JHEP

Published for SISSA by Springer

Received: December 19, 2019
Accepted: March 9, 2020
Published: April 1, 2020

Dirac vs. Majorana HNLs (and their oscillations) at SHiP

J.-L. Tastet[a] **and I. Timiryasov**[b]

[a] *Discovery Center, Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, DK-2010, Copenhagen, Denmark*

[b] *Institute of Physics, Laboratory for Particle Physics and Cosmology, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland*

E-mail: jeanloup@nbi.ku.dk, Inar.Timiryasov@epfl.ch

Abstract: SHiP is a proposed high-intensity beam dump experiment set to operate at the CERN SPS. It is expected to have an unprecedented sensitivity to a variety of models containing feebly interacting particles, such as Heavy Neutral Leptons (HNLs). Two HNLs or more could successfully explain the observed neutrino masses through the seesaw mechanism. If, in addition, they are quasi-degenerate, they could be responsible for the baryon asymmetry of the Universe. Depending on their mass splitting, HNLs can have very different phenomenologies: they can behave as Majorana fermions — with lepton number violating (LNV) signatures, such as same-sign dilepton decays — or as Dirac fermions with only lepton number conserving (LNC) signatures. In this work, we quantitatively demonstrate that LNV processes can be distinguished from LNC ones at SHiP, using only the angular distribution of the HNL decay products. Accounting for spin correlations in the simulation and using boosted decision trees for discrimination, we show that SHiP will be able to distinguish Majorana-like and Dirac-like HNLs in a significant fraction of the currently unconstrained parameter space. If the mass splitting is of order 10^{-6} eV, SHiP could even be capable of resolving HNL oscillations, thus providing a direct measurement of the mass splitting. This analysis highlights the potential of SHiP to not only search for feebly interacting particles, but also perform model selection.

Keywords: Beyond Standard Model, Neutrino Physics

ArXiv ePrint: 1912.05520

https://doi.org/10.1007/JHEP04(2020)005

JHEP04(2020)005

Contents

JHEP04(2020)005

1 Introduction

The experimentally observed non-vanishing neutrino mass differences are among a few firmly established deviations from the Standard Model (SM) predictions. An economic way of generating the light neutrino masses is to introduce heavy singlet fermions with Majorana mass terms into the model [1–6]. The masses of the active neutrinos in this extension of the SM are determined by the type-I seesaw formula and at least two singlet fermions are needed to accommodate the two observed mass differences of light neutrinos. A consequence of this mechanism is the presence of heavy Majorana fermions which mix with active neutrinos. The mass scale of these Majorana fermions — Heavy Neutral Leptons (HNLs) — is not fixed. It can be below the electroweak scale,[1] like in the νMSM [9, 10], where two HNLs are responsible for the light neutrino masses and generating the Baryon Asymmetry of the Universe (BAU) via CP-violating oscillations during their production.

From the FIP (feebly interacting particles) search point of view, HNLs with masses below that of a B meson are the most accessible in the foreseeable future [11]. There is a vast program to search for HNLs at *intensity frontier* experiments, either LHC-based, such as MATHUSLA [12–14], FASER [15–17], CODEX-b [18, 19], AL3X [20, 21] and ANUBIS [22], or at beam-dump facilities, such as DUNE [23–25] (using the near detector), NA62^{++} [26, 27] (in dump mode) and SHiP [28–30]. Comparative studies of the *exclusion* limits expected from these experiments have been performed in refs. [31–34]. If a candidate HNL signal were to be observed, the latter three experiments would be sensitive to both its mass and mixing angles.

SHiP is a proposed beam-dump experiment (represented in figure 1) set to operate at the CERN SPS. It will use an intense, 400 GeV proton beam from the SPS, dumped on a thick target in order to produce a large number of heavy hadrons, which subsequently decay into Standard Model (SM) or feebly-interacting particles. SHiP is designed to provide a background-free environment to look for the decays of these heavy FIPs. To this end, a hadron absorber located right after the target absorbs most SM particles. It is followed by an active muon shield which deflects the muons away from the experimental cavern. The main detector consists of a decay volume — evacuated in order to reduce the neutrino background, and surrounded by vetos — with a tracker and a calorimeter located at its far end, enabling it to reconstruct the decay event.

In order to generate the light neutrino masses via the seesaw mechanism, HNLs must be Majorana fermions, which violate the total lepton number. However, if the mass splitting is small enough, they can pair to form a coherent superposition of two quasi-degenerate Majorana fermions, which behaves almost like a Dirac fermion. Such a combination is dubbed "*quasi-Dirac pair*". In this case, the mixing angles can exceed the naive seesaw limit $U^2 \approx m_\nu/M_N$ [35–37], where m_ν and M_N are respectively the mass scales of light neutrinos and HNLs. This is possible because a quasi-Dirac fermion approximately conserves the total lepton number, hence protecting the light neutrino masses. For instance, the νMSM [9, 10] contains such a quasi-Dirac pair if one requires the mass degeneracy which is needed

[1] An argument in favour of the low-scale seesaw comes from the measured values of the Higgs and top masses. HNLs with masses below the electroweak scale are not destabilising the Higgs mass [7, 8].

JHEP04(2020)005

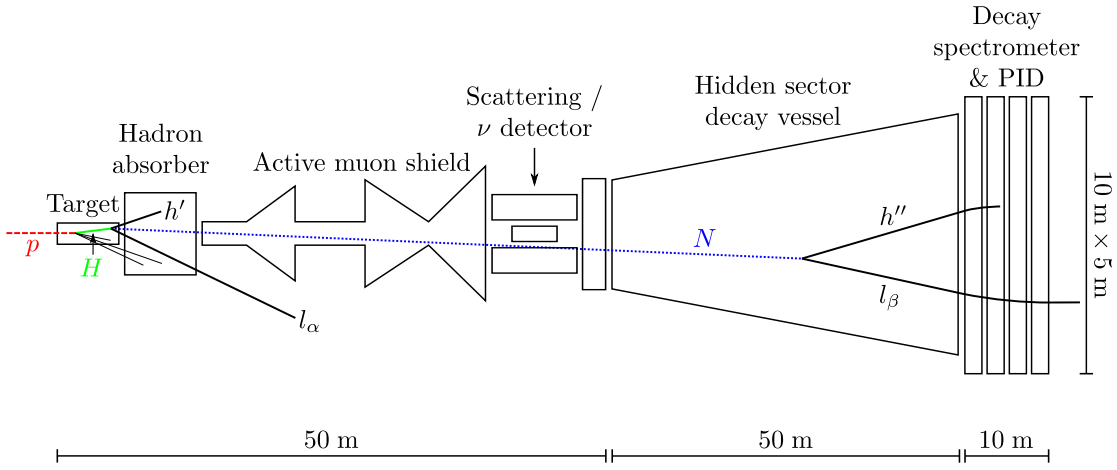


Figure 1. Sketch of the SHiP experiment, with the decay chain $H \to h' l_\alpha (N \to l_\beta h'')$.

for baryogenesis [10, 38] and especially for late-time leptogenesis [39]. Quasi-Dirac pairs also naturally appear in some models of neutrino mass generation, such as the inverse seesaw [40, 41] and the linear seesaw [42, 43]. This near degeneracy of the HNL masses leads to coherent HNL oscillations. In the νMSM, these oscillations in the early Universe are responsible for baryogenesis.

For sufficiently light ($\lesssim 10\,\text{GeV}$) HNLs like the ones accessible at SHiP, LNV may be experimentally observable even when they form a quasi-Dirac pair [44, 45]. We can distinguish three cases,[2] depending on the scale of the oscillation phase $\delta M \tau$, where δM is the mass splitting of the quasi-Dirac pair and τ the typical proper time probed:

1. *Dirac-like HNL:* one Dirac HNL or a quasi-Dirac pair with an oscillation period exceeding the HNL lifetime or detector size ($\delta M \tau \ll 2\pi$).[3] Only LNC processes can be observed.

2. *Majorana-like HNL:* one Majorana HNL or a quasi-Dirac pair with a lifetime and detector size exceeding the oscillation period ($\delta M \tau \gg 2\pi$). Both LNC and LNV processes can be observed, with equal integrated rates (see section 2.2).

3. *Manifestly quasi-Dirac HNLs:* an interesting case occurs when the oscillation period is comparable to the HNL lifetime or to the size of the detector[4] ($\delta M \tau \sim 2\pi$): the experiment may then be sensitive to the coherent oscillations of HNLs.

If HNLs were to be observed at SHiP, the detection or non-observation of lepton number violation and HNL oscillations would allow constraining models and their parameters.

[2] To be generic, we have included the more exotic cases of a single Dirac or Majorana HNL. The limits presented below are for a quasi-Dirac pair, which only differs from those in the number of events produced.

[3] As pointed out in ref. [45], for most experiments, this possibility might be technically unnatural due to the very small mass splitting needed to satisfy the inequality.

[4] Interestingly, the mass difference needed to generate DM in the νMSM, as found in ref. [39], is exactly in this borderline range.

JHEP04(2020)005

The most relevant LNV process at SHiP is the well-studied *same-sign dilepton decay*: $H \to [h']l_\alpha^+(N \to h''l_\beta^+)$, where H, h' and h'' are hadrons (with h' possibly missing), and l_α^+, l_β^+, $\alpha, \beta = e, \mu, \tau$ are charged leptons of potentially different generations. Due to suppressed background, this type of signature is a smoking gun for HNLs in accelerator searches. However, at beam-dump experiments, the heavy hadron decay which produces the HNL takes place inside the target, and therefore the charge of the primary lepton l_α cannot be observed. Naively, it seems that the information about the HNL production is lost, since the charge of the secondary lepton l_β, by itself, is not enough to tell apart LNC and LNV processes. As we shall see in this paper, it turns out that the HNL decay products nevertheless carry important information. Namely, their distribution is different for LNC and LNV processes. Not only does this allow distinguishing Majorana-like from Dirac-like HNLs given sufficiently many events, but the knowledge of these distributions can also be used to *resolve* HNL oscillations and directly measure the mass splitting.

Estimating these two distributions is complicated by the presence of a variety of two- and three-body production channels. In addition, the parent hadrons are produced with a finite spectrum. As we shall see in section 3.3, this smears the distributions, making them look more similar. Therefore, in order to assess whether SHiP will be able to discriminate between Majorana- and Dirac-like HNLs, an accurate treatment of all production channels, including spin correlations, is required. This is accomplished using a Monte-Carlo simulation.

The angular distribution of HNL decay products has been studied in a collider setting for decays which are not fully reconstructible [46–48] (such as trilepton decays), as well as for beam-dump experiments [49, 50]. Our analysis improves on the latter by not relying on HNLs being produced as helicity eigenstates, by handling a larger class of production channels, by considering the full phase-space distribution of the HNL decay products (instead of just their energy) and by producing a concrete sensitivity estimate using a realistic geometry and heavy meson spectrum for SHiP.

This paper is organized as follows. In section 2, we review the Standard Model extended with HNLs, and discuss lepton number violation and coherent HNL oscillations. In section 3, we analyze the different signatures of LNC and LNV processes at the SHiP experiment. In section 4, we propose a strategy to detect LNV and reconstruct HNL oscillations. Finally, in section 5, we present the sensitivity of SHiP to LNV achieved through this method, as well as a possible signature of HNL oscillations. Technical details about the simulation and the statistical analysis are respectively provided in appendices A and B.

2 Model

2.1 Heavy Neutral Leptons

We consider the Standard Model extended with $\mathcal{N}$ HNLs N_I, which are spin-$\frac{1}{2}$ SM singlets with Majorana masses M_I, and new Yukawa couplings $Y^\nu_{\alpha I}$, with $\alpha = e, \mu, \tau$ the lepton flavor index. Using the conventions from ref. [51]:

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + \frac{i}{2} N_I^\dagger (\bar{\sigma} \cdot \partial) N_I - (Y^\nu_{\alpha I})^* (\phi \cdot L_\alpha) N_I - \frac{M_I}{2} N_I N_I + \text{h.c.} \tag{2.1}$$

JHEP04(2020)005

After electroweak symmetry breaking, the Yukawa interaction generates a Dirac mass term $(m_D)_{\alpha I} = \frac{v}{\sqrt{2}}(Y^\nu_{\alpha I})^*$, resulting in a non-diagonal, symmetric Dirac-Majorana mass term for neutrinos [52]:

$$M_{\mathrm{DM}} = -\frac{1}{2}\begin{pmatrix}\nu^T & N^T\end{pmatrix}\begin{pmatrix}0 & m_D^T \\ m_D & M_M\end{pmatrix}\begin{pmatrix}\nu \\ N\end{pmatrix} + \mathrm{h.c.} \tag{2.2}$$

where $M_M = \mathrm{diag}\,(M_I \dots)$. Using a unitary transformation of the fields (Takagi factorization [53]), the mass matrix can be brought to a diagonal form:

$$\nu_\alpha = U_{\alpha i} n_i \quad \text{and} \quad N_I = U_{Ii} n_i \tag{2.3}$$

$$M_{\mathrm{DM}} = -\frac{m_i}{2}(n_i n_i + n_i^\dagger n_i^\dagger) \tag{2.4}$$

In the limit $|M_M| \gg |m_D|$, we can use an approximate block factorization, leading to the mass eigenstates $n_i \cong \nu_i, N_I$ mixing with the flavor fields as:

$$\nu_\alpha \cong U^{\mathrm{PMNS}}_{\alpha i}\nu_i + \Theta_{\alpha I} N_I \tag{2.5}$$

$$\Theta_{\alpha I} \cong M_I^{-1}(m_D)_{\alpha I} \tag{2.6}$$

and the following mass sub-matrices:

$$m_{\alpha\beta} \cong -\sum_I \frac{(m_D)_{\alpha I}(m_D)_{\beta I}}{M_I} \cong -\sum_I M_I \Theta_{\alpha I}\Theta_{\beta I} \tag{2.7}$$

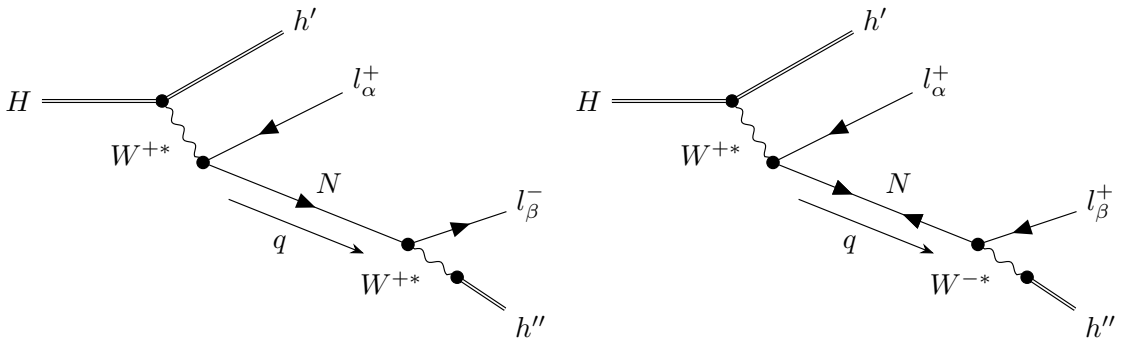
$$m_{IJ} \cong M_I \delta_{IJ} \tag{2.8}$$

The choice of the mass scale M_M and Yukawa couplings $Y^\nu_{\alpha I}$ is not uniquely dictated by low-energy neutrino observables, and should be fixed otherwise.

