Non-unitary three-neutrino mixing in the early Universe

Stefano Gariazzo, Pablo Martínez-Miravé, Olga Mena, Sergio Pastor and Mariam Tórtola

Abstract. Deviations from unitarity in the three-neutrino mixing canonical picture are expected in many physics scenarios beyond the Standard Model. The mixing of new heavy neutral leptons with the three light neutrinos would in principle modify the strength and flavour structure of charged-current and neutral-current interactions with matter. Non-unitarity effects would therefore have an impact on the neutrino decoupling processes in the early Universe and on the value of the effective number of neutrinos, $N_{\text{eff}}$. We calculate the cosmological energy density in the form of radiation with a non-unitary neutrino mixing matrix, addressing the possible interplay between parameters. Highly accurate measurements of $N_{\text{eff}}$ from forthcoming cosmological observations can provide independent and complementary limits on the departures from unitarity. For completeness, we relate the scenario of small deviations from unitarity to non-standard neutrino interactions and compare the forecasted constraints to other existing limits in the literature.

Keywords: cosmological neutrinos, neutrino properties, neutrino theory

ArXiv ePrint: 2211.10522
1 Introduction

In recent years, an unprecedented accuracy has been reached in the extraction of the neutrino mixing parameters [1–3]: solar, atmospheric, reactor and accelerator neutrino oscillation experiments have provided conclusive evidence for neutrino masses in nature. However, the underlying mechanism for generating the neutrino masses remains unknown. Although there is a plethora of possible theoretical frameworks aiming to settle this issue, generally, a common mechanism consists in the generation of the neutrino masses through the mediation of heavy neutral leptons (HNLs). Then, the lepton mixing matrix has to account for the possible admixture between light neutrinos and the HNLs. In seesaw models [4–8], the smallness of neutrino masses would indicate a very high scale for the conventional realisations of this mechanism, leading to a negligible mixing between light neutrinos and the HNLs. In seesaw models [4–8], the smallness of neutrino masses would indicate a very high scale for the conventional realisations of this mechanism, leading to a negligible mixing between light neutrinos and the HNLs [9], rendering a mixing matrix very close to unitary. Nonetheless, in ‘low scale’ realisations of the seesaw mechanism [10–15], much more relevant departures from unitarity can be expected. In this context, two general scenarios can be envisioned: (i) not-so-heavy neutral leptons, with masses such that they can be produced in oscillation experiments, or (ii) canonical HNLs, which are kinematically inaccessible in oscillation experiments. The former is being scrutinised due to its potential to explain short-baseline neutrino oscillation anomalies (see e.g. [16–20]), and its cosmological effects have been studied extensively in the past [21–24]. However, when discussing the non-unitarity (NU) of the three-neutrino mixing matrix, we shall consider exclusively the latter. Direct searches for HNLs and precision flavour observables [25] allow to constrain the light-heavy mixing for a wide range of masses. These measurements rely on the fact that the charged-current leptonic interactions are modified. For neutrino experiments operating at energies below the mass of the HNLs, neutrino neutral-current interactions are
also altered. As a consequence, complementary limits from both the non-unitarity mixing and the changes in the structure of the interactions can also be obtained within this context.

In the early Universe, electroweak processes kept neutrinos in thermal contact with electrons, positrons and photons. Once the collisions became ineffective at temperatures around the MeV, neutrinos decoupled from the cosmic plasma. Consequently, any modification of neutrino interactions would change the picture of neutrino decoupling and alter the resulting background of relic neutrinos. The role of neutrinos is encoded in the effective number of relativistic species \( N_{\text{eff}} \), which quantifies relativistic degrees of freedom different from photons that contribute to the cosmological radiation density. Within the canonical picture of neutrino decoupling, its value after neutrino decoupling is \( N_{\text{eff}} = 3.0440 \) \cite{26–29}, in excellent agreement with the most recent determination from the combination of several cosmological probes, \( N_{\text{eff}} = 2.99^{+0.33}_{-0.33} \) at 95% confidence level (CL) from Planck 2018 (TT, TE, EE + lowE + lensing + BAO) \cite{30}. This measurement is expected to improve significantly with forthcoming data from Cosmic Microwave Background (CMB) measurements from the Simons Observatory (\( \sigma(N_{\text{eff}}) \simeq 0.05–0.07 \)) \cite{31} and CMB-S4 (\( \sigma(N_{\text{eff}}) \simeq 0.02–0.03 \)) \cite{32}.

In this work, we address the impact of a non-unitary three-neutrino mixing on the process of neutrino decoupling and determine the expected complementary bounds on departures from unitarity that will be obtained by near future CMB surveys via their measurement of \( N_{\text{eff}} \). The structure of this manuscript is as follows. In section 2, we introduce the formalism and notation adopted for the study of non-unitarity three-neutrino mixing and in section 3 we discuss the main implications for neutrino cosmology. The obtained results are presented in sections 4 and 5, where we also discuss a mapping to the non-standard interactions scenario in case of small departures from unitarity, as well as the complementarity with other searches. We conclude in section 6.

## 2 Non-unitarity in the three-neutrino mixing scenario

Let us denote the \( n \)-vector representing the neutral lepton states with definite mass, both light and heavy, by \( \nu_L \). In this basis, the charged- and neutral-current (CC and NC, respectively) Lagrangians read

\[
L_{\text{CC}} = \frac{g}{\sqrt{2}} (W^-)_\mu \bar{\nu}_L \gamma^\mu K \nu_L + \text{H.c.} \quad \text{and} \quad L_{\text{NC}} = \frac{g}{2 \cos \theta_W} Z_\mu \bar{\nu}_L (K^\dagger K) \nu_L + \text{H.c.},
\]

(2.1)

where we have introduced the matrix \( K \), which relates the three active flavour eigenstates and the \( n \) mass eigenstates.\(^1\) Here, we have only included the interactions with electrons since they are the only ones relevant for neutrino decoupling. At low energies, it is possible to express the CC and NC interactions in terms of four-fermion interactions, i.e.

\[
L_{\text{CC}} = -2\sqrt{2} G_F \sum_{i,j} (K^\dagger)_i \nu_j (\bar{\nu}_i \gamma^\mu P_L \nu_j) (\bar{\nu}_j \gamma^\mu P_L \nu_j),
\]

(2.2)

and

\[
L_{\text{NC}} = -2\sqrt{2} G_F \sum_{X = L, R} g_X \sum_{i,j} (K^\dagger K)_{ij} (\bar{\nu}_i \gamma^\mu P_L \nu_j) (\bar{\nu}_j \gamma^\mu P_X \nu_j),
\]

(2.3)

where in the case of eq. (2.2) we have performed a Fierz transformation. Here \( G_F \) is the Fermi constant and the index \( X = \{ L, R \} \), so that \( P_X \) denotes the chiral projectors \( P_{R,L} = (1 \pm \gamma_5)/2 \).

