Bounds on light sterile neutrino mass and mixing from cosmology and laboratory searches

Steffen Hagstotz and Pablo F. de Salas

The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Roslagstullsbacken 21A, SE-106 91 Stockholm, Sweden

Stefano Gariazzo and Sergio Pastor

Instituto de Física Corpuscular (CSIC-Universitat de València), Parc Científic UV, c/ Catedrático José Beltrán 2, E-46980 Paterna, Spain

Martina Gerbino

High Energy Physics Division, Argonne National Laboratory, 9700 South Cass Avenue, Lemont, Illinois 60439-4815, USA and Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy

Massimiliano Lattanzi

Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Ferrara, Via Giuseppe Saragat 1, I-44122 Ferrara, Italy

Sunny Vagnozzi

Kavli Institute for Cosmology (KICC) and Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom

Katherine Freese

The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Roslagstullsbacken 21A, SE-106 91 Stockholm, Sweden, Theory Group, Department of Physics, The University of Texas at Austin, 2515 Speedway, C1600, Austin, Texas 78712-0264, USA, and Leinweber Center for Theoretical Physics, Department of Physics, University of Michigan, 450 Church Street, Ann Arbor, Michigan 48109, USA

(Received 17 June 2021; accepted 3 November 2021; published 9 December 2021)

We present a consistent framework to set limits on properties of light sterile neutrinos coupled to all three active neutrinos using a combination of the latest cosmological data and terrestrial measurements from oscillations, β-decay, and neutrinoless double-β-decay (0νββ) experiments. We directly constrain the full 3 + 1 active-sterile mixing matrix elements |Uα4|2, with α ∈ {e, µ, τ}, and the mass-squared splitting Δm241 ≡ m42 − m12. We find that results for a 3 + 1 case differ from previously studied 1 + 1 scenarios where the sterile is coupled to only one of the neutrinos, which is largely explained by parameter space volume effects. Limits on the mass splitting and the mixing matrix elements are currently dominated by the cosmological datasets. The exact results are slightly prior dependent, but we reliably find all matrix elements to be constrained below |Uα4|2 < 10−3. Short-baseline neutrino oscillation hints in favor of eV-scale sterile neutrinos are in serious tension with these bounds, irrespective of prior assumptions. We also translate the bounds from the cosmological analysis into constraints on the parameters probed by laboratory searches, such as mβ or m0ββ, the effective mass parameters probed by β-decay and 0νββ searches, respectively. When allowing for mixing with a light sterile neutrino, cosmology leads to upper
I. INTRODUCTION

The observation of oscillations between different neutrino flavors firmly establishes that at least two out of three neutrino mass eigenstates \( m_i \) (\( i = 1, 2, 3 \)) are nonzero [1–5]. Since in the Standard Model of particle physics neutrinos come with only left-handed chirality, it is not possible to generate a mass term for them by the usual Higgs mechanism as for the other leptons. The discovery of neutrino masses then requires new physics beyond the Standard Model, which may involve the existence of additional right-handed states (see, e.g., [6] for a review). These states are sterile, which means that they do not take part in weak, strong, or electromagnetic interactions, and their number and mass scale depend on the specific model. A wide range of sterile neutrino masses has been considered in the literature (see, e.g., [7] for an overview), ranging from \( O(\text{TeV}) \) that could play an important role in leptogenesis or serve as cold dark matter candidates, over \( O(\text{MeV}) \) neutrinos required by low-scale seesaw models, to \( O(\text{keV}) \) where they could form a warm dark matter component. In this paper, however, we focus on the \( O(\text{eV}) \) scale mostly motivated by experimental anomalies discussed below.

If produced in sufficient numbers in the early Universe via oscillations with active states, sterile neutrinos can lead to observable effects in cosmology, even in the absence of nongravitational interactions. Another way to search for sterile neutrinos is in neutrino oscillation experiments. As we will discuss in this paper, direct laboratory measurements and cosmological observations are complementary, and both avenues have led to tremendous progress in our understanding of the neutrino sector over the past few years.

Neutrino oscillation experiments allow one to constrain the neutrino mass-squared splittings. These mass splittings are responsible for the oscillations between neutrino flavor eigenstates. The majority of experimental results from neutrino oscillation experiments can be explained within the context of the so-called three-neutrino mixing paradigm. In this context, three mass eigenstates mix and produce oscillations between the three neutrino flavor eigenstates associated with the charged leptons of the Standard Model. Neutrino oscillation experiments have provided high-precision measurements of two mass-squared splittings [8–10]:

\[
\Delta m^2_{21} = m_2^2 - m_1^2 = 7.55 \times 10^{-5} \text{ eV}^2, \quad (1)
\]

\[
|\Delta m^2_{31}| = |m_3^2 - m_1^2| = 2.5 \times 10^{-3} \text{ eV}^2. \quad (2)
\]

Given the fact that the sign of \( \Delta m^2_{31} \) is not determined, two mass orderings can be possibly realized in nature: the normal ordering, in which the lightest mass state is \( m_1 \) and \( \Delta m^2_{31} > 0 \), or the inverted ordering, in which \( m_3 \) is the lightest state and \( \Delta m^2_{31} < 0 \). Current neutrino oscillation measurements show a mild statistical preference for the normal ordering [8–11].

For many years, however, a number of experimental results that cannot be explained within the context of the three-neutrino oscillation paradigm have been reported by various short-baseline (SBL) experiments. The first anomalous result was found by LSND [12] and later partly confirmed by MiniBooNE [13], but similar anomalies were also detected by the Gallium Experiment (GALLEX) and Soviet–American Gallium Experiment (SAGE) [14–17] and by a number of observations of reactor antineutrino fluxes at short distances [18–20]. Although the situation is not completely clear, (see, e.g., [21,22]), the anomalies could point to an additional mass splitting, much larger than the other two:

\[
\Delta m^2_{\text{SBL}} \sim O(\text{eV}^2). \quad (3)
\]

The existence of a new mass splitting implies the presence of a new neutrino mass eigenstate, which corresponds to a fourth flavor eigenstate. Since the hypothetical new neutrino has not been found in other weak interaction measurements [23], it cannot take part in Standard Model interactions and is, therefore, denoted “sterile” (see, e.g., [21,24]). Because of the large gap \( O(\text{eV}^2) \gg |\Delta m^2_{31}|, \Delta m^2_{21} \), oscillations with the sterile are not expected to affect standard atmospheric and solar neutrino measurements. One typically assumes that the mixing between the three active neutrinos and the fourth mass eigenstate is small, so that the scenario is commonly referred to as the 3 + 1 model. For recent reviews on experimental searches for eV-scale sterile neutrinos, see also [21,22,25].

In case such a new neutrino exists, its presence can affect the evolution of the Universe as well. Even if the fourth neutrino does not interact with the other Standard Model particles, it would be produced in the early Universe via oscillations with the active flavors. Excluding direct detection of relic neutrinos [26], the cosmological presence of the sterile state can be deduced only indirectly, for instance, from its contribution to the energy density of the Universe [27]. In particular, considering masses of \( O(\text{eV}) \) as motivated by the SBL anomalies, the sterile neutrino is produced as a relativistic particle. Therefore, it contributes at early times to the number of relativistic degrees of freedom in the Universe.
freedom $N_{\text{eff}}$, which quantifies the amount of energy density from relativistic species beyond photons in the early Universe and, therefore, modifies the background expansion rate. This, in turn, sets the time between key transitions in the early Universe, especially the freeze-out of weak interactions and the proton-to-neutron ratio important for big-bang nucleosynthesis (BBN). Measurements of the primordial element abundances, hence, provide constraints on $N_{\text{eff}}$ [28].