The Standard Model features an accidental symmetry — lepton number — which, at tree level, is conserved for massless or Dirac neutrinos, but is violated by the Majorana mass term of HNLs. Charged leptons and neutrinos have lepton number +1, while charged anti-leptons and anti-neutrinos have lepton number −1. If lepton number is conserved (LNC), then the only allowed Feynman diagrams are those with a conserved flow of lepton number (represented by the arrow on the fermion lines of leptons), like the opposite-sign dilepton decay of a heavy hadron shown in figure 2(a). On the other hand, in the presence of lepton number violating (LNV) operators, processes like the same-sign dilepton decay shown in figure 2(b) become possible. Lepton number violation can also manifest itself in neutral-current processes or in neutrinoless double-β decay. Whether such LNV transitions actually happen depends on the specific model.

In the past decade, a class of low-scale seesaw models have risen in popularity, such as the νMSM [10], not least because of their falsifiability at existing or proposed experiments. In these models, M_M is postulated to be below the electroweak scale. The seesaw formula (2.7) requires at least 2 HNLs to explain the two observed mass differences. If their parameters are arbitrary, then the smallness of the light neutrino masses is achieved through small Yukawa couplings of order $Y^\nu \sim \frac{1}{v}\sqrt{|m_\nu||M_M|}$, leading to squared mixing angles $|\Theta|^2 \sim |m_\nu|/|M_M|$. For a typical HNL with $M_M \sim 1\,\mathrm{GeV}$, this gives $|\Theta|^2 \sim 10^{-11}$, a number that is too small to be probed at any current or proposed experiment.

JHEP04(2020)005



(a) LNC (b) LNV

Figure 2. Lepton number conserving and violating decay chains for $H \to h' l_\alpha (N \to l_\beta h'')$.

However, multiple HNLs can have mixing angles well above the seesaw limit, yet at the same time produce the correct neutrino masses in a technically natural way, if a certain symmetry is imposed on their Yukawa couplings. If we consider for simplicity $\mathcal{N} = 2$ nearly degenerate HNLs $N_{1,2}$, their mixing angles should be related by $\Theta_{\alpha 2} \approx \pm i \Theta_{\alpha 1}$ [35, 36]. Such HNLs form a quasi-Dirac fermion, which approximately conserves the total lepton number. This implies that the usual searches for naive LNV effects (e.g. same-sign dilepton decays), may return null results even if HNLs *are there*.

Below we discuss an important consequence of the approximate nature of this lepton number conservation: HNL oscillations, and how quasi-Dirac HNLs can phenomenologically behave either as Majorana or Dirac HNLs depending on their mass splitting δM and the length scale probed at the experiment.

2.2 Coherent oscillations of Heavy Neutral Leptons

The SHiP experiment is only sensitive to GeV-scale HNLs, with mixing angles significantly above the seesaw limit [30]. Therefore it can only probe the quasi-Dirac regime described above. Apart from a small mass splitting $\delta M \ll M$, the two HNLs are otherwise identical. Since these two HNLs cannot be distinguished in any realistic experiment, they both mediate the same processes and each contribute to the total transition amplitude, resulting in interference. Only the initial and final-state particles, which strongly interact with the environment, are measured in the quantum mechanical sense. In order to accurately describe processes involving multiple HNLs, it is therefore necessary to consider them as intermediate particles within a larger process consisting of the HNL production, propagation and decay, and only square the overall transition amplitude between the observed, external particles. This can be formulated rigorously within the framework of the external wave packet model [54, 55] (see also [56–59] and references therein for recent reviews). Let us note in passing that this description automatically takes care of spin correlations between the particles taking part in the HNL production and decay.

In what follows, we consider a typical reconstructible decay chain at SHiP, as depicted in figure 2. We will postpone the detailed discussion of this process to section 3. A heavy hadron H produced in the target decays at space-time coordinates x_P into an HNL N_I, a

JHEP04(2020)005

charged lepton l_α (the *primary* lepton), and an optional hadron h'. If the HNL is sufficiently long-lived, it can propagate a macroscopic distance before decaying at x_D into a charged lepton l_β (the *secondary* lepton) and a hadron h''.

The slightly different masses of the HNLs mediating the process lead to different dispersion relations $q_I^2 = M_I^2$. As a consequence, the space-time-dependent phase $e^{-iq_I\cdot(x_D-x_P)}$ acquired by the HNL between its production and decay will differ slightly for each mass eigenstate. When squaring the amplitude in order to obtain the differential decay rate, the interference terms between the partial amplitudes coming from different mass eigenstates will therefore feature a space-time-dependent modulation: HNL oscillations. The external wave packet model allows one to unambiguously establish the expression for the oscillation phase and check that the entire process remains coherent in all experimentally relevant situations.

The present paper does not aim to be a detailed study of HNL oscillations, which have already been covered in various settings and limits in the literature [10, 44, 48, 60–65]. Therefore, we will only quote the main result. Let $\mathrm{d}\hat{\Gamma}^{\pm\pm}_{\alpha\beta}$ be the differential rate for the above-described process $H \to [h']l^\pm_\alpha(N \to l^\pm_\beta h'')$ mediated by a single Majorana HNL N, in the (unphysical) limit of a unit mixing angle between the HNL and the active flavor α at its production vertex, with flavor β at its decay vertex, and without the absorptive part. The coherent differential rate $\mathrm{d}\Gamma^{\pm\pm}_{\alpha\beta}(\tau)$ in the presence of $\mathcal{N}$ nearly degenerate HNLs mediating the process, as a function of the proper time $\tau = \sqrt{(x_D - x_P)^2}$ between the HNL production and decay vertex, is then:

$$\mathrm{d}\Gamma^{\pm\pm}_{\alpha\beta}(\tau) = \left| \sum_{I=1}^{\mathcal{N}} \Theta^\pm_{\alpha I}\Theta^\pm_{\beta I} e^{-iM_I\tau - \frac{\Gamma_I}{2}\tau} \right|^2 \mathrm{d}\hat{\Gamma}^{\pm\pm}_{\alpha\beta} \tag{2.9}$$

where M_I is the (Majorana) mass of the I-th heavy mass eigenstate, Γ_I its total width, and we have used the shorthand notation $\Theta^+ \overset{\text{def}}{=} \Theta^*$ and $\Theta^- \overset{\text{def}}{=} \Theta$.

In the case of $\mathcal{N} = 2$ HNLs forming a quasi-Dirac pair, i.e. $M_1 = M - \frac{\delta M}{2}$, $M_2 = M + \frac{\delta M}{2}$, $\Theta_{\alpha 2} \cong \pm i\Theta_{\alpha 1}$ and $\Gamma_1 \cong \Gamma_2 \overset{\text{def}}{=} \Gamma$, the coherent differential rate reduces to:

$$\mathrm{d}\Gamma^{\pm\pm}_{\alpha\beta}(\tau) \cong 2\left|\Theta_{\alpha 1}\right|^2\left|\Theta_{\beta 1}\right|^2\left(1 \pm \cos\left(\delta M\tau\right)\right)e^{-\Gamma\tau}\mathrm{d}\hat{\Gamma}^{\pm\pm}_{\alpha\beta} \tag{2.10}$$

where the $+$ sign is for lepton number conserving processes ($\mathrm{d}\Gamma^{+-}_{\alpha\beta}$ and $\mathrm{d}\Gamma^{-+}_{\alpha\beta}$), and the $-$ sign for lepton number violating ones ($\mathrm{d}\Gamma^{++}_{\alpha\beta}$ and $\mathrm{d}\Gamma^{--}_{\alpha\beta}$). Notice how in the quasi-Dirac limit, the oscillation pattern does not explicitly depend on the lepton flavors α and β, but only on whether the process is LNC or LNV. If δM vanishes exactly, HNLs form a Dirac fermion and LNV effects are completely absent. Recently, CP-violating HNL oscillations have attracted some interest [66–69]. However, here we can see that CP-violation is suppressed in the quasi-Dirac limit.

Throughout this paper, we will focus on the case where $\Gamma\tau \ll 1$, which is the most relevant for SHiP, and drop the exponentially decaying factor. Analysing formula (2.10), we see that there are three regimes of interest, depending on the mass splitting δM and proper time scale τ probed at the experiment:

JHEP04(2020)005

- If $\delta M\tau \ll 2\pi$, the HNL pair is observed before the onset of oscillations, and it behaves like a single Dirac HNL, i.e. we cannot observe lepton number violation.
- If $\delta M\tau \gg 2\pi$, fast oscillations are averaged out, and the HNL pair behaves like a single Majorana HNL, with equal integrated decay rates for LNC and LNV channels.[5]
- If $\delta M\tau \sim 2\pi$, oscillations must be accounted for. If it is possible to experimentally reconstruct, for each selected event, the proper time τ between the production and decay vertex of the HNL, then oscillations can be resolved, i.e. the τ-differential event rates for LNC/LNV will show a periodic modulation according to eq. (2.10).

At SHiP, the proper time scale τ is about 2 m for sufficiently long-lived HNLs. It corresponds to the average time between the production and decay of an *observed* HNL, in its rest frame. Therefore, the critical mass splitting separating the three regimes — near which oscillations are resolvable — is about 10^{-6} eV.

3 Probing lepton number violation at SHiP

Many collider searches for Majorana HNLs [70–73] are sensitive to lepton number violation through the charges of the leptons produced at the HNL production and decay vertex. Indeed, due to the chiral nature of the weak interaction, they unambiguously tell the chiral projection through which the HNL interacts at a given vertex. In theory, a same-sign dilepton decay (either prompt or displaced) would thus provide clear evidence for lepton number violation (although, in practice, significant standard model backgrounds exist for prompt decays).

At SHiP, similar numbers of mesons and anti-mesons are expected to be produced.[6] This leads to similar numbers of HNLs being produced along with positively and negatively charged primary leptons. Consequently, the secondary lepton charge contains very little information as to whether the process is LNC or LNV. To lift this degeneracy, it becomes necessary to look at new observables.

Luckily, the HNL lepton number is not the only quantum number conserved by the weak interaction. The HNL also carries spin $\frac{1}{2}$, and the total angular momentum is always conserved. When the HNL is produced, its spin is correlated (opposite if H and h' are pseudoscalar) with that of the primary lepton. Due to chiral suppression, the spin of the primary lepton is itself correlated with its lepton number (see for example the left part of figure 3). This suggests that by looking at the angular distribution of the secondary particles — which may be observable — we should be able to obtain information about the primary interaction, and thus whether the process was LNC or LNV (see the right part

[5] In the rest frame of a single on-shell, Majorana HNL, the only "memory" of the production process is the HNL spin. To perform the phase-space integration for the HNL decay, one can always choose a frame where the HNL is at rest and with a fixed spin projection, hence resulting in the same integrated rates for LNC and LNV processes.

[6] Unless cascade production significantly alters the results from ref. [74]. The charm spectrum will be measured at SHiP prior to data taking [75]. Asymmetries, if present, can only improve the classification accuracy, since the secondary lepton charge would then carry *some* information.

JHEP04(2020)005

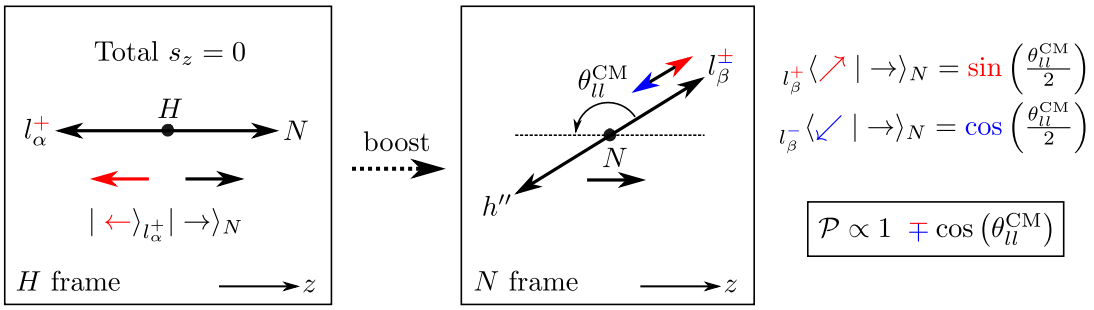


Figure 3. This sketch explains the origin of the different angular correlations for LNC and LNV processes. For simplicity, here we consider two-body primary and secondary decays involving only pseudoscalar mesons, and the masses of the charged leptons and of h'' are neglected. For definiteness, the charge of the primary lepton — which is produced inside the target and thus inaccessible — is also fixed to +. Since the HNL is a Majorana fermion, the secondary lepton l_β can have either charge. However, due to angular momentum conservation, the lepton l_α^+ and the HNL N are produced with opposite spin projections in the rest frame of the heavy meson H. Because of chiral suppression (which is more effective for light fermions), the charge of the primary lepton is correlated with its spin (e.g. in the massless limit, l_α^+ has helicity $+\frac{1}{2}$) and hence with the HNL spin. For the same reason, the angular distribution of the decay products of the resulting HNL spin eigenstate (which is unaffected by a boost along the quantization axis) will therefore depend on the secondary lepton charge. The very same formula for the probability $\mathcal{P}$ also holds for CP-conjugated channels, with the + sign for LNC and the − sign for LNV. The general case (massive, with two- or three-body primary decay) is discussed in section 3.2.

of figure 3). This realization was the starting point of the present work. More generally, we expect LNC and LNV decay chains to have different kinematics due to their different Lorentz structures, potentially allowing us to distinguish them without directly observing the primary decay.

In section 3.1, we describe the relevant HNL production and decay channels at SHiP; in section 3.2, we quantitatively compare the angular distributions for LNC and LNV processes, and in section 3.3 we discuss how this affects the observable momenta in a beam-dump setting.

3.1 HNL production and decay at SHiP

At SHiP, most HNLs are produced in heavy meson decays through flavor-changing charged currents, as discussed in ref. [76]. In addition, for the present analysis, we will only consider fully reconstructible HNL decays such as $N \to l_\beta^\mp \pi^\pm$, producing only charged particles which are sufficiently long-lived to be detected by the tracking station located at the end of the decay vessel. Those are also mediated by the charged-current interaction.

Without losing generality, we can therefore consider the generic lepton number conserving and violating processes $H \to [h']l_\alpha(N \to l_\beta h'')$ represented in figures 2(a) and 2(b), respectively, as well as their CP-conjugates. H denotes a heavy hadron (typically a $D_{[s]}$ or $B_{[c]}$ meson at SHiP), h' and h'' are hadrons (with h' missing for two-body primary decays), and $l_\alpha^\pm$ and $l_\beta^\pm$ are respectively the primary and secondary leptons.

JHEP04(2020)005

Since the heavy hadron H is typically short-lived, the primary decay takes place inside the target and cannot be observed. If the HNL is sufficiently long-lived (we will assume this to be the case throughout this paper), it can propagate a macroscopic distance before decaying, and leave a very displaced vertex inside the SHiP decay vessel. For the selected decay channels $N \to l_\beta^\mp \pi^\pm$, this secondary vertex can be fully reconstructed.

In the present study, we will restrict ourselves to HNL masses between the K and D_s thresholds. Masses below the K threshold have already been heavily constrained [11], while above the D_s mass, HNLs are mainly produced in B meson decays, whose spectrum cannot be directly measured at the beam dump, making our analysis more sensitive to modeling errors.

3.2 Angular correlations in LNC and LNV decay chains

In order to study the angular correlations between all final-state particles, spin correlations between the primary and secondary decay must be accounted for. Those result from the non-observation of the HNL spin, which leads to interference between the two spin eigenstates N_s, $s = \pm\frac{1}{2}$ (similarly to how the non-observation of its precise mass allows for flavor oscillations). To compute the overall transition amplitude, we can therefore use the same trick as for oscillations, i.e. treat the primary and secondary decays as a single process.

