Probing new physics in the laboratory and in space

Ovchynnikov, M.

Probing new physics in the laboratory and in space

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof. dr. ir. H. Bijl
volgens besluit van het college voor promoties
te verdedigen op 14 december 2021
klokke 12:30 uur

door

Maksym Ovchynnikov

geboren te Kiev (Oekraïne)
in 1994

Promotores:	Prof. dr. Alexey Boyarsky
	Prof. dr. Ana Achúcarro
Promotiecommissie:	Prof.dr. J. Aarts
	Dr. D.F.E. Samtleben
	Prof. fr. K.E. Schalm
	Prof. dr. N. Serra (University of Zurich, Switzerland)
	Prof. dr. O. Ruchayskiy (University of Copenhagen, Denmark)

Casimir PhD series Delft-Leiden 2021-37
ISBN 978-90-8593-502-5
An electronic version of this thesis can be found at https://openaccess.leidenuniv.nl

The cover shows the Big Bang Nucleosynthesis reactions framework, which may be affected by currently undiscovered particles, and the SHiP experiment, the proposed flagman to search for heavy new physics particles in laboratory.

Contents

Chapter 1

Introduction

In the twentieth century, attempts to explain subatomic physics phenomena combined with two scientific revolutions – special relativity and quantum mechanics – have resulted in the development of the Standard Model of particle physics (SM). The SM successfully describes all the particle physics processes that we observe at accelerators with the help of only 17 particles and interaction between them, that is based on the $SU(3) \otimes SU(2) \otimes U(1)$ gauge group, see Fig. 1.1. The SM has passed a number of precision tests [1–3]. The last stage of the confirmation of SM was the discovery of the Higgs boson at Large Hadron Collider (LHC) [4, 5] and subsequent tests of its properties, which confirmed that it is exactly the same particle as predicted by the SM [6].

The SM is a closed theory and may be used as an effective theory that makes extremely accurate predictions up to the Planck scale. However, it is not complete. There are a few observational phenomena establishing that SM has to be extended, probably by adding new particles. These phenomena constitute the beyond the SM problems (or BSM problems).

The BSM problems are:

- **Neutrino oscillations**: measurements of solar neutrino flux, as well as observations of atmospheric and collider neutrino interactions, suggest that neutrinos may change their flavor – the phenomenon called neutrino oscillations. The oscillations may occur if neutrinos have different masses. This is not possible in SM, in which neutrinos are massless, but may be resolved by adding interactions with new particles.
- **Dark matter (DM)**: Astrophysical and cosmological observations suggest that most of the mass of the Universe is dominated by a specific type of matter that does not interact with light – the dark matter. The only natural candidate for the role of DM in SM is the SM active neutrino. However, experimental restrictions on neutrino masses make this scenario impossible.

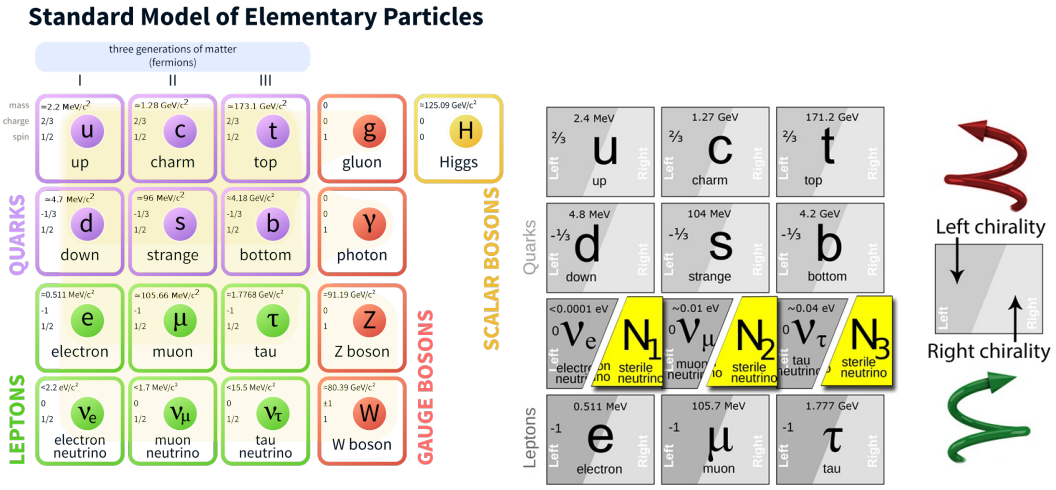

Figure 1.1: *Left panel*: The particle content of the Standard model. Twelve elementary particles of matter – leptons, corresponding lepton neutrinos, and quarks – interact with each other via electromagnetic, weak, and strong interactions. The interaction mediators are, correspondingly, a photon, W, Z bosons, and gluons. All of the matter elementary particles except for neutrinos, and W, Z bosons obtain their mass because of the Higgs mechanism. *Right panel*: an extension of the Standard Model with three electrically neutral fermion particles called Heavy Neutral Leptons N. These particles may be capable of solving all three beyond the Standard model phenomena simultaneously.

- **Baryon asymmetry of the Universe**: within the SM, it is impossible to generate the observed asymmetry between matter and antimatter, which is of an order of $10^9 + 1$ baryons per one 10^9 antibaryons.

In order to resolve the BSM problems, one may construct an extension of SM with some new physics particles that are responsible for these phenomena. Unfortunately, they do not provide unique information about the properties of these hypothetical particles, such as spin, mass, and strength of the interaction with different SM particles. As a result, a number of BSM extensions that may resolve the BSM problems is "degenerate": based on the experimental data, we do not have a clear way to choose a particular extension that is realized in nature. This situation is different from the case of the construction of the SM, for which the guideline was provided by a combination of theoretical arguments such as the unitarity and the gauge symmetry, and signatures from the particle physics experiments, that allow determining properties of particles in a unique way.

