Collider searches for heavy neutral leptons: beyond simplified scenarios

Asmaa Abada\textsuperscript{1,a}, Pablo Escribano\textsuperscript{2,b}, Xabier Marcano\textsuperscript{3,c}, Gioacchino Piazza\textsuperscript{1,d}\textsuperscript{c}

\textsuperscript{1} Pôle Théorie, Laboratoire de Physique des 2 Infinis Irène Joliot Curie (UMR 9012) CNRS/IN2P3, 15 Rue Georges Clemenceau, 91400 Orsay, France

\textsuperscript{2} Instituto de Física Corpuscular (CSIC-Universitat de València), Catedrático José Beltrán 2, 46980 Paterna, Valencia, Spain

\textsuperscript{3} Departamento de Física Teórica and Instituto de Física Teórica UAM/CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain

Received: 7 September 2022 / Accepted: 7 November 2022 / Published online: 15 November 2022 © The Author(s) 2022

Abstract With very few exceptions, the large amount of available experimental bounds on heavy neutral leptons – HNL – have been derived relying on the assumption of the existence of a single (usually Majorana) sterile fermion state that mixes with only one lepton flavour. However, most of the extensions of the Standard Model involving sterile fermions predict the existence of several HNLs, with complex mixing patterns to all flavours. Consequently, most of the experimental bounds for HNLs need to be recast before being applied to a generic scenario. In this work, we focus on LHC searches of heavy neutral leptons and discuss how to reinterpret the available bounds when it comes to consider mixings to all active flavours, not only in the case with a single HNL, but also in the case when more heavy neutral leptons are involved. In the latter case, we also consider the possibility of interference effects and show how the bounds on the parameter space should be recast.

Contents

1 Introduction .................................. 1
2 Status of HNL searches at high-energy colliders .... 2
  Searches at lepton colliders ................. 4
3 Beyond the single mixing assumption ........ 5
4 Beyond the single HNL ..................... 9
  CP violation ................................ 12
5 Conclusions ................................ 12

\textsuperscript{a} e-mail: asmaa.abada@ijclab.in2p3.fr
\textsuperscript{b} e-mail: pablo.escribano@ific.uv.es
\textsuperscript{c} e-mail: xabier.marcano@uam.es
\textsuperscript{d} e-mail: gioacchino.piazza@ijclab.in2p3.fr (corresponding author)

Appendix A: Decay process in presence of 2 HNL ...... 13
A.1 Same sign leptons: \( W^+ \rightarrow \ell_a^+ \ell_b^+ q'\bar{q} \) ............ 13
A.2 Different sign leptons: \( W^+ \rightarrow \ell_a^+ \ell_b^- q'\bar{q} \) ............ 14
References ........................................ 15

1 Introduction

Generating neutrino masses and their mixing as observed in neutrino oscillation phenomena requires to go beyond the Standard Model of Particles (BSM). Many options are presently explored as extensions of the Higgs and/or gauge sectors, most of the time with new fields within the particle content. In particular heavy neutral fermions, such as right-handed neutrinos \( \nu_R \), are often present as building blocks of several neutrino mass generation mechanisms. For instance, at least two \( \nu_R \) are required to accommodate light neutrino masses via the type-I seesaw mechanism [1–5]. Moreover, in several variants of the type-I seesaw realized at low scale, other sterile (from SM gauge interactions) fermions \( \nu_S \) are considered, as in the case for the Inverse [6–8] and Linear [9,10] seesaw mechanisms; these variants allow to have large neutrino Yukawa couplings with a comparatively low seesaw scale, potentially within reach at colliders. From now on, we will refer to these states \( (\nu_R,S) \) as Heavy Neutral Leptons (HNL).

Due to the presence of HNLs, the charged and neutral currents are modified, with the leptonic mixing matrix encoding now not only the PMNS mixing matrix [11,12], but also the active-HNL mixings \( U_{aN}, \alpha = e, \mu, \tau \). With these modifications and depending on the mass scale of these new neutral leptons, one expects an impact on numerous observables and
thus to abundant constraints on the plane \((M_N, |U_{aN}|^2)\) (see Refs. [13–15] and references therein).

In this work we will be interested in HNL searches at high-energy colliders, mainly at the LHC due to its current extensive program dedicated to HNL searches, see for instance [16], and the numerous dedicated works and analyses [17–58]. With very few exceptions, the large amount of available HNL bounds have been derived relying on the assumption of the existence of a single (usually Majorana) HNL that mixes with only one lepton flavour. However, most of the BSM scenarios involving new neutral leptons address the lepton mixing as a whole, as it impacts flavour physics studies and lepton properties – Dirac or Majorana nature for neutrinos – with different lepton number conserving/violating processes. The mixing pattern in these scenarios is expected to be quite complex, so applying the bounds from negative searches on the HNL space of parameters derived in the context of simplified hypotheses (with only one HNL which mixes to only one active flavour) does not seem adequate. Indeed, as we will see, using these limits directly will in general overconstrain the parameter space. Consequently, most of the experimental bounds for HNL need to be recast before being applied to a generic BSM scenario.

The motivation for reinterpreting LHC bounds in general is a well-established topic, see for instance Ref. [59] and references therein. In the context of HNL searches, the reinterpretation of the obtained bounds on the HNL mixings to active flavours has been addressed in previous works, as in for instance Ref. [50] discussing searches for Heavy Neutral Leptons with Displaced Vertices, and more recently in Ref. [58] focusing on HNLs decaying promptly to a tri-lepton final state. In the latter, the single-flavour mixing results obtained by ATLAS have been recast to a low-scale seesaw model with a pseudo-degenerate pair of HNLs, the most minimal and simple extension in order to accommodate neutrino oscillation data (the lightest neutrino being massless). Due to the simplicity of this model, the active neutrino masses and mixings determine the flavour pattern of the HNLs [60,61], and it is possible to define benchmark points beyond the single-flavour scenario [62]. While being a very interesting scenario, as it is probing the parameter space connected to light neutrino masses, this approach has the drawback of being model dependent. For example, considering other sources for light neutrino masses,\(^1\) such as additional HNLs not necessarily within the LHC range, could spoil the correlation between light and heavy sectors that motivated the definition of these scenarios.

For this reason, in this work we will follow a different approach. We will instead work with physical HNL states with independent mixings and masses, with the motivation of covering every scenario that could be realized at generic BSM models. We will also discuss how to go beyond the simplest single-flavour mixing scenario, however we will not attempt to explain light neutrino masses and mixings. This idea is actually the most straightforward extension to what is usually assumed at LHC searches. In doing so, we will discuss what would be the most relevant quantities to be bounded experimentally in order to easily reinterpret the results.

Furthermore, we will also extend the study to the case where more than one HNL is present and, in particular, taking into account possible interference effect when at least two heavy neutral leptons are in the same mass regime (nearly degenerate or possibly forming a pseudo-Dirac neutrino pair). This possible interference effect can lead to different bounds on the active-sterile mixings, see for instance [63–65].

This work is organised as follows: after having thoroughly discussed the status of high-energy collider searches of HNL in Sect. 2, a first insight of going beyond the single active-sterile mixing approximation is examined in Sect. 3. In this section, we also discuss how one would reinterpret experimental data from branching ratios to limits on the model parameter space, i.e. HNL masses and mixings. In Sect. 4, we consider the generic case of having more than one HNL and discuss possible changes in the interpretation of the bounds on HNL, in particular when the interference effects can be relevant. Final comments are collected in Sect. 5. Further details for deriving the relevant amplitudes are collected in the Appendices.

\(^1\) This would actually be needed in order to relax the hypothesis of the lightest active neutrino being massless and thus have in general 3 massive active neutrinos.
to be compared to. Along this work, we will refer as light\(^2\) HNL to those with masses lighter than the W boson, and as heavy HNL to the ones with \(M_N > M_W\). Extensive reviews about HNL searches at colliders can be found for instance in Refs. [66–68]. Here we just summarize and update the list of experimental analyses, introducing at the same time the most relevant aspects that we will use in our discussion in the next sections.

