolution for sterile and active neutrinos at three representative parameter points. In particular, we show the evolution of the density matrix components $\rho_{ee}$ (electron neutrino abundance relative to a fully thermalized species), $\rho_{\mu\mu}$ (muon neutrino abundance), and $\rho_{ss}$ (sterile neutrino abundance). Panel (a), where $M = 10$ MeV, $\alpha_s = 10^{-12}$ was assumed, corresponds to the evolution after the last crossing temperature, while panel (b), with $M = 1$ MeV, $\alpha_s = 10^{-9}$, and panel (c), with $M = 1$ MeV, $\alpha_s = 10^{-6}$, correspond to the evolution during the zero crossing interval. In computing the zero-crossing temperatures and the last crossing temperature, we need to make an assumption on the initial temperature $T_{s,\text{ini}}$, before any $\nu_s$ are produced via oscillations. Here, we assume $T_{s,\text{ini}} = 0.3 T_\gamma$ at $T_\gamma = 1$ TeV. We will motivate this choice below in section III C. Our assumptions on $T_{s,\text{ini}}$ and on the evolution of $T_s$ are of course irrelevant to the actual numerical evolution of the neutrino ensemble as
we set $V_{\text{eff}} = 0$ in the zero-crossing interval and after the last crossing temperature. The grey bands in fig. 2 delimit the temperature range during which we perform the numerical flavor evolution. Within the gray bands, $V_{\text{eff}}$ cannot be considered negligible any more. In the cases shown in panels (a) and (b), sterile neutrinos are copiously produced. In particular, in panel (a), sterile neutrinos are fully thermalized ($\rho_{s} = 1$) by oscillations with active neutrinos occurring at temperatures around 10 MeV and increasing the number of relativistic degrees of freedom $N_{\text{eff}}$. In panel (b) instead, since oscillations and thus sterile neutrino production happen after the neutrino sector has decoupled from the photon bath, the total neutrino number density remains constant. Consequently, the asymptotic values of $\rho_{ee}, \rho_{\mu\mu}$, and $\rho_{ss}$ are all 0.67 in the 2 + 1 scenario. At the parameter points shown in panels (a) and (b), the oscillation rate is much larger than the $\nu_s$ scattering rate, i.e., there is no quantum Zeno suppression, and the $\nu_s$ scattering rate is in turn much larger than the Hubble rate, i.e., scattering-induced production is efficient. In panel (c) of fig. 2, $\nu_s$ are not copiously produced during the zero-crossing interval. The reason is that resonant scattering mediated by an s-channel $A'$ is faster than vacuum oscillations so that $\nu_s$ production is quantum Zeno-suppressed. Note that this is not the typical behavior – for most combinations of $M$ and $\alpha_s$, we find efficient $\nu_s$ production during the zero-crossing interval, except for a few cases where the interval is very short and/or the mediator is very massive ($\sim 1$ GeV). We always find efficient $\nu_s$ production below the last crossing temperature.

C. Results

In fig. 3 we show the main constraints on the parameter space of the secret interactions model in the plane spanned by the secret gauge boson mass $M$ and the corresponding fine structure constant $\alpha_s$. We have assumed a sterile neutrino mass $m_s = 1$ eV, and a vacuum mixing angle $\theta_0 = 0.1$. We distinguish three regimes: in the vertically striped (blue/orange) region labeled “$\sum m_{\nu}$ too large”, $\nu_s$ production is efficiently suppressed down to temperatures $T_s \leq 1$ MeV, so that $N_{\text{eff}}$ limits are evaded. Nevertheless, $\nu_s$ are efficiently produced via collisional decoherence at late times \cite{30}, i.e., around or below the last crossing temperature, so that the constraint on $\sum m_{\nu}$ is violated. Note that for lighter sterile neutrinos, $m_s \lesssim 0.2$ eV, these parameter regions would be experimentally allowed. The dashed line within the vertically hatched region indicates where the recoupling temperature equals the $A'$ mass. In the cross-hatched (brown) region in fig. 3, sterile neutrinos recouple above $T_s \sim 1$ MeV. They can thus fully thermalize with the SM thermal bath, and as a consequence violate constraints on both $N_{\text{eff}}$ and $\sum m_{\nu}$. The red shaded region at the top left of the plots is likely ruled out by CMB data because of insufficient active neutrino free-streaming \cite{12}. The red stars are two benchmark points considered in ref. \cite{40}. We see that both are now disfavored. The boundary between the striped and cross-hatched regions is first based on the value of $T_{\text{rec}}$ calculated using the methods from section III A. These methods, however, do not properly take into account $\nu_s$ production during the short time interval where zero-crossing happens and after the last crossing. Therefore, we use the numerical simulations from section III B to reexamine the zero crossing interval and to determine whether $\nu_s$ production around the zero crossing shifts $T_{\text{rec}}$ to larger values. If so, we set $T_{\text{rec}}$ to the central temperature of the zero crossing interval. We find, however, that this correction never affects the boundary between the striped and cross-hatched regions in fig. 3. We conclude that the sum of neutrino masses constraint and active neutrino free-streaming constraint together rule out all of the parameter space for the model \cite{12}.