Light sterile neutrinos also affect the growth of cosmological perturbations. Their large velocities suppress the formation of overdensities below their free-streaming length so that they would behave as hot dark matter and leave a distinct imprint on structure formation, similar to the one of active neutrinos [29–32]. Cosmological observations are sensitive to the hot dark matter energy density $\Omega_{\text{hot}} = (\Omega_\gamma + \Omega_\nu)$, neglecting the contribution of other nonstandard particles. Recent measurements of the temperature and polarization anisotropies in the cosmic microwave background (CMB) and of large-scale structures are able to constrain both $N_{\text{eff}}$ and $\Omega_{\text{hot}} h^2$ to high precision; see, e.g., [33–47]. Since $N_{\text{eff}}$ is mostly constrained by the Silk damping tail of the CMB while $\Omega_{\text{hot}} h^2$ is measured from the suppression of small-scale structures due to free streaming, both parameters can mostly be determined independently.

There is another motivation to consider additional relativistic particles such as light sterile neutrinos in cosmology, independent of the anomalous results from neutrino experiments. Currently, CMB and baryon acoustic oscillation (BAO) measurements of the Hubble constant $H_0$, based on using the sound horizon at last scattering as a standard ruler, yield a significantly lower value than estimates from Cepheid-calibrated supernovae Ia [48,50–52] and lensed quasar time delay observations [53–55]; see also, e.g., [56,57] for further discussions. While a larger expansion rate due to contributions to the energy density in the early Universe reduces the size of the sound horizon and results in a higher inferred value of the Hubble constant $H_0$ (see, e.g., discussions in Refs. [58–62]), a sterile neutrino alone cannot account for the full discrepancy. The amount of additional radiation required to bring the sound horizon in agreement with the high $H_0$ measurements would be in conflict with the Silk damping tail at small angular scales observed in the CMB, as discussed in Refs. [44,63].

In this work, we study the phenomenology of a light sterile neutrino from various points of view. While the main goal is to compute, for the first time, cosmological bounds on all the mixing angles and the mass splitting between the light sterile state and active neutrinos, we also compare the cosmological results with other probes, such as constraints from direct mass measurements and neutrinoless double-$\beta$ decay, and limits from neutrino oscillation experiments. We will focus on light sterile neutrinos (i.e., $\Delta m^2_{41} < 100$ eV$^2$), since we are motivated by anomalies in oscillation experiments that could be explained by the existence of light sterile states.

The rest of this paper is structured as follows. First, we give a brief overview on the theoretical aspects and the status of experimental sterile neutrino searches in Sec. II. The phenomenology of such neutrinos in the early and late Universe is discussed in Sec. III. In Sec. IV, we discuss the datasets considered and the method used to obtain the bounds on the sterile neutrino mixing parameters, for which we present upper limits in Sec. V. We summarize and conclude in Sec. VI.

II. EXPERIMENTAL SEARCHES FOR LIGHT STERILE NEUTRINOS

In this section, we briefly summarize the theory of active-sterile neutrino oscillations and review the status of light sterile neutrino searches at various experiments. For comprehensive reviews focused on experimental searches for eV-scale sterile neutrinos, see [21,22,25].

A. Active-sterile neutrino mixing

The mixing between active and sterile neutrinos can be parametrized by means of the unitary $4 \times 4$ lepton mixing matrix $U$. There are several ways to write the mixing matrix, so we will briefly explain our choice of parametrization. In the case of mixing between four neutrino states, $U$ is fully characterized by six mixing angles and three $CP$-violating Dirac phases, but for the purposes of this work we set the phases to zero and, therefore, assume that oscillations among neutrinos are identical to those among antineutrinos. Three of the angles characterize the oscillations between the three active neutrinos: $\theta_{13}$, $\theta_{23}$, and $\theta_{14}$. The remaining three mixing angles, namely, $\theta_{14}$, $\theta_{24}$, and $\theta_{34}$, describe the mixing with the sterile state. We choose the following parametrization of the mixing matrix [24]:

$$ U = R^{34} R^{24} R^{14} R^{23} R^{13} R^{12}, $$

where each $R^{ij}$ is a real matrix describing a rotation by an angle $\theta_{ij}$.

Bounds on the mixing matrix can be provided in a way that is independent of the parametrization if instead of quoting limits on the individual mixing angles $\theta_{ij}$ we consider the matrix elements directly. Concerning active-sterile neutrino mixing, the most important elements are those of the fourth column:

$$ |U_{e4}|^2 = \sin^2 \theta_{14}, $$

$$ |U_{\mu4}|^2 = \cos^2 \theta_{14} \sin^2 \theta_{24}, $$

$$ |U_{\tau4}|^2 = \cos^2 \theta_{14} \cos^2 \theta_{24} \sin^2 \theta_{34}, $$

$$ |U_{s4}|^2 = \cos^2 \theta_{14} \cos^2 \theta_{24} \cos^2 \theta_{34}. $$

123524-3
We expect the mixing angles $\theta_{\alpha\beta}$ to be small, in order not to substantially alter the phenomenology of three-neutrino mixing beyond what is allowed by current limits. Therefore, the matrix elements $|U_{e\alpha}|^2$, $|U_{\mu\beta}|^2$, and $|U_{\alpha\beta}|^2$ are expected to be small, while $|U_{e\mu}|^2$ should, in principle, be of the order of unity. For this reason, it is possible to refer to the fourth neutrino mass eigenstate as the “sterile” neutrino, even though strictly speaking a sterile neutrino is a flavor eigenstate and has no definite mass. In the rest of the paper, we will often refer to the mass $m_4 \approx m_3$ as the mass of the sterile neutrino.

In addition to the mixing angles, we need to specify the three mass-squared splittings $\Delta m^2_{31}$, $\Delta m^2_{21}$, and $\Delta m^2_{41}$. For the active neutrinos, we always assume normal ordering $[64]$ ($\Delta m^2_{31} > 0$), and we fix all the standard mixing parameters to the best-fit values obtained in Ref. $[8]$. When considering the full oscillation pattern, the complete expressions for the oscillation probabilities are rather involved and depend on all the mass splittings and mixing angles; see, e.g., $[70]$. When considering SBL oscillations, however, the terms due to the much smaller solar and atmospheric mass-squared differences are suppressed, because they correspond to slower oscillations, and only the effect of $\Delta m^2_{41}$ is relevant. As a consequence, the oscillation probabilities can be well approximated by a two-neutrino mixing formula with appropriate mixing matrix elements. This is the appearance probability to populate a flavor state:

$$P_{\nu_\alpha \rightarrow \nu_\beta}^{SBL} \approx \sin^2(2\theta_{\alpha\beta}) \sin^2 \left( \frac{\Delta m^2_{41} L}{4E} \right) \quad (\alpha \neq \beta),$$

whereas the flux of the initial flavor is modulated by the disappearance probability:

$$P_{\bar{\nu}_e \rightarrow \bar{\nu}_\mu}^{SBL} \approx 1 - \sin^2(2\theta_{\alpha\beta}) \sin^2 \left( \frac{\Delta m^2_{41} L}{4E} \right),$$

where $L$ is the distance traveled by the neutrino, $E$ is its energy, and we use Greek letters to refer to the four flavor states $\alpha, \beta \in [e, \mu, \tau, s]$. The effective angles $\theta_{\alpha\beta}$ and $\theta_{\alpha\beta}$ depend on the elements of the fourth column of the mixing matrix:

$$\sin^2(2\theta_{\alpha\beta}) = 4|U_{e\alpha}|^2|U_{\mu\beta}|^2 \quad (\alpha \neq \beta),$$

$$\sin^2(2\theta_{\alpha\beta}) = 4|U_{e\alpha}|^2(1 - |U_{\alpha\beta}|^2).$$

The last expression is implied by the unitarity of the mixing matrix. From the probabilities in Eqs. (9) and (10), it is clear why the oscillatory behavior manifests when $4E \sim \Delta m^2 L$. If $4E \gg \Delta m^2 L$, the oscillatory term vanishes; if $4E \ll \Delta m^2 L$, the oscillation frequency becomes so high that oscillations cannot be resolved anymore and they are averaged out. The eV scale we focus on arises since various experiments find anomalies pointing toward $(L/E)^{-1} \sim O(eV^2)$. We emphasize that the absolute mass scale of the neutrino states does not enter in the equations, so oscillation experiments depend on the differences of the mass-squared values alone. In other words, oscillations are independent of the lightest neutrino mass $m_1$. Both appearance and disappearance channels can be used to measure the effective mixing angles and to constrain the mixing matrix elements.