As in any collider search looking for heavy unstable particles, we need to consider both the production and the decay channels of the HNLs. At a hadronic collider such as the LHC, the main production channel comes from Drell–Yan, W and Z bosons,

\[
pp \rightarrow W^{(*)} \rightarrow N\ell^\pm \quad \text{and} \quad pp \rightarrow Z^{(*)} \rightarrow N\nu,
\]

where the gauge bosons could be on- or off-shell, depending on whether the HNL is lighter or heavier than the W or Z bosons. Additional production channels could also arise from the Higgs \((H)\) boson decays, which could be motivated in several models providing large neutrino Yukawa couplings. Unfortunately, Higgs bosons are produced less abundantly than weak bosons, so they are usually neglected. Moreover the W channel has the additional prompt charged lepton that can help triggering the process and reducing backgrounds, and thus experimental searches focused mostly on this channel. Nevertheless, it is worth mentioning that for very heavy masses, at around the TeV scale, vector boson fusion channels such as \(WY\) or \(WW\) become important and could even dominate the production of HNLs [25, 27, 57]. Indeed, the latest CMS analysis [69] already included the \(W\gamma\) channel in order to enhance their sensitivity to high HNL masses.

After being produced, a HNL of several GeVs, but still lighter than \(M_W\), will decay dominantly via off-shell \(W\) or \(Z\) bosons to a 3-body final state

\[
\begin{align*}
N &\rightarrow \ell^\pm_j j, \\
N &\rightarrow \ell^+_\alpha \ell^-_\beta \nu, \\
N &\rightarrow v_\alpha j j, \\
N &\rightarrow v_\alpha \ell^\pm_\beta \ell^\mp_\beta, \\
N &\rightarrow 3\nu.
\end{align*}
\]

On the other hand, if \(M_N\) is above the EW scale, the dominant decays will be to on-shell \(W\), \(Z\) and \(H\) bosons, i.e. \(N \rightarrow \ell^\pm W^\mp, vZ, vH\). These 2-body decays will be followed by the decay of the heavy bosons, leading at the end to the same final states as before. Nevertheless, it is important to keep in mind that the kinematics in these two mass regimes will be different.

Combining both production and decay channels, we get a full process such as the example shown in Fig. 1. Depending

\[
n \rightarrow \ell^\pm_\alpha \ell^-_\beta \ell^\mp_\gamma. 
\]

This channel has a trilepton signature, rather clean in a hadronic collider, nevertheless it also has a source of MET,
which might spoil the complete reconstruction of $M_N$. The trilepton channel offers the possibility to search for both LNV and LNC signals, however most of the experimental analyses still focus only on the LNV channels as to reduce backgrounds, again from $Z \to \ell^+\ell^-$. For example, ATLAS searched [84] for $e^\pm e^\pm \mu^\pm$ and $\mu^\pm \mu^\pm e^\pm$ channels\(^3\), but not for $e^\pm e^\pm \mu^\pm$ and $\mu^\pm \mu^\pm e^\pm$. CMS did something similar for light HNLs, although they also included channels with OS but same flavour lepton pairs in the heavy HNL regime, removing only those events with lepton pairs compatible with a decay of a $Z$ boson [69]. As stated before for the SS dilepton channel, searching for HNLs without assuming their Majorana nature will be helpful to probe scenarios compatible with neutrino oscillation data, and thus with potentially suppressed LNV signals.

Finally, it is important to stress that improving the experimental sensitivities to smaller values of mixings implies exploring HNL with longer lifetimes, which can travel macroscopic distances before decaying. Such long-lived HNL would avoid the searches mentioned so far, as they all assumed prompt decaying HNLs, and therefore we need a dedicated search for this kind of topologies. Recently, both ATLAS [85] and CMS [86] have searched for displaced vertex signatures for light HNLs, setting the strongest constraints for GeV masses up to 20 GeV. This can be seen in Fig. 2, where we summarize all the relevant LHC constraints explained in this section.

\[^3\] Having an undetected (anti)neutrino, it is not always possible to define a LNV or LNC process unambiguously. Assuming the presence of a HNL that mixes only to electrons or to muons, as ATLAS did, the $e^\pm e^\pm \mu^\pm$ and $\mu^\pm \mu^\pm e^\pm$ channels are originated only from LNV processes. However, this is not anymore true if the HNL mixes to both flavours [58].

Searches at lepton colliders

Despite the fact that our current most powerful high-energy collider is a hadronic one, it is important to stress that lepton colliders are extremely relevant for HNL searches. Not only due to the impressive sensitivities expected at future leptonic colliders such as the FCCee [93], but also because HNL searches at LEP still provide the most relevant limits for some $M_N$ hypotheses.

The great advantage of a leptonic collider is its clean environment, in contrast with the hadronic ones. In the case of LEP, they combined this cleanliness with the huge amount of $Z$ bosons they collected to search for HNLs produced in $Z \to \nu\bar{\nu}$. Moreover, they considered both visibles and semi-invisible HNL decays, such as monojet final states [94]:

$$e^\pm e^- \to \nu\bar{\nu} \to v v q\bar{q},$$

with the $q\bar{q}$ pair clustered as a single jet due to the large HNL boost (efficient for $M_N \lesssim 30$ GeV). For heavier masses, $M_N \in (30, 80)$ GeV, the signature was composed by two jets with or without a charged lepton. Such a search would be very challenging at a hadronic collider, however it has the advantage of being sensitive to all flavours, including the mixing to the $\tau$ lepton, not explored so far by LHC searches. The DELPHI results [94], derived for both long-lived and prompt light HNLs, were not improved (for mixings to $e$ and $\mu$ flavours) by LHC until very recently, and still dominate for some mass ranges (cf. Fig. 2).

Additionally, the L3 collaboration explored the heavy HNL regime by considering their production via the $t$-channel $W$ diagram [95]. This process dominates the heavy HNL production at a $e^+ e^-$ collider running above the $Z$ pole, however it is sensitive only to mixings to electrons.

Table 1: HNL searches at the LHC, classified according to the type of signal searched for. OS/SS are for opposite/same sign for the charges of final leptons and $n_j$ for a number $n$ of final jets.

| Channel | Lepton flavour | Experiment | $\sqrt{s}$ (TeV) | $L$ (fb$^{-1}$) | $M_N$ (GeV) |
|---------|----------------|------------|------------------|----------------|------------|
| Prompt SS dilepton $pp \to \ell_a^\pm N \to \ell_a^\pm \ell_B^\pm + n_j$ | $ee/\mu\mu$ | CMS’12 [87] | 7 | 4.98 | (50, 210) |
| | $\mu\mu$ | CMS’15 [88] | 9 | 19.7 | (40, 500) |
| | $ee/\mu\mu$ | CMS’16 [89] | 8 | 19.7 | (40, 500) |
| | $ee/\mu\mu$ | ATLAS’15 [90] | 8 | 20.3 | (100, 500) |
| | $ee/\mu\mu$ | CMS’18 [69] | 13 | 35.9 | (20, 1600) |
| Prompt OS dilepton $pp \to \ell_a^\pm N \to \ell_a^\pm \ell_B^\pm + n_j$ | $\mu\mu$ | CMS’18 [86] | 7–8 | 3.0 | (5, 50) |
| Prompt trilepton $pp \to \ell_a^\pm N \to \ell_a^\pm \ell_B^\pm \ell_C^\mp + n_j$ | $ee + e\mu/\mu\mu + \mu\mu$ | CMS’18 [91] | 13 | 35.9 | (1, 1200) |
| | $e\mu/\mu\mu$ | CMS’18 [69] | 13 | 36.1 | (5, 50) |
| Displaced trilepton $pp \to \ell_a^\pm N, N \to \ell_B^\pm \ell_C^\mp + n_j$ | $\mu - e\mu/\mu\mu - \mu\mu$ | ATLAS’19 [84] | 13 | 32.9 | (4.5, 10) |
| | 6 combinations of $e, \mu$ | ATLAS’22 [85] | 13 | 139 | (3, 15) |
| | 6 combinations of $e, \mu$ | CMS’22 [86] | 13 | 138 | (1, 20) |
Fig. 2 Summary of direct HNL searches performed at the LHC so far, either by CMS (dashed), ATLAS (dotted) or LHCb (dot-dashed), and group by colors for different kind of searches as given in Table 1. In the upper (lower) panel a single mixing scenario to electrons (muons) is assumed. Shadowed area cover the area excluded by direct searches at LEP. Notice that below 2 GeV and above (approx.) 100 GeV, bounds from meson decays and from non-unitarity of the lepton mixing [92] dominate respectively over current LHC bounds, although we do not show them explicitly for easier reading of the collider results.

The results by L3 still provide the strongest limits for masses between 100 and 200 GeV.