Currently, there is no signature that tells us uniquely about the properties of new physics particles. Given that a lot of SM extensions may be equally responsible for the resolution of BSM problems, one needs a way to constrain as many models as possible.

1.1 Portals

Non-observation of new physics particles at particle physics experiments may be explained by two reasons: either they are too heavy to be produced at the currently reachable energies, or they are too feebly interacting, such that the intensity of events at the experiment is insufficient to observe them. Further, we will consider the second class of the particles. They are called *Feebly Interacting Particles, or FIPs*.

The FIPs may be directly responsible for the BSM phenomena or serve as mediators between the dark sector and the SM. It is convenient to classify BSM extensions with FIPs by the properties of the particles-mediators:

1. Scalar portal:
$$\mathcal{L} = \alpha_1 H^\dagger H S + \alpha_2 H^\dagger H S^2, \tag{1.1.1}$$
with H being the SM Higgs doublet, S being a new scalar particle, while $\alpha_{1,2}$ real couplings. Phenomenologically, the scalar S interacts with SM particles in the same way as a light Higgs boson, but the matrix element of any process is suppressed by the mixing angle $\theta \ll 1$. S may play a role of a mediator for the dark sector [7], or be responsible for the inflation by playing the role of an inflaton [8], if having mass in GeV scale.

2. Neutrino portal:
$$\mathcal{L} = F_{\alpha I} \bar{L}_\alpha \tilde{H} N_I + \text{N mass term}, \tag{1.1.2}$$
where $\tilde{H} = i\sigma_2 H^*$ is the Higgs doublet in the conjugated representation, $L_\alpha = \begin{pmatrix} l_\alpha \\ \nu_\alpha \end{pmatrix}$ is the left SM lepton doublet, N_I Heavy Neutral Leptons (or HNLs), and $F_{\alpha I}$ complex couplings. Phenomenologically, HNLs interact similarly to a SM neutrino ν_α but suppressed by the mixing angle $U_\alpha \ll 1$. HNLs may resolve all of the BSM problems. In particular, to explain neutrino oscillations and the baryon asymmetry of the Universe, we need at least two HNLs with highly degenerate masses, while for DM one needs a long-lived HNL with mass in $\mathcal{O}(\text{keV})$ range. A model with such three HNLs is called Neutrino Minimal Standard Model, or νMSM [9, 10].

3. Vector portal:
$$\mathcal{L} = \frac{\epsilon}{2} F_{\mu\nu} V^{\mu\nu}, \quad \text{or} \quad \mathcal{L} = \epsilon J^\mu V_\mu \tag{1.1.3}$$
where $F_{\mu\nu}$ is the strength tensor of the gauge field associated with the $U_Y(1)$ gauge group, $V^{\mu\nu}$ is the strength associated with the new vector field V_μ, and J_μ is the conserved SM current (for instance, $B - 3L$ current).

1.2 LHC and dedicated accelerator experiments

During the last few years, particle physics experiments with large events intensity have been proposed to probe FIPs. They are called **Intensity frontier experiments**.

FIPs may interact with SM particles in different ways. In dependence on the type of interaction, different kinds of searching may be preferable. Based on the search type, the Intensity frontier experiments may be classified as follows:

1. **Prompt or displaced visible decays of FIPs:** a possible method to search for unstable FIPs that decays into electrically charged particles is to produce them in collisions of SM particles and then detect their decays into SM particles. Such kind events must be distinguished from pure SM events. For instance, an event with the HNL decay $N \to \mu^- \pi^+$ may be mimicked by a long-lived μ^- and π^+ produced outside the decay volume in some SM process, and whose trajectories closely intersect at one point inside the decay volume. The background reduction is typically reached by preventing the outer particles from reaching the decay volume: either by placing the decay volume far from the collision point (such that SM particles decay before reaching the decay volume), by placing SM particles absorbers/deflectors between the collision point, or by imposing specific events selection criteria that minimize the SM background. Examples include: prompt and displaced searches at the LHC, especially during its high-luminosity phase [11–16]; LHC-based experiments – FASER/FASER2 [17], MATHUSLA [18], Codex-b [19]; experiments at extracted beams – SHiP [20], NA62 in the dump mode (SPS beam at CERN), DUNE near detector [21], DarkQuest [22] (Fermilab).

2. **Rare SM decays:** if decays of FIPs cannot be distinguished from some rare SM processes, one may search for an excess of the corresponding events over the yield predicted by SM. Such rare decays are searched, e.g., at meson factories. Examples include: rare decays of mesons – $B \to K\mu\mu$ at LHCb [23], BaBar [24], and Belle II [25], $K \to \pi\nu\nu$ at NA62 in the kaon mode [26].

3. **Events with missing energy/momentum:** Another class of events is common if FIPs leave detectors invisibly, for instance decaying into uncharged particles. Such type of events may be characterized by a missing energy/momentum. Examples are NA64 [27, 28], Belle, BaBar, which search for the process e + target $\to e$ + missing energy, and events of the type $p + p \to$ jet + missing p_T at the LHC [29–31].