The left and middle panels in fig. 3 correspond to different choices of the initial temperature $T_{\text{s,ini}}$ of $\nu_s$ and $A'$ at very early times. (We arbitrarily define $T_{\text{s,ini}}$ as the value of $T_s$ at photon temperature $T_\gamma = 1$ TeV.) We assume that there exist some additional new interactions between $\nu_s$ and SM particles (for instance in the context of a Grand Unified Theory) that lead to thermalization of $\nu_s$ at a very high temperature $T_\gamma \gg$ TeV. When these interactions freeze out (still at $T_\gamma \gg$ TeV), the sterile and SM sectors decouple. Afterwards, $T_s$ and $T_\gamma$ may drift apart, and the amount by which they do so above $T_\gamma = 1$ TeV is encoded in our choice of $T_{\text{s,ini}}$. Of course, further entropy is produced in the SM sector at $T_s < 1$ TeV, which implies that $T_s$ and $T_\gamma$ will drift further apart as the Universe evolves. This effect is taken into account in our calculations. Even if the sterile neutrino abundance is zero after inflation and reheating, and sterile neutrinos are only produced via oscillations, a non-vanishing $T_{\text{s,ini}}$ is still determined by the equation $T_s = H$. In other words, at any given epoch sterile neutrinos will be produced until $V_{\text{eff}}$ becomes large enough to shut production off (or until full thermal equilibrium between $\nu_s$ and $\nu_\alpha$ is reached). In the right panel of fig. 3, we show also constraints under the hypothesis that $T_{\text{s,ini}} = T_\gamma$, i.e., that the sterile sector temperature follows the active neutrino temperature at all times. This scenario, while difficult to realize in a consistent model, can be considered an upper limit on $T_{\text{s,ini}}$.

Among the various $\nu_s$ production processes listed above, the $W$- and $Z$-mediated ones are dominant at the recoupling time if either $A'$ is heavy (close to 1 GeV), or for $\alpha_s \lesssim 10^{-13}$, as shown by the gray region in fig. 4. The $A'$-mediated s-channel contribution to process $(vi)$, $\bar{\nu}_4 + \nu_1 \rightarrow \bar{\nu}_4 + \nu_4$ (shown in red), is dominant for most of the parameter region shown in fig. 4 largely due to the on-shell resonance. The $A'$-mediated t-channel contributions to processes $(vi)$ ($\bar{\nu}_4 + \nu_1 \rightarrow \bar{\nu}_4 + \nu_4$) and $(vii)$ ($\nu_4 + \nu_1 \rightarrow \nu_4 + \nu_4$), shown in blue, become more important when either the s-channel resonance is Boltzmann-suppressed in the case of heavy $A'$, or when the forward
enhancement of the $t$-channel diagrams becomes significant in the case of very light $A'$.

Note that the $A'$ resonance in the $s$-channel is responsible for ruling out the parameter region in which it was previously thought [40] that no recoupling between $\nu_a$ and $\nu_s$ happens. The calculations in [40] were based on naive dimensional arguments, and the enhancements induced by on-shell resonance and forward scattering were missed. This subtle issue was pointed out by Cherry et al. [41], who in particular explained that while the truly forward scattering of $\nu_a$ only gives a refractive index $V_{\text{eff}}$ (which was included in previous papers), multiple “almost forward” small-angle scatterings, which were incorrectly ignored, eventually add up to give large angle scattering and spatially separate the $\nu_1$ and $\nu_4$ eigenstates causing decoherence.

Let us finally address the potential loopholes in our line of argument so far. We have already argued that our results – in particular fig. 3 – are unaffected by sterile neutrino production during the zero crossing interval. We have explicitly checked this using the simulations described in section IIIB. Similarly, the robustness of our results with regard to possible corrections from the more detailed simulations also addresses the other points raised at the beginning of section IIIB. In particular, it illustrates that the approximation of quasi-instantaneous thermalization after after $V_{\text{eff}}$ drops below $\Delta m^2/(2E)$ is a good one, that a simplified treatment of the quantum Zeno effect is usually justified, and that there are no qualitative differences between the $1 + 1$ and $2 + 1$ scenarios.