Despite the great effort in the search for HNLs by both LEP and LHC, it is important to analyze their implications for realistic models introducing and motivating the existence of HNLs. A common feature of all these searches is the assumption of a simplified scenario, most of the time consisting on a single HNL mixing to a single lepton flavour, which is not the standard hypothesis one would use from the theory side. To our knowledge, the only exceptions to these simplifications are provided by the recent ATLAS search for long-lived HNLs that also considered a minimal but realistic 2HNL scenario [85], and CMS searches for SS $e\mu$ final states [69,89], although still neglecting the mixing to taus. In the two following sections, we discuss the importance of going beyond these simplified scenarios to explore more realistic scenarios.

3 Beyond the single mixing assumption

As explained in the previous section, most of the LHC analyses are done assuming the existence of just one HNL that mixes to a single flavour, which we referred to as the single mixing scenario. In this section, we consider deviations from this simplified hypothesis and discuss their implications for reinterpreting the LHC bounds summarized in Fig. 2. In particular, we focus only on prompt searches, while the implications for long-lived HNL were discussed in, for instance, Ref. [50].

For simplicity, we still consider the presence of a single HNL (we will discuss deviations from this hypothesis in the next section), however we open the room for generic mixing patterns. Moreover, we will follow a bottom-up approach where the SM is extended by ad-hoc masses for the 3 active neutrinos, as required by oscillation phenomena, and by the presence of the additional HNL $N$. This framework is useful to study, in particular, the collider phenomenology of HNLs without assuming any specific underlying model or mechanism of light neutrino mass and leptonic mixing generation.

In such a framework, the lepton mixing matrix is thus enlarged to a $4 \times 4$ unitary matrix

$$U_\nu = \begin{pmatrix} U_{e3}^{3 \times 3} & U_{eN}^{3 \times 1} \\ U_{\mu3}^{1 \times 3} & U_{\muN}^{1 \times 1} \end{pmatrix},$$

so the would-be-PMNS matrix $U_{\nu\nu}$ is no longer a unitary matrix, a feature which is indeed used to constrain these mod-

4 Notice however that reproducing oscillation data in a given framework may introduce relations between the HNL mass and mixings, shrinking the parameter space we will consider.
els [92], and the fourth column contains the HNL mixings to each flavour:

\[ U_{\nu N}^T = (U_{\nu eN}, U_{\nu \mu N}, U_{\nu \tau N}) . \]

For our discussion, it is interesting to parametrize this column as

\[ U_{\nu N}^T = \sqrt{U^2 (\varepsilon_e, \varepsilon_\mu, \varepsilon_\tau)} , \]

where \( U^2 \) represents the total (squared) mixing of the HNL and the \( \varepsilon_a \) its flavour strengths, with \( |\varepsilon_e|^2 + |\varepsilon_\mu|^2 + |\varepsilon_\tau|^2 = 1 \).

Notice that this framework is precisely the one considered by most LHC analyses, the only difference being that they simplify it by setting the a priori non-relevant mixings to zero. Here, we are interested in knowing how these bounds need to be modified in a generic mixing pattern scenario.

The reason why we expect the bounds to be modified is twofold. The first reason is due to the importance of the HNL decay width, which depends on every mixing \( U_{\alpha N} \), and which plays a major role in the resonant searches (on-shell produced HNL) we are interested in. This means that the final cross sections will depend on all of the mixings, even on those flavours that are not explicitly present in the charged leptons involved in the process. The second reason is that for some channels, considering generic mixings could open new contributing diagrams, which could modify the distributions and thus the efficiencies of the searches as discussed thoroughly in Ref. [58]. Complete expressions for the computation of the total decay width \( \Gamma_N \) can be found for instance in Ref. [96], however for our purposes we parameterize it as

\[ \Gamma_N = |U_{\alpha N}|^2 \Gamma_N^\alpha + |U_{\mu N}|^2 \Gamma_N^\mu + |U_{\tau N}|^2 \Gamma_N^\tau , \]

where \( \Gamma_N^\alpha \) stands for the sum of partial decay widths depending on the mixing \( U_{\alpha N} \), after factorizing the \( |U_{\alpha N}|^2 \) dependence itself. Thus, \( \Gamma_N^\alpha \) are independent of the mixings (at leading order) and depend only on the HNL mass. Moreover, when the HNL is heavy enough so that we can neglect charged lepton masses, we get \( \Gamma_N^\alpha \simeq \Gamma_N^\mu \simeq \Gamma_N^\tau \) and thus

\[ \Gamma_N \propto \sum_{\alpha} |U_{\alpha N}|^2 = U^2 . \]

With this discussion in mind, we can now study how the different processes displayed in Table 1 depend on the HNL mixings. Let us start focusing on the dilepton channels, the most straightforward case as we only need to track the effect of the HNL total decay width.

In the narrow width approximation, the processes with SS and same flavour dileptons can be factorized in the production of the HNL together with a charged lepton, times its subsequent decay to the same lepton plus jets. The first part depends only on the mixing to the flavour of that lepton, however the second one involves all the mixings due to the HNL decay width. More explicitly, we have

\[ \sigma(pp \to \ell_\alpha^+ \ell_\alpha^- + nj) \propto |U_{\alpha N}|^2 \]

\[ \operatorname{BR}(N \to \ell_\alpha^+ j) \propto \frac{|U_{\alpha N}|^4}{\Gamma_N} , \]

or, assuming a heavy enough HNL,

\[ \sigma(pp \to \ell_\alpha^+ N \to \ell_\alpha^+ + nj) \propto U^2 |\varepsilon_\alpha|^4 . \]

Then, we clearly see that those bounds obtained in the single mixing benchmark (\( \varepsilon_\alpha = 1 \)) will be relaxed in a general flavour scenario with a fixed \( U^2 \), since in general we will have \( |\varepsilon_\alpha|^2 \leq 1 \). This is actually the expected behaviour, as switching on other mixings opens for new decay channels, so not every produced HNL will decay to the final state we are searching for.

We can repeat the exercise for the different flavour SS dilepton processes. Obviously, the minimal setup in this case requires to have two non-zero mixings, which leads to two diagrams that in principle interfere. Nevertheless, using the narrow width approximation, we can see that both diagrams cannot resonate at the same time, so we can neglect the interference and add both processes incoherently:

\[ \sigma(pp \to \ell_\alpha^+ \ell_\beta^- + nj) \propto \left( |U_{\alpha N}|^2 \operatorname{BR}(N \to \ell_\beta^+ j) \right) + \left( |U_{\beta N}|^2 \operatorname{BR}(N \to \ell_\alpha^+ j) \right) \]

\[ \propto \frac{|U_{\alpha N}|^4 |U_{\beta N}|^4}{\Gamma_N^2} \propto U^2 |\varepsilon_\alpha|^2 |\varepsilon_\beta|^2 . \]

The case of the trilepton channels can be more involved, mainly because we cannot know the lepton number and the flavour carried by the missing (anti)neutrino. Let us consider first the case of same flavour trileptons. There are two contributing diagrams, one with a neutrino and one with an antineutrino, which we can add incoherently. \(^5\) Then, considering for simplicity a \( W^+ \) Drell–Yan channel, we have

\[ \sigma(pp \to \ell_\alpha^+ \ell_\alpha^- + \nu_\alpha) \propto \left( |U_{\alpha N}|^2 \operatorname{BR}(N \to \ell_\alpha^+ \ell_\alpha^- \nu_\alpha) \right) \]

\[ \propto |U_{\alpha N}|^4 \operatorname{BR}(N \to \ell_\alpha^+ \ell_\alpha^- \nu_\alpha) . \]

Notice that the first contribution is mediated by the Majorana nature of the HNL, while the second one is of Dirac type. In principle, both contributions are identical when integrated over the full phase space. However, different spin-correlations induce different angular distributions, which might translate into different acceptances under a given

\(^5\) If light neutrinos are of Majorana nature, then there is an interference term, which is however negligible as it is proportional to light neutrino masses.
experimental analysis. Still, both channels have the same flavour dependence, so we can write,

\[
\sigma(pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + E_T) \propto |U_{\alpha N}|^2 \\
\text{BR}(N \to \ell^-_\alpha \ell^-_\alpha \nu_\alpha) \propto \frac{|U_{\alpha N}|^4}{\Gamma_N} \propto U^2 |\epsilon_\alpha|^4. \tag{19}
\]

When the trilepton signal involves leptons of two different flavours, we need to consider two subcases: the same-sign same-flavour (SSSF) and opposite-sign same-flavour (OSSF). These channels are trickier because in a generic flavour pattern there are new diagrams not present in the single mixing scenario. For instance, in the case of the SSSF, we need to consider two subcases: the same-flavour (SSSF) and opposite-sign same-flavour (OSSF). We summarize our discussion in Table 2. Here, we assume a Majorana HNL, although a similar table can be easily obtained for Dirac HNL by just switching off the LNV channels we discussed above. This is done in Table 3. Notice how some channels that were designed for Majorana HNLs are different, i.e. \( \alpha \neq \beta \neq \gamma \). For trileptons, we are neglecting the effects of differential distributions, as discussed in the text.

