A new and trivial CP symmetry for extended $A_4$ flavor

C. C. Nishi*
Centro de Matemática, Computação e Cognição,
Universidade Federal do ABC - UFABC, 09.210-170, Santo André, SP, Brasil

The combination of $\nu_\mu-\nu_\tau$ exchange together with CP conjugation in the neutrino sector (known as $\text{CP}^\mu\tau$ symmetry or $\mu$-reflection) is known to predict the viable pattern: $\theta_{23} = 45^\circ$, maximal Dirac CP phase and trivial Majorana phases. We implement such a CP symmetry as a new CP symmetry in theories with $A_4$ flavor. The implementation in a complete renormalizable model leads to a new form for the neutrino mass matrix that leads to further predictions: normal hierarchical spectrum with lightest mass and $m_{23}$ ($0\nu2\beta$) of only few meV, and either $\nu_1$ or $\nu_2$ has opposite CP. An approximate $L_\mu - L_\tau$ symmetry arises naturally and controls the flavor structure of the model. The light neutrino masses are generated by the extended seesaw mechanism with 6 right-handed neutrinos (RHNs). The requirement of negligible one-loop corrections to light neutrino masses, validity of the extended seesaw approximation and not too long-lived BSM states to comply with BBN essentially restricts the parameters of the model to a small region: three relatively light right-handed neutrinos at the GeV-scale, heavier neutrinos at the electroweak scale and Yukawa couplings smaller than the electron Yukawa. Such a small Yukawa couplings render these RHNs unobservable in terrestrial experiments.

I. INTRODUCTION

The discovery of nonzero $\theta_{13} \sim 8.5^\circ$ in 2012[1] prompted the neutrino physics community to one of its next experimental goals: measure or discard CP violation in the leptonic sector [2]. As one more parameter in the standard three neutrino paradigm joined the list of known quantities, we are only left with three unknowns in case neutrinos are Majorana: neutrino mass ordering, absolute neutrino mass scale and CP violation in the leptonic sector. The last unknown has three sources: one Dirac CP phase analogous to the CKM phase for quarks and two Majorana phases.

From a theory viewpoint, many symmetries were sought over the years in order to predict the CP violating phases of the leptonic sector. The simplest of them that leads to CP violation and viable mixing angles is known as $\mu$-reflection or $\text{CP}^\mu\tau$ which consists on $\nu_\mu-\nu_\tau$ flavor exchange together with CP conjugation [3]. Often, such a CP symmetry is considered in conjunction with nonabelian discrete symmetries [4–7]. In fact, many studies were devoted to the definition of CP symmetry in that context [5–7]. However, differently from many simple flavor symmetries that predicted vanishing $\theta_{13}$, the $\text{CP}^\mu\tau$ symmetry allows nonzero $\theta_{13}$ but predicts all the presently unknown CP phases: the Dirac CP phase $\delta_{\text{CP}} = \pm 90^\circ$ is maximal while the Majorana phases are trivial [3, 8]. Moreover, $\theta_{23}$ is also predicted to be maximal, the neutrinoless double beta decay effective mass is restricted to narrower bands and, in simple implementations, lepton genesis is only allowed to occur in the intermediate range of $T \sim M_1 \sim 10^9 – 10^{12}$ GeV where flavor effects are important [8]. From current global fits[9, 10], we know in fact there is a slight preference for negative $\delta_{\text{CP}}$ and $\theta_{23} = 45^\circ$ is still allowed.

Two directions were recently pursued to generalize the idea of $\text{CP}^\mu\tau$ symmetry. Firstly, we have shown in Ref. [8] that a minimal setting that allowed distinct symmetries in the charged lepton and neutrino sectors consisted of only one abelian symmetry (the combination of lepton flavors $L_\mu - L_\tau$ or subgroup) and CP symmetry ($\text{CP}^\mu\tau$). This setting was shown to be free from the vev alignment problem that plagues many flavor symmetry models for leptons. In contrast, in Ref. [11], it was shown that maximal $\theta_{23}$ and $\delta_{\text{CP}}$ (the prediction for Majorana phases is lost) could follow from much more general assumptions without the imposition of CP symmetry. The necessary conditions involve the symmetry of the charged lepton sector ($G_l$) to be represented by real matrices in the flavor space and, in the same basis, $M_l$ needs to be diagonalizable by a real matrix. The crucial aspect is the former, which presumably follows from a real flavor symmetry conserved in the charged lepton sector. The neutrino sector cannot be invariant by the same residual symmetry and hence must have a large breaking in the form of misaligned vevs.

Here we try to embed a subgroup of $L_\mu - L_\tau$ into a discrete nonabelian flavor group $G_F$ in order to increase predictivity but, at the same time, retain the successful features of Ref. [8]. We choose the $A_4$ group which is an extensively studied flavor group (see [12] and references therein). In fact, the first $\text{CP}^\mu\tau$ symmetric neutrino mass matrix was obtained with this group [13]. More recent studies involving $A_4$ and CP can be seen in Refs. [14, 15].

*Electronic address: celso.nishi@ufabc.edu.br
We anticipate that the light neutrino mass matrix in our model will have the form

\[
M_\nu = \begin{pmatrix}
a_1 + a_2 + a_3 & k(a_1 + \omega a_2 + \omega^2 a_3) & k(a_1 + \omega^2 a_2 + \omega a_3) \\
(k(a_1 + \omega a_2 + \omega^2 a_3) & k^2(a_1 + \omega^2 a_2 + \omega a_3) & k^2(a_1 + a_2 + a_3) \\
k^2(a_1 + a_2 + a_3) & k^2(a_1 + a_2 + a_3) & k^2(a_1 + \omega a_2 + \omega^2 a_3)
\end{pmatrix},
\]

(1)

where \(a_i, k\) are real parameters and \(k > 0\) can be chosen; \(\omega \equiv e^{i2\pi/3}\) as usual. This mass matrix is \(\mathbb{CP}^{\mu \tau}\) symmetric [3] but has 4 real parameters to describe 5 observables: \(\theta_{12}, \theta_{13}, m_1, m_2, m_3\). Hence, we will have one prediction.

The paper is organized as follows: In Sec. II we describe the new CP symmetry that can be implemented for theories with \(A_4\) symmetry. Section III shows that the mass matrix (1) can fit the present oscillation parameters and additionally give predictions for the absolute neutrino mass and CP parities. A complete renormalizable model is shown in Sec. IV where the light neutrino masses are generated by the extended seesaw (ESS) mechanism [16] with relatively light right-handed neutrinos in its spectrum. The approximate symmetry \(L_\mu - L_\tau\) is presented in Sec. V and shown to constrain the flavor structure of the model. Section VI analyzes the constraints on the model coming from (i) the radiative stability of the tree-level result, (ii) validity of the ESS approximation to fit the light neutrino masses and (iii) sufficiently short-lived BSM states that not spoil Big Bang nucleosynthesis. More phenomenological constraints on the presence of relatively light right-handed neutrinos are analyzed in Sec. VII. The conclusions are shown in Sec. VIII and the appendices contain auxiliary material.

II. ANOTHER GCP FOR \(A_4\)

The group \(A_4 = (\mathbb{Z}_2 \times \mathbb{Z}_2) \rtimes \mathbb{Z}_3\) has one three-dimensional irreducible representation (irrep) 3 and three one-dimensional irreps \(1', 1'', 1\), where the latter is the trivial invariant (singlet). The faithful 3 can be generated by

\[
a = \text{diag}(1, -1, -1), \quad b = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix},
\]

(2)

where \(a\) generates one of the \(\mathbb{Z}_2\) subgroups and \(b\) generates the \(\mathbb{Z}_3\) subgroup. Only \(b\) acts nontrivially on the singlets \(1', 1''\) as

\[
1' \xrightarrow{b} \omega 1', \quad 1'' \xrightarrow{b} \omega^2 1'',
\]

(3)

where \(\omega = e^{i2\pi/3}\).

For generic settings where generic irreps of \(A_4\) (e.g. a 3 and one charged 1') are considered in a model, there is only one possible CP symmetry that can be imposed on the model [6, 7]. As first considered in Ref. [4], CP acts on the representations of (2) and (3) as

\[
\text{CP}_1: \quad 3 \rightarrow X3^*, \quad 1 \rightarrow 1^*, \quad 1' \rightarrow 1'^*, \quad 1'' \rightarrow 1''^*,
\]

(4)

where \(X\) can be chosen as (23) exchange:

\[
X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.
\]

(5)

The complex conjugation denotes the CP transformation operation on the fields which should be adjoined with the appropriate Lorentz factors for e.g. spin 1/2 fermions. We denote the whole flavor group considering \(\text{CP}_1\) as \(A_4 \rtimes \mathbb{Z}_2^{\text{CP}}\) and it gives rise to a group isomorphic to \(S_4\), denoted as \(\tilde{S}_4\) in [4]. Obviously any composition of \(\text{CP}_1\) with an element of \(A_4\) is also a GCP symmetry, so any of the 12 GCP symmetries can be chosen as a residual symmetry [15].

In nongeneric settings where only a specific set of irreps is considered, it is clear that there is one more inequivalent option. If only 3 is considered, we can use the usual CP transformation

\[
\text{CP}_2 : 3 \rightarrow 3^*.
\]

(6)

1 For the triplet 3 only, in the context of the \(A_4\) invariant 3HDM, it was first considered in Ref. [17] (erratum) as an accidental symmetry and in Ref. [18] in the course of symmetry classification.

2 It is important to note that the GCP (4), with symmetric \(X\), can also be cast in the form (6) by basis change, after which the representation (2) changes and is no longer manifestly real.
Given that the representation (2) is real, the whole group including \( CP_2 \) will be denoted as \( A_4 \times \mathbb{Z}_2^{CP} \) where \( \mathbb{Z}_2^{CP} \) is generated by \( CP_2 \), which commutes with \( A_4 \) (3 is real).

Now the question is: What is the transformation law for the other irreps (if any is consistent)? We can deduce them by noting that the transformation (6) acts on the representation (2) trivially, i.e.,

\[
CP_2 : \quad a \rightarrow a, \quad b \rightarrow b,
\]

if we apply on any \( 3 \), in this order, \( CP_2 \), the transformation \( a \) or \( b \) and then \( CP_2^{-1} \). In contrast, for \( CP_1 \), the same set of operations induces

\[
CP_1 : \quad a \rightarrow Xa^*X^{-1} = a, \quad b \rightarrow Xb^*X^{-1} = b^2.
\]

Here we are identifying \( a, b \) with its three-dimensional irrep \( D_3(a), D_3(b) \) in (2). Given that (8) and (7) lead to different rules (map different conjugacy classes), they cannot be equivalent. These mapping rules in the group are called automorphisms and only (8) and (7) are nonequivalent for \( A_4 \). So these are the only possibilities for defining GCP in the presence of \( A_4 \) symmetry [6].

We can now deduce that one transformation law for the singlets \( 1' \) that is compatible with (6) and (7) is the trivial transformation

\[
CP_2 : \quad 1' \rightarrow 1'.
\]

However, this transformation law can only be used if the complex field \( \psi_1 \sim 1' \) is neutral under any other group, including the Lorentz group, i.e., it must be a scalar. In this case, we can split any complex scalar into its real and imaginary parts, \( \varphi = (\varphi_r + i\varphi_i)/\sqrt{2} \), and consider the action of \( b \) of \( A_4 \) as a 120° rotation in the plane of \( (\varphi_r, \varphi_i)^T \), hence a real representation that is trivial under \( CP_2 \), i.e., \( \varphi_r, \varphi_i \), are CP-even real scalar fields.