\(^1\)Note that the \( 3 \times n \) matrix \( K \) also includes the rotation required to define the charged-leptons mass basis.
| NU parameters | $|\alpha_{21}|$ | $|\alpha_{31}|$ | $|\alpha_{32}|$ | $\alpha_{11}$ | $\alpha_{22}$ | $\alpha_{33}$ |
|---------------|----------------|----------------|----------------|----------------|----------------|----------------|
| 3σ bounds    | $< 0.025$      | $< 0.075$      | $< 0.02$       | $> 0.93$       | $> 0.98$       | $> 0.72$       |

**Table 1.** Current constraints on the non-unitarity parameters from ref. [36].

Note that, when computing the amplitude of the decay and scattering processes, the sum over mass eigenstates should be limited to the largest mass eigenstate kinematically accessible. Thus, it is useful to separate the matrix $K$ in two blocks,

$$K = \left( \begin{array}{cc} N & S \end{array} \right),$$

where $N$ describes the mixing among the three lightest states and $S$ accounts for the mixing between the three lightest states and the $n - 3$ heavy ones.

In what follows, we shall assume that only three mass eigenstates lie below the energy scale of interest here. In that case, the sums in eqs. (2.2)-(2.3) only include the three lightest mass eigenstates and only the $3 \times 3$ submatrix $N$ is involved. Notice that the unitarity of the full $n \times n$ lepton mixing matrix does not require the submatrix $N$ appearing in the effective CC and NC Lagrangians to be unitary. In order to account for departures of non-unitarity in $N$, we adopt the following parametrisation [33, 34]

$$N = \begin{pmatrix} \alpha_{11} & 0 & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} U ,$$

where $U$ is the usual unitary three-neutrino mixing matrix, $\alpha_{ii}$ are real parameters, and $\alpha_{ij}$ ($i \neq j$) are complex and provide additional sources of CP violation.\(^2\) The off-diagonal parameters are related to the diagonal ones through the triangular inequality [35],

$$|\alpha_{ij}| \leq \sqrt{(1 - \alpha_{ii}^2)(1 - \alpha_{jj}^2)} .$$

Additionally, from the unitarity of the full $n \times n$ mixing matrix [36], the following conditions apply

$$\alpha_{11}^2 |\alpha_{21}|^2 \leq (1 - \alpha_{11}^2)(1 - \alpha_{22}^2 - |\alpha_{21}|^2) ,$$

$$\alpha_{11}^2 |\alpha_{31}|^2 \leq (1 - \alpha_{11}^2)(1 - \alpha_{33}^2 - |\alpha_{31}|^2 - |\alpha_{32}|^2) ,$$

$$|\alpha_{22}\alpha_{32} + \alpha_{21}\alpha_{31}|^2 \leq (1 - \alpha_{22}^2 - |\alpha_{21}|^2)(1 - \alpha_{33}^2 - |\alpha_{31}|^2 - |\alpha_{32}|^2) .$$

Non-unitary three-neutrino mixing would not only alter CC and NC interactions but also the picture of neutrino oscillations, whose observations at both short- and long-baseline experiments can be used to constrain the non-unitarity parameters. Current bounds are summarized in table 1, from [36].

---

\(^2\)The relation between the $\alpha_{ij}$ parameters and the parametrisation of the full $n \times n$ mixing matrix is detailed in appendix A.
3 The impact of non-unitary three-neutrino mixing in neutrino decoupling

In the following, we shall concentrate on a scenario in which HNLs can not be produced at temperatures $\sim$MeV and, therefore, their distribution is Boltzmann-suppressed. Consequently, only the evolution of the three lightest mass eigenstates needs to be studied. In previous studies of neutrino decoupling, the evolution of the neutrino distributions was mostly addressed in terms of flavour states \cite{27,28}, but an adequate treatment of non-unitarity requires to work in the mass basis (see e.g. \cite{26}). The reason is that the truncation in the sum over accessible states is only well-defined for mass eigenstates.

We adopt the usual definitions for the comoving variables $x = m_e a$, $y = p a$, $z = T_\gamma a$, in terms of the electron mass $m_e$, the cosmological scale factor $a$, the neutrino momentum $p$, and the photon temperature $T_\gamma$ \cite{23,28}. The evolution of the density matrix $\varrho$ for tree neutrinos in the mass basis reads \cite{26}

$$\frac{d\varrho(y)}{dx} = \sqrt{\frac{3m_{Pl}^2}{8\pi\rho}} \left\{ -\frac{x^2}{2y} \left[ \frac{\mathbb{M}}{m_e} - \frac{2\sqrt{2}G_F m_e^6}{m_W^2} \left( \rho_\ell + \rho_\nu \right) \right] , \varrho \right\} + \frac{m_e^3}{x^4} I(\varrho), \quad (3.1)$$

where $m_{Pl}$ denotes the Planck mass, $\rho$ is the energy density of the Universe, $m_W$ is the mass of the $W$ gauge bosons and we have defined the diagonal neutrino mass matrix $\mathbb{M} = \text{diag}(0, \Delta m_{21}^2, \Delta m_{31}^2)$. We have also introduced the energy density of charged leptons, $E_\ell$, and the pressure of charged leptons, $P_\ell$. In this scenario, it is useful to define the following matrices

$$\langle Y_L \rangle_{ij} = g_L \left( N^\dagger N \right)_{ij} + \left( N^\dagger \right)_{ie} N_{ej}, \quad \langle Y_R \rangle_{ij} = g_R \left( N^\dagger N \right)_{ij}, \quad (3.2)$$

with the $i$ and $j$ indices running from 1 to 3, and where we have introduced the coefficients $g_L = -1/2 + \sin^2\theta_W$ and $g_R = \sin^2\theta_W$, which depend on the weak mixing angle, $\theta_W$. Regarding the charged-lepton energy density and pressure,

$$E_\ell = \rho_e E_{NU}, \quad P_\ell = P_e E_{NU} \simeq \rho_e E_{NU}/3, \quad (3.4)$$

we only consider the contribution of positrons and electrons (the presence of muons and taus is Boltzmann suppressed), in terms of the $e^\pm$ energy density ($\rho_{e^\pm} = \rho_{e^-} + \rho_{e^+}$) and pressure, as well as via a matrix accounting for their contribution to the matter potential

$$E_{NU} = Y_L - Y_R, \quad (3.5)$$

which depends on the vector component of the interactions.

Finally, the term $I(\varrho)$ in (3.1) encodes the collision integrals, arising from the neutrino-electron scattering and annihilation processes, as well as from the neutrino-neutrino interactions. The terms accounting for neutrino-electron interactions are proportional to $G_F^2$ and also depend on the matrices $Y_L$ and $Y_R$ defined in eqs. (3.2)–(3.3), see appendix B for details. Neutrino-neutrino interactions are also modified if the three-neutrino mixing is not unitary. The low-energy Lagrangian accounting for neutrino self-interactions reads

$$\mathcal{L}_{\nu SI} = -2\sqrt{2}G_F \sum_{i,j,k,m} \left( N^{\dagger} N \right)_{ij} \left( N^{\dagger} N \right)_{mk} (\bar{\nu}_i \gamma^\mu P_L \nu_j)(\bar{\nu}_k \gamma^\nu P_L \nu_m). \quad (3.6)$$
It has been shown that neutrino self-interactions lead to a refractive term in the Hamiltonian, which does not modify the picture qualitatively and it only gives a sub-leading contribution to $N_{\text{eff}}$,\(^3\) hence we ignore such terms. When doing so, we also assume the absence of an initial asymmetry between neutrinos and antineutrinos.