The case of OSSF is similar, although now the roles of Majorana and Dirac HNLs are flipped:

Majorana-like: \( pp \to \ell^+_\alpha N, N \to \ell^-_\beta \ell^-_\beta \nu_\alpha \),

Dirac-like: \( pp \to \ell^+_\alpha N, N \to \ell^-_\beta \ell^-_\beta \nu_\alpha \). \tag{23}

Moreover, since we are working in the prompt HNL regime, we can also have \( pp \to \ell^+_\alpha N, N \to \ell^-_\beta \ell^-_\beta \nu_\alpha \). This means that a proper recasting of this kind of signals would require to compute the efficiencies for all these diagrams. If, for the sake of this discussion, we neglect these effects again, we get:

\[
\sigma(pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\gamma + E_T) \\
\propto \left( |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\gamma \ell^-_\beta \nu_\beta) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\gamma \ell^-_\beta \nu_\beta) \right) \\
\propto \frac{|U_{\alpha N}|^2 + |U_{\beta N}|^2}{\Gamma_N} \propto U^2 |\epsilon_\alpha|^2 (|\epsilon_\beta|^2 + |\epsilon_\gamma|^2). \
\]

Finally, and even if no LHC searches have been performed so far, we can also consider the case with 3 different flavours. Following the same steps, we get:

\[
\sigma(pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\gamma + E_T) \\
\propto \left( |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\gamma \ell^-_\beta \nu_\beta) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\gamma \ell^-_\beta \nu_\beta) \right) \\
\propto U^2 \left\{ |\epsilon_\alpha|^2 \left(1 - |\epsilon_\beta|^2 \right) + |\epsilon_\beta|^2 \left(1 - |\epsilon_\gamma|^2 \right) \right\}. \tag{25}
\]

We summarize our discussion in Table 2. Here, we assume a Majorana HNL, although a similar table can be easily obtained for Dirac HNL by just switching off the LNV channels we discussed above. This is done in Table 3. Notice how some channels that were designed for Majorana HNLs are different, i.e. \( \alpha \neq \beta \neq \gamma \). For trileptons, we are neglecting the effects of differential distributions, as discussed in the text.

Table 2: Summary table for generic flavour dependences of diphoton and trilepton channels at the LHC assuming a single Majorana HNL with generic mixing patterns. Flavour indices are to be understood as

| Process (prompt) | Relevant parameters (Majorana HNL) | Complete dependence |
|------------------|------------------------------------|---------------------|
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + n j \) | \( U^2 |\epsilon_\alpha|^4 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha j j) \) |
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + n j \) | \( U^2 |\epsilon_\alpha|^4 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha j j) + |U_{\beta N}|^2 \text{BR}(N \to \ell^+_\beta j j) \) |
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + E_T \) | \( U^2 |\epsilon_\alpha|^4 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha \ell^-_\alpha \nu_\alpha) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\alpha \ell^-_\beta \nu_\beta) \) |
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + E_T \) | \( U^2 |\epsilon_\alpha|^4 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha \ell^-_\beta \nu_\beta) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\alpha \ell^+_\beta \nu_\beta) \) |
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + E_T \) | \( U^2 |\epsilon_\alpha|^4 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha \ell^-_\beta \nu_\beta) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\alpha \ell^+_\beta \nu_\beta) \) |
| \( pp \to \ell^+_\alpha \ell^-_\alpha \ell^-_\beta + E_T \) | \( U^2 \sum_{i=\alpha,\beta} |\epsilon_i|^2 \) | \( |U_{\alpha N}|^2 \text{BR}(N \to \ell^+_\alpha \ell^-_\alpha \nu_\alpha) + |U_{\alpha N}|^2 \text{BR}(N \to \ell^-_\alpha \ell^-_\beta \nu_\beta) \) |
Table 3 Same as Table 2, but for Dirac HNL. We give the OS dileptons in this case, since the SS are not sensitive to Dirac HNLs

| Process (prompt) | Relevant parameters (Dirac HNL) | Complete dependence |
|------------------|--------------------------------|---------------------|
| $pp \rightarrow \ell_+^\alpha \ell_+^\beta + nj$ | $U^2 |\ell_\alpha|^4$ | $|U_{eN}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ jj)$ |
| $pp \rightarrow \ell_+^\alpha \ell_+^\beta + nj$ | $U^2 |\ell_\beta|^4 |\ell_\beta|^2$ | $|U_{eN}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ jj) + |U_{\beta N}|^2 \text{BR}(N \rightarrow \ell_\beta^+ jj)$ |
| $pp \rightarrow \ell_+^\alpha \ell_+^\alpha + \ell_T$ | $U^2 |\ell_\alpha|^4$ | $|U_{eN}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ \ell_\alpha^+ v_\alpha)$ |
| $pp \rightarrow \ell_+^\alpha \ell_+^\alpha + \ell_T$ | $U^2 |\ell_\alpha|^4 |\ell_\beta|^2$ | $|U_{eN}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ \ell_\alpha^+ v_\alpha) + |U_{\beta N}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ \ell_\beta^+ v_\beta)$ |
| $pp \rightarrow \ell_+^\alpha \ell_+^\alpha + \ell_T$ | $U^2 |\ell_\alpha|^4 (|\ell_\alpha|^2 + |\ell_\beta|^2)$ | $|U_{eN}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ \ell_\alpha^+ v_\alpha) + |U_{\beta N}|^2 \text{BR}(N \rightarrow \ell_\alpha^+ \ell_\beta^+ v_\beta)$ |

Fig. 3 Rescaling of the bounds on $|U_{eN}|^2$ from CMS [69] (solid red line) and LHCb [83] (solid blue line), fixing the active—sterile mixings to the values corresponding to the green points 1 — 4 in the ternary plot on the left. The orange squared point represents the single mixing case

also sensitive to Dirac HNLs in the case of having generic mixing patterns, as it was already discussed in Ref. [58].

These tables are to be compared with the minimal mixing scenario where all of the processes scale as $|U_{eN}|^2$, or $|U_{eN}|^2|U_{\beta N}|^2/(|U_{eN}|^2 + |U_{\beta N}|^2)$ for $\alpha \neq \beta$. However, we see that in general each process is sensitive to a different combinations of mixing strengths. This means that, in order to generalize the bounds to a generic pattern, it is better to set bounds on the quantity on the last columns of Tables 2 and 3, since then we only need to recompute the new BRs for each mixing hypothesis.

While there are some experimental results also presenting (in the single mixing assumption) bounds in the $(M_N, |U_{eN}|^2 \times \text{BR})$ plane, most of the results are given directly in the $(M_N, |U_{eN}|^2)$ one. Translating the latter to the former is straightforward in most of the cases, since the experimental collaborations usually assume a constant\(^6\) BR for channel under study, although not always specifying the precise value they used. Nevertheless, this recasting to $|U_{eN}|^2 \times \text{BR}$ is not possible when the experimental results on $|U_{eN}|^2$ are presented after combining different channels. This is the case for instance for the latest CMS searches for trileptons [91], where they combined channels like $e^\pm e^\pm e^\mp$ and $e^\pm e^\pm \mu^\mp$ (see Table 1). While in the single mixing scenario both channels depend only on $|U_{eN}|^2$, they have a different dependence in the case of a generic mixing scenario (see Table 2), and thus it is not easy to recast the obtained bounds without a dedicated analysis. For this reason, together with the potential efficiency differentials discussed above, we will focus the rest of our discussion only on the dilepton channels.