### B. Current status of oscillation data and global fits

There is not a single SBL anomaly, but several different experiments with short baselines find anomalous neutrino fluxes of varying degree. We will briefly summarize the current situation in this section.

From the historical point of view, the first anomaly was found by LSND [12], which reported the unexpected appearance of electron antineutrinos in a beam of muon antineutrinos produced from $\pi^+$ decays. Years later, also MiniBooNE [13] confirmed the excess of $\bar{\nu}_e$ events, with a similar experimental setup, in partial agreement with LSND even though the MiniBooNE excess is too large to be explained by a sterile alone.

Another anomaly with comparable $L/E$ was found by $\bar{\nu}_e$ disappearance measurements by the gallium neutrino detectors GALLEX and SAGE [15,16,71,72]. Both experiments measured the electron neutrino flux in proximity to a radioactive source and found a lack of events at significance of $\sim 3\sigma$. A similar effect is observed in measurements of $\bar{\nu}_e$ disappearance in close proximity to nuclear reactors [18]. The lack of electron antineutrinos also reaches a significance of $\sim 3\sigma$ and was noticed after new calculations found a higher expected initial flux [19,20].

The LSND and MiniBooNE $\bar{\nu}_e$ appearance data require a nonzero value of $\theta_{\mu\mu}$, while in order to explain the reactor and gallium disappearance measurements one needs $\theta_{ee} > 0$. Taken together, this also implies a nonzero mixing $\theta_{\mu\mu}$ and $|U_{\mu4}|^2$. However, this matrix element can be constrained independently by muon (anti)neutrino disappearance experiments, and no corresponding anomaly for the relevant $L/E$ values is found by either measurements of the atmospheric muon neutrino flux by IceCube [73,74] or by the MINOS+ [75] Collaboration using an accelerator beam. Therefore, a combination of $\bar{\nu}_e$ appearance data by LSND and MiniBooNE and the disappearance results from electron and muon (anti)neutrinos in a global fit is currently problematic [74,76].

### C. Laboratory searches for the absolute neutrino mass scale

In addition to neutrino flavor oscillation experiments, laboratory searches for massive neutrinos also include kinematic measurements of $\beta$-decay and searches for neutrinoless double-$\beta$ $(0\nu\beta\beta)$ decay events. In this section,
we briefly summarize the consequences that a light sterile neutrino mixing with the active flavors can have for these searches.

One way to probe the absolute neutrino mass scale directly is to observe the cutoff of the electron energy spectrum emitted from $\beta$ decay. It occurs at the effective electron neutrino mass $m_\beta$, given by the incoherent sum of squared masses and mixing parameters:

$$m_\beta^2 = \sum_{j=1}^{4} |U_{ej}|^2 m_j^2,$$

where $j = 4$ is the contribution from the additional sterile neutrino. So far, $\beta$-decay measurements have been able only to set upper limits on the mass scale, with the latest bound $m_\beta < 1.1$ eV at 90% confidence level (CL) recently published by the KATRIN Collaboration [77].

Neutrinoless double-$\beta$ decay ($0\nu\beta\beta$) searches constrain the half-life $T_{1/2}$ of the isotope involved in the decay (see, e.g., [78] for a recent review). Assuming that the mechanism responsible for lepton number violation manifested in $0\nu\beta\beta$ events is the mass mechanism, constraints on the half-life can be translated into constraints on the effective Majorana mass $m_{\beta\beta}$:

$$T_{1/2} = \frac{m_e^2}{G_0 |M_{0e}|^2 m_{\beta\beta}^2},$$

where $m_e$ is the electron mass, $G_0$ is the phase space factor, and $M_{0e}$ is the nuclear transition matrix element for the decay. The effective mass parameter $m_{\beta\beta}$ can be expressed as a coherent sum of mass eigenstates and mixing parameters:

$$m_{\beta\beta} = \sum_{j=1}^{4} |U_{ej}|^2 e^{i\alpha_j} m_j,$$

where $\alpha_j$ are Majorana phases and $j = 4$ again contains the contribution from the sterile neutrino. Since only phase differences are observable, we set $\alpha_1 = 0$. So far, no such event has been detected, and only upper limits on the lifetime $T_{1/2}$ of various isotopes are available, as described in Sec. IV C. These bounds are usually converted into a range of upper limits on the Majorana mass $m_{\beta\beta}$ depending on theoretical uncertainty in the calculation of the nuclear matrix elements.

III. LIGHT STERILE NEUTRINOS IN COSMOLOGY

As explained in Sec. I, sterile neutrinos do not take part in weak, strong, or electromagnetic interactions. As a consequence, they will not be produced by Standard Model scattering or annihilations in the very early Universe. In this work, we consider a production mechanism via nonresonant oscillations with active states, the so-called Dodelson-Widrow production mechanism [79,80]. In the very early Universe, densities are high and weak interactions frequent. This generates an effective matter potential that suppresses neutrino oscillations. Therefore, the sterile state is populated only once densities are low enough for oscillations with the active neutrino eigenstates to occur (see, e.g., [81] for a detailed discussion).

The mass splitting sets the timescale for oscillations with the fourth neutrino and determines the time when flavor oscillations can start to arise. Here, we focus on eV-scale sterile neutrinos and, consequently, consider mass splittings $10^{-3} \leq \Delta m^2_{41}/eV^2 \leq 10^2$. This range corresponds to plasma temperatures between $O(100)$ and $O(1)$ MeV when the sterile starts to be populated.

The thermalization process is described by a set of differential equations that encode the evolution of the momentum distribution functions $f_\alpha(p)$ of the various neutrino flavors $\alpha \in \{e, \mu, \tau, s\}$ and an additional one for the evolution of the photon temperature, as described in detail in Ref. [82]. As far as cosmological effects of neutrinos are concerned, what matters is the momentum distribution function of all the neutrino states after their decoupling from the thermal plasma and at the end of electron-positron annihilations into photons, which happens at temperatures around 0.1 MeV.

The final momentum distribution function for the active neutrinos is very close, but not exactly equal, to a Fermi-Dirac shape with the neutrino temperature $T_\nu$ (see, e.g., [82–86]). The deviation from thermal equilibrium is mostly due to the fact that neutrino decoupling does not occur instantaneously, so there are small distortions at high momenta that come from the energy transferred during electron-positron annihilations to the few neutrinos still coupled to the plasma. For the sterile neutrino, the momentum distribution function $f_s$ depends on the degree of thermalization that it reaches. Initially, the sterile is absent, and if the mass splitting and the mixing angles are not large enough, oscillations either start too late or are not efficient enough to bring the fourth neutrino into equilibrium with the active flavors. The degree of thermalization of the sterile can be expressed in terms of the effective number of relativistic species ($N_{\text{eff}}$), which can be constrained by BBN [28] and more tightly by CMB observations [87]. After electron-positron annihilations, this parameter can be expressed as

$$N_{\text{eff}} = \frac{8}{7} \left( \frac{11}{4} \right)^{4/3} \left( \frac{\rho_e + \rho_\mu}{\rho_\gamma} \right),$$

with the energy density in photons $\rho_\gamma$ and active plus sterile neutrinos $\rho_e + \rho_\mu$. They can be computed from the integrated distribution functions, and for negligible contributions from the sterile we recover the standard value...
where the numerical coefficient is slightly different for electron, muon, or tau flavor neutrinos. If one applies this relation to the $3 + 1$ case, with $\delta m^2_{at} \rightarrow \Delta m^2_{at}$ and $\sin^2 (2 \vartheta_{at}) \approx 4 |U_{at}|^2 |U_{at}|^2$, one gets that, in order to have a fully thermalized sterile neutrino with a mass splitting around $1 \text{ eV}^2$, a mixing matrix element $|U_{at}|^2 \approx 10^{-5}$ is required. From this relation we can also see that a larger mixing matrix element generally increases $N_{\text{eff}}$ toward 4. For larger mass splittings, a smaller mixing is sufficient to generate the same level of thermalization, since the oscillations are faster; see Eqs. (9) and (10). While Eq. (18) is a rough estimate, we will see that it is a quite good approximation of the full calculation, as long as one mixing angle is varied at each time.