To simplify the calculations, in this section we will focus on the case of a single Majorana HNL, which mediates both LNC and LNV decay chains with equal rates, and we will omit the absorptive part of the amplitude (i.e. we will study $\mathrm{d}\hat{\Gamma}^{\pm\pm}_{\alpha\beta}$ instead of $\mathrm{d}\Gamma^{\pm\pm}_{\alpha\beta}(\tau)$). We do not lose generality in doing so, because the effect of multiple nearly degenerate HNLs and their finite lifetime can be factored out, and subsequently recovered, using eqs. (2.9) and (2.10). To keep the notation light, we will from now on drop the HNL index $I = 1$.

Since we are only concerned with long-lived HNLs, which are produced on their mass shell and have well separated, localized production and decay vertices, the momentum q of the HNL is practically fixed, which allows factorizing the transition amplitude as:

$$\mathcal{A}\left(H \to h' l_\alpha l_\beta h''\right)\Big|_{N \text{ long-lived}} \propto \sum_{s=\pm\frac{1}{2}} \mathcal{A}\left(H \to h' l_\alpha N_s(q)\right) \mathcal{A}\left(N_s(q) \to l_\beta h''\right) \tag{3.1}$$

where we have omitted the complex phase $e^{-iq\cdot(x_D - x_P)}$ resulting from the HNL propagation, which is unimportant in the case of one HNL. The sub-amplitudes for the primary and secondary polarized decays are then straightforward to compute using the usual Feynman rules with two-component spinors [51].

Consider now the LNC and LNV processes $H \to [h']l_\alpha(N \to l_\beta h'')$ where H, h', h'' are pseudoscalar mesons and h' may be missing. They are respectively represented in figures 2(a) and 2(b), with the arrows denoting the flow of lepton number. Their CP-conjugates have been omitted, since in the absence of oscillations (as is the case for the incoherent width), CP is conserved. As can be seen in figure 5, the primary decays $H \to [h']l_\alpha N$ with h' a pseudoscalar meson or missing indeed produce the majority of HNLs with masses $\gtrsim 0.7\,\mathrm{GeV}$ and below the D_s mass.[7] Let $J^h_{W\mu}$ be the hadronic charge-lowering

[7] Below $M_N \approx 0.7\,\mathrm{GeV}$, a non-negligible fraction of HNLs is produced along with a *vector* meson. In this case, we expect the angular correlations to reverse compared to the pseudoscalar case.

JHEP04(2020)005

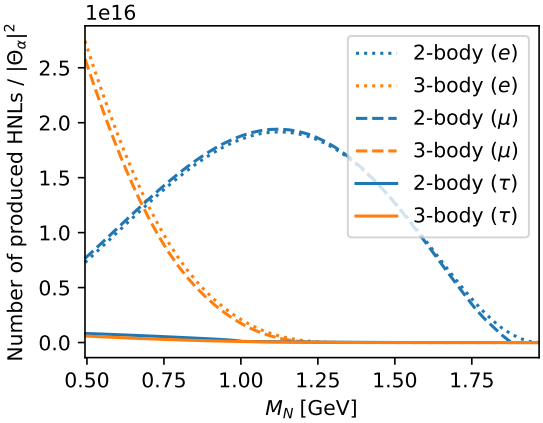


Figure 4. Number of HNLs produced at SHiP as a function of the primary decay multiplicity, for a coupling to one flavor.

current, $j^-_{1\mu} = \langle h'|J^h_{W\mu}|H\rangle$ and $j^\mp_{2\mu} = \langle h''|J^{h(\dagger)}_{W\mu}|0\rangle$ the hadronic matrix elements, $p_{\alpha,\beta}$ the charged lepton momenta, and q the HNL momentum. If the primary decay is purely leptonic, then $|h'\rangle = |0\rangle$. Since SHiP cannot directly measure the spin or helicity of the particles detected, we sum incoherently over all possible spin configurations of final state particles. The spin-summed, squared amplitudes are then, in the Fermi approximation:

$$\overline{\left|\mathcal{A}_{\text{LNC}}(H \to h' l^+_\alpha l^-_\beta h'')\right|^2} = \frac{|\Theta_\alpha|^2 |\Theta_\beta|^2}{v^8} \operatorname{tr}\left(P_R \not{p}_\alpha \not{j}^*_1 \not{q} \not{j}^*_2 \not{p}_\beta \not{j}_2 \not{q} \not{j}_1\right) \tag{3.2}$$

$$\overline{\left|\mathcal{A}_{\text{LNV}}(H \to h' l^+_\alpha l^+_\beta h'')\right|^2} = \frac{|\Theta_\alpha|^2 |\Theta_\beta|^2}{v^8} M_N^2 \operatorname{tr}\left(P_R \not{p}_\alpha \not{j}^*_1 \not{j}^*_2 \not{p}_\beta \not{j}_2 \not{j}_1\right) \tag{3.3}$$

where we have omitted the $^\pm$ for brevity if they can be inferred from context, $\Theta_{\alpha,\beta}$ are the mixing angles, and $v = \langle|\phi|\rangle \approx 246\,\text{GeV}$ is the vacuum expectation value of the Higgs field. These results are consistent with the polarized decay rates from ref. [25], but generalize to the case where the primary decay produces a superposition of HNL helicity eigenstates. The above two expressions differ in the trace, therefore we generically expect them to produce different momentum distributions for LNC and LNV processes. However, in their current form, this difference is not manifest. To understand it, it is interesting to consider the special case where the production process is a two-body decay. As can be seen in figures 4 and 5, it is actually the main production channel for HNLs with masses $\gtrsim 1\,\text{GeV}$ and below the D_s mass.

When both the production and decay process are two-body decays, the hadronic matrix elements are $j^\mu_1 = -iV_{UD} f_H p^\mu_H$ and $j^\mu_2 = +iV_{U'D'} f_{h''} p^\mu_{h''}$, where V_{UD} denotes the relevant CKM matrix element and f_h is the meson decay constant. Neglecting the masses of the final state particles, which give $\mathcal{O}\left(\frac{m^2_{\alpha,\beta,h''}}{M^2_{H,N}}\right)$ corrections, the traces from eqs. (3.2) and (3.3), respectively for LNC and LNV processes, simplify to:

$$\operatorname{tr}\left(P_R \not{p}_\alpha \not{j}^*_1 \not{q} \not{j}^*_2 \not{p}_\beta \not{j}_2 \not{q} \not{j}_1\right) \cong |V_{UD}|^2 |V_{U'D'}|^2 f_H^2 f_{h''}^2 \cdot M_N^6 \left(M_H^2 - M_N^2 - s_{ll}\right) \tag{3.4}$$

$$M_N^2 \operatorname{tr}\left(P_R \not{p}_\alpha \not{j}^*_1 \not{j}^*_2 \not{p}_\beta \not{j}_2 \not{j}_1\right) \cong |V_{UD}|^2 |V_{U'D'}|^2 f_H^2 f_{h''}^2 \cdot M_N^6 s_{ll} \tag{3.5}$$

JHEP04(2020)005

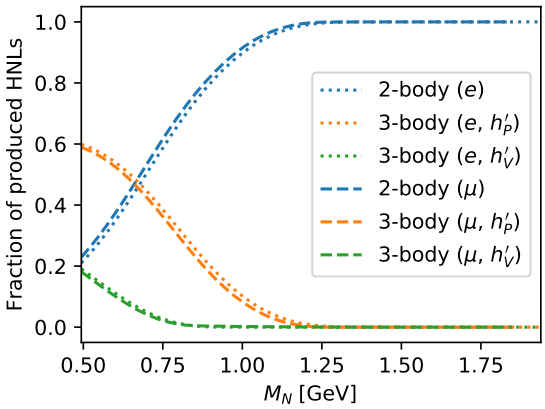


Figure 5. Fraction of HNLs produced at SHiP as a function of the primary decay multiplicity and spin of the outgoing meson, for a coupling to one flavor.

where $s_{ll} \overset{\text{def}}{=} (p_\alpha + p_\beta)^2$ is the invariant dilepton mass. Note the linear and opposite dependences of the LNC and LNV spin-summed squared amplitudes on s_{ll}. To understand their origin, it is enlightening to reexpress s_{ll} in the rest frame of the HNL, in terms of the angle $\theta_{ll}^{\text{CM}} = \angle(\mathbf{p}_\alpha^{\text{CM}}, \mathbf{p}_\beta^{\text{CM}})$ between the two lepton momenta. Still in the massless limit, we find:

$$s_{ll} = \frac{M_H^2 - M_N^2}{2}\left(1 - \cos\left(\theta_{ll}^{\text{CM}}\right)\right) \tag{3.6}$$

Therefore,

$$\overline{\left|\mathcal{A}_{\text{LNC}}\right|^2} \propto 1 + \cos\left(\theta_{ll}^{\text{CM}}\right) \tag{3.7}$$

$$\overline{\left|\mathcal{A}_{\text{LNV}}\right|^2} \propto 1 - \cos\left(\theta_{ll}^{\text{CM}}\right) \tag{3.8}$$

We observe that opposite-sign leptons (LNC) tend to be produced in the same direction, and same-sign leptons (LNV) in opposite directions. As explained in figure 3, this is a consequence of the chirality of the weak interaction and the conservation of the total angular momentum. In the absence of any other dynamics, spin projections lead to the characteristic angular dependence in $\cos\left(\frac{\theta_{ll}^{\text{CM}}}{2}\right)$ and $\sin\left(\frac{\theta_{ll}^{\text{CM}}}{2}\right)$ of the transition amplitude, respectively for LNC and LNV. Eqs. (3.7) and (3.8) then directly follow from squaring the amplitude.

In the massive case, the finite masses of the decay products can result in helicity flips, and in the three-body case, the QCD matrix elements lead to non-trivial correlations between the momenta of the primary decay products. These effects complicate the correlations between the various momenta. Nevertheless, they can be accounted for when *sampling* events. To this end, we have implemented the full matrix elements from eqs. (3.2) and (3.3) in our Monte-Carlo simulation, as discussed in appendix A.4.

3.3 Angular distribution in the laboratory frame

At SHiP, the invariant mass s_{ll} (or angle θ_{ll}^{CM}) cannot be reconstructed. This is because neither the heavy hadron momentum nor the momenta of its decay products (other than

JHEP04(2020)005

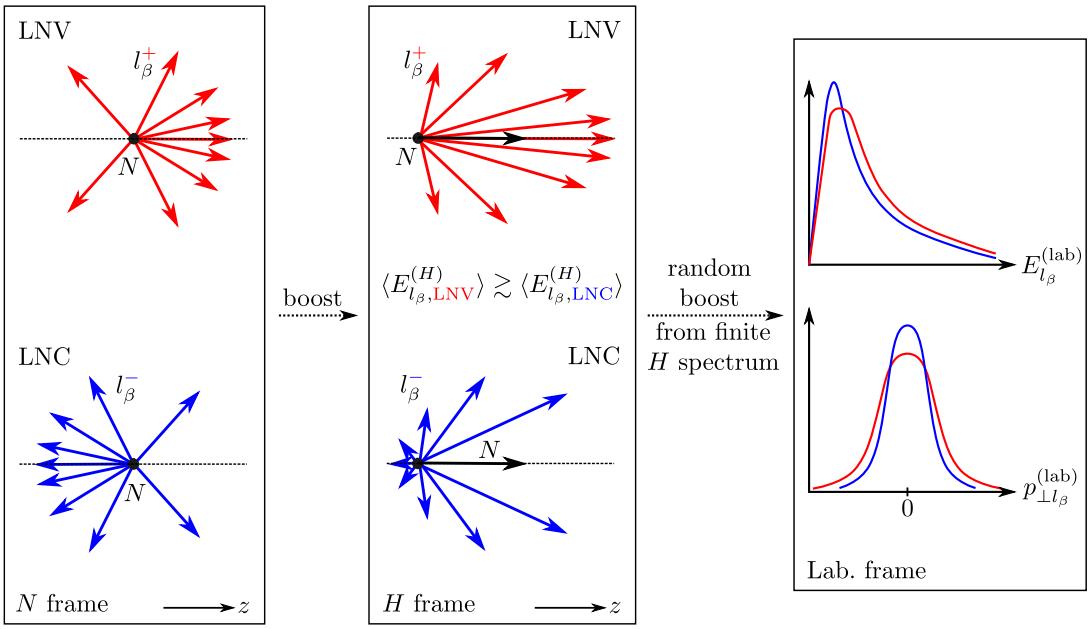


Figure 6. This sketch shows how the different distributions of l_β in the HNL rest frame for LNC vs. LNV processes affect the corresponding distributions in the rest frame of the heavy hadron H and in the laboratory frame. The various momenta shown for l_β represent multiple realizations of the decay. In the H frame, LNV processes typically result in larger momenta for l_β than LNC ones. In the laboratory frame, this effect partly survives the averaging over the heavy hadron spectrum and manifests itself as a broadening of the distribution of the secondary lepton momentum p_β.

the HNL) can be determined. Indeed, the heavy hadrons producing the HNLs do not have a monochromatic spectrum, and the primary decay cannot be observed since it takes place inside the target. One can then reasonably wonder if some difference between the LNC and LNV distributions subsists when looking only at the (observable) secondary decay products, in the laboratory frame, or if it is washed out.

To start answering this question, it is instructive to go back to the simplified case discussed in section 3.2, where the HNL is produced and decays through two-body processes involving pseudoscalar mesons. In the HNL rest frame, we obtained the following correlation: for LNV processes, the direction of the secondary lepton momentum is positively correlated with the boost direction (denoted by z on figures 3 and 6) from the heavy meson rest frame to the HNL rest frame; while for LNC processes it is anti-correlated. This is depicted in the left panel of figure 6. Furthermore, in two-body decays, the magnitudes of all momenta in the rest frame of the parent particle are fixed by four-momentum conservation, and depend only on the particle masses. Consequently, in the heavy meson rest frame, the magnitude of the secondary lepton momentum will on average be larger for LNV processes compared to LNC ones. This argument is still valid for three-body decays involving pseudoscalar mesons. A non-trivial asymmetry thus subsists in the heavy meson rest frame (see the middle panel of figure 6).

As a final step, the momenta must be boosted back to the laboratory frame. Since the heavy hadron momentum is not fixed, this has the potential to wash out the correlations.

JHEP04(2020)005

At SHiP, heavy mesons have a large momentum spread along the beam axis ($\mathcal{O}(10\,\mathrm{GeV})$, much larger than the yield of the meson decay), and a significantly smaller one ($\mathcal{O}(1\,\mathrm{GeV})$) in the transverse direction (see appendix A.3). The asymmetry between the LNC and LNV distributions is therefore more likely to be visible in the transverse plane than along the beam axis. For it to be significant, the HNL kinetic energy in the heavy hadron rest frame should be similar to or exceed the transverse momentum spread of the hadron spectrum. As a result, we expect the p_T spectrum of the secondary lepton l_β to be broader for LNV processes than for LNC ones (see the right panel of figure 6), provided that both of them are broader than the irreducible p_T spread of the heavy meson spectrum.

Alternatively, one could try to approximate the angle $\theta_{ll}^{\mathrm{CM}}$ in the HNL rest frame. If the heavy hadron momentum is fixed, this can be done exactly, and results in the maximal classification accuracy allowed by spin projections (e.g. $a = 3/4$ in the two-body, massless case). It is then equivalent to measuring the (observable) momentum p^{CM} of the secondary lepton l_β in the HNL rest frame. However, when the heavy hadron has a finite spectrum, the boost direction from its rest frame to the HNL rest frame is not fixed any more. This partially decorrelates $\theta_{ll}^{\mathrm{CM}}$ and p^{CM}, hence reducing the discriminating power of the latter.