As a case of study for the use of Table 2, let us consider the CMS [69] and LHCb [83] searches for the LNV dimuon channel $\mu^\pm \mu^\mp$. For a given total active-sterile mixing $U^2$, we can display the full flavour mixing space in a ternary diagram, as in the left panel of Fig. 3. Then, the single mixing scenario constraints by both CMS and LHCb lie in the top corner. As we move along the ternary, we decrease the flavour strength to

\(^6\) This is a well-justified approximation, but still an approximation that could be avoided by setting bounds directly on $|U_{eN}|^2 \times \text{BR}$.
Combination of several bounds in a general mixing pattern for two benchmark points of mass and total mixing $U^2$. Each bound corresponds to the latest SS dilepton searches by CMS (Table 1), which were derived within a single mixing scenario (or assuming $U^\tau N$ for the $e^\pm \mu^\pm$ channel). The white area is still allowed by LHC searches.

As discussed above, the physical reason for the relaxation of these bounds is due to the new HNL decay channels in the generic mixing scenario. On the other hand, this also implies that the experimental searches for the other channels with different flavours might become relevant. In order to show the interplay between different flavour channels, let us consider again a ternary diagram. We can understand it as a subspace of the parameter space with fixed values of $M_N$ and $U^2$, which is dissected in flavour space. Then, searches for dimuon channels will cover the area close to the $\epsilon_\mu = 1$ corner, becoming weaker as we move further away. Equivalently, dielectron searches will cover the ternary from the $\epsilon_e = 1$ corner, while the $e^\pm \mu^\pm$ channel will cover the area in between. This is depicted in Fig. 4 for two benchmark points, one in the light regime ($M_N = 30$ GeV) and one in the heavy regime ($M_N = 300$ GeV).

Figure 4 clearly shows the complementarity of the different dilepton searches, to which we could supplementary add the bounds from trilepton channels if the above mentioned concerns are solved. If the combination of every channel covered all the area of the ternary, we could say that this ($M_N, U^2$) point is excluded no matter which flavour mixing pattern we were considering. This is not the case in any of the two examples in the figure, since the bottom left corner is still allowed by LHC. Notice that this is always the case at present, since there are no LHC searches for HNLs mixing to the tau lepton, so the corner of $\epsilon_\tau = 1$ will always be allowed. Although we already discussed that a single mixing scenario does not seem very natural for a realistic model including HNLs, closing this gap still motivates the need of performing dedicated searches in the tau sector.

On the other hand, it is important to stress that the benchmark points in Fig. 4 are chosen just for illustrative purposes, since they are already excluded by LEP searches or global fit bounds. Indeed, this is actually the case in a large part of the parameter space for the current LHC bounds discussed in Sect. 2. Nevertheless, it is worth emphasizing that LHC sensitivities are expected to improve during the currently ongoing runs, pushing our knowledge about HNLs beyond present limits. Therefore, combining the different LHC channels as we discussed here will become crucial in order to determine whether a heavy neutral lepton with a given mass and mixing is completely excluded or not.

Beyond the single HNL

Should HNL exist in Nature, there is a priori no bound on their number and in general, BSM models involving HNLs do not introduce just one of them. For example, in the standard type-I seesaw model at least two HNL are needed to explain neutrino oscillation data. In this work, in order to explore deviations from the single HNL hypothesis, we extend the framework considered in the previous section to include more
neutral fermions focusing here on the minimal case of SM extension via two HNLs. In this case, the lepton mixing matrix is now a $5 \times 5$ unitary matrix
\[
U_v = \begin{pmatrix}
U_{\nu \nu}^{3 \times 3} & U_{\nu N}^{3 \times 2} \\
U_{N \nu}^{3 \times 3} & U_{NN}^{3 \times 2}
\end{pmatrix},
\]
with the fourth and fifth columns encoding the mixings of both HNLs to the active leptons
\[
U_{\nu N} = \begin{pmatrix}
U_{\nu e N_1} & U_{\nu e N_2} \\
U_{\mu e N_1} & U_{\mu e N_2} \\
U_{\tau e N_1} & U_{\tau e N_2}
\end{pmatrix}.
\]

For the sake of simplicity, we will assume in this section that the two HNLs mix to just a single flavour. In a sense, it can be seen as extending the interpretation of the bounds of Fig. 2 in horizontal to $U_{\nu N}$, to additional columns, while in Sect. 3 we extended them in vertical, to additional flavours. The most general case with several HNLs and arbitrary mixing patterns can then be inferred as a combination of these two discussions.

In the case where only one of these HNLs is within experimental reach or when there are several HNLs but with well-separated mass regimes, our conclusions derived following the single HNL scenario will apply to each of the HNLs. Nevertheless, if two HNLs happen to be close in mass (as motivated by low-scale seesaw models [61,97,98], resonant leptogenesis [99] or ARS leptogenesis [100]), they could lead to interference effects and modify the results and the bounds obtained in the single HNL hypothesis. Moreover, these modifications might affect both LNV and LNC branching ratios, and thus studying their correlation could shed light on the nature of the HNL (see for instance [64] and references therein).

In this section we discuss how these effects could affect the LHC bounds obtained in the single HNL scenario from searches for LNV and LNC channels, and provide a recipe to combine both results (on LNV and LNC searches) in order to bring forth more robust bounds on HNLs parameter space. The recipe is also applicable to the case where HNLs interfere. We mostly focus on the LHCb [83] results for the prompt dimuon channel, since this is the only available analysis addressing both SS and OS dilepton channels (cf. Table 1). More specifically, this search considers light HNLs which are dominantly produced from on-shell DY $W$ bosons in Fig. 1, that is,
\[
pp \rightarrow W^+ \rightarrow \ell_\alpha^+ N_1 \rightarrow \ell_\alpha^+ \ell_\beta^+ q q', \quad pp \rightarrow W^+ \rightarrow \ell_\alpha^+ N_1 \rightarrow \ell_\alpha^+ \ell_\beta^- q q'.
\]

In presence of just a single Majorana HNL, although the angular distributions will be different (see Appendix A for more details), the predictions for the total rates of these two processes are of similar size: equal for channels with $\alpha = \beta$ and twice as large for $\alpha \neq \beta$. The reason for this difference is that the channel with crossed $\ell_\alpha$ and $\ell_\beta$ is also contributing in the case of the LNV process. However, it must be added incoherently to the process since the rate is dominated by on-shell HNLs, fixing the momentum of the first lepton with the 2-body decay kinematics, and thus the interference becomes subdominant. This means that for channels with $\alpha \neq \beta$, the total rate for the LNV process is enhanced by a factor of 2 with respect to the LNC one. On the other hand, in channels with $\alpha = \beta$ there is an additional $1/2$ factor from having two identical particles, and thus we obtain the same total rate for both LNV and LNC.

When assuming the existence of two HNLs, we have two identical contributions to the total amplitude of the processes, one for each $N_i$. The squared amplitude is then given as sum of the individual contributions of each HNL, plus a potential interference between $N_1$ and $N_2$ contributions. For each individual contribution to the amplitudes, we find that they are proportional to $U_{\alpha N_1}^* U_{\beta N_1}$ for the SS process and to $U_{\alpha N_1}^* U_{\beta N_2}$ for the OS one, proving convenient to define
\[
U_{\alpha N_i} = |U_{\alpha N_i}| e^{i \phi_{\alpha i}},
\]
with $\phi_{\alpha i} \in [0, 2\pi]$, $\alpha = e, \mu, \tau$ and $i = 1, 2$. In this way, the interference term resulting from both amplitudes will be proportional to
\[
|U_{\alpha N_1}| |U_{\alpha N_2}| |U_{\beta N_1}| e^{i \phi_{\alpha 1}} |U_{\beta N_2}| e^{i \phi_{\alpha 2}},
\]
where we have defined $\delta \phi^{\pm}$ as follows:
\[
\delta \phi^{\pm} = (\phi_{\alpha 2} - \phi_{\alpha 1}) \pm (\phi_{\beta 2} - \phi_{\beta 1}),
\]
with $+/-$ for the SS/OS channel. Details with the complete analytical amplitude and decay rates, which we have additionally checked with Wizard [101], can be found in Appendix A, both for the case with just one HNL and with two HNLs.