4. **Scatterings of new physics particles:** If being stable, FIPs may still be detected via their scatterings off matter. In this case, the production of FIPs at the laboratory is not necessary, as they could have been generated in the Early Universe and

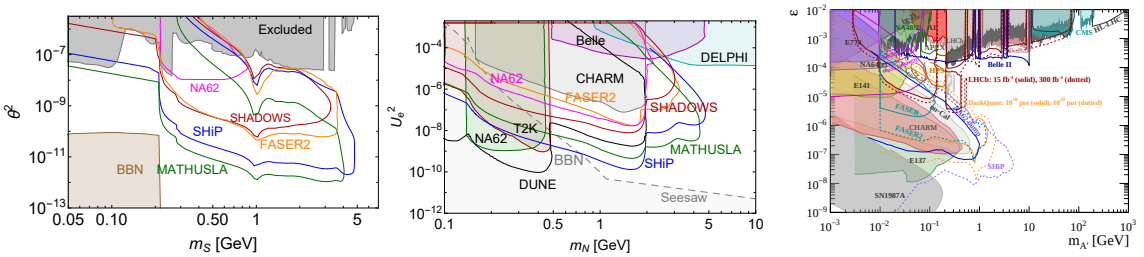

Figure 1.2: Sensitivity reach of Intensity frontier experiment to decays of portal particles: dark scalars (1.1.1) that have zero quartic coupling $HHSS$ (the left panel), Heavy Neutral Leptons (1.1.2) that mix purely with ν_e (the middle panel), and dark photons (1.1.3) (the right panel). The figure for the dark photon and sensitivity contours for HNLs and scalars are given from [36]. The BBN bound for HNLs is reproduced from [37].

surround us today. This may be the case of dark matter particles. Examples of experiments that search for this signature include recently approved SND@LHC [32] and FASERν [33], DUNE near detector, and direct DM detection experiments such as CRESST [34, 35].

The combined sensitivity reach of these experiments to different portal models is shown in Fig. 1.2. We also show there the parameter space excluded by past experiments.

1.2.1 Qualitative comparison of different experiments

Using Fig. 1.2, we may formally compare the potential of different experiments to probe new physics particles. In addition, in order to study principal limitations and advantages of the given experimental setup to probe different new physics models, it would be useful to have a qualitative understanding of the features of the sensitivity.

> The sensitivity of the given experiment to FIPs may be estimated using Monte-Carlo simulations of the number of events. However, this method has several limitations when comparing the sensitivity reach of different experiments.

Indeed, a simulation is typically a "black box" which does not allow us to understand the characteristic features of the sensitivity curve. In addition, it typically costs a huge amount of time and requires a lot of computational resources, which becomes crucial if many simulations are required. This is the case, for instance, during the optimization stage of the experiment, when its design is changed. Another situation is when we change parameters of the FIP model, which requires a new simulation each time.

> A clear, fully controlled estimate of the sensitivity within a factor of few is provided by semi-analytic calculations [9].

Let us consider for instance Intensity frontier experiments that search for visible decays of FIPs X. The number of events at these experiments is given by

$$N_{\text{events}} \approx \sum_i N_i \cdot \text{Br}(i \to X) \cdot \epsilon^{(i)}_{\text{geom}} \cdot P^{(i)}_{\text{decay}} \cdot \text{Br}(X \to \text{vis}) \cdot \epsilon_{\text{decay}} \cdot \epsilon_{\text{rec}} \tag{1.2.1}$$

Here, N_i is the total number of SM particle species i at the experiment (it may be for instance a particle from the incoming beam, or a secondary particle produced in collisions), $\text{Br}(i \to X)$ is the branching ratio of the process with i which leads to the production of X (it may be a decay or a scattering process), $\epsilon^{(i)}_{\text{geom}}$ is the geometric acceptance – the fraction of particles X that fly in the direction of the detector of the experiment. $P^{(i)}_{\text{decay}}$ is the decay probability:

$$\begin{aligned} P_{\text{decay}} &\approx \exp[-l_{\text{min}}/c\tau_X\gamma^{(i)}_X] - \exp\left[-l_{\text{max}}/c\tau_X\gamma^{(i)}_X\right] \approx \\ &\approx \begin{cases} l_{\text{fid}}/c\tau_X\gamma^{(i)}_X, & c\tau_X\gamma^{(i)}_X \gg l_{\text{max}} \\ \exp\left[-l_{\text{max}}/c\tau_X\gamma^{(i)}_X\right], & c\tau_X\gamma^{(i)}_X \lesssim l_{\text{min}} \end{cases} \end{aligned} \tag{1.2.2}$$

with $l_{\text{min}}, l_{\text{max}}$ being the minimal and maximal distance defining the decay volume ($l_{\text{fid}} = l_{\text{max}} - l_{\text{min}}$), τ_X the proper lifetime and $\gamma^{(i)}_X$ the mean γ factor. $\text{Br}(X \to \text{vis})$ is the branching ratio of decays of X into visible states (typically, a pair of charged particles). Finally, ϵ_{decay} is the decay acceptance – a fraction of decay products that travel in the direction of the detector, and ϵ_{rec} is the reconstruction efficiency – the fraction of events that are successfully reconstructed in detectors.

The sensitivity curve is defined by the condition $N_{\text{events}} > N_{\text{min}}$, where N_{min} is the number of events required for the detection. The limiting cases in (1.2.2) define the lower and upper bounds of the sensitivity of the experiments shown in Fig. 1.2. Using (1.2.2) together with (1.2.1), we may estimate these bounds analytically. We consider models in which the production and decays of X are controlled by the same coupling g of X to SM, such that $\text{Br}(i \to X), \tau_X^{-1} \propto g^2$. Assuming for simplicity that the production of X is dominated by one specific channel, the scaling of the lower and upper bounds of the sensitivity is given by the following formulas:

$$\begin{aligned} N_{\text{events,lower bound}} \propto g^4 \Rightarrow g^2_{\text{lower bound}} &\sim \sqrt{\frac{N_{\text{min}}}{N_{\text{events,lower bound}}\big|_{g=1}}} \equiv \\ &\equiv \chi_{\text{lower}} \times \sqrt{\frac{c\tau_X}{\text{Br}(i \to X)\text{Br}(X \to \text{vis})}}\Bigg|_{g=1} \end{aligned} \tag{1.2.3}$$

$$N_{\text{events,upper bound}} \propto g^2_{\text{upper}} \exp\left[-l_{\text{min}}/c\tau_X\gamma^{(i)}_X\right] \Rightarrow g^2_{\text{upper}} \sim \frac{\gamma^{(i)}_X}{l_{\text{min}}c\tau_X}\Bigg|_{g=1} \equiv \chi_{\text{upper}} \cdot \frac{1}{c\tau_X}\Bigg|_{g=1}, \tag{1.2.4}$$