As we shall see in section 4.2, the features discussed above can indeed be used to discriminate between LNC and LNV processes (see for example figure 7). More generally, any difference — in the laboratory frame — between the distributions of the visible decay products of LNC and LNV processes opens up the possibility of measuring their relative rates, given sufficiently many events. Although discriminating between these two classes of events would be very challenging analytically, this problem is well suited to multivariate analysis.

Further complications arise, however, due to HNLs being produced from a mix of various two- and three-body decays, and because of the geometrical acceptance of the experiment, which alters the distribution of visible particles. Generating a training set which faithfully reproduces the angular correlations discussed above while including these effects is therefore best done using a Monte-Carlo simulation. In the next section, we discuss the simulation used to generate the training set (section 4.1), then how we use it to train a binary classifier (section 4.2), and finally how we use the classifier output in order to perform model selection (section 4.3) and reconstruct HNL oscillations (section 4.4). In section 4.5, we discuss the applicability of the method presented here to other proposed experiments.

4 Simulation and analysis

4.1 Simulation

In order to accurately estimate the distribution of the momenta of the HNL decay products, we have devised a simple Monte-Carlo simulation, which generates the primary and secondary decays at once, using the matrix elements presented in section 3.2. The first step is to generate D mesons with a realistic spectrum. Generating these spectra from simulation would be a difficult undertaking, so instead we chose to use experimental data collected by the LEBC-EHS collaboration [74], at the CERN SPS running at 400 GeV with a hydrogen target. We then randomly select a production and decay channel according to the relative abundances of charmed mesons from ref. [29] and the branching fractions from ref. [76].

JHEP04(2020)005

Feature(s)	Description
`Q12`	Charge of the secondary lepton l_β
`E1`, `p1x`, `p1y`, `p1z`	Reconstructed HNL momentum $p_N = p_{l_\beta} + p_\pi$ (lab frame)
`E2`, `p2x`, `p2y`, `p2z`	Secondary lepton momentum p_{l_β} (lab frame)
`E3`, `p3x`, `p3y`, `p3z`	Secondary pion momentum p_π (lab frame)
`pCMx`, `pCMy`, `pCMz`	Secondary lepton momentum p_{CM} (HNL frame)
`xD`, `yD`, `zD`	Decay vertex (lab frame)

Table 1. The 19 features recorded for each event.

Finally, we generate the momenta of both the primary and secondary decay products at once. This is done by first sampling all the momenta according to phase-space, independently for each decay, and finally performing rejection sampling on these momenta using the matrix element for the combined process. As a last step, we simulate the geometrical acceptance by requiring the HNL to decay within the hidden sector decay vessel, into two long-lived, charged particles which both intersect the tracking station. In order to account for the (small) probability of the HNL decaying inside the fiducial volume, each event is weighted by $P_{\mathrm{decay}}(\tau) = \Gamma e^{-\Gamma\tau}$, where τ is the proper time between the HNL production and decay. Throughout this paper, we assume the particle identification to be perfectly efficient, which should be a reasonably good approximation at SHiP [77]. The simulation is described in details in appendix A.

4.2 LNC/LNV classification

For a given choice of relative squared mixing angles $|\Theta_\alpha|^2$ (which are supposed to be known by the time LNV is studied at SHiP), we generate a dataset for a range of HNL masses between the K and D_s thresholds. For each HNL mass, we sample $3 \cdot 10^6$ events with uniform weights, and keep only those passing the acceptance cuts. The HNL is simulated as a single Majorana particle, which ensures that the dataset contains equal numbers of LNC and LNV events, and is also balanced with respect to the primary and secondary lepton charges.

Each event is labelled with a boolean flag set to false for LNC and true for LNV, using the MC truth. The only observable quantities come from the HNL decay in the vacuum vessel. They are: the momenta and charges of the lepton $l_\beta^\pm$ and pion $\pi^\mp$, and the decay vertex x_D. Of these quantities, we record a total of 19 primary or derived features. Their definitions can be found in table 1, and some typical distributions are presented, as an example, in figure 7, for both LNC and LNV processes. Finally, from each dataset, we set aside 30% of events for testing and 20% for validation, leaving us with 50% of events for training the classifier.

For each dataset, we train a binary classifier to discriminate between LNC and LNV decay chains. For this study, we use the LightGBM [78] decision tree boosting algorithm, through the Python interface to the reference implementation [79]. In order to perform simple classification, we choose the `binary` objective. The training is discussed in more

JHEP04(2020)005

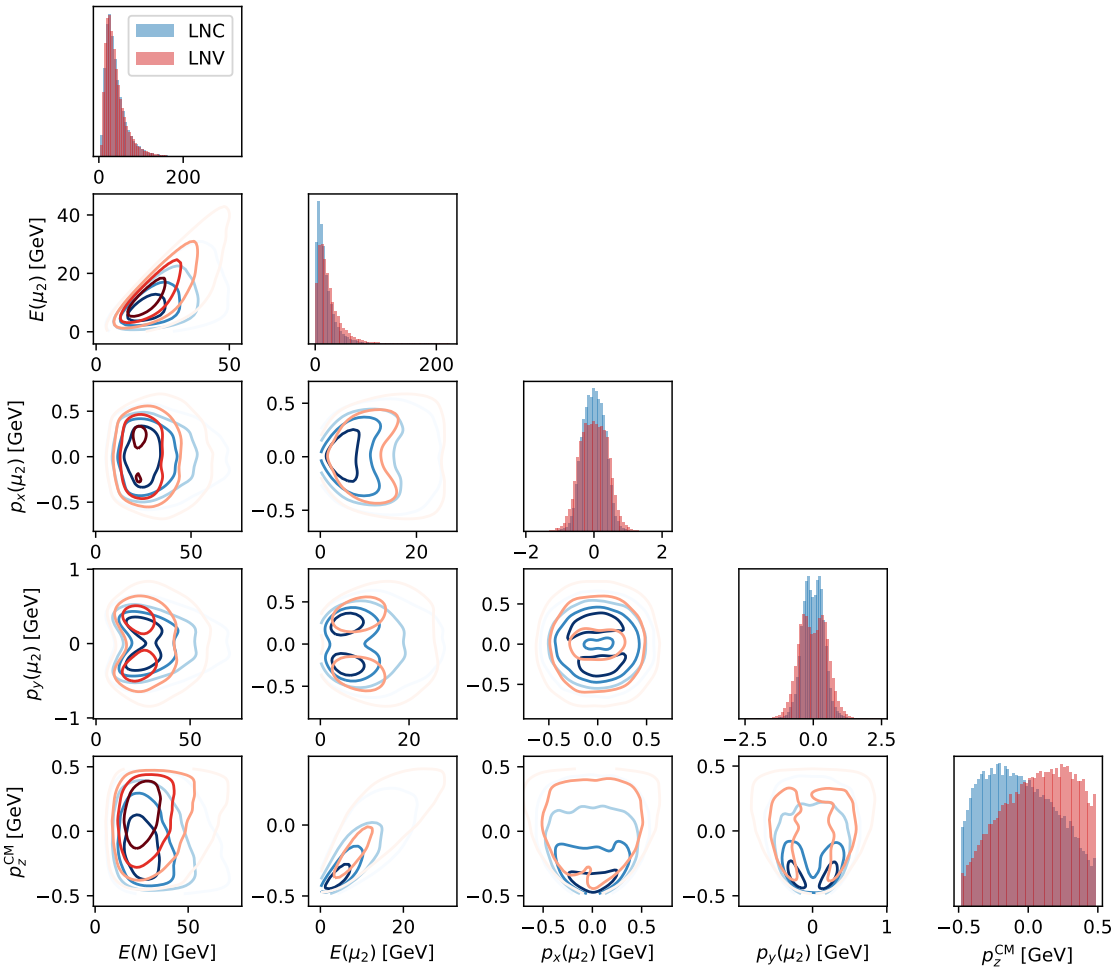


Figure 7. Corner plot showing the correlations between five selected features, for a 1 GeV HNL coupling to the muon. See table 1 for a description of the features. Each subplot shows, on the same scale, the marginal distributions of LNC and LNV events as a function of either one (on-diagonal plots) or two (off-diagonal plots) features. 1d distributions are represented as histograms, and 2d distributions as contour plots of the probability density.

details in appendix B.2. The accuracy of the trained classifier (as evaluated on the test set) is presented in figure 8 as a function of the HNL mass for three scenarios, corresponding to an HNL coupling to electrons, muons, or equally to both.

4.3 Model selection

Assuming the true event distribution to match (or be sufficiently close to) the simulated one, we can then use our trained classifier to classify each event as either LNC or LNV. As stated in section 1, our main goal is to distinguish the following two hypotheses:

- $\mathcal{H}_1$: HNLs are Dirac or quasi-Dirac with $\delta M\tau \ll 1$ (LNC decays only).
- $\mathcal{H}_2$: HNLs are Majorana or quasi-Dirac with $\delta M\tau \gg 1$ (as many LNC/LNV decays).

JHEP04(2020)005

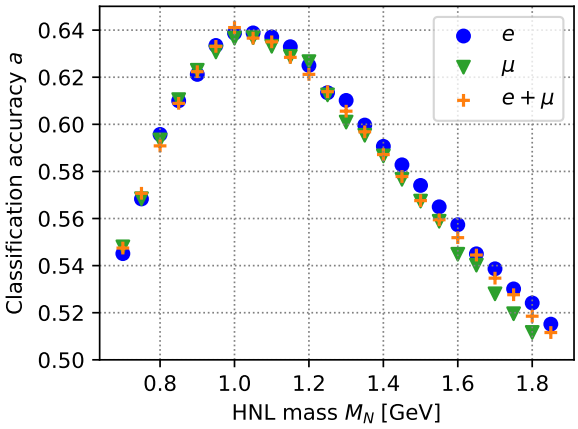


Figure 8. Classification accuracy as a function of the mass, for an HNL coupling to e, μ, or equally to both.

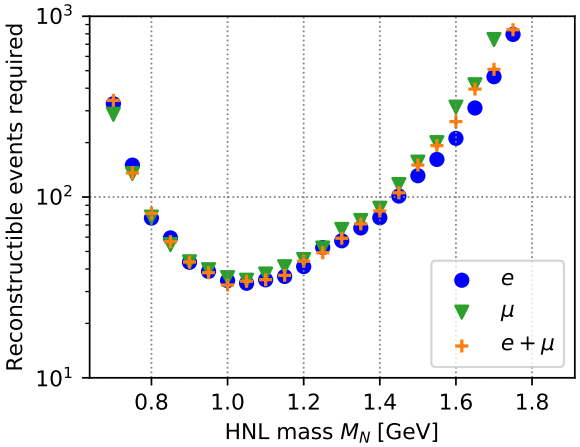


Figure 9. Number of fully reconstructible events required to detect LNV at 90% CL, for an HNL coupling to e, μ, or equally to both.

Since the classifier is not perfectly accurate, its decision cannot be used to directly confirm the presence of LNV processes, or constrain their existence. If we knew the full distribution in feature space $\rho(z)$ for each hypothesis, we could obtain an optimal test statistics by constructing the corresponding likelihood ratio [80]. However, accurately estimating $\rho(z)$ is a non-trivial task and would be error-prone, so we elected to use a less powerful but more reliable, simplified model. Knowing the classification accuracy a for a given binary classifier, we compute the likelihood of classifying k events out of N as LNV, and $N-k$ events as LNC (independently of their specific feature vectors z) assuming that the true fraction of LNV events is f. We then compute the best-fit value for f and use Wilk's theorem [81] in order to determine whether it significantly deviates from either $f=0$ (corresponding to $\mathcal{H}_1$) or $f=\frac{1}{2}$ (corresponding to $\mathcal{H}_2$).

In order to estimate the "model-selection" sensitivity of SHiP, we then compute, under each hypothesis and as a function of the HNL mass M_N and squared mixing angles $|\Theta_\alpha|^2$, the median confidence level at which we can exclude the other hypothesis assuming 5 years

JHEP04(2020)005

of nominal operation (i.e. $2 \cdot 10^{20}$ protons on target). For each true hypothesis, we finally draw the sensitivity limit by plotting, for each M_N, the smallest $|\Theta_\alpha|^2$ for which this median confidence level is at least 0.9. In other words, for mixing angles above this limit, SHiP has a probability of at least 1/2 of disfavouring one hypothesis at CL = 0.9 if the other is realized. The number of fully reconstructible events corresponding to this limit is plotted in figure 9 (when the null hypothesis is taken to be $\mathcal{H}_1$). The construction of these confidence limits is described in details in section B.3, and the resulting sensitivity plots are presented in section 5.1.

4.4 Resolving HNL oscillations

So far we have only considered the two extreme cases ($\mathcal{H}_1$ and $\mathcal{H}_2$), where the HNL(s) behave either as a single Dirac or Majorana particle. However, as discussed in section 2.2, if two nearly degenerate HNLs form a quasi-Dirac pair, both LNC and LNV decay chains will be present, with a non-trivial ratio $\neq 0, 1$, and the corresponding decay rates will feature oscillations as a function of the proper time τ between the HNL production and decay events, with the characteristic $1 \pm \cos(\delta M \tau)$ dependence described by eq. (2.10), where (+) corresponds to LNC and (−) to LNV.

For $\delta M \sim 10^{-6}$ eV, $\delta M \tau$ will be of order 2π at SHiP, leading to potentially resolvable oscillations, provided we can accurately reconstruct the proper time τ between the HNL production and decay. Expressing it as $\tau = \frac{L}{\beta\gamma}$, we see that this can be accomplished if we have sufficiently accurate vertexing and energy reconstruction. At SHiP, the precision on L will be limited by the impossibility of reconstructing the primary vertex within the target. The energy resolution, despite being sufficient for particle identification, is not enough for reconstructing τ (see sections 4.7 and 4.10 in ref. [28]). However, the momentum resolution, combined with the dispersion relation (assuming the HNL mass to be known already with sufficient accuracy) should allow reconstructing γ much more precisely. The high vertexing and momentum resolution permitted by the SHiP tracker, together with our method for (statistically) distinguishing LNC from LNV processes (described in section 4.3), should therefore make it possible to resolve the oscillation pattern in part of the parameter space.

In order to search for HNL oscillations, we first classify the observed events using a model trained (for *one* HNL) at the corresponding mass. We thus assume again that we have sufficiently many events that the HNL mass M_N is well known. The events are then binned in proper time τ, which is the relevant variable for oscillations of massive, relativistic particles. Instead of using the predicted class, here we implement the classifier decision as a weight for the binned events, using the predicted probability p_{LNV}. This weight contains more information than the class does, since it acts as a measure of uncertainty by taking values close to 1/2 for ambiguous events, and closer to 0 or 1 for unambiguous ones. However, without applying further corrections, the sum of these probabilities would average to $N \langle p_{\mathrm{LNV}} \rangle$ for the entire sample of N events. If used directly as weights, they would therefore cause the oscillatory pattern to be hidden among Poisson fluctuations. In order to reveal this pattern, we instead weight the events by $p_{\mathrm{LNV}} - \overline{p_{\mathrm{LNV}}}$, where $\overline{p_{\mathrm{LNV}}}$ is the sample average of the estimated p_{LNV}. This weight averages to zero over the entire sample, which limits the impact of Poisson fluctuations.

JHEP04(2020)005

HNL oscillations are implemented in our simulation by first generating events without taking interference into account then, in a second time, performing rejection sampling based on the proper time τ, following eq. (2.10). The results obtained using this simulated data set are presented in section 5.2.