As already stated, when the mass difference of the two HNLs is too large (compared to their decay width), the contribution of each HNL resonate independently and the interference is negligible. In this case, the total rates for the LNV and LNC rates are related as in the case of the single HNL scenario. Therefore, in the following we will assume the scenario where the interference effects can modify considerably the relative predictions between LNV and LNC rates, which corresponds to the case where both HNLs are close in mass. More specifically, we assume that the individual contribution of each HNL is of similar size, $M_{N_1} \simeq M_{N_2} \equiv M_N$. $\Gamma_{N_1} \simeq \Gamma_{N_2} \equiv \Gamma_N$, and also that $|U_{\alpha N_1}| |U_{\beta N_1}| = |U_{\alpha N_2}| |U_{\beta N_2}|$. However, we consider that the two HNL mass splitting $\Delta M_N \equiv M_{N_2} - M_{N_1}$ could be different from zero as long as it is small compared to the decay width $\Gamma_N$. Notice that this kind of scenario appears naturally in low-scale seesaws due to the approximated lepton number
and equivalently for the \(W\) channel driven by the two HNLs in the case of them free, in the spirit of the bottom-up approach described. The only difference is that in these models the phases of the HNLs are fixed to be opposite,\(\phi^+ = \pi\),\(\phi^- = 0\) (the heavy neutrinos forming a pseudo-Dirac neutrino pair), and that \(\Delta M_N\) is somehow related to light neutrino masses \(81,82\), while in the following we will let them free, in the spirit of the bottom-up approach described before.

Under these conditions, we can write the total decay rate driven by the two HNLs in the case of \(W^+\) channel as

\[
\Gamma \left( W^+ \rightarrow \ell^+ e^\mp q\bar{q} \right) \bigg|_{l_1 \& l_2} = 2 K_+ (y, \delta \phi^\pm)
\]

\[
\Gamma \left( W^+ \rightarrow \ell^+ e^\mp q\bar{q} \right) \bigg|_{l_1},
\]

and equivalently for the \(W^-\) channel with \(K_- (y, \delta \phi^\pm)\) . Here, we have factorized the total rate in presence of only one HNL, and defined the modulation function

\[
K_\pm (y, \delta \phi^\pm) = \left( 1 + \cos \delta \phi^\pm \frac{1}{1 + y^2} \mp \sin \delta \phi^\pm \frac{y}{1 + y^2} \right).
\]

(33)

with

\[
y \approx \frac{\Delta M_N}{\Gamma_N}.
\]

(34)

The function \(K_\pm\) codifies the role of the interference. In the limit of \(\Delta M_N \gg \Gamma_N\), the two HNL are too separated in mass, coherence is lost and the total contribution is just twice the single HNL contribution for both LNV and LNC. On the other hand, for \(\Delta M_N < \Gamma_N\), the modulation function can take values from 0 (maximally destructive interference) to 2 (maximally constructive). Thus, we are maximizing the effects of the interference between the two HNLs. Moreover, these effects will be different for LNV and LNC, breaking the equal size prediction in the single HNL scenario.

These modulation function can be used to simply recast the bounds derived by LHCb \(83\) under the assumption of a single HNL. Noticing that LHCb searched only for the \(W^+\) channel and given Eq. (15), to recast these bounds to our scenario with two HNLs, we need to rescale the mixing as

\[
|U_{\mu N}|^2 \rightarrow |U_{\mu N}|^2 \times 2K_+ (y, \delta \phi^\pm),
\]

(35)

with \(\delta \phi^- = 0\) in this channel with \(\alpha = \beta\). Following this modulation, we show in the left panel of Fig. 5 how the LHCb bounds might vary, depending on the values of \(y\) and the relative phases \(\delta \phi^+\), for both LNV and LNC searches, which defines the green and blue bands, respectively. The vertical axis needs to be understood now as the (squared) mixing of each of the HNLs to muons, which in absence of interference \((y \gg 1)\) is just a factor of two stronger with respect to the single HNL scenario. For \(y \ll 1\), however, constructive interference can strengthen the bounds by up to an additional factor of two, while destructive interference could relax it, even avoid it completely in the case of LNV signals. The latter corresponds to the case where \(\delta \phi^+ = \pi\), which is precisely when the two HNLs have opposite phases (thus forming a pseudo-Dirac pair), as required by low-scale seesaws with approximated lepton number conservation. Interestingly, the same choice of parameters that maximizes the destructive interference of LNV channels also maximizes the constructive interference for the LNC ones, making the bounds from the latter stronger.
This interplay prompts us to consider both channels at the same time, using their complementarity to set absolute bounds on the mixings that could not be avoided even with ad-hoc values of the parameters that maximize the interference. We show this in the right panel of Fig. 5 for a particular example of $M_N = 30$ GeV and opposite phases, $\delta \phi^+ = \pi$, as required by low-scale seesaws. When the contributions of the two HNLs is maximally coherent ($y \ll 1$), the LNV searches are avoided at the price of maximizing the LNC bounds. If coherence is lost ($y \gg 1$), then the stronger LNV bounds always dominate over the LNC ones. We see then that the largest possible mixing could be obtained in between the two cases, when the two tendencies cross over, which we can consider as an absolute bound on the mixing that cannot be avoided even with 2 interfering HNLs.

To summarize, searches for LNC and LNV processes are important and complementary, since they can cover areas of the parameter space even in the case where there is some interference (partially) cancelling any of the two channels. This strongly motivates the need of searching for both LNC and LNV channels in parallel, even if the latter is more challenging experimentally, since combining both of them we could set more robust bounds on generic scenarios including more than one heavy neutral lepton.

CP violation

When a quasi-degenerated pair of HNLs is considered, new CP-violating phases are introduced, which can induce differences in the decays of these particles to leptons over antileptons. If HNLs were discovered in processes such as those in Eq. (28), and provided that enough events were collected, one way of measuring this potential CP asymmetry would be by defining the ratio [65]

$$A_{\text{CP}}^\pm = \frac{\text{BR} \left( W^- \rightarrow \ell_\alpha^- \ell_\beta^+ q \bar{q}' \right) - \text{BR} \left( W^+ \rightarrow \ell_\alpha^+ \ell_\beta^- q \bar{q}' \right)}{\text{BR} \left( W^- \rightarrow \ell_\alpha^- \ell_\beta^+ q \bar{q}' \right) + \text{BR} \left( W^+ \rightarrow \ell_\alpha^+ \ell_\beta^- q \bar{q}' \right)} \, .$$

(36)

Using our previous results for the $W$ decays, see Eq. (32), it is straightforward to see that $A_{\text{CP}}^\pm$ takes the simple form

$$A_{\text{CP}}^\pm = \frac{y \sin \delta \phi^\pm}{1 + y^2 + \cos \delta \phi^\pm} \, ,$$

(37)

where $+/-$ denotes again LNV/LNC processes.

As can be seen, this equation does not depend on the HNL masses, but on their mass difference $\Delta M_N$ (through $y = \Delta M_N / \Gamma_N$). It vanishes for the obvious case of $\delta \phi^\pm = 0$, since then there is no difference in both the $W$ decay and its CP equivalent; notice that it is also the case when $y$ is close to zero or too large. This is shown in Fig. 6 where we display the CP asymmetry for different values of $\delta \phi^\pm$.

![Fig. 6 CP violating ratio $A_{\text{CP}}^\pm$, defined in Eq. (37), as function of the relevant ratio $y = \Delta M_N / \Gamma_N$ and for different choices of the relative phases $\delta \phi^\pm$. Red dots were obtained with Whizard as a cross-check of our analytical results.](image)

Notice that a similar computation was performed in Ref. [65] with which our results agree but with some slight differences. On the one hand, we find a discrepancy when applying the narrow width approximation for the interference term (see Appendix A for more details). On the other hand, we performed the computation considering very prompt heavy neutral leptons, so we did not take into account their time evolution and their possible oscillations before decaying. Our approach is appropriate for heavy HNLs, while one should take into account the evolution effect for very light HNLs, thus with longer decay lifetimes, as it was done in Ref. [65]. Finally, we also point out that, in order to measure $A_{\text{CP}}$ at a proton-proton collider, we must take into account that the production rates for $W^+$ and $W^-$ are not the same [65].

5 Conclusions

In this work, we focused on LHC searches of heavy neutral leptons that decay promptly (short-leaved). In most of the searches, which we summarized in Sect. 2, the obtained bounds were derived under the hypothesis of the existence of a single (usually Majorana) HNL that mixes with only one lepton flavour, while most of the BSM scenarios involving new neutral fermions require more than one HNL. Moreover, unless some specific symmetries are present, the mixing pattern in these BSM scenarios is more complex, each HNL mixing in general with all charged leptons, and thus, the bounds derived from negative searches on the HNL parameter space have to be recast before being applied to a generic BSM scenario.