Figure 4. Dominant scattering channel for collisional $\nu_s$ production at $T_{\text{rec}}$ as a function of $M$ and $\alpha_s$. We have chosen $m_s = 1 \text{ eV}$, and $T_{\text{ini}} = T_s$. The mixing angle suppression is common to all processes.

IV. RECONCILING SECRET INTERACTIONS WITH COSMOLOGY

While fig. 3 shows that the secret interactions model in its vanilla form is difficult to reconcile with cosmological constraints, it is interesting to ask what modifications are necessary to render it viable. In the following we outline some ideas for modifying the scenario in order to reconcile
eV-scale sterile neutrinos with cosmology. Some of the ideas discussed here may seem rather contrived, but we discuss them nevertheless to give the reader a feeling for what it takes to make eV-scale \( \nu_s \) cosmologically viable.

We are not going to comment here on ways of reconciling sterile neutrinos with cosmology that do not involve secret interactions.

### A. Recoupling never happens

If sterile neutrinos are much heavier than the sterile sector temperature \( T_s \) at early times and only become light when \( \nu_s \)-producing processes have decoupled, their production will be suppressed.

One way to implement this idea is to use a scalar mediator \( \phi \) instead of a vector mediator \( A' \). The resulting Yukawa coupling \( \phi \nu_x \nu_y \) makes the mass of \( \nu_y \) dependent on the vacuum expectation value (vev) \( v_\phi \) of \( \phi \). If \( v_\phi \) is large at high temperatures and vanishes at lower temperatures, \( \nu_y \) can be “hidden” until the Universe is cold enough to suppress their production. Scenarios of this type have been known for a long time [57][69].

To construct a toy model based on this idea, we consider two real scalars \( \phi_1, \phi_2 \), enjoying a \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) symmetry under which \( \phi_1 \) carries charges \((-,+), \) while \( \phi_2 \) carries charges \((+,-)\). The tree level scalar potential is then

\[
V = \frac{\lambda_1}{4} \phi_1^4 + \frac{\lambda_2}{4} \phi_2^4 + \frac{\lambda_p}{2} \phi_1^2 \phi_2^2 + \frac{\mu_1^2}{2} \phi_1^2 + \frac{\mu_2^2}{2} \phi_2^2. \tag{15}
\]

Boundeness of the potential from below requires that \( \lambda_{1,2} > 0 \) and \( \lambda_2^2 < \lambda_1 \lambda_2 \). As long as \( \mu_1^2 > 0 \) and \( \mu_2^2 > 0 \), there are no broken symmetries at zero temperature.

At high temperatures the potential receives thermal corrections. At 1-loop order and with \( T_s^2 \gg \mu_{1,2}^2 \), these are [57]

\[
\Delta V(T_s) = \frac{T_s^2}{24} \sum_{i=1,2} \frac{\partial^2 V}{\partial \phi_i^2} \approx \frac{T_s^2}{24} \left[ (3\lambda_1 + \lambda_p)\phi_1^2 + (3\lambda_2 + \lambda_p)\phi_2^2 \right]. \tag{16}
\]

If \( 3\lambda_1 + \lambda_p < 0 \), the field \( \phi_1 \) develops a nonzero vev \( v_{\phi_1} \) at temperatures above \( T_{s,crit} \equiv T_s(T_{crit}) \approx \left[ 12\mu_1^2 / |3\lambda_1 + \lambda_p| \right]^{1/2} \), breaking one of the \( \mathbb{Z}_2 \) symmetries. Here, \( T_s(T_{crit}) \) denotes the sterile sector temperature at the time when the photon temperature is \( T_{crit} \). The subscript “crit” stands for critical temperature. One can see that \( v_{\phi_1} \neq 0 \) will occur for modestly large negative values of the quartic cross-coupling \( \lambda_p \). Because of the boundeness conditions that force \( 3\lambda_2 + \lambda_p > 0 \), the other scalar \( \phi_2 \) cannot develop a vev simultaneously, so one of the \( \mathbb{Z}_2 \) symmetries remains unbroken. It is easy to see that if \( \mu_2^2 \) is very small, \( T_{crit} \) can be quite low, perhaps lower than the temperature \( T_{dec} \) at which active and sterile neutrinos finally decouple for good.