Notice that we have treated the Fermi constant $G_F$ as a fundamental quantity that we can precisely measure experimentally. However, in the presence of non-unitary three-neutrino mixing, what is measured is an effective parameter depending on the mixing between light neutrinos and HNLs [34]. The effective parameter measured in the case of $\beta$-decay is

$$G_F^\beta = G_F \sqrt{(NN^\dagger)_{ee}} = G_F \alpha_{11},$$

(3.7)

whereas for muon decay it is

$$G_F^\mu = G_F \sqrt{(NN^\dagger)_{ee}(NN^\dagger)_{\mu\mu}} = G_F \sqrt{\alpha_{11}^2 (\alpha_{22}^2 + |\alpha_{21}|^2)}.$$

(3.8)

The measurement from muon decay has achieved a higher precision and, consequently, we take it as a reference, and consider $G_F^\mu = 1.1663787(6) \times 10^{-5}$ GeV\(^{-2}\) [38].

## 4 Results

We make use in what follows of the publicly available code FortEPiaNO (Fortran Evolved Primordial Neutrino Oscillations) [23, 28], which has been modified in order to work in the mass basis and to include the effects from a possible non-unitary three-neutrino mixing. The numerical settings adopted here allow to compute the value of $N_{\text{eff}}$ with a precision better than $10^{-3}$, which is safely smaller than the expected sensitivity from future surveys and, therefore, accurate enough for the study of non-unitary three-neutrino mixing. In particular, we compute the evolution from $x \simeq 0.010-0.015$ up to $x = 35$. Regarding the size of the momentum grid, we take 50 nodes between $y = 0.01$ and $y = 20$.

### 4.1 The role of diagonal parameters $\alpha_{ii}$

According to the parametrisation in eq. (2.5), non-unitarity could manifest as a deviation from unity in the parameters $\alpha_{ii}$ (see appendix A for the relation with the parametrisation of the $n \times n$ unitary mixing matrix). To understand their impact, we address the role of each one individually, as well as the interplay between them.

Figure 1 depicts the dependence of $N_{\text{eff}}$ on each of the individual diagonal parameters, $\alpha_{ii}$. In the left panel, we show the exact calculation of $N_{\text{eff}}$ as a function of $\alpha_{ii}$, allowing only one of them to be different from unity simultaneously. It is known that, in other non-standard scenarios, the main contribution of new physics to $N_{\text{eff}}$ comes from the modification of the collision integrals. For illustrative purposes, the top right panel shows the values of $N_{\text{eff}}$ after considering the dependence of $G_F$ on each of the $\alpha_{ii}$ parameters only in the collision terms (and not in the Hamiltonian) whereas the bottom right panel displays the opposite situation (note the different scale for $N_{\text{eff}}$).

Notice that, since the relation between the true value of $G_F$ and the one measured from muon decay does not depend on $\alpha_{33}$, the calculated $N_{\text{eff}}$ is almost insensitive to this

\(^3\)The role of neutrino self-interactions in cosmology has been studied extensively and was proposed as a solution to the so-called Hubble tension, see e.g. ref. [37] and references therein. Note, however, that self-interactions invoked in those scenarios are orders of magnitude larger than the one expected to arise from the non-unitarity of the three-neutrino mixing.
Figure 1. Values of $N_{\text{eff}}$ as a function of the diagonal non-unitarity parameters $\alpha_{ii}$, when only one of them differs from unity. For comparison, we show the exact calculation (left panel), where the vertical dotted lines indicate the existing limits (see Table 1) and the black horizontal line and the shaded region correspond to $(N_{\text{eff}})_0 \pm 2\sigma = 3.044 \pm 0.04$. We also depict the cases when the dependence of $G_F$ on the $\alpha_{ii}$ parameters is considered only in the collision integrals (upper right panel) or in the Hamiltonian (lower right panel).

parameter. Indeed, the main impact of non-unitarity on neutrino decoupling arises from the mismatch between the true and the observed value of $G_F$. The second key point is that, as it can be noticed from the right plots of figure 1, the collision terms dominate the dependence of $N_{\text{eff}}$ on the non-unitarity parameters: changes on the effective $G_F$ alter both the scattering and annihilation processes involving neutrinos, electrons and positrons, leading to a delayed decoupling and hence to a larger value of $N_{\text{eff}}$. Albeit matter effects also depend on the effective $G_F$, this is not as relevant for neutrino decoupling as the interactions with the cosmic plasma. A third aspect that should be noticed is that, although $G_F'$ has the same dependence on $\alpha_{11}$ and $\alpha_{22}$ (see (3.8)), the results in figure 1 are not identical. The difference arises from the fact that the two parameters, $\alpha_{11}$ and $\alpha_{22}$, do not modify the structure of the matrices $Y_L$ and $Y_R$ in the same way. This difference is seen not only when $G_F$ is fixed in the Hamiltonian but also when it is fixed in the collision term. Having a physical intuition is not straightforward due to the fact that neutrino decoupling is normally understood in terms of flavour eigenstates, instead of mass eigenstates. However, in the limit of small departures from unitarity, the scenario is formally analogous to having neutrino non-standard interactions. We discuss this limit in section 5.1. From the current measurement of $N_{\text{eff}} = 2.99^{+0.34}_{-0.33}$ at 95% C.L. [30], it is possible to set weak limits on the non-unitarity of the three-neutrino mixing matrix,

$$\alpha_{11} > 0.07, \quad \alpha_{22} > 0.15,$$

both at 95% C.L. As an exercise, we predict the improvement of these limits from future CMB observations. Let us assume a forecasted sensitivity to $N_{\text{eff}}$ with an uncertainty $\sigma = 0.02$ [32], together with a Gaussian posterior distribution for this parameter. Then, one can define the $\chi^2$ test:

$$\chi^2 = \frac{[N_{\text{eff}} - (N_{\text{eff}})_0]^2}{\sigma^2},$$

(4.2)
Figure 2. The three panels show the $1\sigma$, $2\sigma$ and $3\sigma$ allowed regions expected from the measurement of $N_{\text{eff}}$ when two of the diagonal parameters are allowed to vary simultaneously. The top panel shows the $\alpha_{11} - \alpha_{22}$ plane, whereas the bottom left and bottom right panels correspond to the $\alpha_{11} - \alpha_{33}$ and $\alpha_{22} - \alpha_{33}$ planes, respectively. The isocontour of $N_{\text{eff}} = (N_{\text{eff}})_0 = 3.044$ is shown in white, together with a shaded band indicating the numerical uncertainty, $(N_{\text{eff}})_0 \pm 0.001$.

where $(N_{\text{eff}})_0$ corresponds to the measured value of $N_{\text{eff}}$, assumed to match the standard theoretical prediction, i.e. $(N_{\text{eff}})_0 = 3.044$. Under these assumptions, we find that future CMB measurements will improve considerably the former 95\% C.L. upper limits:

$$\alpha_{11} > 0.29, \quad \alpha_{22} > 0.50.$$ (4.3)

In the following, we shall present an exhaustive study of the parameter correlations. Given the $\chi^2$ test definition in eq. (4.2), one can extract the confidence regions expected when two parameters are allowed to differ from unity simultaneously in the analyses. Figure 2 depicts the expected sensitivity to the non-unitarity diagonal parameters, projected in two-dimensional contours. The three coloured regions correspond to the $1\sigma$, $2\sigma$ and $3\sigma$ deviations from the standard value for $N_{\text{eff}}$, shown by the white line. The white-shaded band corresponds to an estimate of the numerical error in the standard prediction, given the numerical settings chosen in this work.