In this study we discussed how to recast the present experimentally obtained bounds on the parameter space, i.e. active-sterile mixings $U_{a N}$, $\alpha = e, \mu, \tau$, versus the mass of the HNL, to the case of generic mixing to all active flavours as
well as to the case with several HNLs. The former was covered in Sect. 3, where we inspected the flavour dependencies of each of the channels searched for by the LHC, and stressed the importance of setting bounds not only on the mixings, but on the relevant combination of $|U_{eN}|^2 \times BR$ (see Tables 2 and 3). Considering the bounds on this combination, we proposed a method to combine the results in flavour space, using the ternary diagrams in Fig. 4 to conclude whether that area of parameter space is fully excluded, regardless of the assumed mixing pattern.

In the case with several heavy neutrino leptons, we focused on the scenario when two HNLs are in the same mass regime, nearly degenerate or possibly forming a pseudo-Dirac neutrino pair, paying special attention to the non-trivial role of the interference between their contributions. To illustrate its importance, we focused on dilepton channels and moreover considered both channels with same and opposite charge of the final leptons, as it was done by LHCb [83]. We showed the complementary of the LNC and LNV searches and the importance of performing both of them in parallel. We stressed that by doing this, we are not only taking into account the two possible nature of a single HNL, Dirac or Majorana, but also covering the case when, for example, two Majorana HNLs exist and however they interfere destructively, suppressing the expected LNV signature.

To summarize, we have discussed the importance of going beyond simplified scenarios such as the single mixing hypothesis. While they are useful for simplifying experimental analyses, they are not directly applicable to BSM models introducing HNLs. Unfortunately, recasting the bounds of each experimental analysis to a given BSM scenario can be a tedious task. In this work, we have proposed an alternative way of presenting the bounds on the parameter space of the HNLs, which under some approximations can be both directly constrained by experimental analyses and also easily recast to a generic BSM scenario.

Acknowledgements PE is grateful for the hospitality of the Pôle de Physique Théorique of the ICLAB (Orsay) during the first stages of this project. The project has received support from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 860881 (HIDdev network). The work of PE is supported by the Spanish grants PID2020-113775GB-I00 (AEI / 10.13039/501100011033), CIPROM/2021/054 (Generalitat Valenciana) and by the FPI grant PRE2018-084599. XM work. The work of PE is supported by the Spanish grants PID2020-113775GB-I00 (AEI / 10.13039/501100011033), CIPROM/2021/054 (Generalitat Valenciana) and by the FPI grant PRE2018-084599. XM also acknowledges partial support from the Spanish Research Agency (Agencia Estatal de Investigación) through the Grant IFT Centro de Excelencia Severo Ochoa No CEX2020-001007-S and Grant PID2019-108892RB-I00 funded by MCIN/AEI/10.13093/501100011033.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors’ comment: No data generated in this research.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Funded by SCOAP3. SCOAP3 supports the goals of the International Year of Basic Sciences for Sustainable Development.

Appendix A: Decay process in presence of 2 HNL

In this appendix we collect the relevant details for the computation of the $W$ boson decay into a HNL, followed by its semileptonic decay. This is the relevant process for the searches performed by LHCb [83] and that we discussed in Sect. 4 in the scenario with two heavy neutrino leptons.

A.1 Same sign leptons: $W^+ \rightarrow \ell^+_\alpha \ell^+_\beta q' q$  \textcolor{red}{\textbf{2123}}

We start with the decay rate for the process $W^+ \rightarrow \ell^+_\alpha \ell^+_\beta q' q$ mediated by two HNLs $N_{1,2}$ almost degenerate in mass, similar to that shown in Fig. 1, but without the initial quarks. The amplitude reads

$$\mathcal{M}^+ = \sum_{i=1,2} \frac{g^3}{\sqrt{2}} U^*_\alpha N_i U^*_\beta N_i \frac{M_{N_i}}{M_W} \frac{P_L \ell^+_\alpha q' \gamma_\nu P_L \ell^+_\beta q}{p_N^2 - M_{N_i}^2 + i\Gamma_{N_i} M_{N_i}},$$

with $p_N = p_W - p_{\ell_\alpha}$. The channel with crossed $\ell_\alpha$ and $\ell_\beta$ gives rise to the same amplitude, but it must be added incoherently to our process since the rate is dominated by on-shell $N_i$, and thus the momentum of the first lepton is fixed by the 2-body decay kinematics. Therefore the two processes do not interfere and we can neglect for the moment the crossed channel. The only modification results in a factor of 2 in the rate.

Defining $U_{\alpha N_i} = |U_{\alpha N_i}| e^{i\phi_{\alpha i}}$, $\delta \phi^+ = (\phi_{\mu 2} - \phi_{\alpha 1}) + (\phi_{\mu 2} - \phi_{\beta 1})$ and using the narrow width approximation (NWA), the squared matrix element becomes

$$|\mathcal{M}^+|^2 = \left[ \frac{g^3}{\sqrt{2} M_W^2} \right]^2 \pi (p_{\ell_\beta} \cdot p_q) (2 E_{\ell_\alpha} E_{q'} + p_{\ell_\alpha} \cdot p_q) \times \left[ |U_{\alpha N_1}|^2 |U_{\beta N_1}|^2 \frac{M_{N_1}}{\Gamma_{N_1}} \delta (p_N^2 - M_{N_1}^2) \right] + \left[ |U_{\alpha N_2}|^2 |U_{\beta N_2}|^2 \frac{M_{N_2}}{\Gamma_{N_2}} \delta (p_N^2 - M_{N_2}^2) \right] + \left[ 2 |U_{\alpha N_1}| |U_{\alpha N_2}| |U_{\beta N_1}| |U_{\beta N_2}| M_{N_1} M_{N_2} \times \delta (p_N^2 - M_{N_1}^2) \delta (p_N^2 - M_{N_2}^2) \right]$$
where $\Delta M^2_N \equiv M^2_{N_1} - M^2_{N_2}$. Notice that, for the interference term, we used the NWA as follows:

\[
\frac{1}{(p^2_N - M^2_{N_1} + i\Gamma_{N_1}M_{N_1})(p^2_N - M^2_{N_2} - i\Gamma_{N_2}M_{N_2})}
= \frac{\pi}{(\Delta M^2_N)^2 + (\Gamma_{N_1}M_{N_1} + \Gamma_{N_2}M_{N_2})^2} \left[ \delta(p^2_N - M^2_{N_1}) + \delta(p^2_N - M^2_{N_2}) + i\pi \Delta M^2_N \left( \frac{\Delta M^2_N}{(\Delta M^2_N)^2 + (\Gamma_{N_1}M_{N_1} + \Gamma_{N_2}M_{N_2})^2} \delta(p^2_N - M^2_{N_1}) \right) \right. \\
\left. + \delta(p^2_N - M^2_{N_2}) \right],
\]

which differs from the expression in Ref. [65], as discussed in Sect. 4. Assuming $M_{N_1} \approx M_{N_2} \equiv M_N$, $\Gamma_{N_1} \approx \Gamma_{N_2} \equiv \Gamma_N$ and $\Delta M_N \equiv M_{N_1} - M_{N_2} \neq 0$, and considering that $|U_{\alpha N_1}| |U_{\beta N_1}| = |U_{\alpha N_2}| |U_{\beta N_2}| \approx |U_{\alpha N_1}| |U_{\beta N_1}|$, we get

\[
|M^+|^2 \approx \left[ \frac{g^3}{2\sqrt{2}M_W^2} \right]^2 \pi (p_{\ell\beta} \cdot p_q) (2E_{\alpha}E_{q'} + p_{\alpha} \cdot p_q) \delta(p^2_N - M^2_{N_1}) \frac{M_N}{\Gamma_N |U_{\alpha N_1}|^2 |U_{\beta N_1}|^2}
\]

\[
|M^-|^2 \approx \left( \frac{g^3}{2\sqrt{2}M_W^2} \right)^2 \left[ \frac{K_{11}}{(p^2_N - M^2_{N_1})^2 + \Gamma^2_{N_1}M^2_{N_1}} |U_{\alpha N_1}|^2 |U_{\beta N_1}|^2 \times \right.
\left. + \frac{K_{22}}{(p^2_N - M^2_{N_2})^2 + \Gamma^2_{N_2}M^2_{N_2}} |U_{\alpha N_2}|^2 |U_{\beta N_2}|^2 \right]
\]

\[
\times 2 \left\{ 1 + 2 \left[ \frac{2\cos\delta\phi^+}{(\Delta M^2_N)^2 + 4\Gamma^2_{N_1}M^2_{N_1}} \right] \right. \\
\left. - \sin\delta\phi^+ \frac{M_N\Gamma_{N_1}\Delta M^2_{N_1}}{(\Delta M^2_N)^2 + 4\Gamma^2_{N_1}M^2_{N_1}} \right\},
\]

with the $K_{ij}$ factors defined as

\[
K_{ij} = \frac{-16}{M_W} \left( p_{\beta} \cdot p_{q'} \right) \left[ M_W \left( p_{N_1} \cdot p_{N_2} \right) \left( p_{\alpha} \cdot p_{q} + 2E_{\alpha}E_{q} \right) \right]
\]