Here, we have separated the experiment-independent parameters, which cancel out when comparing different experiments, from the experiment-specific quantities:

$$\chi_{\text{lower}} \equiv \sqrt{\frac{N_{\text{min}} \cdot \gamma_X^{(i)}}{N_i \cdot \epsilon_{\text{geom}}^{(i)} \cdot l_{\text{fid}} \cdot \epsilon_{\text{decay}} \cdot \epsilon_{\text{rec}}}}, \quad \chi_{\text{upper}} \equiv \frac{\gamma_X^{(i)}}{l_{\text{min}}} \tag{1.2.5}$$

The lower and upper bounds of the sensitivity of the Intensity frontier experiments may be estimated with the help of several geometric parameters.

We are now ready to compare the lower and upper bounds of different experiments. We consider the following experiments: MATHUSLA and FASER2 at the LHC, and SHiP and SHADOWS at SPS, and restrict the masses of FIPs by the GeV range. Their parameters are summarized in Table 1.1.

Experiment	SHiP	SHADOWS	MATHUSLA	FASER2
$\sqrt{s}$, GeV	28	28	13000	13000
N_{PoT}	$2 \cdot 10^{20}$	$\sim 5 \cdot 10^{19}$	$2.2 \cdot 10^{17}$	$2.2 \cdot 10^{17}$
$\frac{l_{\text{min}}}{\text{m}}$	50	10	40	480
$\frac{\langle l_{\text{fid}} \rangle}{\text{m}}$	50	20	100	5
$\frac{\theta_{\text{det}}}{\text{rad}}$	$(0, 0.025)$	$(0.03, 0.09)$	$(0.48, 0.9)$	$(0, 2.1 \cdot 10^{-3})$

Table 1.1: Parameters of different Intensity frontier experiments: the beam CM energy $\sqrt{s}$, the total number of the proton-proton collisions expected during the working period N_{PoT}, the distance to the beginning of the decay volume l_{min}, the average length of the decay volume $\langle l_{\text{fid}} \rangle$, polar angle coverage of detectors θ_{det}.

We consider three different production channels: proton bremsstrahlung, that is important for dark photons and dark scalars, decays of D mesons, which are important for HNLs, and decays of B mesons, which are important for dark scalars and HNLs [38, 39], see Fig. 1.3. We adopt the description of the probabilities of these channels from [38, 39]. For particles from decays of mesons, we approximate the kinematic quantities such as γ_X and ϵ_{geom} by the corresponding quantities of mesons. This is a meaningful approximation since the angular distribution differs only by the quantity $\Delta\theta \simeq m_{B,D}/E_{B,D}$, which is typically much smaller than the angular coverage of the experiments of interest. We use the total amount and spectra of D, B mesons at the LHC provided by the FONLL package [40–43] and at SPS by [44]. Finally, we estimate ϵ_{decay} with the help of a simple Monte-Carlo simulation.[1]

Let us now highlight important points relevant for the comparison:

[1] We require decay products from two-body decays of particles with masses $m_{B/D}$ for decays from B/D mesons and 1 GeV for the production by bremsstrahlung to point to detectors.

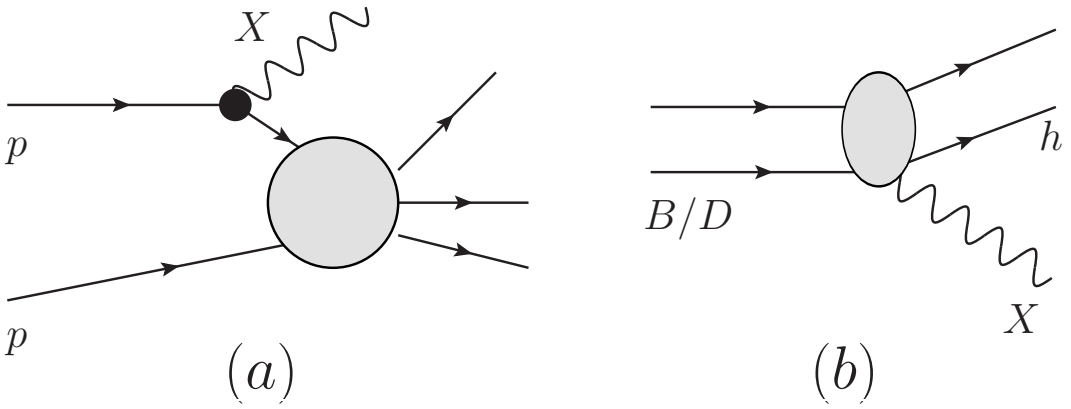

Figure 1.3: Typical production channels for a FIP X that has mass in $\mathcal{O}(\text{GeV})$ range: proton bremsstrahlung (the diagram (a)), and decays of B, D mesons into a FIP and other particle h (the diagram (b)).