The charges of sterile neutrinos under the \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) are chosen as follows: the left-handed component \( \nu_sL \) of the Dirac fermion \( \nu_s \) carries charges \((+,+)\), while its right-handed partner \( \nu_sR \) is a singlet with charges \((-,+). \) The couplings of \( \nu_s \) are then

\[
\mathcal{L}_s = -y_\phi \bar{\nu}_sL \nu_sR - \frac{1}{2} m_{sL} \bar{\nu}_sL \nu_sL - \frac{1}{2} m_{sR} \bar{\nu}_sR \nu_sR + \text{h.c.}. \tag{18}
\]

The two main issues that we discuss now are whether this interaction is sufficient to generate a large \( v_{\phi_1} \)-induced thermal mass for \( \nu_s \) in order to prohibit \( \nu_s \) production until \( T_s < T_{dec} \), and whether sterile neutrino scattering can freeze-out already at \( T_s > T_{crit} \) in order to avoid recoupling below \( T_{crit} \). Of course, we also have to demand that \( T_{crit} < 1 \text{ MeV} \) to prevent collisional sterile neutrino production via \( Z \)- and \( W \)-mediated processes.

At high temperatures, \( T_s \gg T_{crit} \), the temperature-dependent (Dirac) mass for \( \nu_s \) is

\[
m_s(T_s) = y \sqrt{-\frac{12\mu_1^2 - (3\lambda_1 + \lambda_p)T_s^2}{12\lambda_1}} \quad \text{for} \; T_s > T_{crit}. \tag{19}
\]

At \( T_s \ll T_{crit} \), \( \nu_s \) splits into two Majorana fermions of mass \( m_{sL} \) and \( m_{sR} \). We assume that at least one of these is \( \mathcal{O}(\text{eV}) \). To prohibit production of \( \nu_s \), we demand that \( m_s(T_s) \ll T_s \gg T_{crit} \), so that \( \nu_s \) production becomes exponentially suppressed.

As an example, if \( \lambda_2 \simeq 1 \) then \( \lambda_1 \gtrsim \lambda_2^2 \) satisfies the boundeness criterion, and \( m_s(T_s) \simeq y T_s / \sqrt{12} |\lambda_p| \) for \(-1 < \lambda_p < 0 \) and \( T_s \gg T_{crit} \). The requirement \( T_s < T_{dec} \) then implies

\[
|\lambda_p| > 12\mu_1^2/T_{s,dec}^2, \tag{20}
\]

while the condition \( m_s(T_s) > T_{s,dec} \) implies

\[
|\lambda_p| < \frac{y^2}{12}. \tag{21}
\]

To determine \( T_{dec} \), we need to consider \( \phi_1 \)-mediated neutrino–neutrino scattering. We first redefine the field \( \phi_1 \to \nu_{\phi_1} + \rho \) after symmetry breaking, where the mass of the physical scalar \( \rho \) is given by \( m_{\rho}^2(T_s) = 2\lambda_1 \nu_{\phi_1}^2 \approx T_s^2 |\lambda_p| / 6 \) for \( T_s > T_{crit} \) (i.e., \( T_s > T_{s,crit} \)) and \( m_{\rho}^2 \approx \mu_1^2 + (3\lambda_2 + \lambda_p)T_s^2 / 12 \) for \( T_s \ll T_{crit} \). The decoupling temperature is defined as the temperature at which the rate for \( \nu_s - \nu_l \) inelastic scattering drops below the Hubble rate:

\[
n_{\nu_s}(T_{s,dec}) \frac{\sin^2 \theta_{23} y}{m_{\rho}^2} = H(T_{dec}) \approx \frac{T_{dec}^4}{M_{Pl}}, \tag{22}
\]

where \( \theta_{23} \) is the mixing angle between the active and sterile neutrinos.
Therefore, eqs. (20) and (21) are true as long as

\[ \frac{\mu_1}{\lambda_1} \approx \lambda_2 \approx 1, \mu_2 \approx 0 \]

is merely to illustrate the phenomenological viability of independent Coleman-Weinberg corrections \([72, 73]\), the tree level terms given in eq. (15), the temperature-dependent mass of the sterile neutrinos, never recouple in the toy model given by eqs. (15) and (18). Different colors corresponds to different values of the scalar mass parameter \(\mu_1\), as indicated in the legend. We have taken \(\lambda_1 \approx \lambda_2^2, \lambda_2 \approx 1, \mu_2 \approx 0\) to make this figure.