4.5 Applicability of the method to other experiments

The present study crucially relies on the identification of the HNL decay products and the measurement of their momenta. However, a number of proposed experiments to search for HNLs, such as MATHUSLA [12–14], CODEX-b [18, 19] (in its baseline configuration) and ANUBIS [22], cannot measure the momenta of the decay products. Since low-mass HNLs ($M_N < M_{B_c}$) at the LHC are also mostly produced in the decays of heavy mesons, one can wonder to which extent the present analysis would apply to these experiments. Training a classifier using only the *directions* of the tracks of the visible decay products and the same geometry as SHiP reveals that the distributions of LNC/LNV for a given set of HNL parameters can still be distinguished, with an accuracy only slightly lower than the one obtained using the full momenta. There are, however, two caveats. First, training the classifier requires knowing the HNL mass, which cannot be obtained without measuring the momenta of its decay products (or matching the displaced decay to its reconstructed production process in the main detector, if this is feasible). In addition, the large center-of-mass energy at the LHC could result is a very broad heavy meson spectrum, which would smear out the LNC/LNV distributions and make them indistinguishable. It therefore seems unlikely that MATHUSLA, CODEX-b or ANUBIS could benefit from this method.

Other planned or proposed detectors, such as NA62^{++} [26, 27] (in beam-dump mode), the DUNE near detector [23–25], FASER [15–17] and AL3X [20, 21], are in principle capable of reconstructing the HNL mass. The AL3X detector, thanks to its large time projection chamber and its magnetic field, should be able to directly measure both the charges and momenta of the two leptons, making the method described here unnecessary. It is unclear to the authors, however, whether FASER could benefit from it. The answer likely depends on the spectrum of the heavy mesons producing the HNLs which eventually interact with the detector. A Monte-Carlo simulation would provide a definitive answer to this question. The remaining beam-dump experiments: NA62^{++} and DUNE, share a similar geometry with SHiP and face the same challenge (no observation of the primary charged lepton $l_\alpha^\pm$). As such, we generically expect the method presented here to be applicable to these experiments, within the mass range where it is valid, and subject to the heavy meson spectrum being similar to the one at SHiP. This could be ascertained using a Monte-Carlo simulation. Whether these experiments can also resolve HNL oscillations will depend on how accurately they can reconstruct the HNL momentum.

5 Results

5.1 Sensitivity to Lepton Number Violation

In order to easily compare our results to existing exclusion bounds or to the sensitivities of future experiments, let us consider two simplified models where a single HNL exclusively

JHEP04(2020)005

mixes with the electron or muon neutrino.[8] As can be seen in figure 8, more generic mixing patterns with the e and μ flavors do not significantly degrade the classification accuracy; therefore they should leave the limits presented below mostly unchanged. However, if a significant fraction of HNLs is produced through mixing with the τ neutrino, then the present analysis would need to be modified to handle secondary production of HNLs in τ decays, including spin correlation effects.

As discussed in section 4.3, we define the sensitivity to lepton number violation as the smallest mixing angles for which SHiP has a 1/2 probability of either rejecting or detecting LNV, if it is respectively absent or present with the same rate as LNC. The results are presented in figure 10, along with various existing exclusion bounds and detection sensitivity[9] limits for planned or proposed experiments, extracted from the report of the Physics Beyond Colliders working group [11]. We only show the sensitivities of experiments which can not only set exclusion bounds, but also reconstruct the HNL mass, should it be observed. Note that in order to be consistent with the SHiP detection sensitivity, which was computed for one Majorana HNL, we present our results for one HNL as well. In the realistic case of $\mathcal{N} \geq 2$ HNLs, both curves must be scaled down by a factor of $\mathcal{N}^{1/2}$. Above the black dashed line, SHiP should be able to distinguish Dirac-like ($\mathcal{H}_1$) and Majorana-like ($\mathcal{H}_2$) HNLs. We have discarded the HNL masses for which the early stopping criterion returned the first iteration as the best, since it suggests that the classifier has failed to learn anything about the data. Below 0.7 GeV, additional production channels $H \to h'_V l_\alpha N$ (where h'_V denotes a vector meson) become significant, and have not been implemented with spin correlations in our Monte-Carlo simulation. Therefore we also restrict the HNL mass to $M_N \gtrsim 0.7\,\text{GeV}$. Additionally, since the sensitivity is almost identical for excluding $\mathcal{H}_1$ or $\mathcal{H}_2$, we only plot one limit, which corresponds to excluding $\mathcal{H}_1$ at 90% CL if LNV is actually present.

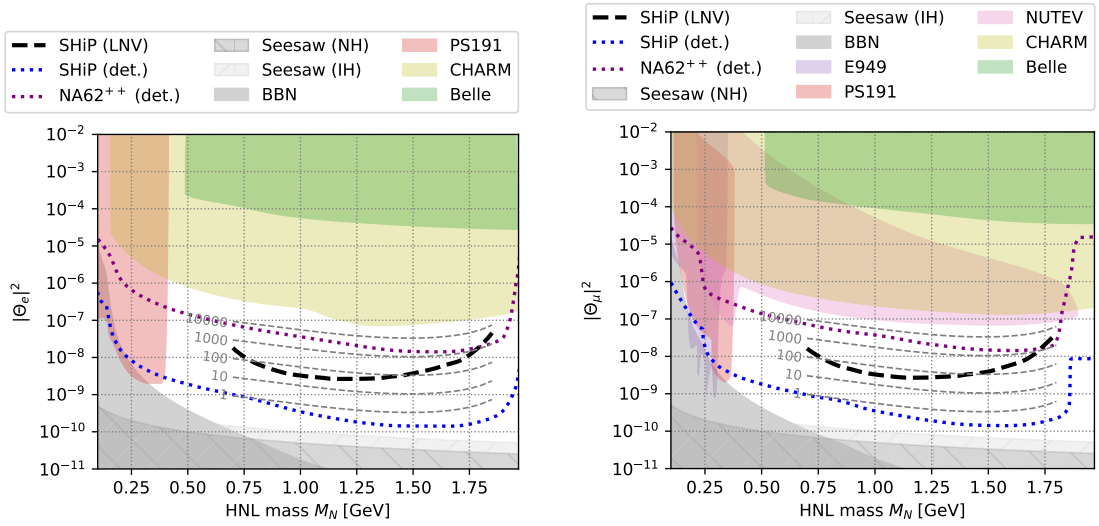
We can see that the larger number of accepted events (indicated in figure 10 by the thin dashed grey lines) at higher masses initially compensates for the worse classification accuracy, but is not sufficient any more as we approach the D threshold. In practice, we expect that systematic uncertainties about the D spectrum and the simulation will decrease the sensitivity at both ends of the mass range, where the classification accuracy is already close to 1/2. Comparing the results to the SHiP detection sensitivity, we see that around 1 GeV, the model-selection sensitivity limit is about one order of magnitude above the detection one, while remaining well below the planned NA62^{++} limit as well as existing bounds.

This leads us to an interesting conclusion: there exists a non-trivial region of parameter space, unconstrained by current or near-future experiments, where SHiP would not only be able to detect HNLs, but also characterize them as either Dirac-like or Majorana-like particles. As discussed in appendices A.3 and B.4, this conclusion is robust with respect to uncertainties on the heavy meson spectrum.

[8] Within the seesaw mechanism, it is impossible to generate the two observed light neutrino mass differences with a single HNL, or if HNLs mix with one generation only [82]. The two benchmarks presented in figures 10(a) and 10(b) are thus simplifications, used here because they are consistent with the parametrization employed by the PBC working group.

[9] The usual sensitivity, by opposition to the sensitivity to lepton number violation discussed here.

JHEP04(2020)005



(a) HNL mixing with ν_e.

(b) HNL mixing with ν_μ.

Figure 10. SHiP sensitivity to lepton number violation. The thick dashed curve is the "model-selection" sensitivity computed in this work. The thin dashed grey lines show the number of fully reconstructible events which would be observed at SHiP for a given mass and mixing angle. Dotted curves are the (lower) detection sensitivities for the proposed or planned experiments which can reconstruct the HNL mass. Coloured, filled areas are regions of parameter space which have been excluded by previous experiments. The grey filled area denoted by BBN indicates the region which is incompatible with Big Bang Nucleosynthesis. Below the seesaw limit[10](hatched region), mixing angles are too small to produce the observed neutrino masses.

5.2 Resolvable quasi-Dirac oscillations

The result of the procedure described in section 4.4 is presented in figure 11 for a new simulated dataset (independent from the training set), corresponding to a quasi-Dirac pair of mass $M_N = 1\,\text{GeV}$, mass splitting $\delta M = 4 \cdot 10^{-7}\,\text{eV}$, and mixing with muon neutrinos only, with a squared mixing angle $|\Theta_{\mu I}|^2 = 2 \cdot 10^{-8}$, $I = 1, 2$. The oscillatory pattern is manifest at $\tau < 5\,\text{m}$, where most of the events fall. At larger τ it is hidden in Poisson fluctuations. The uncertainty on τ at SHiP is dominated by the (boosted) length of the target $\sim 0.1\,\text{m}$, which contains the unresolved primary vertex. It could smear out fast oscillations, in which case an accurate treatment of this uncertainty would be needed in the simulation. However, for longer oscillation periods like the one shown in figure 11, its effect should be negligible. Deriving precise sensitivity limits for HNL oscillations is beyond the scope of this paper, since it is likely that no simple analytical expression exists for them, due to the more complex test statistics required, compared to the detection or model-selection limits. HNL oscillations might for instance be amenable to methods such as maximum likelihood estimation, wavelets, or matched filtering, for which the null distribution can be estimated numerically using a (computationally expensive) bootstrapping procedure.

[10]The seesaw limit can only be rigorously computed if the mixing angles are consistent with the seesaw equation (2.7). This is not possible for HNLs mixing with only one generation, nor for a single HNL. The limits presented here instead correspond to the "naive" estimate $\sum m_\nu \leq M_N \cdot \sum_\alpha |\Theta_\alpha|^2$, where we have assumed the lightest neutrino to be massless.

JHEP04(2020)005

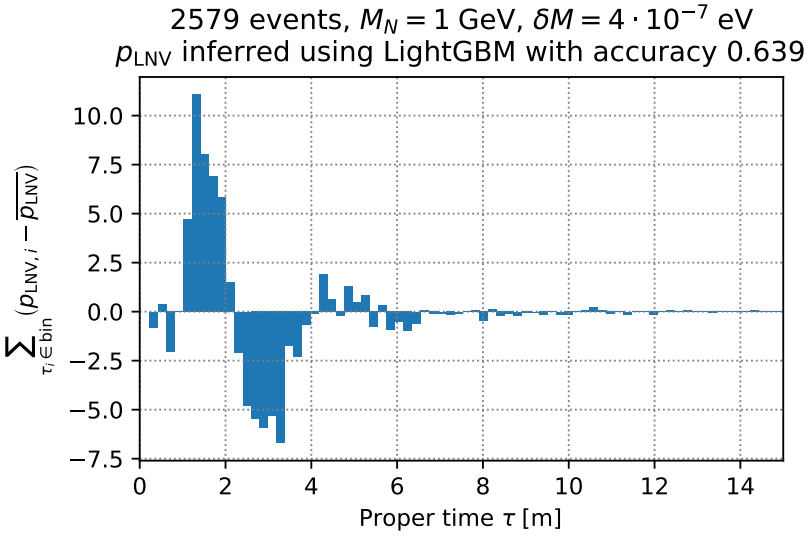


Figure 11. Events binned by proper time τ and weighted by $p_{\mathrm{LNV}} - \overline{p_{\mathrm{LNV}}}$, revealing the oscillatory pattern, for *two* HNLs with $M_N = 1\,\mathrm{GeV}$, $|\Theta_{\mu I}|^2 = 2 \cdot 10^{-8}$, $|\Theta_{eI}|^2 = |\Theta_{\tau I}|^2 = 0$ and $\delta M = 4 \cdot 10^{-7}\,\mathrm{eV}$.

6 Conclusions

The SHiP experiment is set to have an unprecedented detection reach for a variety of models containing feebly interacting particles, such as Heavy Neutral Leptons (HNLs). A distinctive feature of SHiP among other intensity frontier experiments is its decay spectrometer, which allows it to not only place exclusion bounds, but also perform event reconstruction and measure the HNL properties. The simplest consistent HNL model accessible at SHiP contains two nearly degenerate HNLs, which can undergo oscillations. Their mass splitting δM is of particular interest, since it greatly influences their phenomenology as well as early-Universe cosmology (specifically, baryogenesis and dark matter production).

In the present work, we have investigated to which extent SHiP may be able to constrain or even measure δM. Depending on the scale of the oscillation phase $\delta M \tau$ accessible at an experiment, HNLs may or may not exhibit lepton number violation (LNV). The problem thus amounts to distinguishing LNC from LNV decay chains (figure 2) in a beam-dump setting (figure 1), where the primary lepton cannot be observed. We have shown that the angular distribution of the *visible* secondary decay products provides a partial solution to this problem, since, depending on the HNL mass, it can significantly differ between LNC and LNV in the laboratory frame (figure 7). This result has been qualitatively understood in the simplified case of two-body decays in the massless limit (figures 3 and 6). In order to handle more realistic cases, a Monte-Carlo simulation has been employed to generate accurate data sets of LNC and LNV events, including spin correlations and geometrical acceptance. The different distributions of the kinematic variables thus allow discriminating between LNC and LNV events using multivariate analysis; and with sufficiently many events, it becomes possible to statistically detect or exclude lepton number violation.

In order to produce sufficiently accurate training sets, our simulation must satisfy several requirements. It should be able to generate all the relevant two- and three-body

JHEP04(2020)005

meson decays containing an HNL (figure 4), as well as the selected HNL decay channel $N \to \pi^{\mp} l_{\beta}^{\pm}$. It should be accurate for GeV-scale HNLs, and should account for the spin correlations between the primary and secondary decays. Finally, it should run sufficiently fast to allow producing large training sets for various hypotheses and parameters. In order to meet all these requirements, we have written our own Monte-Carlo simulation, the output of which is used to train a binary classifier.

Knowing the accuracy of the classifier decision (figure 8) for a given mass and (relative) mixing angles, we can finally draw a "model-selection" sensitivity limit in the $(M_N, |\Theta|^2)$ plane (shown in figures 10(a) and 10(b)), above which SHiP should be able to either discover or rule out lepton number violation from HNLs. Interestingly, this limit lies below the detection sensitivity of near-future experiments such as NA62^{++}. This leads to a striking conclusion: SHiP might be able to not only *discover* HNLs, but also *characterize* them as either "Dirac-like" or "Majorana-like" fermions (depending on whether they feature LNV) even if previous experiments *see no signal at all*. Better yet, if the mass splitting between the two HNLs is of order $\delta M \sim 10^{-6}$ eV, SHiP should be able to *resolve the oscillations* of HNLs (figure 11), given sufficiently many events. Intriguingly, this mass splitting falls within the range required for producing dark matter in the νMSM [39]. Its measurement — or constraining — would therefore be an important test of cosmological models.

Acknowledgments

The authors would like to thank Mikhail Shaposhnikov for his suggestion of investigating HNL oscillations at SHiP, Oleg Ruchayskiy and Maksym Ovchynnikov for their helpful comments and proofreading of the present manuscript, Sonia Bouchiba and Federico Leo Redi for stimulating discussions about helicity effects in HNL decays, and Kyrylo Bondarenko and Elena Graverini for sharing with us their code to compute HNL decay widths. We thank the SHiP collaboration for their interest in our work. We are grateful to Annarita Buonaura, Richard Jacobsson, and Nicola Serra for their helpful comments and clarifications about the dimensions of the facility. JLT acknowledges support from the Carlsberg foundation. The work of IT was supported in part by the ERC-AdG-2015 grant 694896.