In the top panel, we show how the expected limit on $N_{\text{eff}}$ translates into allowed regions in the $\alpha_{11} - \alpha_{22}$ two-dimensional plane. Although these regions are larger than the ones

where $(N_{\text{eff}})_0$ corresponds to the measured value of $N_{\text{eff}}$, assumed to match the standard theoretical prediction, i.e. $(N_{\text{eff}})_0 = 3.044$. Under these assumptions, we find that future CMB measurements will improve considerably the former 95\% C.L. upper limits:

$$\alpha_{11} > 0.29, \quad \alpha_{22} > 0.50.$$ (4.3)

In the following, we shall present an exhaustive study of the parameter correlations. Given the $\chi^2$ test definition in eq. (4.2), one can extract the confidence regions expected when two parameters are allowed to differ from unity simultaneously in the analyses. Figure 2 depicts the expected sensitivity to the non-unitarity diagonal parameters, projected in two-dimensional contours. The three coloured regions correspond to the $1\sigma$, $2\sigma$ and $3\sigma$ deviations from the standard value for $N_{\text{eff}}$, shown by the white line. The white-shaded band corresponds to an estimate of the numerical error in the standard prediction, given the numerical settings chosen in this work.

In the top panel, we show how the expected limit on $N_{\text{eff}}$ translates into allowed regions in the $\alpha_{11} - \alpha_{22}$ two-dimensional plane. Although these regions are larger than the ones
Figure 3. Expected allowed regions at 1σ, 2σ and 3σ significance from the measurement of $N_{\text{eff}}$ in the $\alpha_{11} - \alpha_{22}$ plane, when $\alpha_{21}$ is allowed to vary while respecting the unitarity condition eq. (2.7) and after marginalizing over $\alpha_{21}$, see main text for details.

derived from the comparison of the measurements of $G_F$ from $\beta$ and muon decays or from neutrino oscillation experiments, they provide independent constraints on the non-unitarity diagonal parameters. The lower panels correspond to the sensitivity in the $\alpha_{11} - \alpha_{33}$ and $\alpha_{22} - \alpha_{33}$ two-dimensional planes. As it was already shown in figure 1, values of $\alpha_{33}$ which deviate significantly from unity do not result in a relevant change in $N_{\text{eff}}$: in particular, for values of $\alpha_{33} < 0.5$, $N_{\text{eff}}$ becomes almost independent of $\alpha_{33}$. The reason for the small dependence of $N_{\text{eff}}$ on $\alpha_{33}$ comes from the fact that the former only affects interaction rates involving the $\tau$ neutrino, which have a minor role in the neutrino decoupling process. On the contrary, $\alpha_{11}$ and $\alpha_{22}$ tend to result in a larger value of $N_{\text{eff}}$ because of their effect on $G_F$.

4.2 Including off-diagonal parameters

Allowing for non-zero off-diagonal parameters, $\alpha_{ij} > 0$ with $i < j$, changes the prediction for the effective number of relativistic species, as these parameters modify the structure of neutrino interactions with the plasma as well. In addition, for non-zero $\alpha_{21}$, the relation between the effective Fermi constant measure from muon decay, $G_F^\mu$, and the corresponding fundamental constant $G_F$ is altered. In what follows, we present the results of adding one extra non-zero off-diagonal parameter to the three diagonal ones, limiting the analysis to real values of $\alpha_{ij}$.

We show in figure 3 how non-zero values of $\alpha_{21}$ satisfying eqs. (2.7)–(2.9) enlarge the parameter space allowed in the $\alpha_{11} - \alpha_{22}$ two-dimensional plane after marginalising over the $\alpha_{21}$ parameter, see the upper left panel of figure 2 for comparison. Notice that non-zero values of the $\alpha_{21}$ parameter can partially compensate the larger value of $N_{\text{eff}}$ expected when $\alpha_{11}$ and $\alpha_{22}$ depart from one, thanks to its effect on $G_F$. Conversely, non-zero values of the off-diagonal parameters $\alpha_{31}$ and $\alpha_{32}$ were found to introduce only sub-leading changes in the planes $\alpha_{11} - \alpha_{33}$ and $\alpha_{22} - \alpha_{33}$, respectively. This implies that $\alpha_{31}$ and $\alpha_{32}$ do not reduce the value of $N_{\text{eff}}$ but rather increase it, leaving therefore the two-dimensional contours in the $\alpha_{11} - \alpha_{33}$ and $\alpha_{22} - \alpha_{33}$ planes almost unchanged.

\footnote{This choice is motivated by the results in [39], where it was shown that the CP phase does not alter the predicted $N_{\text{eff}}$ in the case of unitary mixing.}
5 Discussion

In this section, we proceed to contextualise the results previously presented. Firstly, we discuss the complementarities between the searches for neutrino non-standard interactions with electrons and the signatures expected when departures from unitarity are small. This allows to better understand the role of the non-unitarity parameters in this limit. Then, we review other existing limits on the non-unitarity of the three-neutrino mixing matrix from both neutrino experiments and precision electroweak observables.

5.1 From non-unitarity to non-standard interactions

In the case in which the deviations from unitarity of the three-neutrino mixing matrix are small, one can write the effective Lagrangian from charged- and neutral-current interactions in terms of the active neutrino states, as

\[
\begin{align*}
    \mathcal{L}_{NU} &= -2\sqrt{2}G_F \left[ \sum_{\alpha,\beta} \left( (NN^\dagger)_{\alpha e} (NN^\dagger)_{e\beta} + g_L \left[ NN^\dagger \right]_{\alpha\beta}^2 \left( \bar{\nu}_\alpha \gamma^\mu P_L \nu_\beta \right) \bar{e} \gamma^\mu P_L e \right) \\
    &\quad + g_R \sum_{\alpha,\beta} \left[ NN^\dagger \right]_{\alpha\beta}^2 \left( \bar{\nu}_\alpha \gamma^\mu P_L \nu_\beta \right) \bar{e} \gamma^\mu P_R e \right] \right],
\end{align*}
\]

(5.1)

This means that, if we take into account that the value of \( G_F \) measured from muon decay is given by eq. (3.8), then one can map the scenario of small deviations from unitarity to the one of neutrino-electron neutral current non-standard interactions (NSI). If we define the corresponding Lagrangian including Standard Model contributions and neutral current NSI as

\[
\begin{align*}
    \mathcal{L}_{SM+NSIe} &= -2\sqrt{2}G_F \left( \bar{\nu}_e \gamma^\mu P_L e \right) \left( \bar{e} \gamma^\mu P_L \nu_e \right) \\
    &\quad - 2\sqrt{2}G_F \sum_{X,\alpha} \left[ g_X \left( \bar{\nu}_\alpha \gamma^\mu P_L \nu_\alpha \right) \left( \bar{e} \gamma^\mu P_L e \right) + \sum_{\beta} \varepsilon^{eX}_{\alpha\beta} \left( \bar{\nu}_\alpha \gamma^\mu P_L \nu_\beta \right) \left( \bar{e} \gamma^\mu P_X e \right) \right],
\end{align*}
\]