1. Particles produced by the proton bremsstrahlung have small transverse momenta $p_T \lesssim \Lambda_{\text{QCD}}$ and hence are very collimated with respect to the beam axis. Therefore, off-axis experiments like SHADOWS and MATHUSLA do not have the sensitivity to this channel.

2. For the production from mesons, the invariant mass of collisions $\sqrt{s}$ must be much larger than the doubled mass of meson; otherwise, the meson production probability gets suppressed. As a result, at the SPS beam energy, the fraction of produced B mesons per one proton-proton collision is $\simeq$ few $\cdot\, 10^{-7}$, since $2m_B \approx 10$ GeV is not too far from $\sqrt{s_{\text{SPS}}} \approx 28$ GeV. At the LHC, $\sqrt{s} = 13$ TeV, and the probability is much higher, reaching 10^{-2}. For D mesons, there is no such significant difference, as their mass is significantly lower. However, the suppression at SPS is compensated by a much larger beam intensity.

3. The distribution of B, D mesons is collimated, although not so strong as compared to the bremsstrahlung. This leads to much lower (but non-zero) geometric acceptance for the off-axis experiments. As a result of these factors, the amounts of B mesons at SHiP, FASER2, and MATHUSLA are comparable, while at SHADOWS it is even suppressed.

4. Experiments located close to the beam collision point, such as SHADOWS, may have an advantage at the upper bound of the sensitivity even despite lower average momentum. On-axis experiments at the LHC such as FASER2, although being located far away from the collision point, are still competitive at the upper bound due to large average momenta of particles produced in energetic beams collisions in the far-forward direction.

The resulting parameters (1.2.5) are summarized in Table 1.2.

Experiment	SHiP	SHADOWS	MATHUSLA	FASER2
N_{min}	3	3	3	3
$N_B \cdot \epsilon^B_{\text{geom}}$	$8 \cdot 10^{13}$	$5 \cdot 10^{11}$	$3 \cdot 10^{13}$	10^{13}
$N_D \cdot \epsilon^D_{\text{geom}}$	$8 \cdot 10^{17}$	$2 \cdot 10^{16}$	$5 \cdot 10^{14}$	$2 \cdot 10^{14}$
$N_{\text{PoT}} \cdot \epsilon^{\text{brem}}_{\text{geom}}$	10^{20}	–	–	$2 \cdot 10^{16}$
$\epsilon_{\text{decay, }B}$	0.4	< 0.4	$\mathcal{O}(1)$	$\mathcal{O}(1)$
$\epsilon_{\text{decay, }D}$	0.4	$\simeq 0.3$	$\mathcal{O}(1)$	$\mathcal{O}(1)$
$\epsilon_{\text{decay, brem}}$	$\mathcal{O}(1)$	–	–	$\mathcal{O}(1)$
$\chi_{\text{lower, }B}$	$4 \cdot 10^{-7}$	$6 \cdot 10^{-6}$	10^{-7}	$2 \cdot 10^{-6}$
$\chi_{\text{lower, }D}$	$3 \cdot 10^{-9}$	$8 \cdot 10^{-8}$	$4 \cdot 10^{-8}$	$5 \cdot 10^{-7}$
$\chi_{\text{lower, brem}}$	10^{-10}	–	–	10^{-7}
$\chi_{\text{upper, }B}$	2	6	0.1	3
$\chi_{\text{upper, }D}$	1	2	0.03	2
$\chi_{\text{upper, brem}}$	3	–	–	2

Table 1.2: Potential reach of different Intensity frontier experiments as predicted by Eqns. (1.2.5). The relevant parameters $N_{B,D}, \epsilon_{\text{geom}}, \epsilon_{\text{decay}}$. The minimal number of events required for the discovery, N_{min}, corresponds to the assumption of absence of background at 95% CL.

From the discussion, we conclude that SHiP is the most powerful "non-compromised" experiment proposed to search for FIPs with masses below 5 GeV.

1.3 Astrophysical and cosmological probes: defining the parameter space of interest for accelerator experiments

Searches for FIPs at accelerator experiments are restricted by short-lived particles. The upper bound on lifetimes that may be constrained is model-dependent. For instance, for HNLs it may be as large as $\tau_N \simeq 10^{-2}$ s at masses $m_N > 0.5$ GeV, see Fig. 1.4. It is therefore important to define the target parameter space of FIPs to be probed by the accelerator experiments.

Signatures that may be sensitive to long-lived FIPs are cosmological and astrophysical observables.

In the Early universe or inside a dense medium of a supernova, the intensity of collisions was much higher than at colliders. Therefore, despite tiny couplings to the SM particles, FIPs may be produced in amounts significant enough to potentially alter the cosmological observables.

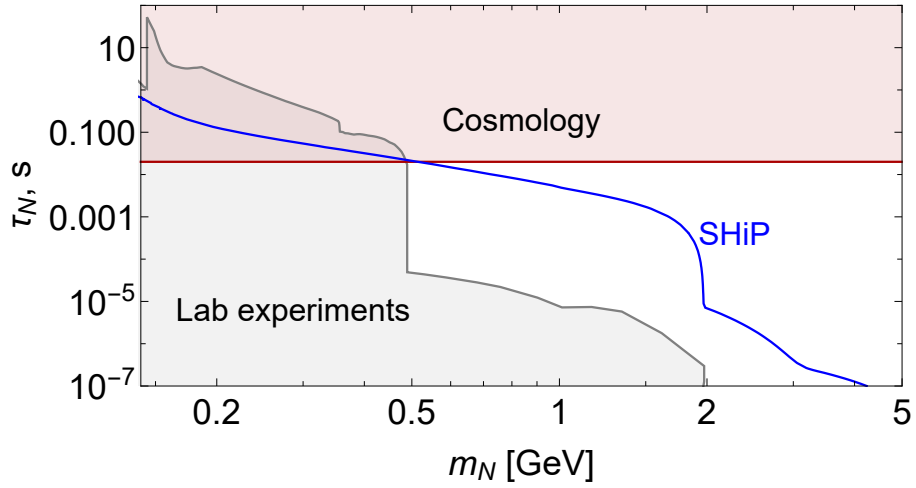

Figure 1.4: The complementarity of FIP signatures coming from past laboratory experiments and cosmological observations, demonstrated using a particular FIP example – HNLs with the pure e mixing. Laboratory experiments are able to probe the parameter space of relatively short-lived particles, whereas cosmological observations such as Big Bang Nucleosynthesis and Cosmic Microwave Background may rule out long-lived HNLs. Together, they define the target parameter space for future experiments such as SHiP.