On the other hand, if \( \psi_1 \) carries other complex quantum numbers (it excludes \( \mathbb{Z}_2 \)) other than \( A_4 \), say a charge \( q \) of \( U(1) \), then (9) is not compatible with the fact that CP should reverse the charge \( q \). Therefore, in this case another field \( \psi_2 \sim 1'' \) with the same charge \( q \) (or any other quantum number) needs to be introduced to define the transformation

\[
CP_2 : \quad \psi_1 \rightarrow \psi_2^*,
\]

so that both sides transform as \( \omega \) by \( b \) but the field of charge \( q \) is mapped to a field of charge \( -q \). This is also the transformation law for fermions. To summarize, the irreps \( 1' \) and \( 1'' \) are exchanged by \( CP_2 \),

\[
CP_2 : \quad 1' \rightarrow 1''^*.
\]

unless \( 1''^* \) can be identified with \( 1' \). Therefore, for charged fields (such as the SM fields or any chiral fermion) the irreps \( 1', 1'' \) need to be introduced in pairs. It is always possible to recast (11) as the usual CP transformation by changing basis; see appendix B of Ref. [8] for the explicit basis change.

Compatibility with the triplet transformation law (6) can also be checked independently by forming an invariant with two triplets \( \psi = (\psi_1, \chi) = (\chi_1) \sim 3 \) (say fermionic and left chiral) and a scalar \( \varphi \sim 1' \), and ensuring that \( CP_2 \) maps an \( A_4 \) invariant to an \( A_4 \) invariant [6]. The only trilinear \( A_4 \) invariant involving \( \psi, \chi \) and \( \varphi \) is

\[
I = (\psi_1 \chi_1 + \omega \psi_2 \chi_2 + \omega^2 \psi_3 \chi_3) \varphi.
\]

It is transformed by (6) (for \( \psi, \chi \)) and (9) (for \( \varphi \)) to

\[
CP_2 : \quad I \rightarrow (\chi_1 \psi_1 + \omega \chi_2 \psi_2 + \omega^2 \chi_3 \psi_3) \varphi,
\]

which remains as an \( A_4 \) invariant.

The symmetry \( CP_2 \) (associated to the trivial automorphism) can be straightforwardly extended for other groups with structure \( H \times \mathbb{Z}_2 \) such as the \( \Delta(3 \cdot N^2) = (\mathbb{Z}_N \times \mathbb{Z}_N) \times \mathbb{Z}_3 \) family (e.g., \( \Delta(27) \) [19]) or some of its subgroups such as \( T_7 \) or \( T_{13} \). The only difference is that the triplet representations would be complex and CP symmetry would act as usual.

We stress that the \( CP_2 \) symmetry for \( A_4 \) has not been considered for flavor model building before. This possibility is raised in the general context of discrete nonabelian symmetries in [6] but no model application was discussed. For \( A_4 \), this possibility was mentioned in [15] but it was not pursued. Ref. [7] discards this kind of CP symmetry dubbing it as CP-like symmetries but—as we will see for the simple case of \( A_4 \)—no theoretical consideration prevents its use. As an added bonus, we will see that the transformation property (9) allows us to avoid the vev alignment problem [8].

---

3 One could also use (9) as charge conjugation for a pair of Majorana fermion fields where \( b \) acts by 120° rotation in the plane.
III. MASS MATRIX

We first analyze our mass matrix (1) in the flavor basis to show that we can correctly fit the oscillation parameters. This is a new form for the neutrino mass matrix that has not been considered so far.

The CP$^T$ symmetry of (1) implies that $\theta_{23} = \pi/4$ and $\delta_{\text{CP}} = \pm \pi/2$ are automatic [3] and the diagonalization

$$U^T M_\nu U = \text{diag}(m'_i),$$

(14)

can be performed by a matrix $U = U_0$ of the form

$$U_0 = \begin{pmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ w_1^* & w_2^* & w_3^* \end{pmatrix},$$

(15)

with $u_i$ conventionally real and positive. The Majorana phases are trivial and possible CP parities appear along with the eigenvalues $m'_i = \pm m_i$, $m_i \geq 0$. We denote the different cases of CP parities by the sign of $(m'_i)$ as

$$ (+ + +), (- + +), (+ - +), (+ + -).$$

(16)

In addition to being CP$^T$ symmetric, the mass matrix in (1) obeys

$$M_\nu|_{a_2 \leftrightarrow a_3} = M_\nu^*,$$

$$M_\nu|_{a_1 \rightarrow a_2 \rightarrow a_3} = \text{diag}(1, \omega^2, \omega) M_\nu \text{diag}(1, \omega^2, \omega).$$

(17)

Thus cyclic permutation of $a_i$ leaves all observables of $M_\nu$ invariant while a transposition $(a_2 \leftrightarrow a_3)$ flips the Dirac CP phase: $\delta_{\text{CP}} \rightarrow -\delta_{\text{CP}}$. Hence, permutations of solutions for $a_i$ are solutions as well.

A. Obtaining the masses

To extract the light neutrino masses, it is more convenient to change to a real basis:

$$M'_\nu = U^T_{23} M_\nu U_{23} = \begin{pmatrix} a_1 + a_2 + a_3 & \frac{k}{\sqrt{2}}(2a_1 - a_2 - a_3) & \sqrt{\frac{3}{2}}k(a_3 - a_2) \\ \frac{k}{\sqrt{2}}(2a_1 - a_2 - a_3) & \frac{1}{2}k^2(4a_1 + a_2 + a_3) & \frac{1}{2}\sqrt{3}k^2(a_2 - a_3) \\ \sqrt{\frac{3}{2}}k(a_3 - a_2) & \frac{1}{2}\sqrt{3}k^2(a_2 - a_3) & \frac{3}{2}k^2(a_2 + a_3) \end{pmatrix},$$

(18)

where

$$U_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} \end{pmatrix}. $$

(19)

Now $M'_\nu$ is real symmetric and can be diagonalized by a real orthogonal matrix.

The eigenvalues of $M'_\nu$ will correspond to the light neutrino masses $m'_i = \pm m_i$ with its CP parities. They are solutions of the characteristic equation

$$\lambda^3 + c_1 \lambda^2 + c_2 \lambda + c_3 = 0,$$

(20)

with coefficients

$$-c_1 = (1 + 2k^2)(a_1 + a_2 + a_3) = m'_1 + m'_2 + m'_3$$

$$-c_3 = 27k^4a_1a_2a_3 = m'_1m'_2m'_3$$

$$c_2 = 3k^2(2 + k^2)(a_1a_2 + a_2a_3 + a_3a_1) = m'_1m'_2 + m'_2m'_3 + m'_3m'_1.$$  

(21)

It is clear that $k = 1$ is a special point where

$$3a_i = m'_i, \quad i = 1, 2, 3,$$

(22)
is a solution; permutation of $a_i$ still leads to a solution. However, our mass matrix (1) with $k = 1$ and with the second and third columns (rows) exchanged is invariant by cyclic permutations which means it is diagonalized by $U_{\text{PMNS}} = U_{\omega}$. This mixing matrix is clearly in contradiction with experiments, a fact that still applies if $k \approx 1$ (for hierarchical $m_i$). Hence, we need to analyze the cases away from $k = 1$.

Generically we can invert (21) and obtain $a_i$ as functions of $m_i$ and $k$. A simplification is achieved for generic $k > 0$ by defining

$$a_i \equiv (1 + 2k^2)a_i.$$  \hfill (23)

Then the equations in (21) can be rewritten as

$$\begin{align*}
\tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3 &= m'_1 + m'_2 + m'_3, \\
\tilde{a}_1\tilde{a}_2\tilde{a}_3 &= g_3(k)m'_1m'_2m'_3, \\
\tilde{a}_1\tilde{a}_2 + \tilde{a}_2\tilde{a}_3 + \tilde{a}_3\tilde{a}_1 &= g_2(k)(m'_1m'_2 + m'_2m'_3 + m'_3m'_1).
\end{align*} \hfill (24)$$

where

$$g_3(k) = \frac{(1 + 2k^2)^3}{27k^4}, \quad g_2(k) = \frac{(1 + 2k^2)^2}{3k^4(2 + k^2)}. \hfill (25)$$

The key relation that can be extracted from (24) is that $\tilde{a}_i$ should now be roots of the cubic equation similar to (20) but with coefficients modified by

$$c_1 \rightarrow \tilde{c}_1 = c_1, \quad c_2 \rightarrow \tilde{c}_2 = g_2(k)c_2, \quad c_3 \rightarrow \tilde{c}_3 = g_3(k)c_3. \hfill (26)$$

This construction gives $\tilde{a}_i$ as functions of $m'_i$ and $k$, except for permutations of $\tilde{a}_i$. The solutions (22) for $k = 1$ are modified as $g_2(k), g_3(k)$ differ from unity when $k \neq 1$. Moving away from $k = 1$, both functions increase monotonically ($g_2$ reaches $4/3$ asymptotically as $k \rightarrow \infty$).

Now, the distortions caused by $g_{2,3}$ cannot be too large because the $\tilde{a}_i$ need to be real. To illustrate this point, compare the two polynomials

$$p_1(x) = x^3 - 2.1x^2 + 1.1x, \quad p_2(x) = x^3 - 2.1x^2 + 1.2x, \hfill (27)$$

where the second polynomial differs from the first just by a small deviation in the third coefficient. The first polynomial has three real and distinct roots while the second polynomial has only $x = 0$ as a real root. This can be confirmed by calculating the discriminant of the factored second-degree polynomials: $\Delta = (2.1)^2 - 4 \times 1.1 = 0.01$ and $\Delta = (2.1)^2 - 4 \times 1.2 = -0.39$ for $p_1$ and $p_2$ respectively. We can see that two quasidegenerate eigenvalues are specially sensitive to deviations by $k$. This is the case of IH with CP parities $(+++) \text{ or } (++-)$.

The values for $k$ that allow real solutions for $\tilde{a}_i$ can be extracted from the discriminant of the cubic polynomial (20) for which

$$\Delta = \tilde{c}_1^2\tilde{c}_2^2 - 4\tilde{c}_2^3 - 4\tilde{c}_1\tilde{c}_2\tilde{c}_3 + 18\tilde{c}_1\tilde{c}_2\tilde{c}_3 - 27\tilde{c}_3^2 \geq 0. \hfill (28)$$

In Fig. 1 we show the values of $k$ as a function of the lightest mass $m_0$ where the discriminant above is non-negative; we use the current best fit values for the mass differences [9]. The figure on the left (right) corresponds to NH (IH) and the various possibilities for CP parities are depicted in different colors. For IH, only the case of CP parities $(-++) \text{ and } (+--)$ have wide regions for $k$ for a given mass $m_0$; the remaining cases only have very narrow ranges of possible $k$, including $k \approx 1$ which is phenomenologically excluded. The other possible narrow range for $k$ for IH-$(+++) \text{ (e.g. } k \approx 7 \text{ for } m_0 = 10^{-3}\text{eV})$ is also phenomenologically excluded because it leads to $a_1 \approx a_2 \approx a_3$ and two mixing angles are vanishing.

We also illustrate in Fig. 2 the deviations from $\tilde{a}_i = m_i$ when $k$ moves away from $k = 1$. $k$ varies only in the range where the discriminant (28) is non-negative, as shown in Fig. 1. Note that close to the critical values of $k$ ($\Delta = 0$) two (or more) $\tilde{a}_i$ tend to be quasidegenerate. This is a generic phenomenon.