(5.2)

then, the NSI-like parameters, after taking into account the results in eq. (3.8), read as

\[
\begin{align*}
    \varepsilon^{eL}_{\alpha\beta} &= -\left( \delta_{\beta e} \delta_{ae} + g_L \delta_{a\beta} \right) + \frac{(NN^\dagger)_{\alpha e} (NN^\dagger)_{e\beta} + g_L \left[ NN^\dagger \right]_{\alpha\beta}^2}{\sqrt{\alpha_{11}^2 + \alpha_{22}^2}},
\end{align*}
\]

(5.3)

and

\[
\begin{align*}
    \varepsilon^{eR}_{\alpha\beta} &= -g_R \delta_{a\beta} + g_R \frac{\left[ NN^\dagger \right]_{\alpha\beta}^2}{\sqrt{\alpha_{11}^2 + \alpha_{22}^2}}.
\end{align*}
\]

(5.4)

Note that the effective Lagrangians of the fundamental theory, eqs. (5.1)–(5.2), depend on the Fermi constant, \( G_F \). However, if we want to map our computation of the neutrino decoupling process when the three-neutrino mixing is not unitary to the scenario with NSI, one needs to account for the fact that what is experimentally accessible is \( G_F^2 \), see eq. (3.8).
\[
\alpha_{11} < 1 \\
\epsilon_{\alpha \beta}^{eL} = (1 + g_L)(-1 + \alpha_{11}^3)\delta_{\beta e}\delta_{\alpha e} + g_L\left(\frac{1}{\alpha_{11}} - 1\right)(\delta_{\alpha \mu}\delta_{\beta \mu} + \delta_{\alpha \tau}\delta_{\beta \tau}) \\
\epsilon_{\alpha \beta}^{eR} = -g_R(1 - \alpha_{11}^3)\delta_{\alpha e}\delta_{\beta e} + g_R\left(\frac{1}{\alpha_{11}} - 1\right)(\delta_{\alpha \mu}\delta_{\beta \mu} + \delta_{\alpha \tau}\delta_{\beta \tau}) \\
\epsilon_{\alpha \beta}^{eL} = -(1 + g_L)(\frac{1}{\alpha_{22}} - 1)\delta_{\alpha e}\delta_{\beta e} \\
-g_L(1 - \frac{1}{\alpha_{22}})\delta_{\alpha e}\delta_{\beta e} - g_L(1 - \alpha_{22}^3)\delta_{\alpha e}\delta_{\beta e} - g_R(1 - \frac{1}{\alpha_{22}})\delta_{\alpha e}\delta_{\beta e} \\
\epsilon_{\alpha \beta}^{eR} = -g_R\left(\frac{1}{\alpha_{22}} - 1\right)(\delta_{\alpha e}\delta_{\beta e} + \delta_{\alpha \tau}\delta_{\beta \tau}) - g_R(1 - \alpha_{22}^3)\delta_{\alpha e}\delta_{\beta e} \\
\alpha_{33} < 1 \\
\epsilon_{\alpha \beta}^{eL} = g_L(-1 + \alpha_{33}^4)\delta_{\alpha e}\delta_{\beta e} \\
\epsilon_{\alpha \beta}^{eR} = g_R(-1 + \alpha_{33}^4)\delta_{\alpha e}\delta_{\beta e}
\]

Table 2. Mapping between the diagonal non-unitarity (NU) parameters and effective NSI with electrons.

The involved matrix elements, \((NN^\dagger)_{\alpha \beta}\), can be computed at leading order [35] and then, a deeper physical understanding of the results becomes possible. The equivalent NSI parameters resulting from mapping departures from unity in the diagonal parameters are shown in table 2. If the interactions between neutrinos and the cosmic plasma become weaker, one expects an earlier decoupling, whereas stronger interactions would keep neutrinos in thermal contact for a longer time, and, consequently, neutrinos could receive more energy from the electron-positron pair annihilation.

From table 2, it is possible to notice that the behaviour of the predictions in figure 1 result from the modification of the interactions, which, depending on the values of \(\alpha_{11}\) and \(\alpha_{22}\), can shift to earlier or later times the neutrino decoupling process. Notice, in particular, that the effects induced by \(\alpha_{11} < 1\) and \(\alpha_{22} < 1\) are not the same. Conversely, the results obtained for \(\alpha_{33}\) are clearly understood from the fact that a value \(\alpha_{33} < 1\) reduces the strength of the neutral-current interactions, keeping tau neutrinos in thermal contact with the plasma, and leading to an earlier decoupling.

When two or more parameters are needed to parametrise the non-unitarity of the three-neutrino mixing, the picture becomes more complicated since the different effects can compensate. Nevertheless, one can get an idea of the results obtained when three parameters were allowed to vary simultaneously. For instance, \(\alpha_{31} > 0\) and \(\alpha_{32} > 0\) result in the appearance of flavour-changing neutral currents, which lead to a delayed decoupling [40]. The scenario of \(\alpha_{21}\) is different since it modifies both the structure of interactions and their strength.

Note that the treatment above is only valid for small departures from unitarity. Otherwise, the calculations must be performed in the mass basis in order to consider only the kinematically accessible states. A discussion on the similarities and differences between non-unitarity, light sterile neutrinos and non-standard interactions in the context of neutrino oscillations can be found in [41].

5.2 Comparison to other limits

At Earth, several precision and flavour observables can also be used to constrain deviations from unitarity in the three-neutrino mixing matrix. Whereas a detailed global analysis of all the existing constraints is beyond the scope of this work, for completeness, we discuss some of the most restrictive bounds from references [25, 35, 42–44].
At tree level, the weak mixing angle is related to the Fermi constant by
\[
\sin^2 \theta_W \cos^2 \theta_W = \frac{\alpha \pi}{\sqrt{2} G_F m_Z^2} = \frac{\alpha \pi}{\sqrt{2} G_F^\mu m_Z^2} \sqrt{(NN^\dagger)_{ee}(NN^\dagger)_{\mu\mu}},
\]
where \( \alpha \) is the fine structure constant. Requiring the agreement between measurements of the weak mixing angle and its indirect inference using \( G_F \) allows to set limits on departures from unitarity for the matrix \( N \).

The prediction of weak decay widths of the Z boson also depends on the assumptions about the unitarity of the lepton mixing matrix. At tree level, the decay width to two fermions depends linearly on \( G_F \), whereas for the invisible decay width,
\[
\Gamma_{\text{inv}} \propto G_F \sum_{i,j=1,2,3} |(N^\dagger N)_{ij}|^2.
\]

(5.6)

Also at tree level, the mass of the W boson can be expressed in terms of the Fermi constant, \( G_F \), as
\[
m_W^2 = \frac{\alpha \pi}{\sqrt{2} G_F \sin^2 \theta_W}.
\]

(5.7)

Again, bounds on departures from unitarity arise from requiring the agreement between direct measurements and the value of \( m_W \) derived using \( G_F^\mu \).