Including also the mixing among the active neutrino states has also been considered before, from the early simplified analyses [96,97] to more recent multirangle studies [101,102,107] performed within the averaged-momentum approximation. Although some authors have studied the multirangle $2 + 1$ scenario) and multimomment problem [103], only very recently was the evolution in the early Universe of the momentum-dependent kinetic equations for the full $4 \times 4$ density matrix of neutrinos calculated with a dedicated numerical code, FontEPluto [108], as described in Ref. [82].

At late times, the sterile neutrino becomes nonrelativistic, and it contributes to the matter energy density. At such point in the evolution, its contribution to the energy density is [80,81]

$$\Omega_{s} h^2 = \frac{m_s \Delta N_{\text{eff}}}{93.14 \text{ eV}^2} \equiv \frac{m_{s,\text{eff}}}{93.14 \text{ eV}^2},$$

where we have introduced the effective mass $m_{s,\text{eff}} \equiv m_s \Delta N_{\text{eff}}$. As mentioned in Sec. I, light sterile neutrinos might behave as a hot dark matter component and affect the evolution of matter perturbations in a similar way as active neutrinos [27]. The form of the distribution function (17) implies that the sterile neutrinos have the same average momentum as the active species. Consequently, the maximum free-streaming length of the sterile is equal to the one of an active neutrino with mass $m_s = m_s$, corresponding to a comoving wave number $k_{fs}$ [81]:

$$k_{fs} = 0.018 \Omega_{m}^{1/2} \left( \frac{m}{1 \text{ eV}} \right)^{1/2} h \text{ Mpc}^{-1},$$

if neutrinos become nonrelativistic during the matter-dominated era, or

$$k_{fs} = 0.776 \Omega_{r}^{1/2} \left( \frac{m}{1 \text{ eV}} \right)^{1/2} h \text{ Mpc}^{-1},$$

if they become nonrelativistic during the radiation-dominated era. Here, $m$ can be either the mass of an active
neutrino or the mass of a sterile neutrino produced through nonresonant oscillations. Their large velocities prevent neutrinos from clustering at scales smaller than the free-streaming length, so the collective effect of active and sterile neutrinos is a steplike suppression of the amplitude of matter perturbations below the free-streaming scale [80,81]:

\[
\frac{P_\nu}{P} \approx 1 - 8 f_{\text{hot}},
\]

where \( P_\nu \) and \( P \) refer to the matter power spectrum with or without neutrinos, respectively. The size of the suppression at small scales depends on the fraction of matter density provided by neutrinos: \( f_{\text{hot}} = (\Omega_\nu + \Omega_s)/\Omega_m \). For sterile neutrinos, the onset of the suppression is determined by the free-streaming scale and, therefore, \( m_s \), while the size of the suppression is proportional to \( m_s \Delta N_{\text{eff}} \). The cosmological effects of the sterile neutrino are, therefore, completely determined once \( m_s \) (or, equivalently, \( \Omega_s \)) and \( \Delta N_{\text{eff}} \) are specified.

In Fig. 1, we show the predicted contribution of the sterile neutrino to \( N_{\text{eff}} \) (upper row) and the fraction of total dark matter energy density it represents (lower row, assuming a total \( \Omega_s h^2 = 0.119 \) from the latest Planck measurements [44]), for two different choices for the mixing matrix elements. From the top row in Fig. 1, it is clear that if at least one of the mixing matrix elements is much larger than \( 10^{-3} \), one has \( \Delta N_{\text{eff}} \approx 1 \) for \( m_s \sim 1 \text{ eV} \) in agreement with the expectation from Eq. (18) and the results of previous analyses of active-sterile neutrino oscillations in the early Universe (see, e.g., [82,98]). Such a value of \( \Delta N_{\text{eff}} \) is at odds with what is inferred from Planck measurements of CMB anisotropies or from astrophysical determinations of the abundances of light elements. In general, we note that in some parts of the parameter space probed by our analysis, corresponding to large mixing angles and/or large masses, the contribution of sterile neutrinos to both \( N_{\text{eff}} \) and the dark matter energy density is large. The latter is also a problem, since cosmological observations also constrain the fraction of hot dark matter to be small. As we shall see in the next section, these regions of parameter space will be indeed excluded by data, in agreement with our expectations.

As we will discuss in detail in Sec. V, large contributions to \( N_{\text{eff}} \) are in tension with cosmological constraints, so in order to allow a sterile neutrino to form all the dark matter

FIG. 1. Top: \( N_{\text{eff}} \) as a function of the mixing parameters \( (\Delta m_{41}^2, |U_{e4}|^2, |U_{\mu 4}|^2, |U_{\tau 4}|^2) \), considering three equal varying mixing matrix elements \( (|U_{e4}|^2 = |U_{\mu 4}|^2 = |U_{\tau 4}|^2) \) in the left panel or one varying \( (|U_{e4}|^2) \) and two fixed \( (|U_{\mu 4}|^2 = |U_{\tau 4}|^2 = 10^{-3}) \) matrix elements in the right panel. Bottom: fraction of sterile dark matter compared to the total dark matter density \( \Omega_{\text{DM}} h^2 = 0.119 \) for the same mixing matrix elements assumptions as in the respective plots above.
one needs to move to higher masses \( m \sim \text{keV} \) and very small mixing angles. In this mass range, a sterile neutrino produced by oscillations would behave as warm (or even cold) dark matter, so limits on the hot component would not apply. Then even modest contributions to \( N_{\text{eff}} \) are sufficient to provide the required dark matter density, although other astrophysical limits apply, as reviewed, for instance, in Refs. [7,27,109]. The analysis of the mass range \( \Delta m^2_{41} > 10^2 \text{eV}^2 \), however, is beyond the scope of this paper, as we are interested in light sterile neutrinos motivated by anomalies in neutrino oscillation experiments. The main question we want to address here is whether such a sterile state is compatible with cosmological bounds.

IV. DATASETS AND METHODS

In this section, we present the datasets employed in this analysis, and we discuss the method adopted in this work, including the setup adopted for the FontEpiaNO code. Since we perform our analysis in a Bayesian framework, we also report the prior choices for the parameters varied in the analysis.

A. Cosmological data

We consider measurements of the CMB anisotropies in temperature and polarization [110], as well as determinations of the CMB lensing power spectrum [87], as provided by the latest Planck 2018 data release [44,111]. The datasets are incorporated by using the publicly available CLIK likelihood code described in Ref. [110].

To the CMB dataset, we add late-time distance and expansion rate BAO measurements from the 6dFGS [112], SDSS-MGS [113], and BOSS DR12 [114] surveys. These measurements are fully consistent with the Planck results and greatly help to break parameter degeneracies.

B. Oscillation experiments

As described in Sec. II B, various oscillation experiments find conflicting results for the active-sterile mixing parameters, and it is not possible to combine the measurements in a self-consistent way. Therefore, we group the nonconflicting oscillation measurements and present their results separately.

The \( \nu_e \) flux from nuclear reactors is measured by Bugey-3 [115], DANSS [116,117], NEOS [118], and PROSPECT [119]. Note that these experiments use differential measurements at a varied distance from the source, so a calibration of the intrinsic neutrino flux is not necessary. A combined frequentist analysis results in a preference for nonzero mixing with a sterile at about \( \sim 2.5 \sigma \) with two nearly degenerate best-fit points, located at \( (\Delta m^2_{41} \sim 0.4 \text{ eV}^2, |U_{e4}|^2 \sim 0.01) \) and \( (\Delta m^2_{41} \sim 1.3 \text{ eV}^2, |U_{e4}|^2 \sim 0.01) \) [74] (see also [76,120,121] for previous analyses).