**B. Seeking solutions**

After an exhaustive numerical search we conclude that the mass matrix (1) is only compatible with oscillation data for normal hierarchy (NH) and CP parities $(+-+)$ and $(+--)$. The cases of IH and quasidegenerate masses are excluded. The lightest neutrino mass is restricted to

$$\begin{align*}
(-++) : \quad & 1.6 \text{ meV} \lesssim m_1 \lesssim 3 \text{ meV}, \\
(+--): \quad & 3.5 \text{ meV} \lesssim m_1 \lesssim 7.7 \text{ meV}. \hfill (29)
\end{align*}$$
FIG. 1: Left (right): Regions in the $k$-$m_1$ ($k$-$m_3$) plane where solutions for $a_i$ are real for NH (IH). The mass-squared differences are fixed to their best-fit values of [9]. A hole is only present for the case NH-$(+++)$, the regions for IH-$(−++)$ and IH-$(++++)$ are almost overlapping, and the regions for IH-$(+++)$ and IH-$(++−)$ can be seen only as lines.

FIG. 2: Solutions of $\tilde{a}_i$ (in solid blue, red and green, respectively) as functions of $k$ and $m_i’$ (in dashed blue, red and green, respectively) for fixed values ($−2.5, 9.0, 44.9$) meV for $m_i’$ with ordering defined by (32). The gray dashed line corresponds to the average $\frac{1}{3}(m_1′ + m_2′ + m_3′) = \frac{1}{3}(\tilde{a}_1 + \tilde{a}_2 + \tilde{a}_3)$.

The predictions for the contribution for neutrinoless double-beta decay coming from light neutrinos is given by

\begin{align*}
(-++): & \quad 1.9 \text{ meV} \lesssim |m^\nu_{\beta\beta}| \lesssim 2.6 \text{ meV}, \\
(+−+): & \quad 1.1 \text{ meV} \lesssim |m^\nu_{\beta\beta}| \lesssim 2.05 \text{ meV},
\end{align*}

They fall inside the regions denoted by NH-$(−++)$ and NH-$(+−+)$ in Ref. [8]. Note that $m^\nu_{\beta\beta} = (M_\nu)^{ee}_\nu$. For future use, we also list

\begin{align*}
(-++): & \quad 26 \text{ meV} \lesssim (M_\nu)^{\mu\tau}_\nu \lesssim 28 \text{ meV}, \\
(+−+): & \quad 20.5 \text{ meV} \lesssim (M_\nu)^{\mu\tau}_\nu \lesssim 23 \text{ meV}.
\end{align*}

The parameter distribution for the two sets of solutions is shown in Fig. 3 for $|a_i|$ as functions of $m_1$ (left), and $k$ as a function of $m_1$ (right). The values $\theta_{23} = \pi/4$ and $\delta_{\text{CP}} = \pm \pi/2$ are fixed from symmetry and we only consider values for $\theta_{12}, \theta_{13}, \Delta m^2_{21}$ and $\Delta m^2_{23}$ within 3-σ of the global fit in Ref. [9] by varying $a_1$ and $k$ independently. Approximate values are obtained from the procedure below.
FIG. 3: \textbf{Left}: Scatter plot of $|a_i|$ as a function of the lightest mass $m_1$ (NH) for the two possible CP parities for the light $\nu_{iL}$: ($-++$) (darker colors) and ($+--$) (lighter colors). Other solutions are related by permutations of $a_i$, cf. (17). The depicted ordering of $a_i$ leads to $\delta_{CP} = -\pi/2$. \textbf{Right}: $k$ as a function of $m_1$; black dots and gray dots denote the cases ($-++$) and ($+--$) respectively.

We use the following procedure to exclude solutions and search for approximate solutions:

1. For each lightest mass $m_0$, we find $\tilde{a}_i$ through Eq. (24) (or Eq. (20) with (26)) for a given $k$, restricted to the range specified by Fig. 1. We keep $\Delta m^2_{12}$ and $\Delta m^2_{23}$ fixed to their best-fit values of $[9]$. An illustration of this procedure is given in Fig. 2.

2. Then, we diagonalize (1) to extract the mixing matrix $U = U_{PMNS}$. We adopt the ordering of eigenvectors to satisfy

$$|U_{e1}| > |U_{e2}| > |U_{e3}|.$$  \hspace{1cm} (32)

The ordering of $m_i$ follows. This means that our mass eigenstates $\nu_1, \nu_2, \nu_3$ are in the order of decreasing contribution to $\nu_e$ ($\nu_1$ contributes the most and so on) and not in a specific mass ordering. This definition explains the color flipping in Fig. 2 for $k \leq 1$.

3. At last, we check if the mass ordering is correct and if the mixing angles fall inside the 3-\(\sigma\) ranges. An illustration of this step is shown in Fig. 4.

One remark on this procedure is in order: to correctly fit the oscillation parameters we need that (i) the mixing angles are correct and (ii) the mass ordering is correct. The condition (ii) arises because mass eigenstates $\nu_i$ are defined by (32) and mass orderings that do not correspond to NH or IH are excluded. For example, we can read from Fig. 4 that the correct values for both $s^2_{12}$ and $s^2_{13}$ are only achieved for $k \approx 3.5$, as can also be confirmed in Fig. 3. Correct values for $s^2_{13}$ can also be obtained for $k \approx 0.4$ but $s^2_{12}$ as well as the mass ordering in Fig. 2 is not correct: for $k < 1$, $\nu_e$ has a greater contribution from the heaviest state ($\nu_2$ in red) than the second heaviest state ($\nu_3$ in green).

IV. EXTENDED SEESAW MODEL

Here we present a low-scale seesaw model where the light neutrino mass matrix has the form (1). The model will retain the successful predictions of $U(1)_{\mu-\tau} \times \mathbb{Z}_2^{CP}$ [8] for the low-energy neutrino observables but additional predictions arise due to the more constrained nature of the group $A_4$. Two sets of heavy neutrinos – one at the GeV-scale and another at the electroweak scale – arise naturally due to the extended seesaw mechanism (ESS) [16]. The combination of lepton flavor numbers $L_{\mu} - L_{\tau}$ will be approximately conserved in the model.

The flavor symmetry of the model will be $A_4 \times \mathbb{Z}_2^{CP}$ ($\mathbb{CP}_2$), explained in Sec. II \footnote{Note that the combined group is a direct product because both factors commute [8].}. The SM lepton fields are, however,
all singlets of $A_4$ and only feel the $Z_3$ subgroup\(^5\), thus entirely avoiding the need of any vev alignment in this sector. An auxiliary $Z_3^D$ will also be necessary in the neutrino sector. The SM lepton fields are assigned to $L_i \sim l_i \sim (1,1',1'')$ while the Higgs doublet $H$ is invariant; $L_i = (\nu_{iL}, l_{iL})^\dagger$ are lepton doublets while $l_i \equiv l_{iR}$ are the charged lepton singlets. Thus $\mathbb{C}P_2$ in (11) can be identified with $\mathbb{C}P^{\mu\tau}$ \[^8\]. There are also two sets of SM singlets (right-handed neutrinos) $N_i \equiv N_{iR}$ and $S_i \equiv S_{iR}$, $i = 1,2,3$, assigned to $(1,1',1'')$ and 3 respectively. Hence, only the neutrino sector feels the full $A_4$ group through $S_{iR}$. We also need complex flavons $\eta \sim 3$ and $\varphi_1 \sim 1'$, and a real $\varphi_0 \sim 1$. The full assignment can be seen in table I. Additional fields necessary to break $\mathbb{C}P^{\mu\tau}$ in the charged lepton sector are not shown since they can just be adapted from \[^8\].

The charged lepton sector at the electroweak scale will effectively be the SM one\(^6\)

\begin{equation}
- \mathcal{L}_{\text{eff}}^l = y_1 \bar{L}_1 H l_1 + y_2 \bar{L}_2 H l_2 + y_3 \bar{L}_3 H l_3 + h.c.,
\end{equation}

where the $Z_3$ subgroup is unbroken but $\mathbb{C}P^{\mu\tau}$ is broken at a higher scale by a CP-odd scalar \[^8\] so that the correct splitting for $y_\mu = y_2$ and $y_\tau = y_3$ is generated ($y_e = y_1$).

The neutrino sector at the high scale is given by

\begin{equation}
- \mathcal{L}_\nu = f_1 \tilde{N}_1 \tilde{H}^\dagger L_1 + f_2 \tilde{N}_2 \tilde{H}^\dagger L_2 + f_3 \tilde{N}_3 \tilde{H}^\dagger L_3 \\
+ f_1' (\tilde{S}\eta)_1 N_1^c + f_2' (\tilde{S}\eta)_1' N_2^c + f_3' (\tilde{S}\eta)_1'' N_3^c \\
+ \frac{1}{2} M_{11} \tilde{N}_1 N_1^c + M_{23} \tilde{N}_2 N_3^c \\
+ \frac{1}{2} k_0 \varphi_0 (\tilde{S} S^c)_1 + \frac{1}{2} k_1 \varphi_1 (\tilde{S} S^c)_1' + \frac{1}{2} k_1' \varphi_1' (\tilde{S} S^c)_1'' + h.c.,
\end{equation}

\(^5\) This contrasts with most of the $A_4$ models for leptons where at least the lepton doublets form triplets\[^{12}\].

\(^6\) For simplicity we are considering the UV completion by heavy leptons but the multi-Higgs version can be equally considered with the difference that the Higgs that couples to the $\mu - \tau$ flavors is distinct \[^8\].
where we have defined singlet combinations of two triplets of $A_4$ as

$$(xy)_{1}^1 \equiv x_1 y_1 + x_2 y_2 + x_3 y_3,$$
$$\quad (xy)_{1}^\prime \equiv x_1 y_1 + x_2 y_2 \omega + x_3 y_3 \omega^2,$$
$$\quad (xy)_{1}^{\prime \prime} \equiv x_1 y_1 + x_2 y_2 \omega^2 + x_3 y_3 \omega.$$

Note that $\text{CP}^{\mu \tau}$ acts as

$$\text{CP}^{\mu \tau} : \begin{array}{c} L_1 \rightarrow L_1^{\mu \tau}, \quad L_2 \rightarrow L_2^{\mu \tau}, \quad L_3 \rightarrow L_3^{\mu \tau}, \\ H \rightarrow H^*, \quad S_i \rightarrow S_i^{\mu \tau}, \quad \eta_i \rightarrow \eta_i^*, \\ \phi_0 \rightarrow \phi_0, \quad \phi_1 \rightarrow \phi_1, \end{array}$$

and $l_i$ and $N_i$ transform like $L_i$ and $\psi_{cp}$ denotes the usual CP conjugate of the chiral fermion $\psi$. Therefore, $f_1, f_1', M_{11}, M_{23}, k_0$ are real and $f_3 = f_2^*, f_3' = f_2'$ due to $\text{CP}^{\mu \tau}$. The parameters $f_2, f_3, f_2', f_3'$ can be further chosen real and positive by rephasing $L_2, 3$ and $N_2, 3$.

The mass matrix for $(\nu^c, N^c_i, S_i^c)$ after EWSB will be

$$M = \begin{pmatrix} 0 & M_D^T \\ M_D & M_R \end{pmatrix} = \begin{pmatrix} 0 & m_T^D \\ m_D & M_N & \Lambda^T \end{pmatrix},$$

where

$$m_D = \text{diag}(m_{D_i}) = \frac{v}{\sqrt{2}} \text{diag}(f_1, f_2, f_3),$$
$$\Lambda = \text{diag}(u_1, u_2, u_3) \sqrt{3} U^\omega \text{diag}(f_1', f_2', f_3'),$$
$$M_N = \begin{pmatrix} M_{11} & \ast \\ M_{23} \end{pmatrix},$$
$$\mu = \text{diag}(\mu_1, \mu_2, \mu_3),$$

where

$$U^\omega \equiv \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{pmatrix}.$$

In this model, we are considering that $\varphi_{0,1}$ acquire very small vevs which lead to the real Majorana masses $\mu_i$ for $S_i$ and also

$$\langle \eta_i \rangle = u_i, \quad \text{all real.}$$

We justify the hierarchy of vevs in appendix B.