As mentioned above, the values of \( G_F \) derived from \( \beta \) and muon decays would receive different corrections. As a consequence, limits on the unitarity of the Cabibbo-Kobayashi-Maskawa matrix can be recast, since
\[
\sum_{i=1}^{3} |V_{ui}|^2 = \left( \frac{G_F^\beta}{G_F^\mu} \right)^2 = \frac{1}{\alpha_{22}^2 + |\alpha_{21}|^2}.
\]

(5.8)

In the presence of non-unitarity, couplings between leptons and gauge bosons are no longer flavour independent, implying that semileptonic decay rates can be flavour-dependent. Thus, measurements of pion and kaon decays also restrict deviations from unitarity. One can also consider universality tests in the leptonic sector and W-boson decays. For instance,
\[
R_{\alpha\beta} = \frac{\Gamma(W \to l_\alpha \bar{\nu}_\alpha)}{\Gamma(W \to l_\beta \bar{\nu}_\beta)} = \frac{(NN^\dagger)_{\alpha\alpha}}{(NN^\dagger)_{\beta\beta}}.
\]

(5.9)

These and other probes have been studied extensively in the literature and provide stringent limits to possible deviations from unitarity of the three-neutrino mixing matrix [35, 42–44]. In scenarios with only one heavy neutral lepton, stronger limits were found from matrix theory [45]. Table 3 reports the constraints on the \( \alpha_{ij} \) parameters obtained in [35] when considering constraints from neutrinos and charged leptons. Constraints from neutrinos alone [36] are complementary albeit not completely competitive.
6 Concluding remarks

In this work, we have considered the impact of non-unitarity in the three-neutrino mixing matrix on the process of neutrino decoupling. The presence of non-unitarity, a smoking-gun low-energy effect for the presence of heavy neutral leptons, alters the structure of both charged and neutral current interactions, thus affecting how neutrinos decouple from the thermal plasma in the early Universe.

In light of the forecasted accuracy expected from forthcoming cosmological surveys, we have derived the deviations on $N_{\text{eff}}$ induced by a non-unitarity mixing in the three-neutrino sector. In this regard, future cosmological measurements can serve as a novel high-redshift tool to provide model-independent and laboratory-complementary constraints on the non-unitarity parameters arising from new physics. Furthermore, they can be used as a cross-check to either falsify (or confirm) neutrino mass-generation scenarios.

We employ the commonly exploited parameterisation for the non-unitarity parameters in eq. (2.5). If the diagonal parameters $\alpha_{11}$ and $\alpha_{22}$ are different from unity, they will delay neutrino decoupling, resulting in a larger value of $N_{\text{eff}}$. The lower limits we obtained here when considering the future CMB-S4 mission combined with large-scale structure observations are $\alpha_{11} > 0.29$ and $\alpha_{22} > 0.50$, both at 95% C.L. Concerning the third non-unitarity diagonal parameter $\alpha_{33}$, it would barely alter the overall picture. We have also explored the possible interplay between the diagonal and off-diagonal non-unitarity parameters when either two or three of them are allowed to vary simultaneously. Interestingly, a non-zero value $\alpha_{21}$ relaxes considerably the constraints in the $\alpha_{11} - \alpha_{22}$ plane.

We conclude here emphasizing the fact that, even if the constraints on the non-unitarity parameters expected from cosmological observations are among the weakest when compared to other terrestrial probes, they provide an independent limit and a consistency test of the neutrino decoupling picture. Furthermore, they allow to set constraints on the putative non-unitary three-neutrino mixing at a completely different epoch in the evolution of the universe. Therefore, they should not be regarded as competitive limits but as supportive searches.

Acknowledgments

SG and OM acknowledge financial support from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754496 (project FELLINI) and the ITN grant H2020-MSCA-ITN-2019/860881-HIDDeN, respectively. In Valencia, this work was supported by the Spanish grants PID2020-113775GB-I00 and PID2020-113644GB-I00 (MCIN/AEI/10.13039/501100011033), as well as projects PROMETEO/2019/083 and CIPROM/2021/054 (Generalitat Valenciana). PMM is also supported by the grant FPU18/04571 (MICIU). He is grateful for the hospitality of the Particle and Astroparticle Physics Division of the Max-Planck-Institut für Kernphysik (Heidelberg), the Centro de Física Teórica de Partículas (Instituto Superior Técnico, Lisbon) and the AstroNu group at Niels Bohr Institute (Copenhagen) during the development of this project.
A Parametrisation of the non-unitary three-neutrino mixing matrix

The non-unitarity of the three-neutrino mixing matrix can be parametrised by a triangular matrix [33, 46, 47]. The parameters introduced can be related to the full \( n \times n \) unitary lepton mixing matrix, \( U_{n \times n} \). If we define the following complex rotation matrix,

\[
\omega_{ij} = \begin{pmatrix}
1 & 0 & \cdots & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 & \cdots & 0 \\
\vdots & c_{ij} & \cdots & 0 & \cdots & \eta_{ij} \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 1 & 0 & \cdots & 0 & \eta_{ij} \\
\eta_{ij} & \cdots & 0 & \cdots & 1 & 0 \\
0 & \cdots & 0 & \cdots & 0 & 1
\end{pmatrix},
\]

(A.1)

where we use the notation \( c_{ij} = \cos \theta_{ij}, \eta_{ij} = \sin \theta_{ij} e^{-i\phi_{ij}} \) and \( \tilde{\eta}_{ij} = -\sin \theta_{ij} e^{i\phi_{ij}} \), the full lepton mixing matrix can be parametrised as

\[
U_{n \times n} = \omega_{n-1 \times n} \omega_{n-2 \times n} \cdots \omega_1 \omega_{n-2 \times n-1} \omega_{n-3 \times n-1} \cdots \omega_1 \omega_{n-1 \times n} \omega_{n-2 \times n} \omega_{n-3 \times n-1} \cdots \omega_1 \omega_{n-1 \times n} \omega_{n-2 \times n} \omega_{n-3 \times n-1} \cdots \omega_1 \omega_{n-1 \times n} \omega_{n-2 \times n} \omega_{n-3 \times n-1} \cdots \omega_1 ,
\]

(A.2)

The diagonal parameters \( \alpha_{ii} \) in eq. (2.5) are real and expressed as [34, 36]

\[
\alpha_{ii} = c_{1i} c_{1i-1} \cdots c_{i4} ,
\]

(A.3)

and the off-diagonal ones are

\[
\alpha_{21} = c_{2n} c_{2n-1} \cdots c_{25} \eta_{24} \tilde{\eta}_{14} + c_{2n} \cdots c_{26} \eta_{25} \tilde{\eta}_{15} c_{14} + \cdots + \eta_{2n} \tilde{\eta}_{1n} c_{1n-1} \cdots c_{14} ,
\]

(A.4)

\[
\alpha_{31} = c_{3n} c_{3n-1} \cdots c_{35} \eta_{34} \tilde{\eta}_{24} c_{14} + c_{3n} \cdots c_{36} \eta_{35} c_{15} \tilde{\eta}_{14} + \cdots + \eta_{3n} \tilde{\eta}_{2n} c_{2n-1} \cdots c_{14} ,
\]