A Simulation

A.1 Overview

It is not obvious whether the different angular correlations of LNC and LNV events lead to an observable effect in a realistic beam-dump experiment. To answer this question, we have devised a toy Monte-Carlo simulation, inspired from the one used in ref. [30], to simulate the production and decay of HNLs at the SHiP experiment [28, 29] (represented on figure 1).

The simulation of rare BSM processes with spin correlations entails two main requirements. First, we cannot afford to simulate all the possible processes, since, due to the small HNL mixing angles, the decay chains mediated by an HNL only represent a tiny fraction of all decays. Instead, we only simulate the BSM processes, and use importance

JHEP04(2020)005

sampling (i.e. introduce weights) in order to obtain the correct absolute number of events and expectation values (appendix A.2).

Secondly, we cannot sample the primary and secondary decays separately, since they are not independent. Instead, we construct all possible decay chains for the production and decay processes of interest, and sample the entire chain at once, with a probability proportional to its combined branching fraction. The momenta of all the decay products are then sampled simultaneously, using the matrix element for the entire chain (appendix A.4).

In addition, in order to accurately model the SHiP experiment, we need to sample the heavy meson momenta from a realistic spectrum (appendix A.3) and take into account the finite size of SHiP and its geometrical acceptance (appendix A.5). Finally, since most machine learning algorithms take unweighted data points as input, it is necessary to perform a last step of rejection sampling in order to produce a training set consisting of events with equal weights (appendix A.6).

A.2 Decay chains

As discussed in ref. [76], the dominant HNL production process at SHiP is from weak decays of the lightest charmed or beauty mesons. In the present study, we focus on HNL masses below the D_s mass, and only select the fully reconstructible secondary decays $N \to \pi^{\pm} l_{\beta}^{\mp}$, By producing long-lived, charged particles which can be measured by the decay spectrometer located at the end of the decay vessel, they allow the HNL momentum to be reconstructed. The efficiency of particle identification at SHiP is high enough [77] that we can approximate it as one for the present estimate. Therefore we do not need to simulate decay chains containing any other secondary decays.

For the mixing angles of interest (i.e. below existing bounds), the fraction of all decays which are mediated by an HNL is tiny. We therefore need to use importance sampling in order to efficiently simulate only the processes of interest. For every proton on target (POT), the probability of producing a charmed hadron of species H is:

$$P(H) = \frac{\sigma_{cc}}{\sigma_{pN}} \cdot A_H \tag{A.1}$$

where σ_{cc} is the production cross-section for charmed hadrons, σ_{pN} the interaction cross-section for protons hitting the target nuclei, and A_H is the relative abundance of the charmed hadron species H (as given in appendix A of [29]). The *nominal* (i.e. physical) probability of producing an HNL which mediates a given decay chain $H \to [h']l_{\alpha}(N \to l_{\beta}h'')$ (irrespective of whether the decay is observed in the detector) is then:

$$\begin{aligned} P\left(H \to [h']l_{\alpha}(N \to l_{\beta}h'')\right) &= P(H) \cdot P(h'l_{\alpha}N|H) \cdot P(l_{\beta}h''|h'l_{\alpha}N) \\ &= \frac{\sigma_{cc}}{\sigma_{pN}} \cdot A_H \cdot \mathrm{Br}_{\mathrm{prod}}(H \to [h']l_{\alpha}N) \cdot \mathrm{Br}_{\mathrm{decay}}(N \to l_{\beta}h'') \end{aligned} \tag{A.2}$$

where the last two terms are the production and decay branching ratios for HNLs in the considered decay chain. The *importance* distribution P' is defined as a uniform scaling for

JHEP04(2020)005

decay chains involving an HNL, and as zero for all other outcomes:

$$\begin{cases} P'\left(H \to [h']l_\alpha(N \to l_\beta h'')\right) = \frac{1}{w_{\text{prod}}} P\left(H \to [h']l_\alpha(N \to l_\beta h'')\right) \\ P'(\text{no HNL}) = 0 \end{cases} \tag{A.3}$$

where w_{prod} is the weight to be applied to all the chains sampled from the importance distribution, and corresponds to the total probability of producing an HNL according to the nominal distribution:

$$w_{\text{prod}} = \sum_{\text{chains}} P\left(H \to [h']l_\alpha(N \to l_\beta h'')\right) \tag{A.4}$$

When computing expected numbers of events over the entire duration of the SHiP experiment, which represents an integrated $N_{\text{POT}} = 2 \cdot 10^{20}$ protons on target for 5 years of nominal operation, we must further multiply by N_{POT} the expectation values obtained for one event. This is most easily done by simply multiplying the total weights by N_{POT}.

A.3 Heavy meson spectrum

Once a chain is selected, we sample the momentum of the corresponding charmed meson from the spectrum measured by the LEBC-EHS collaboration [74] at the CERN SPS running at 400 GeV with a hydrogen target. The differential cross-section is parametrized as the product of a β distribution in x_F and an exponential distribution in p_T^2:

$$\frac{\mathrm{d}^2\sigma}{\mathrm{d}x_F \mathrm{d}p_T^2} = \sigma \frac{(n+1)b}{2}(1 - |x_F|)^n e^{-bp_T^2} \tag{A.5}$$

with the best-fit values $n = 4.9 \pm 0.5$ and $b = (1.0 \pm 0.1)$ GeV^{-2}. We thus implicitly assume the spectrum to be separable. Due to their very similar mass, and to compensate for the lack of data, we assume D_s mesons to share the same spectrum as D mesons.

By using the spectrum for a hydrogen target, we effectively neglect cascade production of heavy hadrons inside the target, leading us to underestimate the number of hadrons produced at the low-energy end of the spectrum. This could be problematic if their p_T spectrum happens to be significantly different from that of primary hadrons produced in pp collisions. However, the lower acceptance for these softer hadrons should help mitigate the issue. In figure 12, we show how varying the width of the heavy meson p_T spectrum affects the final sensitivity. As expected, a larger p_T spread reduces the sensitivity, while a narrower spectrum improves it.

A.4 Decay product momenta

In order to preserve spin correlations between the HNL siblings and its decay products, we simulate both the HNL production and decay processes at once. For the masses and mixing angles of interest, the HNL is long-lived and can be assumed to be on its mass shell. Therefore the phase-space sampling can be performed independently for the primary and secondary decays. We use the m-generator algorithm [83] for that, as described in ref. [84]. In order to sample events with a probability proportional to the squared transition

JHEP04(2020)005

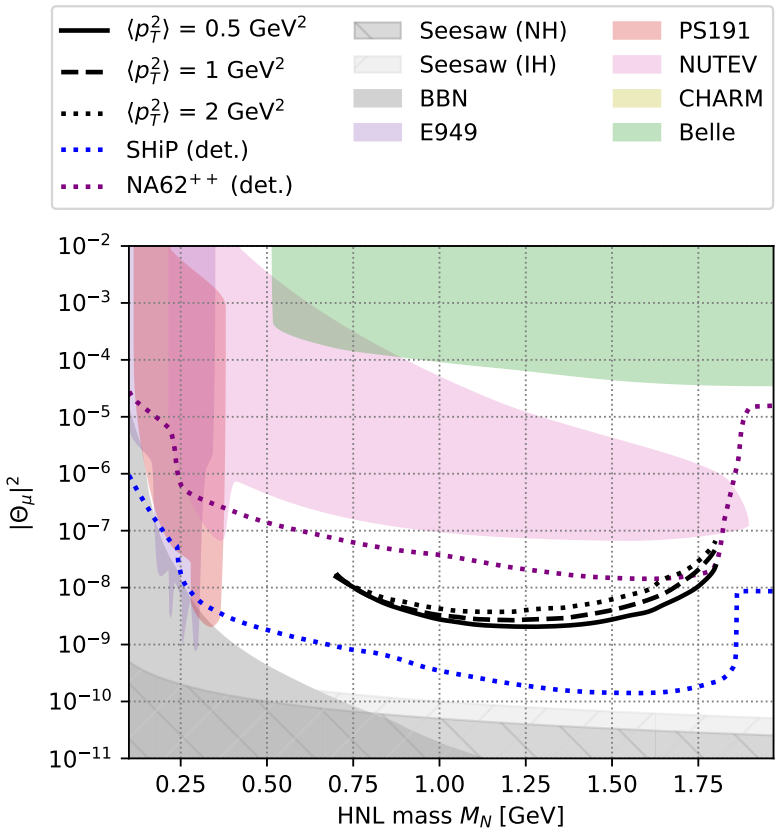


Figure 12. Effect of varying the width of the heavy meson p_T spectrum on the sensitivity to lepton number violation (90% CL), for an HNL coupling to the muon. Black lines represent the model-selection sensitivity of SHiP for various values of $\langle p_T^2 \rangle$. The dashed line corresponds to the best-fit value $\langle p_T^2 \rangle = 1\,\text{GeV}^2$ from the LEBC-EHS collaboration [74].

amplitude, we then perform rejection sampling, taking the phase-space distribution as proposal distribution, and an acceptance probability proportional to the spin-summed, squared matrix elements (3.2) and (3.3) for the entire decay chain. Only the spin states of the external particles (which interact with the detector and are thus "measured" in the quantum mechanical sense) are summed over.

A.5 Geometry

In order to model the geometry of the SHiP experiment, we must account for the finite size of the detector and its geometrical acceptance. In the current SHiP design (represented on figure 1), the fiducial volume consists of an evacuated right pyramidal frustum of length 50 m, located at a distance of 50 m from the target, and with horizontal and vertical sides 5 m and 10 m respectively at the far end. It is followed by a 10 m long tracking station.

To estimate the probability of the HNL decaying within the fiducial volume and passing the acceptance cuts, we use once again importance sampling for sampling the decay vertex. This is required in order to overcome the potentially very long lifetime of HNLs, which could cause most of them to decay away from the experiment. We choose an importance

JHEP04(2020)005

distribution (approximately) covering the fiducial volume, by sampling the decay vertex uniformly along the HNL momentum, at a distance such that it falls inside the decay vessel. The nominal decay probability density is, as a function of the proper time τ (or boost factor γ and distance L) between the HNL production and decay:

$$P_{\text{decay}}(\tau) = \Gamma e^{-\Gamma\tau} \quad \Longrightarrow \quad P_{\text{decay}}(L|\gamma) = \frac{\Gamma}{\beta\gamma} e^{-\frac{\Gamma L}{\beta\gamma}} \tag{A.6}$$

The partial weight resulting from this importance sampling step is therefore:

$$w_{\text{decay}}(L|\gamma) = \frac{\Gamma L_{\text{DV}}}{\beta\gamma\cos(\theta)} e^{-\frac{\Gamma L}{\beta\gamma}} \tag{A.7}$$

where $L_{\text{DV}} = 50\,\text{m}$ is the length of the decay vessel and θ the angle between the HNL momentum and the beam axis. In the linear regime, where $\Gamma\tau \ll 1$, this partial weight reduces to $w_{\text{decay}}(L|\gamma) \cong \frac{\Gamma L_{\text{DV}}}{\beta\gamma\cos(\theta)}$.

We finally apply acceptance cuts by requiring the HNL to decay within the decay vessel, and the trajectories of its two decay products ($l_\beta^\mp$ and $\pi^\pm$) to intersect the tracking station located at its far end.

A.6 Unweighting

As a last step, we perform again rejection sampling on the weighted events in order to obtain a set of events with equal weights, which are easier to analyse and process with machine learning algorithms. This is done by accepting events with a probability proportional to their weight, and can be justified as follows.

Let X denote a random variable representing the simulated event, and x a concrete realization of it. Let $f(x) = P(X = x)$ be the nominal (i.e. true) distribution and $g(x)$ the importance distribution, such that $g(x) > 0$ for all outcomes x in the domain of interest Ω (i.e. all relevant observables must have their support in Ω). If x is sampled from the importance distribution $g(x)$, its associated weight will be $w(x) = f(x)/g(x)$. Let M be an upper bound on $w(x)$, i.e. $M \geq w(x), \forall x \in \Omega$. If we choose the acceptance probability to be $a(x) \stackrel{\text{def}}{=} w(x)/M \leq 1$, then it immediately follows that the accepted events, effectively drawn from the new importance distribution $g(x) \cdot a(x)$, will have uniform weight M.

It is therefore possible to perform rejection sampling a posteriori in order to produce uniformly weighted events. However, storing all the generated events, many of which will eventually be rejected, would be inefficient from a memory perspective. A more economical solution, which we decided to use, consists in performing rejection sampling directly as events are being generated. This requires estimating an upper bound M on the weights, during an initial burn-in phase.

B LNC/LNV classification

At leading order in the light lepton and hadron masses, the matrix elements for LNC and LNV decay chains have a straightforward analytical dependence on the invariant mass s_{ll} of the charged lepton pair. However, unlike in collider experiments, this variable is not

JHEP04(2020)005

readily available in a beam-dump setting, due to the primary lepton being unobservable. As we saw in section 3.2, the different angular correlations between the charged leptons can nevertheless lead to residual correlations between the visible HNL decay products. The absence of an obvious test statistics, along with the almost background-free conditions and highly efficient PID at SHiP [77], makes the task of distinguishing LNC from LNV ideally suited for multivariate analysis. In the following subsections, we describe how we generate the training set (appendix B.1), the classifier used to discriminate between LNC and LNV events (appendix B.2), how to produce a sensitivity limit from its output (appendix B.3), and finally how sensitive is the classification to systematic uncertainties on the heavy meson spectrum (appendix B.4).

B.1 Dataset

As mentioned in section 4.2, we need to generate datasets for various HNL masses M_N and rays in $|\Theta_\alpha|^2$ space, where $\alpha = e, \mu$ (the overall normalization does not matter). In practice, we choose a mass range spanning the region between the K and D_s thresholds, and consider several benchmark models with fixed $|\Theta_e|^2 : |\Theta_\mu|^2$ ratios.[11] For each choice of physical parameters, we sample $3{\cdot}10^6$ events with uniform weights. This is done by sampling sufficiently many weighted events and, as they are being generated, "unweighting" them by performing rejection sampling with an acceptance probability proportional to their weight. Only events which pass the acceptance cuts are used for training. In the simulation, the HNL is taken to be a single Majorana particle, such that the dataset contains equal numbers of LNC and LNV events and is balanced with respect to the primary and secondary lepton charges. We select only the fully reconstructible HNL decays $N \to \pi^\mp l_\beta^\pm$, which do not contain an unobservable light neutrino, and produce long-lived charged particles which can be measured by the decay spectrometer. For the sake of simplicity, we will assume the PID to be perfectly efficient throughout this analysis. Non-trivial efficiencies are expected to slightly reduce the final sensitivity reach. As explained in section 4.2, each event is labelled as being either LNC or LNV, and we record the 19 observable features listed in table 1. The dataset is split into training/validation/test sets with respective proportions $0.5 : 0.2 : 0.3$.

B.2 Classifier

We employ the LightGBM [78] gradient boosting algorithm, accessed through the Python interface to the reference implementation [79]. For classification, we choose the `binary` objective. We use early stopping based on the binary log-loss (`binary_logloss`) and the area-under-curve (`auc`) metrics, with a 10 round threshold. The hyperparameters `num_leaves` and `learning_rate` are manually optimized by maximizing the above two metrics on the validation set. The classification accuracy is presented in figure 8 as a function of the HNL mass M_N for two orthogonal scenarios, corresponding to the HNL coupling exclusively to

[11] We do not consider HNL production through τ mixing in this work, since it would have required to implement secondary production from τ decays. It is negligible in the considered mass range unless the Θ_τ mixing angle is significantly larger than the others, as can be seen in figure 4. In addition, visible HNL decays through τ mixing are forbidden below the τ threshold.