This condition can be satisfied if

\[ 12 \sin^4 \theta_0 \frac{\mu_2^2 M^2_{P1}}{m_p^4} < |\lambda_p| < \frac{y^2}{12}. \]

This condition can be satisfied if \(y\) is tiny, and \(|\lambda_p|\) is even tinier.

The parameter space which satisfies all the above conditions is shown in fig. 5 for various choices of \(\mu_1\). We see that in this minimal toy model the parameters \(\lambda_p\) and \(y\) need to have rather extreme values. Nevertheless, the model serves as a proof of principle that inverse symmetry breaking provides a viable mechanism for preventing \(\nu_s\) production in the early Universe.

To constrain the favored parameter regions of this model more quantitatively, it would be necessary to compute the effective temperature-dependent potential \(V_{\text{eff}}\) and follow its evolution through cosmological history, for instance using the public software package CosmoTransitions \([70, 71]\). For a given set of model parameters, this would lead to a prediction for the temperature-dependent mass of the sterile neutrinos, which could be plugged into the Boltzmann equations governing their production to determine their final abundance. The relevant contributions to \(V_{\text{eff}}\) are, besides the tree level terms given in eq. (15), the temperature-independent Coleman-Weinberg corrections \([72, 73]\), the one-loop finite-temperature corrections \([74]\), and the resumed higher-order "daisy" terms \([75]\). As our goal here is merely to illustrate the phenomenological viability of sterile neutrino models with inverse symmetry breaking, this computation is far beyond the scope of the present work.

This model is an example for a more general class of models exhibiting inverse symmetry breaking, where there is greater symmetry at lower temperatures as opposed to the usual scenario where symmetries are restored at higher temperatures \([57, 58, 76]\). There are other implementations of this mechanism, for instance Weinberg’s \(O(N_1) \times O(N_2)\) scalar models \([57]\), which break to \(O(N_1 - 1) \times O(N_2)\) at high temperature. At the non-perturbative level the symmetry may get restored at very high temperatures and the parameter space available for such inverse symmetry breaking is smaller than what is suggested by a 1-loop perturbative treatment \([59, 77, 78]\). However, for our purposes it is sufficient that a phase of broken symmetry exists at intermediate temperatures.

### B. Recoupling happens below MeV but CMB bounds on neutrino mass are avoided

An alternative way of reconciling sterile neutrinos with cosmology is to tolerate their production at \(T_s < 1\) MeV, but to invoke extra degrees of freedom to evade constraints on \(\sum m_\nu\).

(i) *Extra relativistic degrees of freedom to avoid structure formation bounds.* At intermediate couplings (blue/orange vertically striped region in fig. 5), eV-scale sterile neutrinos with secret interactions are constrained only by structure formation bounds on \(\sum m_\nu\). One way to avoid these is to introduce several additional sterile states, also charged under \(U(1)_s\) and with not too small mixing with active neutrinos, but with masses \(< 1\) eV. When secret interactions recouple active and sterile states at temperatures \(<\) MeV, the energy density in the neutrino sector is evenly distributed among all neutrino states. If the number of nearly massless states is sufficiently large, only a small fraction of energy will remain for the eV-scale state. More precisely, by adding \(n\) massless states in addition to the three active neutrinos and the one eV-scale sterile neutrino, the energy density after recoupling will be \(3 \rho_{\text{SM}}/(4 + n)\) in each state, where \(\rho_{\text{SM}}\) is the energy density of each active neutrino flavor in the SM. Correspondingly, the effective bound on \(\sum m_\nu\) is weakened by a factor \(3/(4 + n)\). We see that, in order to reconcile a 1 eV sterile neutrino with the limit \(\sum m_\nu < 0.23\) eV \([24]\), we need to add \(n \geq 9\) massless states.

(ii) *Extra relativistic degree of freedom for enhanced free-streaming.* At large \(\alpha_s\) (red region in fig. 5), where the \(\sum m_\nu\) constraint is avoided because \(\nu_s\) start to free-stream only very late, the main problem faced by the vanilla model is that also at least one of the active neutrinos will start to free-stream too late. Again, this problem could be avoided by adding one extra relativistic species. This could be for instance a second, nearly massless, ster-
ile neutrino that partially thermalizes before neutrino decoupling, or it could be the $A'$ boson itself, provided it is nearly massless. This scenario would predict $N_{\text{eff}}$ slightly larger than the SM value, but possibly still consistent with constraints. On the other hand, the extra free-streaming provided by the additional species is likely to improve the fit to CMB and structure formation data nevertheless. A detailed investigation of the viability of such a scenario requires a full fit to CMB and large scale structure data, which is left for future work.