We also include bounds on the mixing with muon neutrinos from measurements of the atmospheric flux by IceCube [73,74] and from the MINOS + experiment [75], which uses an accelerator beam with two detectors, one close (~500 m) and one far (~800 km) from the source, to put strong constraints covering a wide range of \( \Delta m^2_{34} \) and \( |U_{\mu 4}|^2 \). Both are combined in a frequentist analysis [74], and we refer to the combination of both datasets as “\( \nu_\mu \) disappearance.” Neither IceCube nor MINOS + find anomalous events and, therefore, provide upper bounds on the mixing parameters. IceCube also has a limited sensitivity on \( |U_{\tau 4}|^2 \) thanks to the low-energy data from DeepCore [73].

C. Decay experiments

In this work, we consider the latest measurements of the tritium \( \beta \)-decay spectrum released by the KATRIN Collaboration [77] that sets a limit \( m_\beta < 1.1 \text{ eV} \) at 90% CL. We include the constraint by using the approximated analytical likelihood function presented in Eq. (B.3) in Ref. [122]. Since the modification to the decay energy spectrum induced by an extra light sterile species is below the resolution of KATRIN, this likelihood is also valid for the additional sterile neutrino.

Concerning \( 0\nu\beta\beta \) searches, we consider the results from the KamLAND-ZEN Collaboration phase I [123], and phase II [124], from the GERDA Collaboration [125], and from the EXO Collaboration [126]. The constraints are included by using the approximated likelihoods given in Ref. [127]. A summary of all datasets used, the isotopes employed by each collaboration, and the respective bounds on the lifetime \( T_{1/2} \) can be found in Table I, and we show the limits on \( m_{\beta\beta} \) from the individual experiments in Fig. 2.

For completeness, we also present the individual constraints on \( m_{\beta\beta} \) from GERDA, KamLAND ZEN I and II, and EXO200 in Fig. 2. The width of each curve represents the uncertainty in the nuclear matrix element given in Table II. A modest excess of events observed at EXO200 makes the corresponding probability distribution for \( m_{\beta\beta} \) to peak at slightly nonzero values of \( m_{\beta\beta} \). The combination of the individual constraints presented in Fig. 2 gives the

<table>
<thead>
<tr>
<th>Collaboration</th>
<th>Isotope</th>
<th>( T_{1/2} ) (10^{25} \text{ yr})</th>
<th>( m_{\beta\beta} ) (eV)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>GERDA</td>
<td>^{76}\text{Ge}</td>
<td>&gt;5.3</td>
<td>&lt;0.25</td>
<td>[125]</td>
</tr>
<tr>
<td>KamLAND ZEN I</td>
<td>^{136}\text{Xe}</td>
<td>&gt;1.9</td>
<td>&lt;0.21</td>
<td>[123]</td>
</tr>
<tr>
<td>KamLAND ZEN II</td>
<td>^{136}\text{Xe}</td>
<td>&gt;10.7</td>
<td>&lt;0.10</td>
<td>[124]</td>
</tr>
<tr>
<td>EXO200</td>
<td>^{136}\text{Xe}</td>
<td>&gt;1.1</td>
<td>&lt;0.24</td>
<td>[126]</td>
</tr>
</tbody>
</table>
probability distribution reported as a red solid line labeled $0v2\beta$ in the right panel in Fig. 5.

Recently, new results from $0v\beta\beta$ collaborations have been published from the GERDA Collaboration [128], the EXO200 Collaboration [129], and the CUORE Collaboration [130]. We note that the tightest constraints on $m_{\beta\beta}$ are still those by KamLAND-ZEN phase II. As a base model, we assume a $\Lambda$CDM cosmology, with its six parameters [131], complemented with the lightest neutrino mass $m_1$, the sterile mass splitting $\Delta m_{41}$, and the three mixing matrix elements $|U_{e4}|^2$, $|U_{\mu4}|^2$, and $|U_{\tau4}|^2$.

In order to use cosmological data to constrain the sterile neutrino properties, we have to map the fundamental parameters $(\Delta m_{41}^2, |U_{e4}|^2, |U_{\mu4}|^2, |U_{\tau4}|^2)$ onto $N_{\text{eff}}$ and $m_1$, which determine its impact on cosmological observables. As described in Sec. III, we use FortEPiNo to calculate the evolution of the sterile distribution function including mixing with all three active neutrinos. We precompute an $N_{\text{eff}}$ table on a four-dimensional grid spanning the parameter space we are interested in. For the mass splitting, we consider nine logarithmically spaced samples over the range $\Delta m_{41}^2 \in [10^{-2}, 10^{2}] \text{ eV}^2$, while for each mixing matrix element we take 11 logarithmically spaced samples in the range $|U_{i4}|^2 \in [10^{-6}, 10^{-1}]$. The lower boundary chosen for the mass splitting ensures that the hierarchy $\Delta m_{31}^2 \ll \Delta m_{41}^2 \ll \Delta m_{21}^2$, always holds, so the sterile does not affect the oscillation patterns among active neutrinos, which we keep fixed to their standard values. As a base model, we assume a $\Lambda$CDM cosmology, with its six parameters [131], complemented with the lightest neutrino mass $m_1$, the sterile mass splitting $\Delta m_{41}$, and the three mixing matrix elements $|U_{i4}|^2$, $i = e, \mu, \tau$.

We calculate the evolution of cosmological perturbations using the numerical Einstein-Boltzmann solver CLASS [132–134]. Since we map the four active-sterile mixing parameters into the two cosmological parameters $N_{\text{eff}}$ and $m_1$, a very large number of samples are necessary to obtain a well-converged posterior in the full sterile parameter space. To speed the calculations up, we therefore proceed in two steps. First, we use MontePython [135,136] to run a standard Markov chain Monte Carlo (MCMC) with CMB + BAO data. For this run, we vary all $\Lambda$CDM parameters and include three massless neutrinos with $N_{\text{eff}} = 3.045$ and one additional massive state with varying mass and $\Delta N_{\text{eff}}$. Note that with the current sensitivities the data cannot distinguish between various combinations of neutrino parameters as long as they result in the same $\Delta N_{\text{eff}}$ and $\Sigma m_\nu$, so distributing the mass in different ways over the four available states does not affect the resulting constraints in a significant way (this reflects the fact that cosmology is not yet able to constrain the mass hierarchy; see, e.g., [36,37,66,67,137–139]). Given the prior $N_{\text{eff}} > 3.045$, the cosmological data set a 95% limit of $\Sigma m_\nu < 0.15 \text{ eV}$ and $N_{\text{eff}} < 3.44$. We then use a Gaussian kernel density estimate (KDE) to interpolate the marginalized $N_{\text{eff}} - \Sigma m_\nu$ likelihood surface. Evaluating this estimated likelihood is $\sim 10^5$ faster than running the full PLIK CMB likelihood which dominates the computation time, and we explicitly make sure that sampling from it results in the same constraints as obtained from the MCMC.

In a second step, we use emcee [140] together with the estimated likelihood to sample the neutrino parameter space as follows: We vary the lightest neutrino mass $m_1$, the mass splitting $\Delta m_{41}^2$, and the three mixing matrix elements $|U_{i4}|^2$. From the last four, we interpolate $\Delta N_{\text{eff}}$ from the precomputed FortEPiNo grid. Since the mass splittings $\Delta m_{31}^2$ and $\Delta m_{21}^2$ are much smaller than the sensitivity of current constraints on neutrino masses, we assume three degenerate active neutrinos $m_1 = m_2 = m_3$, so we get a total $\Sigma m_\nu = 3m_1 + \Delta N_{\text{eff}} m_4$. We then evaluate the KDE of the cosmological likelihood for $N_{\text{eff}} = 3.045 + \Delta N_{\text{eff}}$ and $\Sigma m_\nu$. In order to compare the constraints to direct experiments, we also vary the nuclear matrix elements for $^{136}$Xe and $^{76}$Ge defined in Eq. (14) and the three Majorana phases $\alpha_j$ in Eq. (15) to derive limits on $m_{\beta\beta}$ and $m_{\beta\beta}$.