Cosmological and astrophysical probes include: abundances of primordial light elements synthesized during Big Bang Nucleosynthesis (BBN); the spectrum of Cosmic Microwave Background (CMB); galactic X-ray lines; supernova explosion observation; primordial magnetic fields. The experimental status of these observations is different. For instance, while the CMB spectrum is measured very accurately, the supernova evolution is probed poorly, as it is based on only one observation of the explosion of SN1987A.

Predictions for these observables based on the standard cosmological model (Λ Cold Dark Matter, or ΛCDM) and only SM particles populating the plasma do not contradict the measurements. By adding a new particle, we may break this agreement.

Cosmological probes that may constrain the shortest FIPs lifetimes are BBN and CMB. They may be affected by FIPs with lifetimes as small as $\tau_{\rm FIP} \simeq 0.01$ s.

The earliest probe from the Early Universe is BBN. BBN is sensitive to the evolution of the Universe at cosmological times as small as $t \simeq \mathcal{O}(1\text{ s})$, when neutrons decouple from the primordial plasma (see subsection 1.3.1 below). As their decoupling is not instantaneous, the population of neutrons at decoupling is sensitive to the dynamics at even earlier times. In particular, it may constrain particles with lifetimes as small as $\tau_{\rm FIP} \gtrsim 0.01$ s (see Sec. 1.3.1). Another probe is CMB: although it is associated to the epoch at which electrons and protons get bounded into the Hydrogen atom, its characteristic features (i.e., the angular horizon or the damping scale) depend on the primordial helium abundance and the effective number of

relativistic degrees of freedom, $N_{\rm eff}$, which, in their turn, are determined by the evolution of the Universe at time scales as small as $t \simeq 0.1$ s. Below, we discuss them in more detail.

1.3.1 BBN

There are two cosmic sources of chemical elements: the evolution of stars, and the primordial evolution of the Universe, which is assumed initially to be hot and/or ultra-dense, thus having conditions for nucleosynthesis. The primordial abundances may be measured in star-poor regions. If the primordial nuclear synthesis was in equilibrium, we would expect that the present Universe is dominated by iron, as it has the largest binding energy per nucleon and therefore is thermodynamically favorable. However, the measurements have indicated the existence of only light elements up to ^{7}Li. This leads us to the conclusion that nuclear reactions were not in thermodynamical equilibrium.

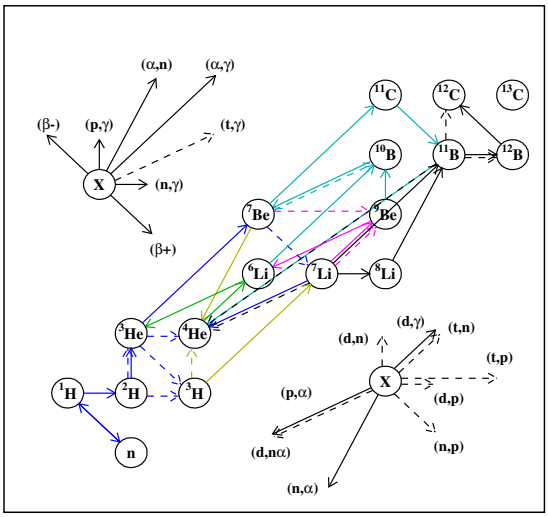

Figure 1.5: The chain of primordial nuclear reactions adopted from [45].

A natural explanation of this feature has been given by Gamow in 40-ties. He assumed that the Universe was hot and radiation dominated during BBN rather than cold and ultradense, which implies a small baryon-to-photon ratio, $\eta_B = n_B/n_\gamma \ll 1$. In the thermal medium, two-particle reactions are much more probable than other multi-body processes. Therefore, the first reaction in the nuclear chain is the deuterium synthesis process $p + n \to d + \gamma$, see Fig. 1.5. At temperatures larger than the binding energy of deuterium nuclei, $T > \Delta_D$, BBN was not efficient because synthesized nuclei were immediately destroyed by photons (that have average energy $E_\gamma \simeq T$). The temperature of the onset of nuclear reactions, $T_{\rm BBN}$, may be estimated as a temperature at which the number density of photons with energies higher than the deuterium binding energy, $E_\gamma > \Delta_D = 2.2$ MeV, became comparable with the number density of nucleons:

$$n_{\gamma, E_\gamma > \Delta_D}(T_{\rm BBN}) \simeq n_B(T_{\rm BBN}) = \eta_B n_\gamma(T_{\rm BBN}), \tag{1.3.1}$$

which implies $T_{\rm BBN} \simeq \Delta_D/\ln(\eta_B^{-1}) \sim 0.1\Delta_D \simeq 100\,{\rm keV}$. At such low temperatures, however, heavy elements cannot be formed because of the Coulomb barrier. Indeed, the probability of the synthesis process $X_1 + X_2 \to X_3$, with X_i being some nucleus with mass and charge m_i and Z_i, is suppressed by the Coulomb repulsion factor

$$\sigma v \propto e^{-\eta}, \quad \eta = \frac{Z_1 Z_2 \alpha_{\rm EM}}{v(T)} = \frac{Z_1 Z_2 \alpha_{\rm EM}}{\sqrt{T}} \frac{\sqrt{m_1 m_2}}{\sqrt{m_1} + \sqrt{m_2}}, \tag{1.3.2}$$