(iii) Fast sterile neutrino decay. Cosmological constraints on the secret interactions scenario could also be avoided if the eV-scale sterile neutrino decays fast enough. In particular, if $\nu_s$ decays to nearly massless states before the onset of structure formation at $T \sim 1$ eV, large scale structure observations can hardly probe the impact of $\nu_s$ mass. An appealing possibility is to introduce two additional (nearly) massless particles: one pseudo-Goldstone boson $\phi$, and a second sterile neutrino $\nu'_s$. For an interaction vertex of the form $y\phi(\bar{\nu}_s\gamma^5\nu'_s)$ with $y \gtrsim 10^{-13}$, the lifetime corresponding to the decay $\nu_s \rightarrow \nu'_s + \phi$ is shorter than the time scale of recombination, avoiding the CMB bounds on both $\sum m_\nu$ and $N_{\text{eff}}$. In scenarios of this type, the strong $\nu_s$ self-interaction induced by the coupling $\phi(\bar{\nu}_s\gamma^5\nu_s)$ is in itself enough to suppress $\nu_s$ production before BBN [28]. In other words, $\phi$ can take the place of the gauge boson $A'$ in mediating secret interactions. Alternative decay scenarios, such as three-body decays $\nu_s \rightarrow 2\nu'_s + \gamma$ (via an off-shell massive $A'$) or $\nu_s \rightarrow \nu'_s + \gamma$ cannot generate sufficiently fast $\nu_s$ decays without violating cosmological constraints [42].

\[ \text{V. SUMMARY} \]

To summarize, we have assessed the status of models featuring light sterile neutrinos $\nu_s$ with “secret” self-interactions mediated by a new gauge boson $A'$. Such models had originally been introduced as a way of reconciling light sterile neutrinos (as motivated by the short baseline oscillation anomalies) with cosmological constraints. Indeed, the effective temperature-dependent potential generated by secret interactions can efficiently suppress active-sterile neutrino mixing in the early Universe down to temperatures $\ll$ MeV. At that time, SM weak interactions have frozen out and the neutrino sector is fully decoupled from the photon bath, so that the number of relativistic species $N_{\text{eff}}$ cannot change any more. However, efficient collisional production of $\nu_s$ (at the expense of active neutrino $\nu_\alpha$) will occur as soon as the mixing angle suppression is lifted. Of particular importance in this context are $A'$-mediated scattering processes which can be strongly enhanced by the s-channel $A'$ resonance (for $M \sim T_\nu$), and by collinear enhancement in the forward direction (for $M \ll T_\nu$). For $\nu_s$ masses around 1 eV and vacuum mixing angles of order 0.1 (as motivated by the short-baseline oscillation anomalies), the resulting population of $\nu_s$ is large enough to violate the cosmological constraint on $\sum m_\nu$. Thus, for $m_s = 1$ eV and $\theta_0 = 0.1$, all values of the $A'$ mass $M$ and the corresponding fine structure constant $\alpha_s$ are disfavored. Our results confirm earlier findings from ref. [41]. A possible loophole to these arguments exists at very large values of $\alpha_s$. Namely, if secret interactions are so strong that $\nu_s$ cannot free-stream, measurements of $\sum m_\nu$ are not sensitive. However, it has been shown in ref. [42] that in this case also active neutrino free-streaming is reduced, in conflict with CMB bounds.

In the second part of the paper we have discussed several new mechanisms for reconciling eV-scale sterile neutrinos with cosmology. We have outlined a toy model in which sterile neutrinos have initially a very large mass generated by the vacuum expectation value (vev) of a new scalar field, so that their production is kinematically forbidden. Only very late in cosmological history, a phase transition reduces the scalar vev to zero and the $\nu_s$ mass to $O(eV)$. At that time, collisional production is no longer possible, so the cosmological $\nu_s$ abundance remains negligible to this day. We have shown that this scenario is viable, but requires rather extreme values for some of its coupling constants.

We have in addition discussed scenarios with several new relativistic degrees of freedom with masses $\ll$ eV. As far as cosmological bounds are concerned, these degrees of freedom behave like active neutrinos. They can either serve to deplete the $\nu_s$ abundance by equilibrating with them, or they can act as an extra free-streaming species, compensating for the reduced free-streaming of active neutrinos in the regime of very large $\alpha_s$.