TABLE II. Priors for all noncosmological parameters varied in our analysis. We comment on different prior choices for $\Delta m_{41}^2$ in detail in Sec. V B.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior shape</th>
<th>Prior bounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lightest neutrino mass</td>
<td>$m_1/\text{eV}$</td>
<td>Flat</td>
</tr>
<tr>
<td>Mass splitting</td>
<td>$\Delta m_{41}^2/\text{eV}^2$</td>
<td>Log</td>
</tr>
<tr>
<td>Mixing matrix elements</td>
<td>$</td>
<td>U_{i4}</td>
</tr>
<tr>
<td>Majorana phases</td>
<td>$\alpha_j$ (j $\in$ [2, 3, 4])</td>
<td>Flat</td>
</tr>
<tr>
<td>Nuclear matrix elements</td>
<td>$</td>
<td>M_{0j}(^{136}\text{Xe})</td>
</tr>
<tr>
<td></td>
<td>$</td>
<td>M_{0j}(^{76}\text{Ge})</td>
</tr>
</tbody>
</table>
A summary over all non-$\Lambda$CDM parameters varied in the analysis, their prior bounds, and shapes is given in Table II.

For constraints involving the $\beta$-decay and $0\nu\beta\beta$ likelihoods described in Sec. IV C, we marginalize over all parameters listed in Table II as well. The constraints from reactors and $\nu_\mu$ appearance measurements from MINOS + and IceCube (see Sec. IV B), on the other hand, are derived in a frequentist framework and do not depend on the prior choices made here.

V. RESULTS

In this section, we present the results of our analysis. First, we discuss the complementarity of the various datasets either in terms of the fundamental parameters of the sterile mixing matrix or in terms of the derived parameters $m_\beta$ and $m_{\beta\beta}$ directly constrained by terrestrial experiments. Then, we explore the effect of assuming other priors on the sterile mass parameter for the cosmological datasets. In the same context, we compare our constraints with the ones previously obtained in a simplified scenario where the sterile is coupled to only one neutrino species.

A. Constraints from cosmology and direct experiments

Cosmology provides tight limits on the sum of neutrino masses $\Sigma m_\nu$, and the effective number of relativistic species $N_{\text{eff}}$, that can be translated to tight bounds on the parameter space of the mixing matrix with significant relativistic energy contributions from the sterile neutrino. After marginalizing over all other cosmological and nuisance parameters, we obtain the constraints on the mixing matrix elements in Fig. 3 from CMB + BAO. As any mixing reaching a value of $|U_{e\alpha}|^2 \approx 10^{-3}$ starts to populate the sterile state and leads to detectable $N_{\text{eff}}$ contributions as seen in Fig. 1, this is where cosmological data become constraining. There are only small differences between mixing for the various flavors, so the limits on all of the matrix elements are very similar. Since the constraints on both the mass $m_\alpha$ derived from the total sum $\Sigma m_\nu = 3m_1 + \Delta N_{\text{eff}} m_4$ and $N_{\text{eff}}$ vanish for small amplitudes of the sterile distribution function set by $\Delta N_{\text{eff}}$, the cosmological data, in principle, allow large sterile masses as long as the mixing is small and the sterile state is not thermalized or populated to a significant level. As mentioned in Sec. III, much higher mass ranges in the keV range and larger are, in principle, possible if the mixing angles are small enough. In such a case, the free-streaming length $k_{\text{fs}}$ of the sterile neutrino would be pushed to smaller scales, and it would form a warm or even a cold dark matter component [7,27,109].

In Fig. 4 (left panel), we compare the cosmological results with limits obtained from neutrinoless double-$\beta$-decay and tritium decay measurements by KATRIN [77] in the $\Delta m^2_{41} - |U_{e\alpha}|^2$ plane, since the decay experiments are not sensitive to the other matrix elements. While sensitivity of the neutrinoless double-$\beta$-decay experiments and $\beta$-decay measurements from KATRIN on the sterile parameters are comparable, cosmological bounds are orders of magnitude stronger. To address the sterile neutrino interpretation of the reactor anomaly, we also show the parameter space preferred by a combined fit of short-baseline measurements of the reactor antineutrino flux. As explained in Sec. I, such experiments observe a preference in favor of a sterile with $\Delta m^2_{41} \sim 0(\text{eV}^2)$ and $|U_{e\alpha}|^2 \approx 10^{-2}$. While these parameter values are compatible with current measurements from $0\nu\beta\beta$ and KATRIN, they are in severe tension with the CMB + BAO data. A mixing of the size needed to explain the reactor data would be more than sufficient to bring the sterile in thermal equilibrium in the early Universe. On the right in Fig. 4, we show a similar comparison for the mixing angle $|U_{\mu\alpha}|^2$ and present cosmological constraints together with bounds from $\nu_\mu$ disappearance measurements from IceCube and MINOS + described in Sec. IV B. While the experimental sensitivity is stronger than on $|U_{e\alpha}|^2$, we still find the CMB + BAO dataset to be more constraining. While the cosmological limits rely on model assumptions and can be slightly relaxed in extended parameter spaces, accommodating $N_{\text{eff}} \approx 4$ is very challenging [141,142].

We also map the cosmological bounds onto the parameter space $m_\beta$ or $m_{\beta\beta}$ directly probed by decay experiments in Fig. 5. On the left side, we show limits from CMB + BAO together with the latest results from KATRIN [77]. Since $m_\beta$ defined in Eq. (13) receives contributions from the sterile, the resulting limit from cosmology in the extended $3 + 1$ parameter space is slightly higher compared to the expectation $m_\beta \approx m_1 \approx \Sigma m_\nu / 3$ from standard neutrinos. The CMB + BAO data together with the prior assumption $N_{\text{eff}} > 3.045$ lead to an upper bound on the neutrino mass sum of $\Sigma m_\nu < 0.15$ eV, corresponding to a limit $m_{\beta\text{cosmo}} < 0.05$ eV assuming only three active neutrinos, which is slightly degraded to $m_{\beta\text{cosmo}} < 0.09$ eV if the sterile is included. However, this constraint is still a factor of $\sim 10$ stronger than current experimental limits from KATRIN.

On the right-hand side in Fig. 5, we show a similar comparison of the derived bound on $m_{\beta\beta}$ from cosmology and direct $0\nu\beta\beta$ searches. Note that the posterior from the combined $0\nu\beta\beta$ data has a maximum at slightly nonzero values due to an excess of events observed by EXO200 compared to the background expectation [129]. In Table III, we present a summary of all 95% bounds on the sterile mass $m_4$, the mixing matrix elements $|U_{e\alpha}|^2$ for each flavor, $m_\beta$ and $m_{\beta\beta}$, comparing the cosmological bound to the respective strongest bound from direct searches. Cosmology provides the tightest limits on all parameters, for most of them by orders of magnitude.
B. Priors and parameter space volume effects

Since all cosmological bounds presented in Fig. 3 and discussed in the previous Sec. VA provide only upper bounds on the sterile parameters in a Bayesian framework, the choice of priors affects the limit. In this section, we focus on various priors for the mass splitting $\Delta m^2_{41}$. While we adopt a logarithmic prior as a standard case, since cosmological data are ignorant of the order of magnitude of the mass splitting, one can also argue that the parameter of interest is either the mass-squared difference itself—or the sterile mass scale, since cosmological data are approximately sensitive to the energy density parameter $\Omega_s \propto m_s \approx m_4$. We, therefore, repeat the previous cosmological analysis using a flat prior either on $\log_{10} \Delta m^2_{41}$, or on $\Delta m^2_{41}$, or on $m_4$, always considering the same range specified before in Table II. Note that while sampling over...
we enforce the additional constraint $\Delta m_{41}^2 > 10^{-2}$ eV$^2$ to make sure that the neutrino mass hierarchy is unchanged.