JHEP04(2020)005

Feature	p2y	p3y	p2x	p3x	pCMz	zD	xD	yD	p1x	pCMy
# splits	302	282	243	238	141	114	105	97	91	85

Feature	pCMx	p1y	E1	E2	E3	p3z	p2z	p1z	Q12
# splits	77	74	69	67	61	53	34	14	9

Table 2. Feature importance for a 1 GeV HNL coupling to μ.

electrons ($|\Theta_e|^2 : |\Theta_\mu|^2 : |\Theta_\tau|^2 = 1 : 0 : 0$) or muons ($|\Theta_e|^2 : |\Theta_\mu|^2 : |\Theta_\tau|^2 = 0 : 1 : 0$), and a third one where it couples equally to both ($|\Theta_e|^2 : |\Theta_\mu|^2 : |\Theta_\tau|^2 = 1 : 1 : 0$).

It is instructive to understand the origin of this dependence, if only to make sure that it corresponds to a physical effect. LightGBM provides a way to estimate the feature importance, by counting the number of times a feature is used to split a tree. Those are listed in table 2 for a 1 GeV HNL coupling to muons (which results in a classification accuracy of 63.5%). They reveal that the most important features are the transverse components of the momenta of the HNL decay products. Indeed, it is possible to successfully train a model using a single feature such as the transverse momentum $p_{T,\mu}$ of the secondary muon, while still obtaining a classification accuracy of 61.5% (for the same dataset).

Inspecting the results more closely (see figure 7) shows that LNV events have on average a slightly larger transverse momentum than LNC ones. This is consistent with our discussion from section 3.2, and allows us to understand the mass dependence. At large HNL masses, as we approach the closing mass of D meson leptonic decays, the kinetic energy of the HNL in the heavy meson rest frame decreases, until it becomes so small that the difference between LNC and LNV becomes negligible compared to the transverse momentum spread of the heavy meson spectrum. As the HNL mass decreases, 3-body semileptonic decay channels open, and become dominant at lower masses. The additional meson takes away part of the energy from the HNL, leaving it with insufficient kinetic energy to "escape" the transverse momentum spread of the heavy meson spectrum. Finally, the large boost of the heavy mesons along the beam axis washes out most of the information contained in the longitudinal part of all laboratory frame momenta, which explains their low importance.

B.3 Sensitivity to lepton number violation

As stated in section 4.3, our main goal is to distinguish between the following two hypotheses using exclusively the classifier decision (i.e. not the underlying feature vector z):

- $\mathcal{H}_1$: HNLs are Dirac or quasi-Dirac with $\delta M\tau \ll 1$ (LNC decays only).
- $\mathcal{H}_2$: HNLs are Majorana or quasi-Dirac with $\delta M\tau \gg 1$ (LNC and LNV decays).

Those can be expressed as special cases of a more general hypothesis $\mathcal{H}(f)$, $f \in [0, 1]$, parametrized by the relative frequency f of LNV events:

- $\mathcal{H}(f)$: (LNV rate) = $f \times$ (total rate).

such that $\mathcal{H}_1 = \mathcal{H}(f = 0)$ and $\mathcal{H}_2 = \mathcal{H}(f = 1/2)$.

JHEP04(2020)005

We model the classifier decisions using a 2×2 confusion matrix $C_{ij} = P(i \text{ classified as } j)$, where $i, j = 1, 2$ correspond to the two classes, respectively LNC and LNV. The confusion matrix can be expressed in terms of the classification accuracies as:

$$C = \begin{pmatrix} a_1 & 1 - a_1 \\ 1 - a_2 & a_2 \end{pmatrix} \tag{B.1}$$

Suppose we observe N events passing the selection cuts, k of which are classified as LNV. Then, under $\mathcal{H}(f)$, the likelihood of classifying $N - k$ events in class 1 (LNC) and k in class 2 (LNV) is given by the following binomial distribution:

$$\mathcal{L}(k; f) = \binom{N}{k} \big(a_2 f + (1 - a_1)(1 - f)\big)^k \big(a_1(1 - f) + (1 - a_2) f\big)^{N-k} \tag{B.2}$$

Under hypothesis $\mathcal{H}_1$, i.e. all events are LNC, this likelihood reduces to:

$$\mathcal{L}_1(k) = \mathcal{L}(k; f = 0) = \binom{N}{k} (1 - a_1)^k a_1^{N-k} \tag{B.3}$$

while under hypothesis $\mathcal{H}_2$, i.e. events come from either class with equal probability, it becomes:

$$\mathcal{L}_2(k) = \mathcal{L}(k; f = 1/2) = \binom{N}{k} \frac{(1 + a_2 - a_1)^k (1 + a_1 - a_2)^{N-k}}{2^N} \tag{B.4}$$

For many models, including LightGBM (with a balanced training set), $a_1 \approx a_2 \stackrel{\text{def}}{=} a$. In this limit, $\mathcal{L}_2(k)$ simplifies to $\binom{N}{k} 2^{-N}$.

Since $\mathcal{H}_{1,2}$ and $\mathcal{H}(f)$ are nested, then, assuming we have sufficiently many events, we can use Wilk's theorem[12] to try to exclude $\mathcal{H}_{1,2}$. To this end, we construct the two likelihood ratios $\Lambda_{1,2}(k)$ as:

$$\Lambda_i(k) = \frac{\mathcal{L}_i(k)}{\mathcal{L}(k; \hat{f})}, \quad i = 1, 2 \tag{B.5}$$

where $\hat{f}$ is the maximum likelihood estimator for f:

$$\hat{f} = \frac{1 - a - k/N}{1 - 2a} \tag{B.6}$$

Wilk's theorem states that if $\mathcal{H}_i$ ($i = 1$ or 2) is realized, then $-2 \ln(\Lambda_i(k))$ follows a χ^2 distribution with one degree of freedom. Conversely, if we observe $-2 \ln(\Lambda_i(k)) > 2.7$, then $\mathcal{H}_i$ will be disfavoured at 90% CL. If both hypotheses $\mathcal{H}_{1,2}$ were disfavoured simultaneously, this would suggest $\delta M \tau \sim 2\pi$ and potentially resolvable HNL oscillations.

[12] A potential issue in the case of $\mathcal{H}_1$ could be that the null value $f = 0$ lies on the boundary of the domain $[0, 1]$ of f, while Wilk's theorem requires the true value to be in the interior of the parameter space. However, $\ln(\mathcal{L}(k; f))$ has a well-behaved analytical continuation over a domain larger than $[0, 1]$. As long as the estimator $\hat{f}$ has a sufficiently small variance, this boundary effect can therefore be ignored and Wilk's theorem still applies. See [85] for a comprehensive discussion of the validity conditions of Wilk's theorem.

JHEP04(2020)005

If hypothesis $\mathcal{H}_1$ is actually realized, we expect k to take a value around the expected number of events misclassified as LNV: $(1-a)N$, which, for large N, is approximately equal to the median. The median of the log-likelihood-ratio when testing for $\mathcal{H}_2$ is therefore:

$$\mathrm{med}_1\left(\ln(\Lambda_2)\right) \approx -N \underbrace{\left(\ln(2) + a\ln(a) + (1-a)\ln(1-a)\right)}_{\stackrel{\text{def}}{=} l_1(a)>0} \tag{B.7}$$

If, instead, $\mathcal{H}_2$ is realized, then we expect k to take a median value of approximately $N/2$, such that:

$$\mathrm{med}_2\left(\ln(\Lambda_1)\right) \approx N \underbrace{\left(\ln(2) + \frac{1}{2}\ln(a) + \frac{1}{2}\ln(1-a)\right)}_{\stackrel{\text{def}}{=} l_2(a)<0} \tag{B.8}$$

For a fixed confidence level, we can invert these two formulas to estimate, for each true hypothesis $\mathcal{H}_i$, $i = 1, 2$, the median number of events $N_i(a)$ required to exclude the other hypothesis:

$$N_i(a) = \left|\frac{\ln(\Lambda_{\mathrm{cr}})}{l_i(a)}\right| \tag{B.9}$$

with $-2\ln(\Lambda_{\mathrm{cr}}) \approx 2.7$ for a 90% CL. The higher the classification accuracy, the less events are required to reach the target, while accuracies close to 1/2 do not allow distinguishing the two hypotheses, as $N_i(1/2) \to \infty$. So far we have only considered the two extreme cases $f = 0$ or 1/2, i.e. $\delta M\tau \lesssim 2\pi$. We can generalize this analysis to the case where the true hypothesis or the null hypothesis have a non-trivial LNV fraction f. A larger number of events will then be required to reach the same confidence level. We will not discuss these cases further in this paper, in order to avoid making the discussion unnecessarily complicated.

As a final step, for each HNL mass M and ratio $|\Theta_e|^2 : |\Theta_\mu|^2 : |\Theta_\tau|^2$, we compute the squared mixing angles $|\Theta_\alpha|_i^2\,(M)$ required to produce $N_i(a(M))$ events, thus producing for each true hypothesis $\mathcal{H}_i$ a sensitivity limit, above which SHiP should be able to exclude the other hypothesis with a probability of at least 1/2. The resulting sensitivity plots are presented in section 5.1.

B.4 Systematic uncertainties coming from the heavy meson spectrum

For a classifier to generalize well out of sample, i.e. on real-world data, the distribution used for training should match the true, physical distribution of features. This is in general not the case, since a simulation never perfectly represents reality. We can, however, work around this requirement by explicitly evaluating the classification accuracy over a set of test distributions which is likely to encompass the true distribution. This requires knowing and parametrizing the uncertainties coming from the simulation. We can then obtain a conservative estimate for the classification accuracy by varying the unknown parameters within their uncertainties, and taking a lower bound. If this lower bound is high enough, we should still be able to probe lepton number violation on real data.

At SHiP, the main uncertainty affecting the LNC/LNV classification accuracy comes from the transverse momentum spread of the heavy meson spectrum, which is only known with limited accuracy. In order to estimate the actual sensitivity of SHiP to LNV for a

JHEP04(2020)005

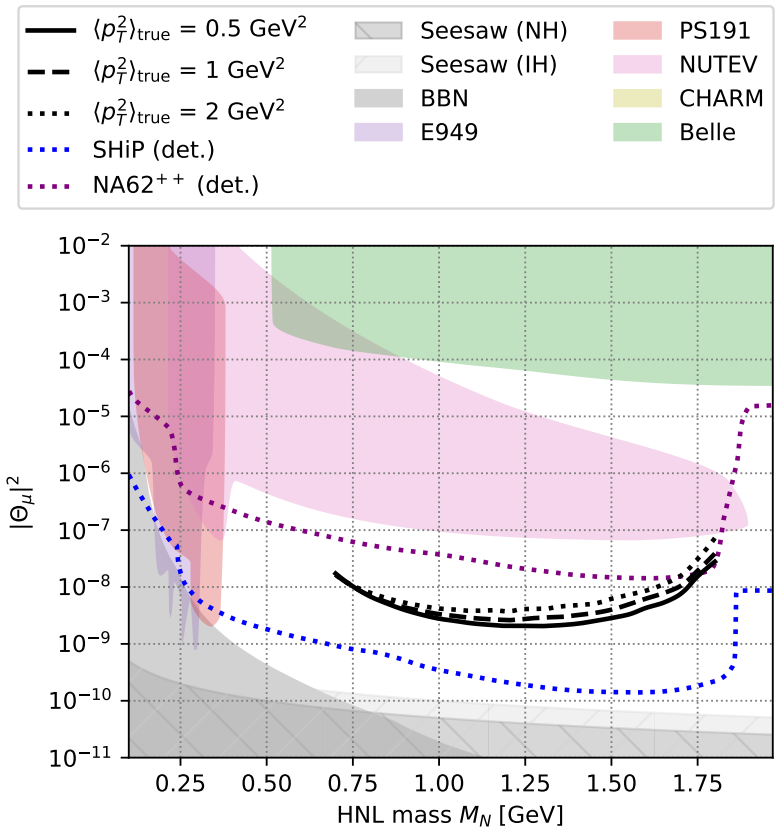


Figure 13. Effect on the LNV sensitivity (90% CL) of computing the classification accuracy on a test set generated with a different p_T spectrum compared to the training set, for an HNL coupling to the muon. Black lines represent the model-selection sensitivity of SHiP for various true $\langle p_T^2 \rangle$. Here, the training set is always generated with $\langle p_T^2 \rangle = 1\,\text{GeV}$.

realistic dataset, we therefore compute the classification accuracy for a family of test sets generated using slightly different p_T spectra, and we take the lowest value as our estimate. The change in the sensitivity resulting from varying $\langle p_T^2 \rangle$ by a factor of two up and down with respect to the best-fit value from LEBC-EHS [74] is shown in figure 13. The planned charm spectrum measurements at SHiP should be able to constrain $\langle p_T^2 \rangle$ to a much better accuracy than the range displayed in the figure.

Interestingly, when comparing this result with figure 12, we observe that the classification accuracy seems to mostly depend on the $\langle p_T^2 \rangle$ of the test set, but not much on the one used for training. This suggests that we might be able to safely use the best-fit spectrum for training without worrying about biasing the results should the true spectrum turn out to be different, provided that we use a conservative estimate for the accuracy. In a more comprehensive study, one would likely want to vary additional parameters related to the spectrum, geometry and simulation.

JHEP04(2020)005

References

[1] P. Minkowski, *$\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?*, *Phys. Lett.* **B 67** (1977) 421.

[2] M. Gell-Mann, P. Ramond and R. Slansky, *Complex spinors and unified theories*, *Conf. Proc.* **C 790927** (1979) 315 [arXiv:1306.4669] [INSPIRE].

[3] R.N. Mohapatra and G. Senjanović, *Neutrino mass and spontaneous parity nonconservation*, *Phys. Rev. Lett.* **44** (1980) 912 [INSPIRE].

[4] T. Yanagida, *Horizontal symmetry and masses of neutrinos*, *Prog. Theor. Phys.* **64** (1980) 1103 [INSPIRE].

[5] J. Schechter and J.W.F. Valle, *Neutrino masses in* SU(2) × U(1) *theories*, *Phys. Rev.* **D 22** (1980) 2227 [INSPIRE].

[6] J. Schechter and J.W.F. Valle, *Neutrino decay and spontaneous violation of lepton number*, *Phys. Rev.* **D 25** (1982) 774 [INSPIRE].

[7] F. Vissani, *Do experiments suggest a hierarchy problem?*, *Phys. Rev.* **D 57** (1998) 7027 [hep-ph/9709409] [INSPIRE].

[8] F. Bezrukov et al., *Higgs boson mass and new physics*, *JHEP* **10** (2012) 140 [arXiv:1205.2893] [INSPIRE].

[9] T. Asaka, S. Blanchet and M. Shaposhnikov, *The νMSM, dark matter and neutrino masses*, *Phys. Lett.* **B 631** (2005) 151 [hep-ph/0503065] [INSPIRE].

[10] T. Asaka and M. Shaposhnikov, *The νMSM, dark matter and baryon asymmetry of the universe*, *Phys. Lett.* **B 620** (2005) 17 [hep-ph/0505013] [INSPIRE].

[11] J. Beacham et al., *Physics beyond colliders at CERN: beyond the standard model working group report*, *J. Phys.* **G 47** (2020) 010501 [arXiv:1901.09966] [INSPIRE].