The synthesis effectively does not occur if

$$\eta(T_{\rm BBN}) \gtrsim 1 \Rightarrow T_{\rm BBN} \lesssim T_{\rm coulomb} = \frac{m_1 m_2 Z_1^2 Z_2^2}{(\sqrt{m_1} + \sqrt{m_2})^2} \tag{1.3.3}$$

As a result, using $T_{\rm BBN} \simeq 100$ keV, we conclude that no nucleus heavier than ^{12}C may be synthesized efficiently (^{12}C is produced in the process ^{4}He $+^8$ Be $\to^{12}$ C $+ 2\gamma$, for which $T_{\rm coulomb} \simeq T_{\rm BBN}$).

The only robust measurements of primordial abundances are those of d and ^{4}He.

Indeed, the ^{3}He isotope is measured only in regions with high metallicity, and it is not possible to estimate the effect of the stellar evolution on its abundance [46]. The measurements of ^{7}Li [47] set the lower bound on the primordial abundance since ^{7}Li might be destroyed in low-metallicity stars.

The abundance of the primordial ^{4}He, Y_p, is measured by three ways: *(i)* the low-metallicity extragalactic method, according to which the helium abundance is measured in low-metallicity regions and then extrapolated to zero metallicity; *(ii)* the intergalactic method – measurements of Y_p in the low-metallicity extragalactic gas; and *(iii)* the CMB method, for which the helium abundance is extracted from the CMB damping tail. The current error in the determination of ^{4}He at 1σ is around 4% (see Sec. 3.1.1).

In the Standard Model BBN (or SBBN), the only free parameter is η_B.

Using the value of η_B measured from CMB, we find that the predictions of SBBN agree with measurements of d and ^{4}He [48]. Therefore, to understand the impact of FIPs on BBN, we should first learn SBBN.

The SBBN proceeds as follows. Above $T_{\rm BBN}$, we have only $p, n, e^\pm, \nu, \bar{\nu}, \gamma$ particles in the plasma. During the whole BBN, ultrarelativistic particles dominate the energy density of the Universe and determine the dynamics of the expansion of the Universe and hence the Hubble rate H. $e^\pm, \nu, \bar{\nu}$, in addition, keep protons and neutrons in thermal equilibrium by weak reactions

$$p + \bar{\nu} \leftrightarrow n + e^+, \quad n + \nu_e \leftrightarrow p + e \tag{1.3.4}$$

at large temperatures $T \gg 1$ MeV, such that $n_n/n_p = e^{-(m_n - m_p)/T}$. However, neutrons decouple at some temperature of order $T_{\rm dec} \simeq 1$ MeV, determined by the condition

$$\Gamma_{p \leftrightarrow n}(T_{\rm dec}) \simeq H(T_{\rm dec}), \tag{1.3.5}$$

where $\Gamma_{p \leftrightarrow n}$ is the $p \leftrightarrow n$ conversion rate. Afterwards, the population of neutrons evolves due to free decays only, $n \to e + p + \bar{\nu}_e$, with the neutron lifetime $\tau_n \approx 880$ s.

At temperatures around $T_{\rm BBN}$, the number density of photons with $E_\gamma < \Delta_D$ drops below n_B, and the synthesis starts. Among all light elements which may be synthesized, remind Eq. (1.3.3), helium has the largest binding energy per nucleon. Therefore, its abundance may be estimated from the assumption that all free neutrons got bound in ^{4}He:

$$Y_p = \left.\frac{m_{^4\rm He} n_n}{m_B (n_n + n_p)}\right|_{T=T_{\rm BBN}} \simeq \left.\frac{n_n/n_p}{n_n/n_p + 1}\right|_{T=T_{\rm dec}} \cdot e^{-t(T_{\rm BBN})/\tau_n} \tag{1.3.6}$$

The amounts of other elements is not possible to estimate analytically in accurate way, as the BBN dynamics is not equilibrium. In addition, the estimate of the helium abundance based on Eq. (1.3.5) is very sensitive to $T_{\rm dec}$, since it assumes the instant decoupling of neutrons and therefore $T_{\rm dec}$ enters the exponent in $n_n/n_p|_{T=T_{\rm dec}} = e^{-(m_n - m_p)/T_{\rm dec}}$. Therefore, in order to obtain precise values predicted by SBBN, one has to solve the system of Boltzmann equations for nuclear abundances (see, e.g. [48] and references therein).