Finally, we have argued that bounds on $\sum m_\nu$ can also be avoided in scenarios in which the sterile neutrino has a fast decay mode to active neutrinos plus a light boson. Such models are of particular interest in the context of the LSND and MiniBooNE anomalies [17].

\[ \text{ACKNOWLEDGEMENTS} \]

It is a pleasure to thank Alessandro Mirizzi for useful discussions. X.C. is supported by the ‘New Frontiers’ program of the Austrian Academy of Sciences. BD is partially supported by the Dept. of Science and Technology of the Govt. of India through a Ramanujan Fellowship and by the Max Planck-Gesellschaft through a Max Planck-Partnergroup. JK and NS have been supported by the German Research Foundation (DFG) under Grant Grant Nos. KO 4820/11, FOR 2239, EXC-1098 (PRISMA) and by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 637506, “νDirections”). JK would like to thank CERN for hospitality and support during the final months of this project.
Table I. Dominant processes and cross sections for production of the mostly sterile mass eigenstate $\nu_s$.

<table>
<thead>
<tr>
<th>process</th>
<th>relativistic cross section</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) $e^- + e^+ \rightarrow \bar{\nu}_s + \nu_4$</td>
<td>$\sigma_{\text{exp4}} = \sin^2 \theta_m \frac{\pi e^2}{24 s \nu_{\text{conf}} c_w^4} \frac{\sqrt{s}}{\sqrt{s-4 m_e^2}} \frac{(13-20 c_{W}^2 + 8 c_{W}^4) + (23 - 40 c_{W}^2 + 16 c_{W}^4) m_{\nu_s}^2}{m_{\nu_s}^2}$</td>
</tr>
<tr>
<td>(ii) $e^- + \nu_e \rightarrow e^- + \nu_4$</td>
<td>$\sigma_{\text{el4}} = \sin^2 \theta_m \frac{\pi e^2}{24 s \nu_{\text{conf}} c_w^4} \frac{(s-2 m_e^2)^2}{m_e^2} \frac{(31-44 c_{W}^2 + 16 c_{W}^4) s^2 - 2(7-11 c_{W}^2 + 4 c_{W}^4) s m_{\nu_s}^2 + 4(1-c_{W}^2) m_{\nu_s}^4}{s^2}$</td>
</tr>
<tr>
<td>(iii) $\nu^+ + \nu_e \rightarrow \nu^+ + \nu_4$</td>
<td>$\sigma_{\text{p1p4}} = \sin^2 \theta_m \frac{\pi e^2}{24 s \nu_{\text{conf}} c_w^4} \frac{(s-2 m_e^2)^2}{m_e^2} \frac{(21-36 c_{W}^2 + 16 c_{W}^4) s^2 - (9-18 c_{W}^2 + 8 c_{W}^4) s m_{\nu_s}^2 + 3(3-c_{W}^2) m_{\nu_s}^4}{s^2}$</td>
</tr>
<tr>
<td>(iv) $\nu_1 + \nu_e \rightarrow \nu_1 + \nu_4$</td>
<td>$\sigma_{\text{1141}} = \sin^2 \theta_m \frac{\pi e^2}{s \nu_{\text{conf}} c_w^2} \frac{s}{2 m_e^2}$</td>
</tr>
<tr>
<td>(v) $\bar{\nu}_1 + \nu_e \rightarrow \bar{\nu}_1 + \nu_4$</td>
<td>$\sigma_{\text{1144}} = \sin^2 \theta_m \frac{\pi e^2}{s \nu_{\text{conf}} c_w^2} \frac{s}{3 m_e^2}$</td>
</tr>
<tr>
<td>(vi) $\nu_4 + \nu_e \rightarrow \nu_4 + \nu_4$</td>
<td>$\sigma_{\text{3144}} = \sin^2 \theta_m \frac{\pi e^2}{3} \frac{(3 s^2 + 10 M^2 s - 12 M^4) s + 12 M^2(s - M^4) \log(1 + s/M^2)}{M^2 s(s - M^4) + M^4 \Gamma_{\nu_e}}$</td>
</tr>
<tr>
<td>(vii) $\nu_4 + \nu_e \rightarrow \nu_4 + \nu_4$</td>
<td>$\sigma_{\text{4144}} = \sin^2 \theta_m \frac{\pi e^2}{4 \alpha_s} \frac{2 (s + 2 M^2 s + 2 M^2(s + M^4) \log(1 + s/M^2))}{M^2(s + M^4) + s + 2 M^2}$</td>
</tr>
</tbody>
</table>

Appendix A: Production Rate of Sterile Neutrinos

As discussed in section III, we have taken into account seven different processes in the computation of the sterile neutrino recoupling temperature. Here, in Table I we list the corresponding cross sections in the $1+1$ flavor approximation and to leading order in the mixing angle. Of course, the corresponding CP-conjugate processes contribute with the same cross sections. Note that process (iv) has identical initial state particles, a fact that needs to be properly taken into account when computing the rate for this process by integrating over the distribution of initial state $\nu_e$. In process (vi), which can be mediated by an $s$-channel $A'$, we need to take into account the non-zero width of $A'$, which is given by $\Gamma_{A'} = \alpha_s M/3$ for Dirac $\nu_s$.