[12] J.P. Chou, D. Curtin and H.J. Lubatti, *New detectors to explore the lifetime frontier*, *Phys. Lett.* **B 767** (2017) 29 [arXiv:1606.06298] [INSPIRE].

[13] D. Curtin et al., *Long-lived particles at the energy frontier: the MATHUSLA physics case*, *Rept. Prog. Phys.* **82** (2019) 116201 [arXiv:1806.07396] [INSPIRE].

[14] MATHUSLA collaboration, *A letter of intent for MATHUSLA: a dedicated displaced vertex detector above ATLAS or CMS*, arXiv:1811.00927 [INSPIRE].

[15] J.L. Feng, I. Galon, F. Kling and S. Trojanowski, *ForwArd Search ExpeRiment at the LHC*, *Phys. Rev.* **D 97** (2018) 035001 [arXiv:1708.09389] [INSPIRE].

[16] F. Kling and S. Trojanowski, *Heavy Neutral Leptons at FASER*, *Phys. Rev.* **D 97** (2018) 095016 [arXiv:1801.08947] [INSPIRE].

[17] FASER collaboration, *Letter of intent for FASER: ForwArd Search ExpeRiment at the LHC*, arXiv:1811.10243 [INSPIRE].

[18] V.V. Gligorov, S. Knapen, M. Papucci and D.J. Robinson, *Searching for long-lived particles: a compact detector for exotics at LHCb*, *Phys. Rev.* **D 97** (2018) 015023 [arXiv:1708.09395] [INSPIRE].

[19] G. Aielli et al., *Expression of interest for the CODEX-b detector*, arXiv:1911.00481 [INSPIRE].

[20] V.V. Gligorov et al., *Leveraging the ALICE/L3 cavern for long-lived particle searches*, *Phys. Rev.* **D 99** (2019) 015023 [arXiv:1810.03636] [INSPIRE].

JHEP04(2020)005

[21] D. Dercks, H.K. Dreiner, M. Hirsch and Z.S. Wang, *Long-lived fermions at AL3X*, *Phys. Rev.* **D 99** (2019) 055020 [arXiv:1811.01995] [INSPIRE].

[22] M. Bauer, O. Brandt, L. Lee and C. Ohm, *ANUBIS: Proposal to search for long-lived neutral particles in CERN service shafts*, arXiv:1909.13022 [INSPIRE].

[23] LBNE collaboration, *The 2010 interim report of the long-baseline neutrino experiment collaboration physics working groups*, arXiv:1110.6249 [INSPIRE].

[24] I. Krasnov, *DUNE prospects in the search for sterile neutrinos*, *Phys. Rev.* **D 100** (2019) 075023 [arXiv:1902.06099] [INSPIRE].

[25] P. Ballett, T. Boschi and S. Pascoli, *Heavy neutral leptons from low-scale seesaws at the DUNE near detector*, arXiv:1905.00284 [INSPIRE].

[26] NA62 collaboration, *The beam and detector of the NA62 experiment at CERN*, 2017 *JINST* **12** P05025 [arXiv:1703.08501] [INSPIRE].

[27] M. Drewes, J. Hajer, J. Klaric and G. Lanfranchi, *NA62 sensitivity to heavy neutral leptons in the low scale seesaw model*, *JHEP* **07** (2018) 105 [arXiv:1801.04207] [INSPIRE].

[28] SHiP collaboration, *A facility to Search for Hidden Particles (SHiP) at the CERN SPS*, arXiv:1504.04956 [INSPIRE].

[29] S. Alekhin et al., *A facility to search for hidden particles at the CERN SPS: the SHiP physics case*, *Rept. Prog. Phys.* **79** (2016) 124201 [arXiv:1504.04855] [INSPIRE].

[30] SHiP collaboration, *Sensitivity of the SHiP experiment to Heavy Neutral Leptons*, *JHEP* **04** (2019) 077 [arXiv:1811.00930] [INSPIRE].

[31] J.C. Helo, M. Hirsch and Z.S. Wang, *Heavy neutral fermions at the high-luminosity LHC*, *JHEP* **07** (2018) 056 [arXiv:1803.02212] [INSPIRE].

[32] I. Boiarska et al., *Probing baryon asymmetry of the Universe at LHC and SHiP*, arXiv:1902.04535 [INSPIRE].

[33] K. Bondarenko, A. Boyarsky, M. Ovchynnikov and O. Ruchayskiy, *Sensitivity of the intensity frontier experiments for neutrino and scalar portals: analytic estimates*, *JHEP* **08** (2019) 061 [arXiv:1902.06240] [INSPIRE].

[34] E.J. Chun et al., *Sensitivity of lepton number violating meson decays in different experiments*, *Phys. Rev.* **D 100** (2019) 095022 [arXiv:1908.09562] [INSPIRE].

[35] M. Shaposhnikov, *A possible symmetry of the nuMSM*, *Nucl. Phys.* **B 763** (2007) 49 [hep-ph/0605047] [INSPIRE].

[36] J. Kersten and A.Yu. Smirnov, *Right-handed neutrinos at CERN LHC and the mechanism of neutrino mass generation*, *Phys. Rev.* **D 76** (2007) 073005 [arXiv:0705.3221] [INSPIRE].

[37] K. Moffat, S. Pascoli and C. Weiland, *Equivalence between massless neutrinos and lepton number conservation in fermionic singlet extensions of the Standard Model*, arXiv:1712.07611 [INSPIRE].

[38] E.K. Akhmedov, V.A. Rubakov and A.Yu. Smirnov, *Baryogenesis via neutrino oscillations*, *Phys. Rev. Lett.* **81** (1998) 1359 [hep-ph/9803255] [INSPIRE].

[39] L. Canetti, M. Drewes, T. Frossard and M. Shaposhnikov, *Dark matter, baryogenesis and neutrino oscillations from right handed neutrinos*, *Phys. Rev.* **D 87** (2013) 093006 [arXiv:1208.4607] [INSPIRE].

JHEP04(2020)005

[40] R.N. Mohapatra, *Mechanism for understanding small neutrino mass in superstring theories*, *Phys. Rev. Lett.* **56** (1986) 561 [INSPIRE].

[41] R.N. Mohapatra and J.W.F. Valle, *Neutrino mass and baryon number nonconservation in superstring models*, *Phys. Rev.* **D 34** (1986) 1642 [INSPIRE].

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

[43] E.K. Akhmedov, M. Lindner, E. Schnapka and J.W.F. Valle, *Dynamical left-right symmetry breaking*, *Phys. Rev.* **D 53** (1996) 2752 [hep-ph/9509255] [INSPIRE].

[44] G. Anamiati, M. Hirsch and E. Nardi, *Quasi-Dirac neutrinos at the LHC*, *JHEP* **10** (2016) 010 [arXiv:1607.05641] [INSPIRE].

[45] M. Drewes, J. Klarić and P. Klose, *On lepton number violation in heavy neutrino decays at colliders*, *JHEP* **11** (2019) 032 [arXiv:1907.13034] [INSPIRE].

[46] C. Arbelaéz, C. Dib, I. Schmidt and J.C. Vasquez, *Probing the Dirac or Majorana nature of the Heavy Neutrinos in pure leptonic decays at the LHC*, *Phys. Rev.* **D 97** (2018) 055011 [arXiv:1712.08704] [INSPIRE].

[47] C.O. Dib, C.S. Kim and K. Wang, *Signatures of Dirac and Majorana sterile neutrinos in trilepton events at the LHC*, *Phys. Rev.* **D 95** (2017) 115020 [arXiv:1703.01934] [INSPIRE].

[48] P. Hernández, J. Jones-Pérez and O. Suarez-Navarro, *Majorana vs. Pseudo-Dirac Neutrinos at the ILC*, *Eur. Phys. J.* **C 79** (2019) 220 [arXiv:1810.07210] [INSPIRE].

[49] G. Cvetič, C. Dib and C.S. Kim, *Probing Majorana neutrinos in rare* $\pi^+ \to e^+e^+\mu^-\nu$ *decays*, *JHEP* **06** (2012) 149 [arXiv:1203.0573] [INSPIRE].

[50] A.B. Balantekin, A. de Gouvêa and B. Kayser, *Addressing the Majorana vs. Dirac question with neutrino decays*, *Phys. Lett.* **B 789** (2019) 488 [arXiv:1808.10518] [INSPIRE].

[51] H.K. Dreiner, H.E. Haber and S.P. Martin, *Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry*, *Phys. Rept.* **494** (2010) 1 [arXiv:0812.1594] [INSPIRE].

[52] C. Giunti and C. W. Kim, *Fundamentals of neutrino physics and astrophysics*, (2007).

[53] T. Takagi, *On an algebraic problem related to an analytical theorem and Fejer and on an allied theorem of Landau*, *Japanese J. Math.* **1** (1927) 83.

[54] R. Sachs, *Interference phenomena of neutral K mesons*, *Ann. Phys.* **22** (1963) 239.

[55] C. Giunti, C.W. Kim, J.A. Lee and U.W. Lee, *On the treatment of neutrino oscillations without resort to weak eigenstates*, *Phys. Rev.* **D 48** (1993) 4310 [hep-ph/9305276] [INSPIRE].

[56] M. Beuthe, *Oscillations of neutrinos and mesons in quantum field theory*, *Phys. Rept.* **375** (2003) 105 [hep-ph/0109119] [INSPIRE].

[57] E.K. Akhmedov and J. Kopp, *Neutrino Oscillations: quantum Mechanics vs. Quantum Field Theory*, *JHEP* **04** (2010) 008 [*Erratum ibid.* **10** (2013) 052] [arXiv:1001.4815] [INSPIRE].

[58] E.K. Akhmedov and A.Yu. Smirnov, *Neutrino oscillations: Entanglement, energy-momentum conservation and QFT*, *Found. Phys.* **41** (2011) 1279 [arXiv:1008.2077] [INSPIRE].

[59] E. Akhmedov, *Quantum mechanics aspects and subtleties of neutrino oscillations*, in the proceedings of the *International Conference on History of the Neutrino: 1930–2018*, September 5–7, Paris, France (2018), arXiv:1901.05232 [INSPIRE].

JHEP04(2020)005

[60] S. Eijima, M. Shaposhnikov and I. Timiryasov, *Parameter space of baryogenesis in the νMSM*, *JHEP* **07** (2019) 077 [arXiv:1808.10833] [INSPIRE].

[61] D. Boyanovsky, *Nearly degenerate heavy sterile neutrinos in cascade decay: mixing and oscillations*, *Phys. Rev.* **D 90** (2014) 105024 [arXiv:1409.4265] [INSPIRE].

[62] G. Cvetič, C.S. Kim, R. Kogerler and J. Zamora-Saa, *Oscillation of heavy sterile neutrino in decay of $B \to \mu e \pi$*, *Phys. Rev.* **D 92** (2015) 013015 [arXiv:1505.04749] [INSPIRE].

[63] S. Antusch, E. Cazzato and O. Fischer, *Resolvable heavy neutrino–antineutrino oscillations at colliders*, *Mod. Phys. Lett.* **A 34** (2019) 1950061 [arXiv:1709.03797] [INSPIRE].

[64] A. Das, P.S.B. Dev and R.N. Mohapatra, *Same sign versus opposite sign dileptons as a probe of low scale seesaw mechanisms*, *Phys. Rev.* **D 97** (2018) 015018 [arXiv:1709.06553] [INSPIRE].

[65] G. Cvetič, A. Das and J. Zamora-Saá, *Probing heavy neutrino oscillations in rare W boson decays*, *J. Phys.* **G 46** (2019) 075002 [arXiv:1805.00070] [INSPIRE].

[66] P.S. Bhupal Dev, R.N. Mohapatra and Y. Zhang, *CP violating effects in heavy neutrino oscillations: implications for colliders and leptogenesis*, *JHEP* **11** (2019) 137 [arXiv:1904.04787] [INSPIRE].

[67] G. Cvetič, A. Das, S. Tapia and J. Zamora-Saá, *Measuring the heavy neutrino oscillations in rare W boson decays at the Large Hadron Collider*, *J. Phys.* **G 47** (2020) 015001 [arXiv:1905.03097] [INSPIRE].

[68] S. Tapia and J. Zamora-Saá, *Exploring CP-Violating heavy neutrino oscillations in rare tau decays at Belle II*, *Nucl. Phys.* **B 952** (2020) 114936 [arXiv:1906.09470] [INSPIRE].

[69] A. Abada, C. Hati, X. Marcano and A.M. Teixeira, *Interference effects in LNV and LFV semileptonic decays: the Majorana hypothesis*, *JHEP* **09** (2019) 017 [arXiv:1904.05367] [INSPIRE].

[70] LHCb collaboration, *Search for Majorana neutrinos in $B^- \to \pi^+ \mu^- \mu^-$ decays*, *Phys. Rev. Lett.* **112** (2014) 131802 [arXiv:1401.5361] [INSPIRE].

[71] ATLAS collaboration, *Search for heavy Majorana neutrinos with the ATLAS detector in pp collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **07** (2015) 162 [arXiv:1506.06020] [INSPIRE].

[72] CMS collaboration, *Search for heavy Majorana neutrinos in $\mu^\pm \mu^\pm$+ jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *Phys. Lett.* **B 748** (2015) 144 [arXiv:1501.05566] [INSPIRE].

[73] CMS collaboration, *Search for heavy Majorana neutrinos in $e^\pm e^\pm$+ jets and $e^\pm$ $\mu^\pm$+ jets events in proton-proton collisions at $\sqrt{s} = 8$ TeV*, *JHEP* **04** (2016) 169 [arXiv:1603.02248] [INSPIRE].

[74] LEBC-EHS collaboration, *D meson production from 400 GeV/cpp Interactions*, *Phys. Lett.* **B 189** (1987) 476 [*Erratum ibid.* **B 208** (1988) 530] [INSPIRE].

[75] SHiP collaboration, *Measurement of associated charm production induced by 400 GeV/c protons*, CERN-SPSC-2017-033 (2017).

[76] K. Bondarenko, A. Boyarsky, D. Gorbunov and O. Ruchayskiy, *Phenomenology of GeV-scale Heavy Neutral Leptons*, *JHEP* **11** (2018) 032 [arXiv:1805.08567] [INSPIRE].

[77] SHiP collaboration, *Particle identification tools and performance in the SHiP experiment*, CERN-SHiP-NOTE-2017-002 (2017).

JHEP04(2020)005

[78] G. Ke et al., *Lightgbm: a highly efficient gradient boosting decision tree*, in *Advances in neural information processing systems 30*, I. Guyon et al. eds., Curran Associates, U.S.A. (2017).

[79] Microsoft Corporation, *LightGBM*, GitHub repository (2016).

[80] J. Neyman and E.S. Pearson, *On the problem of the most efficient tests of statistical hypotheses*, *Phil. Trans. Roy. Soc. Lond.* **A231** (1933) 289.

[81] S. S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, *Annals Math. Statist.* **9** (1938) 60.

[82] M. Drewes, *On the minimal mixing of heavy neutrinos*, arXiv:1904.11959 [INSPIRE].

[83] F. James, *Monte-Carlo phase space*, CERN-68-15 (1968).

[84] P. Ilten, *Electroweak and Higgs measurements using tau final states with the LHCb Detector*, Ph.D. thesis, University Collge Dublin, Dublin, Ireland (2013), arXiv:1401.4902 [INSPIRE].

[85] S. Algeri, J. Aalbers, K. Dundas Moråand J. Conrad, *Searching for new physics with profile likelihoods: Wilks and beyond*, arXiv:1911.10237 [INSPIRE].

JHEP04(2020)005