Considering that $M_N$ is composed of bare masses, the ESS limit is naturally achieved [16]: $M_N \gg \{\Lambda, m_D\} \gg \mu$ and also $\mu \ll \Lambda^2/M_N$. We can see that there are two sources of lepton number violation (LNV) in (34)\(^7\): (a) large scales $M_N$ and (b) low-scales $\mu_i \sim \langle \varphi_{0,1} \rangle$.

At tree level and leading order we obtain

$$\nu_i : \quad M_\nu = m_D^{T} \Lambda^{-1} \mu (\Lambda^T)^{-1} m_D,$$
$$S_i^c : \quad M_S = -\Lambda M_N^{-1} \Lambda^T,$$
$$N_i^c : \quad M_N,$$

\(^7\) If $L(N_i) = L(L_i) = -L(S_i) = 1.$
with light-heavy mixing

\[ \theta_{\nu S}^* = \frac{m_{D}^{T} \Lambda^{-1}}{M_{S}}, \]

\[ \theta_{\nu N}^* = \frac{m_{D}^{T} M_{N}^{-1}}{M_{S}}. \]

Additional mixings can be seen in appendix A. We can see that the small LNV scale \( \mu \) only enters \( M_{l} \) while the large LNV scale \( M_{N} \) contributes only to heavier masses. Given that the mass matrix for the heavier states \( N_{i} \) are approximately unchanged, we can define

\[ M_{N_{1}} \equiv M_{11}, \quad M_{N_{2}} = M_{N_{3}} \equiv M_{23}, \]

assuming positive quantities. The leading correction can be seen in appendix A.

Explicitly, the light neutrino mass matrix is

\[ M_{\nu} = \frac{1}{3} \text{diag}(m_{D,1}/f_{1}) U_{\omega} \text{diag}(\mu_{i}/\omega_{i}^{2}) U_{\omega} \text{diag}(m_{D,1}/f_{1}), \]

which has the desired form (1) with

\[ a_{i} = \frac{1}{3} \mu_{i} \frac{m_{D,1}^{2}}{\omega_{i}^{2} f_{1}^{2}}, \quad k = \frac{|m_{D,22}/f_{1}|}{|m_{D,11}/f_{2}|}. \]

We have used the shorthand \( m_{D,11} \equiv f_{1} v/\sqrt{2} \); cf. (38). The fitting of the light neutrino parameters in Fig. 3 implies

\[ |a_{i}| \approx 0.2 - 2.1 \text{meV}, \]

\[ k \approx 3.2 - 4.4. \]

Also, the sign change of one of the \( a_{i} \) needs to be generated by \( \mu_{i} \) and not by \( u_{i}^{2} \) which is always positive.

Although the heavier states \( N_{i} \) are frequently chosen to lie above the TeV scale [16, 20], in our case (i) a negligible one-loop contribution for light neutrino masses, (ii) validity of the ESS approximation and (iii) BBN constraints will essentially restrict \( M_{N} \) to the electroweak scale; see Sec. VI and (72) for a benchmark point.

V. APPROXIMATE U(1)_{\mu-\tau} LIMIT

We consider first the limit where \( Z_{3} \) of \( A_{4} \) is only broken by the small quantities in \( \mu \). This means that below the scale of \( \langle \eta \rangle \) \( Z_{3} \) is only broken by light neutrino masses. This approximate \( Z_{3} \) symmetry corresponds to the lepton flavor triality (LFT) [21] where lepton fields carry the discrete charges

\[ \text{LFT} : \quad L_{i} \sim l_{i} \sim N_{i} \sim S_{i}^{l} \sim (1, \omega, \omega^{2}); \]

\[ S_{i}^{l} \text{ is related to } S_{i} \text{ by change of basis } S_{i}^{l} = (U_{\omega})_{ij} S_{j}^{e}. \]

The heavy vevs of \( \eta \) conserve LFT when

\[ \langle \eta_{i} \rangle = u_{i} \approx u_{0}(1,1,1). \]

This feature is justified in appendix B. In this case, after \( \eta_{i} \rightarrow \langle \eta_{i} \rangle \) and in the limit \( k_{1} \rightarrow 0 \), the Lagrangian (34) is in fact invariant by the continuous version of (47) with charges [8]

\[ \text{U}(1)_{\mu-\tau} : \quad L_{i} \sim l_{i} \sim N_{i} \sim S_{i}^{e} \sim (0,1,-1). \]

It corresponds to the combination \( L_{\mu} - L_{\tau} \) of family lepton numbers. The approximate conservation of \( \text{U}(1)_{\mu-\tau} \) will lead to a number of consequences.

In this limit the mass matrix (41) and mixing (42) of the heavy neutrinos \( S_{i} \) yield

\[ M_{S} = -\frac{1}{3} \begin{pmatrix} M_{S_{1}} + 2M_{S_{2}} & M_{S_{1}} - M_{S_{2}} & M_{S_{1}} - M_{S_{2}} \\ * & M_{S_{1}} + 2M_{S_{2}} & M_{S_{1}} - M_{S_{2}} \\ * & * & M_{S_{1}} + 2M_{S_{2}} \end{pmatrix}, \]

\[ \theta_{\nu S}^{*} = \frac{m_{D,11}}{\sqrt{3} f_{1} u_{0}} \begin{pmatrix} 1 \\ k \\ k \end{pmatrix} U_{\omega}, \]
where the masses read

\[ M_{S_1} \equiv \frac{(\sqrt{3}u_0 f'_1)^2}{M_{N_1}}, \quad M_{S_{2,3}} \equiv \frac{(\sqrt{3}u_0 f'_2)^2}{M_{N_2}}. \] (51)

These relations allows us to trade \( f'_1 u_0 \) and \( f'_2 u_0 = f'_3 u_0 \) for physical masses:

\[ |\sqrt{3} f'_1 u_0| = \sqrt{M_{S_1} M_{N_1}}, \quad |\sqrt{3} f'_2 u_0| = \sqrt{M_{S_2} M_{N_2}}. \] (52)

The mass matrix \( M_S \) is invariant by cyclic permutations and then \((1, 1, 1)\) is an eigenvector. We can diagonalize it by

\[ V_3^* U_23 = \omega (1, 0, 0). \] (53)

giving

\[ V_3^T M_S V_S = \text{diag}(M_{S_1}, M_{S_2}, M_{S_3}). \] (54)

The matrix \( U_{23} \) was defined in (19). Therefore, \( S_1 \) is a Majorana fermion of \( U(1)_{\mu - \tau} \) charge 0 and \( S_{2,3} \) are degenerate Majorana fermions that form a (pseudo-)Dirac pair of fields with charge \( \pm 1 \). The latter implies that LNV effects induced by \( S_{2,3} \) exchange will vanish in this limit.

The active-sterile \( \nu - S \) mixing reduces to

\[ (\theta_{\nu S} V_S)^* = (-i) \frac{m_{D1}}{\sqrt{3}u_0 f'_1} \text{diag}(1, k, k) \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{i}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} & 0 \end{pmatrix}. \] (55)

It is important to note that in this approximation

\[ (\theta_{\nu S} V_S)_{ei} = 0, \quad \text{for } i = 2, 3, \] (56)

and the electron flavor is only coupled to \( S_1 \).

VI. ONE-LOOP CONTRIBUTIONS AND BBN CONSTRAINTS

Now we should compute the one-loop contributions to light neutrino masses. When the lightest heavy RHN mass lies below 100 MeV, the one-loop contributions to 0νβ can be sizable [22], although such a sterile neutrinos are severely constrained by cosmological data [23]. Heavy neutrinos with electroweak-scale masses can still induce sizable contributions [20, 24] and the dominant (and finite) one comes from light neutrino self-energies with Higgs or \( Z \) exchange [25–27].

We can write the self-energy contribution as

\[ M_{\nu v}^{1-1} = \frac{1}{(4\pi v)^2} M_D^T \left( M_R^{-1} F(M_R M_R^T) + F(M_R M_R^T) M_R^{-1} \right) M_D, \] (57)

where the loop function \( F(x) \) is given by

\[ F(x) = \frac{x}{2} \left[ 3 \frac{\ln(x/M_Z^2)}{x/M_Z^2 - 1} + \frac{\ln(x/M_h^2)}{x/M_h^2 - 1} \right], \] (58)

with \( M_Z \) and \( M_h \) being the Z and Higgs boson masses, respectively; \( v = 246 \text{ GeV} \) is the electroweak scale. This contribution should be added to the tree-level contribution (44) coming from the ESS mechanism. We should note that heavy neutrino masses \( M_R \) at the electroweak scale leads to a contribution (57) functionally similar to the tree-level contribution \( M_D^2/M_R \), but smaller only by the loop factor \( 1/16\pi^2 [25] \) [notice \( F((100 \text{ GeV})^2)/v^2 \approx 1.5 \)].

---

8 We keep using the same name \( S_i \) for the heavy neutrino fields although they have a small component of \( \nu_{iL}^c \) and \( N_{iR} \).
Therefore, the one-loop contribution in the ESS mechanism can possibly be large since the cancellation that occurs in the tree-level mass matrix is not expected to carry over to the one-loop contribution.

We can adapt the one-loop contribution for generic type-I seesaw (57) to the extended seesaw with mass matrix (37) as

\[
M_{\nu}^{1-1} = \frac{1}{(4\pi)^2} m_D^T \left\{ M_N^{-1} \Lambda^T V_S \hat{M}_S^{-1} 2F(\hat{M}_S^2)V_S^T \Lambda M_N^{-1} + V_N \hat{M}_N^{-1} 2F(\hat{M}_N^2)V_N^T \right\} m_D . \tag{59}
\]

We have first block diagonalized \(M_R\) (see appendix A) and then used the basis where \(M_S\) and \(M_N\) is diagonal (\(\hat{M}_S\) and \(\hat{M}_N\)). It is also possible to write the expression in terms of the light-heavy mixing angles as

\[
M_{\nu}^{1-1} = \frac{1}{(4\pi)^2} \left\{ (\theta_{\nu S} V_S)^* \hat{M}_S \dot{2} F(\hat{M}_S^2)(\theta_{\nu S} V_S)^\dagger + (\theta_{\nu N} V_N)^* \hat{M}_N \dot{2} F(\hat{M}_N^2)(\theta_{\nu N} V_N)^\dagger \right\} . \tag{60}
\]

We can see that generically the contribution from the heavier states \(N_i\) dominates over the contribution from \(S_i\) because the smaller mixing angle \(\theta_{\nu S}^2/\theta_{\nu N}^2 \sim \Lambda^2/M_N^2\) is compensated by \(M_N/M_S \sim M_N^2/\Lambda^2\) and \(F(x)\) grows with \(x\).