We observe that the squared amplitude in the case of only one sterile neutrino factorizes out. Integrating over the phase space and after factorizing the decay width in the single HNL framework, it is straightforward to obtain

\[
\Gamma \left( W^+ \rightarrow \ell^+_\alpha \ell^+_\beta q' q \right) = 2 \left( 1 + \cos\delta\phi^+ \frac{1}{1 + y} \right) \\
- \sin\delta\phi^+ \frac{y}{1 + y} \Gamma \left( W^+ \rightarrow \ell^+_\alpha \ell^+_\beta q' q \right)|_{\delta = 0},
\]

where we have defined

\[
y \equiv \frac{\Delta M^2_{N_1}}{2M_N\Gamma_N} \simeq \frac{M^2_{N_2} - M^2_{N_1}}{(M_{N_1} + M_{N_2})\Gamma_N} = \frac{\Delta M^2_N}{\Gamma_N}.
\]

Notice that we obtain the same result of Ref. [63]. This is because we are using the NWA, which corresponds to consider on-shell HNL, leading to the same result after integrating over $t$, the time evolution of the intermediate HNL, from 0 to $\infty$. In fact the factor $1/\Gamma$ coming from the NWA is equivalent to the factor $\int_0^{\infty} |e^{-\Gamma t/2}|^2 = 1/\Gamma$, and the interference part of Eq. (41) coincides with the finding of Ref. [63].
\[ -2 E_\alpha \left[ (p_{N_1} \cdot p_W) (p_{N_2} \cdot p_q) \right] + (p_{N_1} \cdot p_W) (p_{N_2} \cdot p_q) - M_W \left[ (p_{N_1} \cdot p_q) (p_{N_2} \cdot p_{N_1}) \right] + (p_{N_1} \cdot p_q) (p_{N_2} \cdot p_{N_1}) \right]. \]

Assuming as before that \( M_{N_1} \approx M_{N_2} \equiv M_N \) and \( \Gamma_{N_1} \approx \Gamma_{N_2} \equiv \Gamma_N \), then \( K_{11} = K_{22} = K_{12} \equiv K \). Also, considering that \( U_{\alpha N_1} \left| U_{\beta N_1} \right| = \left| U_{\alpha N_2} \right| \left| U_{\beta N_2} \right| = \left| U_{\alpha N} \right| \left| U_{\beta N} \right| \), we can simplify the expression notably

\[ |M^2| = \left( \frac{g^3}{2 \sqrt{2} M_W^3} \right)^2 |U_{\alpha N}|^2 |U_{\beta N}|^2 \frac{K}{\Gamma_N M_N^2} \frac{1}{\Gamma_N M_N^2} \frac{1}{\Gamma_N M_N^2} \frac{1}{\Gamma_N M_N^2} \sqrt{\frac{\pi}{2}} \frac{\delta (p_N^2 - M_N^2)}{\left( p_N^2 - M_N^2 + \Gamma_N M_N^2 \right)^2} \left\{ 1 + \text{Re} \left[ 2 \Gamma_N M_N^2 \frac{2 \Gamma_N M_N^2 + i \Delta M_N^2}{\Delta M_N^2} e^{i \delta \phi} \right] \right\}. \]

Finally, it is straightforward to obtain the total decay width in terms of the decay width mediated by just one HNL,

\[ \Gamma \left( W^+ \rightarrow \ell^+_\alpha \ell^-_\beta q \bar{q}' \right) = 2 \left( 1 + \cos \delta \phi - \frac{1}{1 + y^2} \right) - \sin \delta \phi \frac{y}{1 + y^2} \times \Gamma \left( W^+ \rightarrow \ell^+_\alpha \ell^-_\beta q \bar{q}' \right), \]

with \( y \) defined in Eq. (43).
81. M. Drewes, J. Klarić, P. Klose, On lepton number violation

82. E. Fernández-Martínez, X. Marcano, D. Naredo, To appear soon

83. LHCb collaboration, R. Aaij et al., Search for heavy neutral leptons in

84. ATLAS collaboration, G. Aad et al., Search for heavy neutral

85. ATLAS collaboration, M. Aaboud et al., Search for heavy neutral

87. CMS collaboration, S. Chatrchyan et al., Search for heavy Majorana neutrinos in e+e− annihilations at LEP. Phys. Rev. Lett. B 517, 67–74 (2001). arXiv:hep-ex/0107014

88. CMS collaboration, V. Khachatryan et al., Search for heavy Majorana neutrinos in c±c±+ jets and e±μ±+ jets events in proton–proton collisions at √s = 8 TeV. JHEP 04, 169 (2016). arXiv:1603.02248

89. CMS collaboration, V. Khachatryan et al., Search for heavy Majorana neutrinos in c±c±+ jets and e±μ±+ jets events in proton–proton collisions at √s = 8 TeV. JHEP 04, 169 (2016). arXiv:1603.02248

90. ATLAS collaboration, G. Aad et al., Search for heavy Majorana neutrinos with the ATLAS detector in pp collisions at √s = 8 TeV. JHEP 07, 162 (2015). arXiv:1506.06020

91. CMS collaboration, A.M. Sirunyan et al., Search for heavy neutral leptons in events with three charged leptons in proton–proton collisions at √s = 13 TeV. Phys. Rev. Lett. 120, 221801 (2018). arXiv:1802.02965

92. E. Fernandez-Martinez, J. Hernandez-Garcia, J. Lopez-Pavon, Global constraints on heavy neutrino mixing. JHEP 08, 033 (2016). arXiv:1605.08774

93. FCC-ee study Team collaboration, A. Blondel, E. Graverini, N. Serra, M. Shaposhnikov, Search for Heavy Right Handed Neutrinos at the FCC-ee. Nucl. Part. Phys. Proc. 273-275, 1883–1890 (2016). arXiv:1411.5230

94. DELPHI collaboration, P. Abreu et al., Search for neutral heavy leptons produced in Z decays. Z. Phys. C 74, 57–71 (1997)

95. L3 collaboration, P. Achard et al., Search for heavy isosinglet neutrino in e+e− annihilation at LEP. Phys. Lett. B 517, 67–74 (2001). arXiv:hep-ex/0107014

96. K. Bondarenko, A. Boyarsky, D. Gorbunov, O. Ruchayskiy, Phenomenology of GeV-scale heavy neutral leptons. JHEP 11, 032 (2018). arXiv:1805.08567

97. A. Ibarra, E. Molinaro, S.T. Petcov, TeV scale see-saw mechanisms of neutrino mass generation, the Majorana nature of the heavy singlet neutrinos and ββ0ν-decay. JHEP 09, 108 (2010). arXiv:1007.2378

98. A. Abada, M. Lucente, Looking for the minimal inverse seesaw realisation. Nucl. Phys. B 885, 651–678 (2014). arXiv:1401.1507

99. A. Pilaftsis, T.E.J. Underwood, Resonant leptogenesis. Nucl. Phys. B 692, 303–345 (2004). arXiv:hep-ph/0309342

100. E.K. Akhmedov, V.A. Rubakov, A.Y. Smirnov, Baryogenesis via neutrino oscillations. Phys. Rev. Lett. 88, 1359–1362 (1998). arXiv:hep-ph/9803255

101. S. Bräss, W. Kilian, O. Mittag, P. Stienemeier, Precision Monte Carlo simulations with WHIZARD. CERN Yellow Report Monogr 3, 205–210 (2020)

102. V. Shatovenko, R. Mertig, F. Orellana, New developments in FeynCalc 9.0. Comput. Phys. Commun. 207, 432–444 (2016). arXiv:1601.01167