(A.5)

\[
\alpha_{32} = c_{3n} c_{3n-1} \cdots c_{35} \eta_{34} \tilde{\eta}_{24} + c_{3n} \cdots c_{36} \eta_{35} \tilde{\eta}_{25} c_{24} + \cdots + \eta_{3n} \tilde{\eta}_{2n} c_{2n-1} \cdots c_{24} .
\]

(A.6)

B Collision integrals in the mass basis

Following the notation and procedure for integral reduction in [28], the neutrino-electron collision integral is separated in the terms accounting for scattering and annihilations,

\[
\mathcal{I}_{ve}[\varrho(x, y)] = \frac{G_F^2}{(2\pi)^3 y^2} \left\{ I_{ve}^{\mathrm{scatt}}[\varrho(x, y)] + I_{ve}^{\mathrm{ann}}[\varrho(x, y)] \right\} ,
\]

(B.1)

using the comoving variables defined in section 3. The term accounting for scattering reads

\[
I_{ve}^{\mathrm{scatt}}[\varrho(x, y)] = \int dy_2 dy_3 \sum_{a=L, R} \sum_{b=L, R} A_{ab}(x, y_2, y_3) F_{ve}^{ab}\left(\varrho^{(1)}, f_e^{(2)}, \varrho^{(3)}, f_e^{(4)}\right) ,
\]

(B.2)

whereas the one related to annihilations is

\[
I_{ve}^{\mathrm{ann}}[\varrho(x, y)] = \int dy_2 dy_3 \sum_{a=L, R} \sum_{b=L, R} B_{ab}(x, y_2, y_3) F_{ab}^{ann}\left(\varrho^{(1)}, \varrho^{(2)}, f_e^{(3)}, f_e^{(4)}\right) .
\]

(B.3)
The exact dependence on the comoving variables of the scattering kernels, $A_{ab}$ and $B_{ab}$, can be found in [28]. Their expressions do not depend on the basis in which the evolution of the density matrix is computed (mass basis here, flavour basis in [28]) nor on the strength and structure of the interactions.

The fact that we are working in the mass basis only manifests in the phase-space factors $F_{ab}^{\text{ann}}$ and $F_{ab}^{\text{scatt}}$, which depend on the matrices $Y_L$ and $Y_R$ defined in eqs. (3.2)–(3.3). Although not stated explicitly, the density matrix $\varrho$ is also in the mass basis. Then, the expressions of the factors accounting for the phase-space of scattering and annihilation processes read

\begin{align}
F_{ab}^{\text{scatt}} \left( \varrho^{(1)}, f_{e}^{(2)}, \varrho^{(3)}, f_{e}^{(4)} \right) &= f_{e}^{(4)}(1-f_{e}^{(2)}) \left[ Y_a \varrho^{(3)} Y_b (1-\varrho^{(1)}) + (1-\varrho^{(1)}) Y_b \varrho^{(3)} Y_a \right] \\
&\quad - f_{e}^{(2)}(1-f_{e}^{(4)}) \left[ \varrho^{(1)} Y_b (1-\varrho^{(3)}) Y_a + Y_a (1-\varrho^{(3)}) Y_b \varrho^{(1)} \right], \quad (B.4)
\end{align}

\begin{align}
F_{ab}^{\text{ann}} \left( \varrho^{(1)}, \varrho^{(2)}, f_{e}^{(3)}, f_{e}^{(4)} \right) &= f_{e}^{(3)} f_{e}^{(4)} \left[ Y_a (1-\varrho^{(2)}) Y_b (1-\varrho^{(1)}) + (1-\varrho^{(1)}) Y_b (1-\varrho^{(2)}) Y_a \right] \\
&\quad - (1-f_{e}^{(3)})(1-f_{e}^{(4)}) \left[ Y_a \varrho^{(2)} Y_b \varrho^{(1)} + \varrho^{(1)} Y_b \varrho^{(2)} Y_a \right], \quad (B.5)
\end{align}

where $a, b = L, R$. These expressions depend both on the electron momentum distribution function, $f_e$, and on the density matrix $\varrho^{(i)} = \varrho(y_i)$, where $y_i$ is the comoving momentum of particle $i$.

References

[1] P.F. de Salas et al., 2020 global reassessment of the neutrino oscillation picture, JHEP 02 (2021) 071 [arXiv:2006.11237] [INSPIRE].

[2] I. Esteban et al., The fate of hints: updated global analysis of three-flavor neutrino oscillations, JHEP 09 (2020) 178 [arXiv:2007.14792] [INSPIRE].

[3] F. Capozzi et al., Unfinished fabric of the three neutrino paradigm, Phys. Rev. D 104 (2021) 083031 [arXiv:2107.00532] [INSPIRE].

[4] P. Minkowski, $\mu \to e\gamma$ at a Rate of One Out of $10^9$ Muon Decays?, Phys. Lett. B 67 (1977) 421 [INSPIRE].

[5] T. Yanagida, Horizontal gauge symmetry and masses of neutrinos, Conf. Proc. C 7902131 (1979) 95 [INSPIRE].

[6] J. Schechter and J.W.F. Valle, Neutrino Masses in SU(2) $\times$ U(1) Theories, Phys. Rev. D 22 (1980) 2227 [INSPIRE].

[7] J. Schechter and J.W.F. Valle, Neutrino Decay and Spontaneous Violation of Lepton Number, Phys. Rev. D 25 (1982) 774 [INSPIRE].

[8] R.N. Mohapatra and G. Senjanovic, Neutrino Masses and Mixings in Gauge Models with Spontaneous Parity Violation, Phys. Rev. D 23 (1981) 165 [INSPIRE].

[9] W. Grimus and L. Lavoura, The Seesaw mechanism at arbitrary order: Disentangling the small scale from the large scale, JHEP 11 (2000) 042 [hep-ph/0008179] [INSPIRE].

[10] R.N. Mohapatra and J.W.F. Valle, Neutrino Mass and Baryon Number Nonconservation in Superstring Models, Phys. Rev. D 34 (1986) 1642 [INSPIRE].

[11] E.K. Akhmedov, M. Lindner, E. Schnapka and J.W.F. Valle, Left-right symmetry breaking in NJL approach, Phys. Lett. B 368 (1996) 270 [hep-ph/9507275] [INSPIRE].

[12] E.K. Akhmedov, M. Lindner, E. Schnapka and J.W.F. Valle, Dynamical left-right symmetry breaking, Phys. Rev. D 53 (1996) 2752 [hep-ph/9509255] [INSPIRE].
[13] M. Malinsky, J.C. Romao and J.W.F. Valle, Novel supersymmetric SO(10) seesaw mechanism, *Phys. Rev. Lett.* 95 (2005) 161801 [hep-ph/0506296] [SPIRE].

[14] M. Malinsky, T. Ohlsson, Z.-Z. Xing and H. Zhang, Non-unitary neutrino mixing and CP violation in the minimal inverse seesaw model, *Phys. Lett. B* 679 (2009) 242 [arXiv:0905.2889] [SPIRE].