The production rate of sterile neutrinos, normalized to the volume element occupied by an active neutrino as explained in section III A, is given by

$$\Gamma_s = \frac{c_{QZ}}{n_{\nu_s}} \sum_i \int d^3 p_1 \int d^3 p_2 \bar{f}_i(\vec{p}_1) f_i(\vec{p}_2) \times \sigma_i(s) v_{\text{Mol}},$$

(A1)

where the sum runs over the seven production processes listed above, the integral is over the 3-momenta of the initial state particles, and the prefactor $c_{QZ}$ accounts for quantum Zeno suppression (see section III A for details). The Møller velocity $v_{\text{Mol}}$ reduces to the relative velocity of the two colliding particles in the non-relativistic limit (see ref. [79] for details). The momentum distribution functions of the initial state particles $f_i(\vec{p}_i)$ have a Fermi-Dirac shape as we only consider fermionic processes. The number density of active neutrinos $n_{\nu_e}^{\text{eq}}$ is the integral over the corresponding Fermi-Dirac distribution for one massless degree of freedom. Abreviating the integral by introducing the notation $\langle \cdot \rangle$ for the momentum-averaged cross section, we obtain eq. (4).

The characteristic temperatures of the initial state particles are $T_\gamma$ for electrons and positrons, $T_\nu$ for active neutrinos, and $T_s$ for sterile neutrinos. As an approximation, we assume the same relation between $T_\gamma$ and $T_\nu$ as in the SM. The only exception is that for processes (ii) and (iii), which involve both a charged lepton and a neutrino in the initial state, we set $T_\nu = T_\gamma$ despite the temperature difference between them after $e^+e^-$ annihilation. This approximation, which makes the numerical evaluation of $\Gamma_s$ easier, is justified by the fact that after BBN electrons quickly become decoupled, so that processes mediated by SM $W$ and $Z$ bosons should not play an important role any longer. In the case of process (iv), which is initiated by two identical particles, care must be taken to restrict the integration domain such that double-counting of initial state phase space is avoided. (Or, alternatively, an extra factor of $1/2$ needs to be included in the integrand.)
The integrals in eq. (A1) can be partially evaluated [79]. With the definitions $E_{\pm} = E_1 \pm E_2$, the result is

$$
\Gamma_s = \frac{c Q Z}{n_{\nu_a}} \frac{2 \pi^2}{s} \sum_i \int dE_+ dE_- f_i \left( \frac{E_+ - E_-}{2} \right) f_i \left( \frac{E_+ - E_-}{2} \right) \sigma_i(s) \sqrt{\frac{[s - (m_1 + m_2)^2][s - (m_1 - m_2)^2]}{s}} , \quad (A2)
$$

with the integration boundaries

$$
 s \geq (m_1 + m_2)^2, \quad E_+ \geq \sqrt{s} , \quad (A3)
$$

$$
|E_- - E_+| \frac{m_1^2 - m_2^2}{s} \leq \frac{\sqrt{(E_+^2 - s)[s - (m_1 + m_2)^2][s - (m_1 - m_2)^2]}}{s} . \quad (A4)
$$

We use the condition

$$
\Gamma_s(T_{\text{rec}}) = H(T_{\text{rec}}) \quad (A5)
$$

to numerically decide whether and at what recoupling temperature, $T_{\text{rec}}$, the sterile $\nu_s$ can be brought into the thermal equilibrium with active neutrinos.


[2] MiniBooNE Collaboration, A. Aguilar-Arevalo et al., A Combined $\nu_x \rightarrow \nu_e$ and $\bar{\nu}_x \rightarrow \bar{\nu}_e$ Oscillation Analysis of the MiniBooNE Excesses, 1207.4809.


[22] N. Palanque-Delabrouille et al., Neutrino masses and cosmology with Lyman-alpha forest power spectrum, JCAP 1511 (2015), no. 11 011, 1506.05976.