For our purposes, it is useful to define the adimensional function \(g(x)\) as

\[
g(M_{i}/100 \text{ GeV}) \equiv \frac{2F(M_i^2)}{M_i \times 100 \text{ GeV}} . \tag{61}
\]

A slightly different definition can be seen in [28]. This function peaks at the electroweak scale \(M_i \approx 93.3\) GeV with maximum 3.64 and decreases away from the peak with rate slower than \(M_i^{-1}\) for \(M_i \gtrsim 100\) GeV; see behaviour in Fig. 5. This function allows us to rewrite (59) as

\[
M_{\nu}^{1-1} = \frac{100 \text{ GeV}}{(4\pi)^2} m_D^T \left\{ M_N^{-1} \Lambda^T V_S g(\hat{X}_S) V_S^T \Lambda M_N^{-1} + V_N g(\hat{X}_N) V_N^T \right\} m_D . \tag{62}
\]

We have used the shorthand \(\hat{X}_S \equiv \text{diag}(M_{S_i})/100\) GeV and similarly for \(\hat{X}_N\).

**Fig. 5:** Plot of the function in (61) where \(g_{\text{max}} = 3.63547\) and the maximum occurs at \(M \approx 93.3\) GeV.

Computing (62) in our model in the U(1)\(_{\mu-\tau}\) symmetric limit, we obtain the texture

\[
M_{\nu}^{1-1} = \begin{pmatrix}
* & 0 & 0 \\
0 & 0 & * \\
0 & * & 0
\end{pmatrix}, \tag{63}
\]

whose nonzero entries correspond to \(L_\mu - L_\tau = 0\). Explicitly,

\[
(M_{\nu}^{1-1})_{ee} \approx 10 \text{ keV} \times \frac{m_{D_{11}}^2}{\text{GeV}^2} \left[ - \frac{g(x_1)}{M_{N_1}/M_{S_1}} + g \left( \frac{M_{N_1}}{M_{S_1}} x_1 \right) \right] ,
\]

\[
(M_{\nu}^{1-1})_{\mu\tau} \approx 10 \text{ keV} \times \frac{m_{D_{22}}^2}{\text{GeV}^2} \left[ - \frac{g(x_2)}{M_{N_2}/M_{S_2}} + g \left( \frac{M_{N_2}}{M_{S_2}} x_2 \right) \right] , \tag{64}
\]
where $x_i \equiv M_{S_i}/100 \text{ GeV}$. We have used Eqs. (38), (53) and $V_N = U_{23}$. We note that indeed the one-loop contribution can lead to an unacceptably large contribution. For example, for $m_D \sim 1 \text{ GeV}, M_N \sim 10 \text{ TeV}, M_S \sim 100 \text{ GeV}$, the one-loop contribution leads to a few keV. From Fig. 5 we also see that to lower the contributions from (64) to acceptable values by increasing $M_N$ requires very large values of the order of $10^7 \text{ GeV}$. Therefore, to have TeV-scale (or lower) right-handed neutrinos, we need to lower the scale of $m_D$ or arrange some cancellation between either the various one-loop contributions or between the tree and one-loop ones [27]. We consider this possibility unappealing and do not pursue it any further.

In order to preserve our predictions of Sec. III we confine ourselves to the case where the loop-induced contributions (64) are negligible compared to the tree level ones in (1). To visualize the possible regions in parameter space, we show in Fig. 6 the regions (blue) in the $M_{N_1}-m_{D_{11}}$ plane (left) and $M_{N_2}-m_{D_{22}}$ plane (right) where the one-loop contribution is at most 10% of the tree-level contribution for the $ee$ (left) and $\mu\tau$ (right) entries. For definiteness we fix the tree-level values to

$$(M^\text{tree})_{ee} = 2 \text{ meV}, \quad R_{N_1} \equiv M_{N_1}/M_{S_1} = 10^2,$$

$$(M^\text{tree})_{\mu\tau} = 24 \text{ meV}, \quad R_{N_2} \equiv M_{N_2}/M_{S_2} = 10^2. \quad (65)$$

These values are in agreement with (30) and (31). We choose to plot the dependence on $M_{N_i}$, because the one-loop contributions depend dominantly on $M_{N_i}$ (rather than on the lighter $M_{S_i}$) in the ESS approximation. For example, if we increase the ratios $R_{N_i}$, the blue regions shrinks down only slightly for large $M_{N_i}$. For completeness, we also show the curves for unit ratio (dashed).

![FIG. 6: The blue regions satisfy $|(M^\text{true})_{ee}|/(M^\text{tree})_{ee} \leq 0.1$ (left) or $|(M^\text{true})_{\mu\tau}|/(M^\text{tree})_{\mu\tau} \leq 0.1$ (right) with the reference values (65). The blue dashed curves obey unit ratios. The points inside the orange regions are the ones necessary to fit the $ee$ (left) or $\mu\tau$ (right) tree-level entries of the light neutrino mass matrix through (66) or (67) restricted to (68) and (65). The orange dashed curves correspond to the subset of points for $R_{N_1} = 0.03$ (left) and $R_{N_2} = 0.076$ (right). The green regions cover the points where $\tau_{S_1} \leq 0.1$ (left) or $\tau_{S_2} \leq 0.1$ (right) for $R_{N_1} = R_{N_2} = 100$. The green dashed curves yields the lifetime of 0.1 s but with $R_{N_1} = 270$ (left) or $R_{N_2} = 400$ (right). The crosses mark the benchmark points in (72). See text for details.]

The next step is to ensure that the tree-level contribution themselves – as they depend on the model parameters as in (45) – lie in the necessary ranges of (30) and (31) (also Fig. 3). For that purpose, we rewrite the sum of all relations for $a_i$ in (45) as

$$\left( \frac{m_{D_{11}}}{10 \text{ keV}} \right)^2 \left( \frac{100 \text{ GeV}}{M_{N_i}} \right) \left( \frac{\bar{\mu}_i}{M_{S_1}} \right) = \frac{|(M^\text{true})_{ee}|}{\text{meV}}, \quad (66)$$

where $\bar{\mu}_i \equiv \sum_i \mu_i/3$. We have also used (52) to eliminate $f_1' u_0$. An analogous relation is valid for the $\mu\tau$ entry:

$$\left( \frac{M_{D_{22}}}{10 \text{ keV}} \right)^2 \left( \frac{100 \text{ GeV}}{M_{N_2}} \right) \left( \frac{\bar{\mu}_i}{M_{S_2}} \right) = \frac{|(M^\text{true})_{\mu\tau}|}{\text{meV}}. \quad (67)$$

As $\bar{\mu}_i \ll M_{S_1}$ in order to satisfy the ESS approximation, we require

$$R_{\mu 1} \equiv \frac{\bar{\mu}_i}{M_{S_1}} \leq 0.1 \quad \text{and} \quad R_{\mu 2} \equiv \frac{\bar{\mu}_i}{M_{S_2}} \leq 0.1. \quad (68)$$
These conditions define allowed regions for $M_{N_i}$-$m_{D_{1i}}$ and $M_{N_2}$-$m_{D_{22}}$ which are shown as orange regions in Fig. 6. We also show in dashed orange curves the values where the above ratios assume the values $R_{\mu 1} = 0.03$ (left) and $R_{\mu 2} = 0.0076$ (right). We use the same reference values in (65).

The conclusion is that the overlapping (allowed) regions impose upper bounds on the heavy RHN states:

$$M_{N_1}, M_{N_2} \lesssim 340 \text{ GeV}. \tag{69}$$

This constraint puts the RHN states $S_i$ at the GeV-scale. We also note that had we allowed $M_{\nu_i}^{1-1} \sim M_{\nu_i}^{\text{tree}}$, $M_{N_1}$ would be unbounded but restricted to a narrow band $M_{D_{1i}}^2/M_{N_1} \sim 10^{-11}$ GeV for $M_{N_1} \gtrsim 1 \text{ TeV}$. A similar consideration applies to $M_{N_2}$.

As the last constraint, we note that $m_{D_{1i}}$ cannot be pushed to arbitrarily low values because it necessarily makes the lighter BSM states $S_i$ very long-lived. In order to not spoil the successful prediction of Big Bang nucleosynthesis (BBN), we require that the lifetimes of all the BSM states do not exceed 0.1 s. It is enough to require that for the lighter $S_i$ states. As their masses lie at the GeV-scale or lower, the main decay modes involve $W$ or $Z$ exchange through active-sterile mixing with decay into light neutrinos, electrons or pions [29]; see appendix C for more details.

The allowed regions are shown in green in Fig. 6 where the border is determined by the fixed $N - S$ ratios of (65); the interior refers to $R_{N_i} > 10^2$ (left) or $R_{N_2} > 10^2$ (right) in accordance to the ESS approximation. For completeness, we also show as dashed green curves the points where $\tau = 0.1$ s and $R_{N_i} = 270$ (left) or $R_{N_2} = 400$ (right).

The combination of all the constraints discussed above, leads to the overlapping regions of Fig. 6. The parameters are restricted to the values listed in Table II. The restriction means that points outside the overlapping region violate some constraint above for the reference values (65). Points inside the overlapping regions need to be further checked for all the constraints as they depend on other parameters not shown in the figures. Moreover, the parameters are not all independent as one ratio is fixed through (46) and

$$m_{D_{22}}/m_{D_{11}} M_{N_1} \sqrt{R_{N_2}/R_{N_1}} = k. \tag{70}$$

To use tree-level values different from (65) but restricted to (30) and (31), we just need to reread Fig. 6 with the vertical axis relabeled as

$$m_{D_{11}} \rightarrow m_{D_{11}} \sqrt{2 \text{ meV}/(M_{\nu_i}^{\text{tree}})};$$

$$m_{D_{22}} \rightarrow m_{D_{22}} \sqrt{24 \text{ meV}/(M_{\nu_i}^{\text{tree}})}. \tag{71}$$

This is possible because all the defining relations, Eqs. (64), (66), (67) and the active-sterile mixing $\theta_{eS}$ in the decay rates (ap. C) depends on $m_{D_{11}}^2$ or $m_{D_{22}}^2$. For the same reason, the blue and orange curves of the right figure of Fig. 6 are identical to the ones on the left if we identify $m_{D_{22}} = m_{D_{11}} \sqrt{24/2}$, where $\sqrt{24/2}$ is basically the factor $k$.

| $m_{D_{11}}/10^{-5}$ GeV | $5 - 8$ | $m_{D_{22}}/10^{-5}$ GeV | $12 - 28$ |
|-------------------------|---------|-------------------------|---------|
| $M_{N_1}/$GeV           | 80 - 340| $M_{N_2}/$GeV           | 25 - 340|
| $M_{Ni}/M_{S_1}$        | 100 - 270| $M_{N_2}/M_{S_2}$      | 100 - 400|
| $\bar{\mu}_i/M_{S_1}$  | 0.03 - 0.1| $\bar{\mu}_i/M_{S_2}$ | 0.0076 - 0.1|

**TABLE II:** Approximate parameter values extracted from Fig. 6.

As an example, the following values pass all the constraints and are also marked in Fig. 6 by crosses:

$$M_{D_{11}} = 7 \times 10^{-5} \text{ GeV}, \quad M_{N_1} = 200 \text{ GeV}, \quad M_{S_1} = 1.33 \text{ GeV}, \quad \bar{\mu}_i \sim 100 \text{ MeV},$$

$$M_{D_{22}} = 2.1 \times 10^{-4} \text{ GeV}, \quad M_{N_2} = 100 \text{ GeV}, \quad M_{S_2} = 1 \text{ GeV}. \tag{72}$$

---

9 We assume all the scalars to be heavier than $S_i$.

10 The actual green regions may lie slightly to the left for two reasons: (i) we only include the dominant decay modes for $S_i$ listed in appendix C and (ii) the strict lifetime limit for successful BBN may be slightly relaxed depending on the details of the model at the BBN era [30].
The intermediate scales $\sqrt{3}u_0 f'_1 \approx 16\text{ GeV}$ and $\sqrt{3}u_0 f'_2 = 10\text{ GeV}$ can be obtained from (52). They set a lower bound for the scales $\langle \eta_i \rangle \sim u_0 \gtrsim 10\text{ GeV}$ and $\langle \varphi_0 \rangle \sim \langle \varphi_1 \rangle \lesssim \mu_i \sim 0.1\text{ GeV}$ while the masses can be chosen $M_R \sim u_0 \gtrsim M_0 \sim 10\text{ GeV}$. Using the values in (72) as a benchmark, we plot in Fig. 7 the ratio of the one-loop contribution to the tree-level value of $|m^{\nu}_{\beta\beta}| = |(M^{\nu}_{\text{tree}})_{\beta\beta}| = 2\text{ meV}$ where now we vary $M_{S_1}$ and rescale $M_{N_i}$ simultaneously by fixing $R_{N_i} = 150$. For the benchmark values (72), the one-loop contribution is indeed less than 10% of the tree-level value. We also show the ratio of the lifetime to the limit of 0.1 s (solid gray) and confirm that $M_{S_1}$ needs to be larger than around 1 GeV.