We present the resulting limits on the mixing matrix elements in Fig. 6 and in Table III. The different choices indeed affect the results, and bounds on the mixing matrix elements are stronger for the flat priors on $m_4$ or $\Delta m_{41}^2$. This is a consequence of the mapping between parameter spaces: The most important constraint from the data is on $N_{\text{eff}}$ (which also contributes to $\Sigma m_\nu$ through $\Delta N_{\text{eff}} m_4$), and there is a strong degeneracy between the mass splitting and the mixing matrix elements as explained in Sec. II and seen from the upper left plot in Fig. 1. For a fixed higher sterile mass splitting, the mixing angles have to be smaller to stay within the allowed $N_{\text{eff}}$ region. Therefore, the more weight the priors give to large masses $m_4$ or $\Delta m_{41}^2$, the smaller the matrix elements $|U_{\alpha 4}|^2$ have to be to fulfill the $\Delta N_{\text{eff}}$ constraint. We want to emphasize that the choice of priors does not affect the degree of tension with reactor and $\bar{\nu}_e$

FIG. 4. Left: marginalized 68% and 95% constraints on the mass splitting $\Delta m_{41}^2$ and mixing matrix element $|U_{e4}|^2$ from cosmology (blue), from the tritium $\beta$-decay measurements by KATRIN (green), and from neutrinoless double-$\beta$-decay experiments ($0\nu\beta\beta$, red), compared with the preferred frequentist regions by the joint fit [74] of reactor experiments discussed in Sec. IV B, which are in strong tension with cosmological bounds. Right: cosmological 68% and 95% marginalized constraints on the mixing matrix element $|U_{\mu 4}|^2$ compared to (frequentist) $\nu_\mu$ disappearance results [74] from IceCube and MINOS+ (gray).

FIG. 5. Left: one-dimensional probability distribution for the neutrino effective mass $m_\beta$ from cosmology and $\beta$-decay measurements from KATRIN. Right: one-dimensional probability distribution for the mass parameter $m_{\beta\beta}$ from $0\nu\beta\beta$ and CMB + BAO. The combined $0\nu\beta\beta$ probability distribution is peaked at slightly nonzero values of $m_{\beta\beta}$ due to a modest excess of events observed at EXO200.
FIG. 6. Cosmological marginalized constraints on the mixing matrix elements $|U_{\ell 4}|^2$ for flat priors on either log $\Delta m_{41}^2$ (blue line), $m_4$ (orange line), or $\Delta m_{41}^2$ (green line), from CMB + BAO data. Off-diagonal panels show 68% and 95% confidence level probability contours. Panels along the diagonal show one-dimensional probability distributions. The more weight the prior gives to higher sterile masses, the lower the allowed $|U_{\ell 4}|^2$ values in order to stay within the $N_{\text{eff}}$ range allowed by the cosmological data.

TABLE III. Upper limits (95%) for the sterile neutrino mass, the parameters of the mixing matrix, and $m_\beta$ and $m_{\beta\beta}$.

For the limits from direct experiments, we take a conservative approach and consider only probes that are not in tension with cosmology. For those, we quote the strongest bound on each parameter. We present cosmological limits for the different prior choices for the mass splitting either on log $\Delta m_{41}^2$ used in Sec. VA, or on either $m_4$ or $\Delta m_{41}^2$ discussed in Sec. VB, but note that the prior choice barely affects the resulting constraints on $m_\beta$ and $m_{\beta\beta}$.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Experimental upper limit (95%)</th>
<th>$P(\log \Delta m_{41}^2)$</th>
<th>$P(m_4)$</th>
<th>$P(\Delta m_{41}^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_4$ (eV)</td>
<td>⋮</td>
<td>1.6</td>
<td>4.4</td>
<td>6.8</td>
</tr>
<tr>
<td>$\log_{10}</td>
<td>U_{e4}</td>
<td>^2$</td>
<td>⋮</td>
<td>-3.04</td>
</tr>
<tr>
<td>$\log_{10}</td>
<td>U_{\mu4}</td>
<td>^2$</td>
<td>-2.2 ($\nu_\mu$)</td>
<td>-3.17</td>
</tr>
<tr>
<td>$\log_{10}</td>
<td>U_{\tau4}</td>
<td>^2$</td>
<td>-0.8 ($\nu_\tau$)</td>
<td>-3.18</td>
</tr>
<tr>
<td>$m_\beta$ (eV)</td>
<td>0.9 (KATRIN)</td>
<td></td>
<td></td>
<td>0.09</td>
</tr>
<tr>
<td>$m_{\beta\beta}$ (eV)</td>
<td>0.08 (0$\beta\beta$)</td>
<td></td>
<td></td>
<td>0.07</td>
</tr>
</tbody>
</table>
appearance measurements. A different choice of priors changes the weight of the mapping \( \Delta m^2 \rightarrow N_{\text{eff}} \), but the region in parameter space necessary to explain the anomalies leads to an almost thermal sterile with \( \Delta N_{\text{eff}} \approx 1 \). This is excluded by cosmological data independent of the assumptions made in the analysis. The limits also imply

\[
|U_{\text{eff}}|^2 = \sum_{i} |U_{\alpha i}|^2,
\]

that, in the parameter space considered here, the sterile neutrino can form only a negligible contribution to the total dark matter density.

The resulting limits on the mixing matrix elements summarized in Table III are robustly constrained to be \( |U_{\alpha i}|^2 < 10^{-3} \) for all flavors. While mixing with electron neutrinos is the most important channel and the resulting bounds on \( |U_{\alpha i}|^2 \) are slightly more stringent, there is overall only a minor difference between mixing with the different active neutrinos. It is, therefore, interesting to understand the difference between a simplified case where the sterile is coupled to only one active neutrino (often assumed to be \( \nu_e \)) and the full mixing with all flavors explored in this paper. We present the comparison between the cosmological constraints on \( \Delta m^2_{41} \) and \( |U_{\alpha i}|^2 \) assuming either a 1 + 1 or a 3 + 1 scenario on the left side in Fig. 8.

The results clearly differ, and the 1 + 1 mixing allows for higher mass splittings of the sterile. However, due to parameter space volume effects, it is expected that the limits look different. For every point in the \( (\Delta m^2_{41}, |U_{\alpha i}|^2) \) plane, in the 3 + 1 scenario there are more ways to achieve a higher \( \Delta N_{\text{eff}} \) and \( \Sigma m_{\nu} \) by increasing the other mixing matrix elements. Higher mass splittings with small \( |U_{\alpha i}|^2 \) that lead to negligible cosmological effects in the 1 + 1 model are then still unlikely in the 3 + 1 case, since the other mixing angles have to be small as well. We explicitly test this explanation by adding another comparison case where the sterile is still coupled only to electron neutrinos, but we consider an effective mixing matrix element

\[
|U_{\text{eff}}|^2 = \sum_{i} |U_{\alpha i}|^2,
\]

for a single sterile coupled only to one neutrino species \( \nu_e \) (red line, same as left), \( \nu_\mu \) (green line), or \( \nu_\tau \) (gray line). The differences between coupling to the different flavors are very small.

FIG. 7. Marginalized 68% and 95% CMB + BAO constraint in the \( N_{\text{eff}}-\Omega_s \) plane using a linear prior on the mass scale. Gray bands are excluded by the range of mass splittings considered in this work: Very small \( \Omega_s \) and large \( N_{\text{eff}} \) values require \( \Delta m^2_{41} \) to be comparable to the ones of the active neutrinos, while small \( N_{\text{eff}} \) contributions and large cosmic densities are possible only for sterile masses of keV or higher. For the parameter region considered here, the sterile can form only a negligible contribution to the total dark matter density.