[15] M. Malinsky, T. Ohlsson and H. Zhang, Non-unitarity effects in a realistic low-scale seesaw model, *Phys. Rev. D* 79 (2009) 073009 [arXiv:0903.1961] [SPIRE].

[16] S. Gariazzo et al., Light sterile neutrinos, *J. Phys. G* 43 (2016) 033001 [arXiv:1507.08204] [SPIRE].

[17] C. Giunti and T. Lasserre, eV-scale Sterile Neutrinos, *Ann. Rev. Nucl. Part. Sci.* 69 (2019) 163 [arXiv:1901.08330] [SPIRE].

[18] A. Palazzo, Exploring Light Sterile Neutrinos at Long Baseline Experiments: A Review, *Universe* 6 (2020) 41 [SPIRE].

[19] S. Schoppmann, Status of Anomalies and Sterile Neutrino Searches at Nuclear Reactors, *Universe* 7 (2021) 360 [arXiv:2109.13541] [SPIRE].

[20] M. Archidiacono and S. Gariazzo, Two Sides of the Same Coin: Sterile Neutrinos and Dark Radiation, Status and Perspectives, *Universe* 8 (2022) 175 [arXiv:2201.10319] [SPIRE].

[21] S. Hannestad, I. Tamborra and T. Tram, Thermalisation of light sterile neutrinos in the early universe, *JCAP* 07 (2012) 025 [arXiv:1204.5861] [SPIRE].

[22] N. Saviano et al., Multi-momentum and multi-flavour active-sterile neutrino oscillations in the early universe: role of neutrino asymmetries and effects on nucleosynthesis, *Phys. Rev. D* 87 (2013) 073006 [arXiv:1302.1200] [SPIRE].

[23] S. Gariazzo, P.F. de Salas and S. Pastor, Thermalisation of sterile neutrinos in the early Universe in the 3+1 scheme with full mixing matrix, *JCAP* 07 (2019) 014 [arXiv:1905.11290] [SPIRE].

[24] S. Hagstotz et al., Bounds on light sterile neutrino mass and mixing from cosmology and laboratory searches, *Phys. Rev. D* 104 (2021) 123524 [arXiv:2003.02289] [SPIRE].

[25] S. Antusch and O. Fischer, Non-unitarity of the leptonic mixing matrix: Present bounds and future sensitivities, *JHEP* 10 (2014) 094 [arXiv:1407.6607] [SPIRE].

[26] K. Akita and M. Yamaguchi, A precision calculation of relic neutrino decoupling, *JCAP* 08 (2020) 012 [arXiv:2005.07047] [SPIRE].

[27] J. Froustey, C. Pitrou and M.C. Volpe, Neutrino decoupling including flavour oscillations and primordial nucleosynthesis, *JCAP* 12 (2020) 015 [arXiv:2008.01074] [SPIRE].

[28] J.J. Bennett et al., Towards a precision calculation of $\Omega_{\nu}$ in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED, *JCAP* 04 (2021) 073 [arXiv:2012.02726] [SPIRE].

[29] K. Akita and M. Yamaguchi, A Review of Neutrino Decoupling from the Early Universe to the Current Universe, *Universe* 8 (2022) 552 [arXiv:2210.10307] [SPIRE].

[30] PLANCK collaboration, Planck 2018 results. VI. Cosmological parameters, *Astron. Astrophys.* 641 (2020) A6 [Erratum ibid. 652 (2021) C4] [arXiv:1807.06209] [SPIRE].

[31] SIMONS OBSERVATORY collaboration, The Simons Observatory: Science goals and forecasts, *JCAP* 02 (2019) 056 [arXiv:1808.07445] [SPIRE].

[32] CMB-S4 collaboration, CMB-S4 Science Book, First Edition, arXiv:1610.02743 [FERMILAB-FN-1024-A-E] [SPIRE].

[33] Z.-Z. Xing, Correlation between the Charged Current Interactions of Light and Heavy Majorana Neutrinos, *Phys. Lett. B* 660 (2008) 515 [arXiv:0709.2220] [SPIRE].
[34] F.J. Escrihuela et al., On the description of nonunitary neutrino mixing, *Phys. Rev. D* 92 (2015) 053009 [Erratum ibid. 93 (2016) 119905] [arXiv:1503.08879] [INSPIRE].

[35] F.J. Escrihuela et al., Probing CP violation with non-unitary mixing in long-baseline neutrino oscillation experiments: DUNE as a case study, *New J. Phys.* 19 (2017) 093005 [arXiv:1612.07377] [INSPIRE].

[36] D.V. Forero, C. Giunti, C.A. Ternes and M. Tórtola, Nonunitary neutrino mixing in short and long-baseline experiments, *Phys. Rev. D* 104 (2021) 075030 [arXiv:2103.01998] [INSPIRE].

[37] E. Di Valentino et al., In the realm of the Hubble tension — a review of solutions, *Class. Quant. Grav.* 38 (2021) 153001 [arXiv:1612.07377] [INSPIRE].

[38] E. Tiesinga, P.J. Mohr, D.B. Newell and B.N. Taylor, CODATA recommended values of the fundamental physical constants: 2018, *Rev. Mod. Phys.* 93 (2021) 025010 [INSPIRE].

[39] J. Froustey and C. Pitrou, Primordial neutrino asymmetry evolution with full mean-field effects and collisions, *JCAP* 03 (2022) 065 [arXiv:2110.11889] [INSPIRE].

[40] P.F. de Salas et al., Cosmological radiation density with non-standard neutrino-electron interactions, *Phys. Lett. B* 820 (2021) 136508 [arXiv:2105.08168] [INSPIRE].

[41] M. Blennow et al., Non-Unitarity, sterile neutrinos, and Non-Standard neutrino Interactions, *JHEP* 04 (2017) 153 [arXiv:1609.08637] [INSPIRE].

[42] S. Antusch and O. Fischer, Probing the nonunitarity of the leptonic mixing matrix at the CEPC, *Int. J. Mod. Phys. A* 31 (2016) 1644006 [arXiv:1604.00208] [INSPIRE].

[43] F.J. Escrihuela, L.J. Flores and O.G. Miranda, Neutrino counting experiments and non-unitarity from LEP and future experiments, *Phys. Lett. B* 802 (2020) 135241 [arXiv:1907.12675] [INSPIRE].

[44] E. Fernandez-Martinez, J. Hernandez-Garcia and J. Lopez-Pavon, Global constraints on heavy neutrino mixing, *JHEP* 08 (2016) 033 [arXiv:1605.08774] [INSPIRE].

[45] W. Flieger, J. Gluza and K. Porwit, New limits on neutrino non-unitary mixings based on prescribed singular values, *JHEP* 03 (2020) 169 [arXiv:1910.01233] [INSPIRE].

[46] Z.-Z. Xing, A full parametrization of the 6 × 6 flavor mixing matrix in the presence of three light or heavy sterile neutrinos, *Phys. Rev. D* 85 (2012) 013008 [arXiv:1110.0083] [INSPIRE].

[47] H.-C. Han and Z.-Z. Xing, A full parametrization of the 9 × 9 active-sterile flavor mixing matrix in the inverse or linear seesaw scenario of massive neutrinos, *Nucl. Phys. B* 973 (2021) 115609 [arXiv:2110.12705] [INSPIRE].