![Fig. 7: Ratio of one-loop to tree-level contribution (2 meV) to the $ee$ entry (0/2/3 parameter) of the light neutrino mass matrix (solid black) as a function of $M_{S_1}$ ($M_{N_i}$ scales with $M_{S_1}$ through $M_{N_i} = R_{N_i} M_{S_1}$). We also show the contribution coming only from $N_i$ (dashed) and $S_1$ (dotted) exchange; the latter is multiplied by 10 for visualization. The contribution for 0/2/3 parameter from $S_1$ exchange (green dashed) relative to the limit 0.3 eV is shown as well; we use the expression in (81). The solid gray curve shows the lifetime for $S_1$ relative to 0.1 s. The other parameters are fixed as $m_{D_{11}} = 7 \times 10^{-5}\text{ GeV}$ and $R_{N_i} = 150$.](image)

Finally, we can estimate the amount of cancellation that is built-in in our ESS mechanism implementation. Rewriting (66) in the form of the naive seesaw relation,

$$|(M^\nu_{\text{tree}})_{ee}| = \epsilon_{ee} \frac{m^2_{D_{11}}}{M_{S_1}},$$

we extract

$$\epsilon_{ee} = \frac{\bar{\mu}_i}{M_{N_1}} \approx 10^{-3} - 10^{-4},$$

if we use Table II. Analogously, for the $\mu\tau$ entry, we obtain $\epsilon_{\mu\tau} \approx 10^{-3} - 2 \times 10^{-6}$. These values are in agreement with the radiative stability conditions discussed in Ref. [20] that estimated a lower bound of $\epsilon > 10^{-6}$ for a GeV-scale right-handed neutrino mass.

### VII. OTHER PHENOMENOLOGICAL CONSTRAINTS AND $U(1)_{\mu-\tau}$ BREAKING

We analyze here other phenomenological constraints coming from the existence of GeV-scale heavy neutrino $S_1$ with mixing with the light neutrinos at the order of

$$|(\theta V)_{e1}|^2 = \left(\frac{m_{D_{11}}}{\sqrt{3}u_0 f'_1}\right)^2 \frac{m^2_{D_{11}}}{M_{S_1} M_{N_1}} = 10^{-12} \times \left(\frac{m_{D_{11}}}{10\text{ keV}}\right)^2 \left(\frac{100\text{ GeV}}{M_{N_1}}\right) \left(\frac{1\text{ GeV}}{M_{S_1}}\right),$$

where we have used (52) and simplified the notation for $\theta_{\nu S} V_S$. For the values (72),

$$|(\theta V)_{e1}|^2 = \frac{m^2_{D_{11}}}{M_{S_1} M_{N_1}} \approx \frac{(7 \times 10^{-5}\text{ GeV})^2}{200\text{ GeV} \times 1.33\text{ GeV}} \sim 2 \times 10^{-11}.$$
The other mixing angles are either of the same order or vanishing in the limit of $U(1)_{\mu-\tau}$ conservation; cf. (55). At the same time, the Yukawa couplings to the RHN in our model are even more suppressed,

$$f_1 \sim \frac{7 \times 10^{-5} \text{ GeV}}{174 \text{ GeV}} \sim 4 \times 10^{-7},$$

$$f_{2,3} \sim \frac{2.1 \times 10^{-4} \text{ GeV}}{174 \text{ GeV}} \sim 10^{-6}. \quad (77)$$

They are smaller than the electron Yukawa coupling and thus the Higgs couplings to the RHN are very much suppressed (their are smaller than the mixing $\theta_{\nu S}$). Hence, the main interactions of the RHN to the SM fields occur through active-sterile mixing in (75).

However, it is clear that indirect detection constraints such as lepton universality violation or electroweak precision tests are not able to restrict or probe such a small mixing angles [31, 32]. They are also unobservable through direct detection in meson decays [29, 31, 32] or in colliders [33]. Note that this scenario contrasts with models where Higgses charged under $Z_3$ [or $U(1)_{\mu-\tau}$] may induce large lepton-flavor-violating Higgs decays [34].

For the same reason, lepton flavor violation (LFV) constraints are very weak in our model. The suppression is even larger because LFV processes such as $\mu \rightarrow e\gamma$ or $\mu \rightarrow eee$ are forbidden in the limit of $U(1)_{\mu-\tau}$ conservation. One can also see this in (55) as $(\theta V_S)_{\epsilon i} (\theta V_S)_{\mu i}^*$ always vanish. Being a larger group, $U(1)_{\mu-\tau}$ is more constraining than lepton flavor triality [21] and the former only allows $\tau^- \rightarrow \mu^+ e^- e^-$. However, when this process is mediated only by heavy neutrinos, it occurs through box diagrams that are very much suppressed [35]. These conclusions are not modified when $U(1)_{\mu-\tau}$ breaking effects are considered. See appendix D.

At last, we can analyze the limits coming from neutrinoless double beta decay, which are the strongest involving the electron flavor. Since the active-sterile mixings are all vanishing or of the same order in the $U(1)_{\mu-\tau}$ symmetry limit, cf. (55), we expect that this process will pose the strongest constraint on the mixings.

The half-life of the process is proportional to [20, 31]

$$\frac{1}{T_{1/2}^{\nu\nu}} \propto \frac{|m_{\beta\beta}^{\nu}|^2}{(p^2)} + \frac{n_n}{2} \left[ (\theta V_S)_{\epsilon i}^2 M_i - \left| (\theta V_S)_{\mu i} \right|^2 \right],$$

where $(p^2) \sim -(200 \text{ MeV})^2$ quantifies the effective momentum transfer inside the nucleus and $M_i$ represent the masses of the additional heavy neutrino states that mix with the three active ones. The light neutrino contribution depends on

$$m_{\beta\beta}^{\nu} = \sum_{i} U_{\epsilon i}^2 m_i, \quad (79)$$

with contributions arising from tree and loop contributions

$$(m_{\beta\beta}^{\nu})^* = (M_{\nu})_{\epsilon\epsilon} = (M_{\nu}^{\text{tree}})_{\epsilon\epsilon} + (M_{\nu}^{1\text{st}})_{\epsilon\epsilon} + \cdots. \quad (80)$$

For CP$^{\mu\tau}$ symmetric theories, it is confined to bands depending on the CP parities of the light neutrinos [8]. For a review on generic aspects of $0\nu2\beta$ see Ref. [36]. We are assuming we are confined to the parameter space where the one-loop contributions are negligible compared to the tree-level one.

Considering (78), we can define, in analogy to the light neutrino contribution [22],

$$m_{\beta\beta}^{S} \equiv |(p^2)| \sum_{i=1}^{3} \frac{(\theta V_S)_{\epsilon i}^2 M_{S_i}}{M_{S_i}^2 + |(p^2)|}. \quad (81)$$

where $|(p^2)| \approx (253 \text{ MeV})^2$ (corresponding to $0.079 \times (0.9 \text{ GeV})^2$ in Ref. [27]) and we have already specialized to $^{76}\text{Ge}$. We disregard the subdominant contribution from the heavier states $N_i$. If the heavy neutrino masses are much larger than the typical momentum transfer in the nucleus, $M_i \gg 200 \text{ MeV}$, we can approximate

$$m_{\beta\beta}^{S} \sim |(p^2)| \sum_{i=1}^{3} \frac{(\theta V_S)_{\epsilon i}^2}{M_{S_i}}. \quad (82)$$

Taking the GERDA+Helderberg-Moscow limit, $T_{1/2}^{\nu\nu}(^{76}\text{Ge}) \geq 3 \times 10^{25} \text{ yr}$ at 90% C.L. [37], it translates into

$$|m_{\beta\beta}^{\nu} + m_{\beta\beta}^{S}| \lesssim 0.3 \text{ eV}. \quad (83)$$
We can see that the contribution from light neutrinos predicted in our model (30) is at least two orders of magnitude smaller than the limit above. It remains to be checked if the contribution from $S_i$ exchange can give a larger contribution.

In the limit where $U(1)_{\mu-\tau}$ (or LFT) in (49) is conserved, only $S_1$ couples to the $e$ flavor and thus to $0\nu2\beta$; cf. (56). We can then write

$$m_{S_1}^{\beta\beta} = -2.67 \times 10^{-5} \text{eV} \times \left( \frac{m_{D_{12}}}{10 \text{keV}} \right)^2 \frac{10^2}{R_{N_1}} \left( \frac{\text{GeV}}{M_{S_1}} \right)^3 ,$$

(84)

where we have assumed that $M_{S_1} \gg 200 \text{MeV}$. For the values (72), this contribution is negligible. One could lower the $M_{S_1}$ mass to increase this contribution (including the correction in (81)) but it hits the BBN constraint rather quickly. Such a feature is illustrated in Fig. 7 where the ratio of the $0\nu2\beta$ contribution from $S_1$ exchange to the limit of 0.3 eV is shown in dashed green. Note that we use the expression (81) to account for $M_{S_1}$ larger than 1.33 GeV. Even if we allow the lifetime of $S_1$ to be around 1 s, it will still be unobservable in future $0\nu2\beta$ experiments. It is possible, however, that $m_{S_1}^{\beta\beta} \sim 30 \text{meV}$ for $M_{S_1} \sim 300 \text{MeV}$ and much larger than the light neutrino contribution.

VIII. CONCLUSIONS

We have presented a new CP symmetry applicable to models with $A_4$ flavor symmetry and other groups with the structure $H \propto \mathbb{Z}_3$ such as $\Delta(27)$. To implement this type of CP symmetry, the singlets $Y'$ that are fermions or carry other quantum numbers should appear in pair with another $1'$ with the remaining quantum numbers identical to those of $1'$. This new CP symmetry allows us to avoid the vev alignment problem in close analogy to the construction using $L_\mu - L_\tau$ and $CP^{\mu\tau}$ symmetries [8]. This feature partly follows because the SM lepton fields are singlets of $A_4$ and only feel the $\mathbb{Z}_3$ subgroup which is contained in $L_\mu - L_\tau$.

We have constructed an explicit renormalizable model that leads to a new form for the light neutrino mass matrix, cf. (1). It retains the successful predictions of $CP^{\mu\tau}$ – namely maximal $\theta_{23}$, maximal Dirac CP phase and trivial Majorana phases – but because of the $A_4$ structure it also predicts normal hierarchy with the lightest neutrino of only few meV; see (29). The CP parities are also restricted to two possibilities which effectively fix the effective parameter $m_{\beta\beta}$ contributing to neutrinoless double beta decay.