FIG. 8. Left: cosmological 68% and 95% marginalized constraints assuming either a 1 + 1 case with the sterile coupled only to \( \nu_e \) (purple line), compared to the full 3 + 1 scenario including the full mixing matrix (blue line) from Fig. 3. The difference is mostly a parameter space volume effect, since the limits obtained with a modified coupling only to \( \nu_e \) with an effective mixing \( |U_{\text{eff}}|^2 = \Sigma_{\alpha} |U_{\alpha i}|^2 \) (\( \alpha \in \{e, \mu, \tau\} \), red line) are almost the same as for the 3 + 1 case. Right: cosmological 68% and 95% marginalized constraints on \( |U_{\text{eff}}|^2 \) for a single sterile coupled only to one neutrino species \( \nu_e \) (red line, same as left), \( \nu_\mu \) (green line), or \( \nu_\tau \) (gray line). The differences between coupling to the different flavors are very small.
and we sample over three distinct contributions $|U_{aj}|^2$ to make up for the larger parameter space volume in the full $3+1$ mixing scenario. As can be seen on the left side in Fig. 8, this parameter space volume effect almost completely accounts for the difference between the scenarios. On the right-hand side in Fig. 8, we test this effective mixing scenario by accounting for the parameter space effect in the same way but coupling the sterile to either one of electron, muon, or tau neutrinos, respectively. Again, differences between the individual flavors are very small, and an effective coupling to $\nu_e$ provides a very good approximation to the full dynamics.

As a consequence, we find that the production of light sterile neutrinos via mixing with three active states can be modeled within a $1+1$ model with a single sterile mixing with $\nu_e$ as long as the effective mixing matrix element in Eq. (23) is used. Therefore, the abundance of a single sterile computed in the extended $3+1$ model is, to an excellent approximation, very similar to that found in the effective $1+1$ scenario, provided that the $1+1$ squared mixing matrix element is the sum of the individual squared mixing matrix elements in the $3+1$ case.

VI. CONCLUSION

Several anomalies observed in short-baseline oscillation experiments hint at a new neutrino mass state at the eV scale. In this paper, we have provided a consistent framework to constrain the additional neutrino mass splitting and the active-sterile mixing angles with cosmological data. This also allows us to compare the bounds from cosmological data with results from oscillations, $\beta$ decay, and $0\nu\beta\beta$ measurements in a common parameter space.

For the first time, we performed this analysis in a full $3+1$ scenario where the sterile state is mixed with all three active neutrinos. In order to map the sterile neutrino mass-squared splitting and mixing matrix elements for each flavor $\Delta m^2_{41}$ and $|U_{a4}|^2$ onto the cosmological observables, we have solved the evolution of the $4 \times 4$ neutrino density matrix in the early universe with the FortEPRANo [82] code and find that the resulting distribution function of the fourth neutrino state is well approximated by a Dodelson-Widrow form [79].

A combination of CMB and BAO data currently provides the strongest available bounds on the sterile mass splitting $\Delta m^2_{41}$ and the mixing matrix elements $|U_{a4}|^2$. While the limit on the mass scale depends on prior assumptions, we robustly find that, once any mixing matrix element reaches a level of $|U_{a4}|^2 \approx 10^{-3}$, the new state would give rise to a detectable relativistic energy contribution in the early Universe not seen in cosmological data. The parameter space needed to explain short-baseline anomalies with a sterile neutrino leads to a fully thermalized relativistic species with $N_{\text{eff}} \approx 4$ and is in strong tension with the CMB bounds. We also derive limits on the effective electron neutrino mass $m_\nu$ and the Majorana mass parameter $m_{\nu\nu}$, measurable in $\beta$- and $0\nu\beta\beta$-decay experiments, from cosmological data, finding $m_\nu < 0.09$ eV and $m_{\nu\nu} < 0.07$ eV at 95% CL. These constraints are tighter than the ones obtained from the latest direct laboratory measurements. Our main results in terms of limits on the sterile mass scale $m_4$, the mixing matrix parameters for each flavor $|U_{a4}|^2$, $m_\nu$, and $m_{\nu\nu}$, are summarized in Table III.

We have explored the effect of prior choices on the cosmological bounds repeating the analysis with different priors on the mass splitting. Since a sizable sterile contribution to $N_{\text{eff}}$ can be produced either by a large mass splittings or large mixing angles, the main effect of priors is to shift the weight between the two quantities. The maximal contribution to $N_{\text{eff}}$ allowed by the data is fixed, so the more weight the prior gives to high mass splittings $\Delta m^2_{41}$, the lower the mixing angles have to be to stay within the allowed region.

Almost all limits on sterile neutrinos from cosmology previously reported in the literature were derived in a simplified $1+1$ scenario where the sterile is coupled to only one active neutrino. Even though the mixing with the different active flavors is almost equivalent, we show that the results differ significantly from the ones obtained with a $3+1$ mixing scheme assumed for this work. This can be largely attributed to parameter space volume effects and can be accounted for by coupling the sterile with one active neutrino with effective mixing as given by Eq. (23). In any case, our results show that allowing for more than one active-sterile mixing does not relax the tension between the active-sterile neutrino parameters favored by the oscillation anomalies and present cosmological observations.

Since cosmological data currently dominate the constraints, we do not perform a joint analysis with laboratory experiments at this time. However, the framework provided here can be the basis for new global constraints on light sterile neutrino properties once new data from laboratory searches become available.

ACKNOWLEDGMENTS

S. H., P. F. d. S., and K. F. acknowledge support from the Vetenskapsrådet (Swedish Research Council) through Contract No. 638-2013-8993 and the Oskar Klein Centre for Cosmoparticle Physics. S. G. acknowledges support from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie individual Grant Agreement No. 796941. S. G. and S. P. acknowledge support from the Spanish Grants No. FPA2017-85216-P (AEI/FEDER, UE), No. PROMETEO/2018/165 (Generalitat Valenciana), and the Red Consolider MultiDark FPA2017-90566-REDC. M. G. acknowledges support from Argonne National Laboratory (ANL). ANL’s work was supported by the DOE under Contract No. W7405-ENG-36. M. G. and
M. L. acknowledge support from the ASI Grant No. 2016-24-H.0 COSMOS "Attività di studio per la comunità scientifica di cosmologia" and from INFN through the InDark and Gruppo IV fundings. S. V. acknowledges support from the Isaac Newton Trust and the Kavli Foundation through a Newton-Kavli Fellowship and acknowledges a College Research Associateship at Homerton College, University of Cambridge. K. F. acknowledges support from the Jeff and Gail Kosdosky Endowed Chair in Physics, DOE Grant No. DE-SC007859, and the LCTP at the University of Michigan. This work is partly based on observations obtained with Planck, an ESA science mission with instruments and contributions directly funded by ESA Member States, NASA, and Canada.

[14] GALLEX and SAGE were designed to find solar neutrinos, but the anomalies appear in the short-baseline calibration runs.
[48] Note that local supernova measurements using a distance calibration with the tip of the red giant branch [49] find results compatible with sound-horizon-based methods.
[64] While not yet conclusive, recent cosmological and oscillations probes slightly favor normal ordering over inverted ordering, with varying degree of preference [8,11,37,46,65–69].


[123] A. Gando et al. (KamLAND-Zen Collaboration), Limit on Neutrinoless $\beta\beta$ Decay of $^{136}$Xe from the First Phase of KamLAND-Zen and Comparison with the Positive Claim in $^{76}$Ge, Phys. Rev. Lett. 110, 062502 (2013).


[125] M. Agostini et al., Background-free search for neutrinoless double-$\beta$ decay of $^{76}$Ge with GERDA, Nature (London) 544, 47 (2017).


[131] The $\Lambda$CDM six parameters are the energy density in baryonic matter $\Omega_b h^2$ and in cold dark matter $\Omega_c h^2$, the angular size of the sound horizon at recombination $\theta$, the reionization optical depth $\tau$, the amplitude $A_s$, and the spectral index $n_s$ of the primordial scalar perturbations.