The model itself is based on the extended seesaw mechanism which naturally leads to relatively light right-handed neutrinos $S_i$ and heavier $N_i$. After enforcing negligible one-loop contributions to light neutrino masses, ensure the ESS approximation and require fast enough decay rate of the BSM states to avoid BBN constraints we only find a small allowed region in the parameter space: $N_i$ neutrinos lie at the electroweak scale and the lighter $S_i$ lie at the GeV scale; see Fig. 6. To suppress the one-loop contributions, it is required that their Yukawa interactions with the SM fields should be smaller than the electron Yukawa coupling. Consequently the active-sterile mixing is largely suppressed, rendering the right-handed neutrinos practically unobservable in terrestrial experiments.

The flavor structure of the model is largely determined by the approximate conservation of the combination $L_\mu - L_\tau$ of lepton flavors, which suppresses various flavor changing processes such as $\mu \rightarrow e\gamma$. Moreover, only $S_1$ mixes appreciably to $\nu_e$ and the mixing of $S_{2,3}$ to the $\mu\tau$ flavors are of the same order of magnitude.

Acknowledgments

The author would like to thank the Maryland Center for Fundamental Physics of the University of Maryland at College Park, USA, for the hospitality where this work initiated. The author also thanks Rabindra Mohapatra for discussion on several points. Partial support by Brazilian Fapesp grant 2013/26371-5 and 2013/22079-8 is acknowledged.

Appendix A: Block diagonalization of $M_R$

The ESS mechanism naturally leads to two disparate scales for the right-handed neutrinos: the lighter $M_S$ ($S_{1R}$) and the heavier $M_N$ ($N_{1R}$). So it is useful to write the complete neutrino mass matrix (37) in a basis where $M_R$ is
The mass matrix $M_S$ is given by (41). The subleading correction to $M_N$ is

$$M'_N = M_N + \frac{1}{2}(A^T \Lambda^* M_N^{-1} + \text{tr.}),$$  

where tr. indicates the transpose of the previous matrix.

The block diagonalization is performed by

$$U_R \approx \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 - \theta_R \theta_R^\dagger/2 & -\theta_R^\dagger \\
-\theta_R & 1 - \theta_R^\dagger \theta_R/2 \end{pmatrix},$$  

with

$$\theta_R^\dagger = \Lambda M_N^{-1}.$$

Further block diagonalization leads to the results in (41) and (42). The complete diagonalization is performed by

$$\nu_i \rightarrow (U_\nu)_{ij} \nu_j + (\theta_{\nu S} V_S)_{ij} \eta_i + (\theta_{\nu N} V_N)_{ij} \eta_j,$$

$$S_{ij}^c \rightarrow (V_S)_{ij} S_{ij}^c + (-\theta_{\nu S} U_\nu)_{ij} \nu_j + (\theta_{\nu N} U_N)_{ij} \eta_j,$$

$$N_{ij}^c \rightarrow (V_N)_{ij} N_{ij}^c + (m_D^{-1} M_D U_\nu)_{ij} \nu_j + (-\theta_{\nu S} V_S)_{ij} S_{ij}^c.$$  

The fields on the left-hand side are in the flavor basis and appear in (34); the ones on the right-hand side are the mass eigenfields and $U_\nu$ is the PMNS matrix in the flavor basis. We have neglected nonunitary effects and the small mixing angles $\theta_{\nu S}, \theta_{\nu N}, \theta_R$ were already given in Eqs. (42) and (A4).

**Appendix B: Comments on the potential**

Here we justify the approximate conservation of $U(1)_{\mu-\tau}$ that follows from the $\mathbb{Z}_3$ conserving vevs for $\eta$ in (48).

We start by observing that when the potential for $\eta$ is invariant by global rephasing, the potential is identical to a potential with three Higgs doublets with $A_4$ symmetry and we know that (48) can be exactly a global minimum [38].

The addition of the two independent quartic terms that breaks $U(1)$ but conserves $\mathbb{Z}_3^D$,

$$I_1 = \eta_1^4 + \eta_2^4 + \eta_3^4 \quad \text{and} \quad I_2 = (\eta_1 \eta_2)^2 + (\eta_2 \eta_3)^2 + (\eta_3 \eta_1)^2,$$

can be chosen to maintain such alignment and also to make $u_0$ real and positive. We stress that these and other quartic terms are not invariant by $U(1)_{\mu-\tau}$ but only the $\mathbb{Z}_3$ subgroup. These terms also help to maintain the deviations of $\langle \eta \rangle$ in the real direction since the coefficients are real because of CP$_{\mu\tau}$.

Now we add the interactions of $\eta$ with $\varphi_0$ and $\varphi_1$. The relevant terms are

$$V \supset \frac{1}{2} M_0^2 \varphi_0^2 + \mu_{0\eta} \varphi_0 (\eta^T \eta + h.c.),$$  

and

$$V \supset M_1^2 |\varphi_1|^2 + \left\{ \mu_{1\eta} [(\eta \eta)^{\dagger} \varphi_1 + (\eta \eta)^{\dagger} \varphi_1 + h.c.] \right\},$$  

where $\mu_{1\eta}$ can be complex and the singlet combination was defined in (35). Clearly there is no $U(1)$ rephasing symmetry for $\varphi_1$ and no Goldstone will be generated.

The mild hierarchy of ESS scales

$$100 \text{ MeV} \sim \mu_i \ll f_i' \sqrt{3} u_0 \sim 10 \text{ GeV},$$  

implies a mild hierarchy between $u_0 \sim \langle \eta \rangle$ and $\langle \varphi_0 \rangle, \langle \varphi_1 \rangle$. We can choose $f_i' \sim 0.1$ so that $u_0 \sim 100 \text{ GeV} \sim M_N$. For an order one $k_0$, the small $\langle \varphi_0 \rangle \sim 100 \text{ MeV}$ can be generated from (B2) by a vev seesaw analogous to type-II
seesaw [39]. In this case $M_{\phi_0} \sim u_0$ is electroweak scale. For $\langle \varphi_1 \rangle$ a vev seesaw cannot be implemented because $(\eta \eta)^{\nu \nu}$ vanishes for the minimum (48). But we can always take $k_1 \sim 10^{-2}$, adjust the potential parameters to obtain $\langle \varphi_1 \rangle \sim 10$ GeV and make $\mu_{\eta \eta}$ in (B3) small enough so that (48) is only slightly disturbed. The mass of the lightest physical states of $\varphi_1$ will be around $\langle \varphi_1 \rangle$ and heavier than $S_i$. Note that $k_0 \langle \phi_0 \rangle$ and $k_1 \langle \phi_1 \rangle$ should be comparable because they lead to $\mu_i$.

At last, in principle the new scalars could be produced in Higgs decays through the Higgs portal but the current limits on the invisible Higgs decays are still weak [40] and can be avoided by decreasing the portal interactions.

**Appendix C: Decay rates for $S_i$**

In our theory the RHN heavy states $S_i$ are the lightest new states beyond the SM which lies at the GeV-scale. The dominant decay channels involve $Z$ or $W$ exchange through mixing with light neutrinos or charged leptons [29]. The decays $S_i^c \rightarrow S_i^c + \cdots$ are highly suppressed.

To ensure that the production of light nuclear elements in the early Universe (Big Bang nucleosynthesis) are not disturbed by the presence of new particles, we require that the lifetimes of the new states are shorter than 0.1 second. In this case these new particles are thermalized much before the BBN era and they decay fast enough. RHNs lighter than around 100 MeV conflict with direct detection constraints and are excluded [30, 41].

Assuming the $U(1)_{\mu - \tau}$ symmetry, the active-sterile mixing (55) leads to the dominant decay channels [29]

$$S_1^c \rightarrow \pi^0 \nu_e, \nu_e \bar{\nu}_e, \pi^+ e^-, \quad S_2^c, S_3 \rightarrow \pi^0 \nu_{\mu, \tau}, \nu_{\mu, \tau} \bar{\nu}_e, \pi^+ \mu^-.$$  \hspace{1cm} (C1)

We neglect the decay to other channels. The decay rates for these processes can be taken from Ref. [29]:

$$\Gamma(S_1^c \rightarrow \pi^0 \nu_e) = \frac{|(\theta V_S)_{e1}|^2 G_F \bar{f}^2 M^3}{32 \pi} \left(1 - \frac{m_{\pi^0}^2}{M^2}\right)^2,$$

$$\Gamma(S_2^c \rightarrow \pi^0 \nu_{\mu, \tau}) = \left(|(\theta V_S)_{\mu2}|^2 + |(\theta V_S)_{\tau2}|^2\right) \frac{G_F \bar{f}^2 M^3}{32 \pi} \left(1 - \frac{m_{\pi^0}^2}{M^2}\right)^2,$$

$$\Gamma(S_3^c \rightarrow \nu_e \bar{\nu}_e) = \frac{|(\theta V_S)_{e1}|^2 G_F \bar{f}^2 M^3}{192 \pi^3},$$

$$\Gamma(S_4^c \rightarrow \nu_{\mu, \tau} \bar{\nu}_e) = \left(|(\theta V_S)_{\mu2}|^2 + |(\theta V_S)_{\tau2}|^2\right) \frac{G_F \bar{f}^2 M^3}{192 \pi^3},$$

$$\Gamma(S_1^c \rightarrow \pi^+ e^-) = \frac{|(\theta V_S)_{e1}|^2 G_F \bar{f}^2 |V_{ud}|^2 M^3}{16 \pi} \left(1 - \frac{m_{\pi^+}^2}{M^2}\right)^2 - \frac{m_{\pi^0}^2}{M^2} \left(1 + \frac{m_e^2}{M^2}\right),$$

$$\times \sqrt{\left(1 - \frac{(m_{\pi^+} - m_e)^2}{M^2}\right)^2 - \left(1 - \frac{(m_{\pi^+} + m_e)^2}{M^2}\right)^2)},$$

$$\Gamma(S_2^c \rightarrow \pi^+ \mu^-) = \frac{|(\theta V_S)_{\mu2}|^2 G_F \bar{f}^2 |V_{ud}|^2 M^3}{16 \pi} \left(1 - \frac{m_{\pi^0}^2}{M^2}\right)^2 - \frac{m_{\pi^+}^2}{M^2} \left(1 + \frac{m_\mu^2}{M^2}\right),$$

$$\times \sqrt{\left(1 - \frac{(m_{\pi^+} - m_\mu)^2}{M^2}\right)^2 - \left(1 - \frac{(m_{\pi^+} + m_\mu)^2}{M^2}\right)^2}).$$

In each expression, $M$ refers to the mass of the decaying particle and each decay rate contributes twice due to the charge conjugate mode. Moreover, the expression for $S_3^c$ are identical to the expressions for $S_2^c$ and note that we can write

$$|(\theta V_S)_{e1}|^2 = \frac{m_{D_{21}}^2}{M_{S_1} M_{N_1}},$$

$$|(\theta V_S)_{\mu2}|^2 + |(\theta V_S)_{\tau2}|^2 = \frac{m_{D_{22}}^2}{M_{S_2} M_{N_2}}$$

$$= 2 |(\theta V_S)_{\mu2}|^2 = 2 |(\theta V_S)_{\tau2}|^2.$$  \hspace{1cm} (C3)
We are also assuming that $U(1)_{\mu-\tau}$ is slightly broken so that $S_{2,3}$ are distinct Majorana fermions. In the exact $U(1)_{\mu-\tau}$ limit, $(S_2^c + iS_3^c + S_2 + iS_3)/\sqrt{2} = S_{\mu\tau}$ forms a Dirac heavy neutrino with $U(1)_{\mu-\tau}$ charge unity while its conjugate carries charge $-1$. In this case, the decay rates of $S_{\mu\tau}$ are the same as $S_2^c$ without the factor two multiplication (the last one would be doubled due to diagonal mixing).

Appendix D: Deviations of $U(1)_{\mu-\tau}$

In the fermion sector our model is approximately invariant by $U(1)_{\mu-\tau}$, which includes $Z_4$ (47) of $A_4$. In the first approximation considered $U(1)_{\mu-\tau}$ is only broken in the neutrino sector by small $\mu_i \sim 100\text{ MeV}$ in (37). Identical $U(1)_{\mu-\tau}$ charges (49) can be assigned to all the lepton fields (49) if we change basis to

$$S_{iR}^c = (U_\omega)_{ij} S_{jR}^c,$$

where $\omega$ carries charge

$$\tau$$

limit, $(S_2^c + iS_3^c + S_2 + iS_3)/\sqrt{2} = S_{\mu\tau}$ forms a Dirac heavy neutrino with $U(1)_{\mu-\tau}$ charge unity while its conjugate carries charge $-1$. In this case, the decay rates of $S_{\mu\tau}$ are the same as $S_2^c$ without the factor two multiplication (the last one would be doubled due to diagonal mixing).

The combination

$$M_S^f = M_S^{(0)} + \delta M_S'$$

is now diagonalized by $V_S' = i U_{23} C_\tau$, where $C_\tau = \frac{1}{\sqrt{2}} (\bar{\epsilon} \epsilon \bar{\epsilon} \bar{\epsilon})$.
where $U_{23}$ denotes the maximal mixing matrix in (19). One can check that $O_\epsilon$ is a real orthogonal matrix given by

$$O_\epsilon \approx \begin{pmatrix} 1 & -d_1' & -d_2' \\ d_1 & c_\theta & -s_\theta \\ d_2 & s_\theta & c_\theta \end{pmatrix}.$$  

(D9)

The small parameters $d_i$ are combinations of the small quantities in (D7) and are defined by

$$-U_{23}^T \delta M_S U_{23} = M_{S_1}^{(0)} \begin{pmatrix} 0 & d_1 & d_2 \\ d_1 & c_1 & c_2 \\ d_2 & c_2 & -c_1 \end{pmatrix};$$  

(D10)

all $d_i, c_i$ are real. The primed $d_i'$ are rotated as

$$\begin{pmatrix} d_1' \\ d_2' \end{pmatrix} = \begin{pmatrix} c_\theta & s_\theta \\ -s_\theta & c_\theta \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix},$$  

(D11)

with angle $\tan 2\theta = c_2/c_1$. One can note that the angle $\theta$ depends only on the deviation parameters $c_i$ and does not need to be small due to the degeneracy $M_{S_2}^{(0)} = M_{S_3}^{(0)}$. The formula (D9) is valid as long as $d_i, c_i \ll 1$ and covers the case where $M_{S_1}^{(0)} \gg M_{S_2}^{(0)} \sim M_{S_3}^{(0)} d_i \sim M_{S_1}^{(0)} c_i$ so that the mass splitting for $S_{2,3}$ can be substantial:

$$M_{S_2} = M_{S_2}^{(0)} + M_{S_1}^{(0)} \sqrt{c_1^2 + c_2^2},$$

$$M_{S_3} = M_{S_2}^{(0)} - M_{S_1}^{(0)} \sqrt{c_1^2 + c_2^2}.$$

(D12)

We are adopting $M_{S_3} < M_{S_2}$.

Putting all together we find the deviation from (55):

$$(\theta V_S) \approx |(\theta V_S^{(0)} e_1) \text{diag}(1, k, k) \begin{pmatrix} 1 & -\epsilon_3^* & -\epsilon_2^* \\ -\epsilon_3^* & 1 & -\epsilon_3^* \\ -\epsilon_2^* & -\epsilon_3^* & 1 \end{pmatrix} \times iU_{23}O_\epsilon, \quad (\theta V_S^{(0)}) e_1 \text{diag}(1, k, k) \times \begin{pmatrix} 1 & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & x_{22} & x_{23} \\ \epsilon_{31} & x_{32} & x_{33} \end{pmatrix},$$

(D13)

where $\epsilon_{ij}$ are small parameters that depend on the small parameters $\epsilon_{2,3}$ while $x_{ij}$ are order one, approximately unitary, quantities. The deviation from maximal (23) mixing in (55) can be large due to $S_{2,3}$ mass degeneracy in the U(1)$_{\mu-\tau}$ limit. Again the superscript (0) denotes the U(1)$_{\mu-\tau}$ limit. Note that $(\theta V_S)$ has the structure

$$\begin{pmatrix} u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \\ w_1^\ast & w_2^\ast & w_3^\ast \end{pmatrix},$$

(D14)

characteristic of CP$^{\mu\tau}$ invariance [8, 11].

Considering the deviation (D13) in $\nu - S$ mixing, we can include the effects of $M_{S_{2,3}}$ exchange in $0\nu2\beta$ as

$$m_{S_{2,3}}^{\beta\beta} = m_{S_{2,3}}^{S_1} \left\{ 1 + \epsilon_{12}^2 M_{S_2}^{(0)} + \epsilon_{13}^2 M_{S_3}^{(0)} \right\},$$

(D15)

where $M_{S_{2,3}}$ are now nondegenerate and include the U(1)$_{\mu-\tau}$ breaking effects. It is clear that the contribution of $S_2$ ($S_3$) exchange can be comparable to $S_1$ exchange only if

$$M_{S_1}/M_{S_{2,3}} \sim O(1/\epsilon^2).$$

(D16)

This cannot happen in our theory.
We can also confirm that U(1)$_{\mu-\tau}$ breaking is not enough to induce observable lepton flavor violating processes such as $\mu \to e\gamma$. The vanishing rate is now proportional to the U(1)$_{\mu-\tau}$ breaking effects. Considering only $S_i$ in the loop, the branching ratio yields [24]

$$B(\mu \to e\gamma) \sim 2 \times 10^{-30} \times \frac{e_{21}}{0.11}^2 \times \left(\frac{\theta V_{eS_i}}{10^{-6}}\right)^4 \tilde{G}_i^2,$$

where $\tilde{G}_i = G(M_2^2/M_\nu^2) - G(0)$ and $G(x)$ is defined in Ref. [24]. For example, $G(1/2^2) - G(0) \approx -10^{-4}$. Therefore, the predicted rate is much below the current MEG limit $B(\mu \to e\gamma) < 2.4 \times 10^{-12}$ [42] and there is no constraint even if $(\theta V)^{S_2,3}$ is as large as $1\%$. One can also check that $S_{2,3}$ contributions lead to similar results. Future $\mu \to e$ conversion experiments in nuclei [43] can improve the limit by few orders of magnitude but our model predictions are still suppressed. Hence, LFV processes constraints are much weaker than $0\nu2\beta$ in our model.

[1] K. Abe et al. [T2K collaboration], Phys. Rev. Lett. 107 (2011) 041801 [1106.2822]; P. Adamson et al. [MINOS Collaboration], Phys. Rev. Lett. 107, 181802 (2011) [1106.0015]; Y. Abe et al. [DOUBLE-CHOOZ Collaboration] Phys. Rev. Lett. 108, 131801 (2012) [1112.6353]; F. F. An et al. [DAYA-BAY Collaboration], Phys. Rev. Lett. 108, 171803 (2012) [1203.1609]; J. K. Ahn et al. [RENO Collaboration], Phys. Rev. Lett. 108, 191802 (2012) [1204.0426].

[2] A. de Gouvea et al. [Intense Frontier Neutrino Working Group Collaboration], 1310.4340 [hep-ex].
[31] F. F. Deppisch, P. S. Bhupal Dev and A. Pilaftsis, New J. Phys. 17 (2015) 7, 075019 [1502.06541 [hep-ph]]; M. Drewes and B. Garbrecht, 1502.00477 [hep-ph]. S. Antusch and O. Fischer, JHEP 1505 (2015) 053 [1502.05915 [hep-ph]]; JHEP 1410 (2014) 094 [1407.6607 [hep-ph]].

[32] A. de Gouvêa and A. Kobach, 1511.00683 [hep-ph].

[33] F. del Aguila and J. A. Aguilar-Saavedra, Nucl. Phys. B 813 (2009) 22 [0808.2468 [hep-ph]]; Phys. Lett. B 672 (2009) 158 [0809.2096 [hep-ph]]; F. del Aguila, J. A. Aguilar-Saavedra and R. Pittau, JHEP 0710 (2007) 047 [hep-ph/0703261]; A. Das, P. S. Bhupal Dev and N. Okada, Phys. Lett. B 735 (2014) 364 [1405.0177 [hep-ph]]; A. Das and N. Okada, Phys. Rev. D 88 (2013) 113001 [1207.3734 [hep-ph]]; A. Das and N. Okada, arXiv:1510.04790 [hep-ph].

[34] Q. H. Cao, A. Dumanik, E. Ma and D. Wegman, Phys. Rev. D 83 (2011) 093012 [1103.0008 [hep-ph]].

[35] A. Ilakovac and A. Pilaftsis, Nucl. Phys. B 437 (1995) 491 [hep-ph/9403398].

[36] W. Rodejohann, Int. J. Mod. Phys. E 20 (2011) 1833 [1106.1334 [hep-ph]]; C. Aalseth et al., hep-ph/0412300.

[37] M. Agostini et al. [GERDA Collaboration], Phys. Rev. Lett. 111 (2013) 12, 122503 [1307.4720 [nucl-ex]]; J. B. Albert et al. [EXO-200 Collaboration], Nature 510 (2014) 229 [1402.6956 [nucl-ex]]. Y. Huang and B. Q. Ma, The Universe 2 (2014) 65 [1407.4357 [hep-ph]].

[38] I. P. Ivanov and C. C. Nishi, JHEP 1501 (2015) 021 [1410.6139 [hep-ph]].

[39] J. Schechter and J. W. F. Valle, Phys. Rev. D 22 (1980) 2227; M. Magg and C. Wetterich, Phys. Lett. B 94 (1980) 61; T. P. Cheng and L. F. Li, Electroweak Phys. Rev. D 22 (1980) 2860. G. Lazarides, Q. Shafi and C. Wetterich, Nucl. Phys. B 181 (1981) 287; R. N. Mohapatra and G. Senjanovic, Phys. Rev. D 23 (1981) 165;

[40] The ATLAS collaboration [ATLAS Collaboration], “Search for an Invisibly Decaying Higgs Boson Produced via Vector Boson Fusion in pp Collisions at $\sqrt{s} = 8$ TeV using the ATLAS Detector at the LHC,” ATLAS-CONF-2015-004.

[41] A. D. Dolgov, S. H. Hansen, G. Raffelt and D. V. Semikoz, Nucl. Phys. B 590 (2000) 562 [hep-ph/0008138]; Nucl. Phys. B 580 (2000) 331 [hep-ph/0008223]; A. Kusenko, S. Pascoli and D. Semikoz, JHEP 0511 (2005) 028 [hep-ph/0405198].

[42] J. Adam et al. [MEG Collaboration], Phys. Rev. Lett. 107 (2011) 171801, [1107.5547 [hep-ex]].

[43] R. Alonso, M. Dhen, M. B. Gavela and T. Hambye, “Muon conversion to electron in nuclei in type-I seesaw models,” JHEP 1301 (2013) 118 [1209.2679 [hep-ph]].