1.3.1.1 How short-lived FIPs affect BBN

Let us now assume that in addition to SM particles we also have FIPs in the plasma. We are interested in the lower bound on lifetimes for which BBN may be affected. Therefore, we consider "small" lifetimes $\tau_{\rm FIP} \ll 1$ s, where the time scale is given by the rough time of the neutron/neutrino decoupling. The most part of such short-lived FIPs decay at times $t \lesssim \tau_{\rm FIP}$ (i.e. at temperatures when neutrinos and neutrons are in perfect equilibrium), and thus does not affect BBN. The BBN is changed by the residual population of FIPs that survive at times $t \gtrsim \tau_{\rm FIP}$, which is exponentially suppressed.[2] FIPs may affect the dynamics of BBN via the following mechanisms:

1. Change the dynamics of the expansion of the Universe. Before FIPs decay, the energy density of heavy FIPs with $m_{\rm FIP} \gg T$ may contribute significantly to the total energy density of the Universe, as the ratio of the energy density of non-relativistic relics to the energy density of SM species scales as $\rho_{\rm FIP}/\rho_{\rm SM} \propto m_{\rm FIP}/T$. The largeness of the ratio $m_{\rm FIP}/T$ may partially compensate the exponential suppression of the population of FIPs at times $t \gtrsim \tau_{\rm FIP}$, and effects of the energy density may not be neglected. Decays of remaining FIPs into neutrinos and EM particles reheat them, in general

[2] Further, we assume that the value of η_B after the disappearance of FIPs from the plasma is given by the value predicted by CMB. This assumption is reasonable since we cannot extract η_B from earlier measurements.

differently. This leads to a change of the number of the ultrarelativistic degrees of freedom, N_{eff}:

$$N_{\text{eff}} = \frac{4}{7}\left(\frac{11}{4}\right)^{\frac{4}{3}}\frac{\rho_\nu}{\rho_\gamma} \tag{1.3.7}$$

and thus the Hubble rate at times $t > \tau_{\text{FIP}}$, which affects the dynamics of BBN.

2. Change the $p \leftrightarrow n$ conversion reactions. This may happen for instance if FIPs decay into neutrinos at temperatures $\mathcal{O}(1\text{ MeV})$. Since neutrinos are not in perfect equilibrium at these temperatures [49], decays of FIPs change the shape of their distribution function, which affects the neutron-to-proton ratio via the conversion processes (1.3.4). Another example is long-lived mesons such as $\pi^\pm/K$, which, being produced by FIPs, convert $p \leftrightarrow n$ via strong interactions:

$$\pi^- + p \to n + \gamma, \quad \pi^+ + n \to p + \gamma, \quad K^- + p/n \to n/p + X \tag{1.3.8}$$

If FIPs decay into mesons, the lower bound on lifetimes that may be constrained by BBN comes from the effect of the meson-driven $p \leftrightarrow n$ conversion.

Indeed, the cross sections the processes (1.3.8) are many orders of magnitude larger than the weak $p \leftrightarrow n$ conversion cross-section:

$$\frac{\sigma^{\text{strong}}_{p\leftrightarrow n}}{\sigma^{\text{weak}}_{p\leftrightarrow n}} \sim \frac{m_p^{-2}}{G_F^2 T^2} \sim 10^{-16}\left(\frac{1\text{ MeV}}{T}\right)^2, \tag{1.3.9}$$

Therefore, if FIPs decay into mesons, even their tiny amount comparable with the baryon number density (i.e., much smaller than the neutrino number density) may significantly affect the dynamics of the n/p ratio. The situation is different for the effects via neutrino spectral distortions and the expansion of the Universe. In the latter cases, the energy density of FIPs has to contribute non-negligible to the energy density of the Universe, which requires a much larger number density of FIPs than in the case of mesons.

1.3.2 CMB

In the primordial Universe, protons, electrons and photons were connected to each other via EM interactions, constituting the EM plasma. The hydrogen atom synthesis process,

$$p + e \to H + \gamma, \tag{1.3.10}$$

was much less efficient than the dissociation process

$$\gamma + H \to p + e \tag{1.3.11}$$

driven by the plasma photons. However, as temperature dropped to values around $T \simeq 1$ eV (or redshifts $z_* \simeq 1100$), the amount of photons with energies large enough to dissociate the hydrogen dropped below the baryon number density n_B. As a result, electrons and protons got bounded into H, and the primordial plasma became transparent for photons. These primordial photons survive until our times in the form of CMB.

At large scales, the CMB spectrum is nearly homogeneous and isotropic and well-described by the Planck distribution with the temperature $T_{\text{CMB}} = 2.7255 \pm 0.0006$ K [50], which is one of the confirmations of the Big Bang theory. Interesting physics is hidden in its perturbations.

Let us introduce the variation of the temperature $\delta T(\mathbf{n}) = (T(\mathbf{n}) - T_{\text{CMB}})/T_{\text{CMB}}$, and expand it in terms of spherical harmonics Y_{lm}:

$$\delta T(\mathbf{n}) = \sum_{l,m} a_{lm} Y_{lm}(\theta, \phi), \tag{1.3.12}$$

where $\mathbf{n}(\theta, \phi)$ is the unit vector defining the direction on the sky. To characterize the inhomogeneities, we introduce the autocorrelation function

$$C(\theta) = \langle |\delta T(\mathbf{n}_1) \delta T(\mathbf{n}_2)| \rangle, \quad \cos(\theta) = \mathbf{n}_1 \cdot \mathbf{n}_2 \tag{1.3.13}$$

Using Eq. (1.3.12), we obtain

$$C(\theta) = \frac{T_{\text{CMB}}^2}{4\pi} \sum_l (2l+1) C_l P_l(\cos(\theta)) \tag{1.3.14}$$

Here, we have assumed that the coefficients a_{lm} satisfy the relation

$$\langle a_{lm}^* a_{l'm'} \rangle = C_l \delta_{mm'} \delta_{ll'} \tag{1.3.15}$$

The monopole component of the CMB, C_0, gives us the information about the CMB temperature T_{CMB}. The dipole component C_2 comes from the relative motion of the Solar system with respect to the CMB radiation frame, which results in the Doppler shift (with $\beta = v/c$ being the β factor of the relative motion)

$$T(\theta) \approx T_\gamma (1 + \beta \cos(\theta) + \mathcal{O}(\beta^2)) \tag{1.3.16}$$

Multi-poles $C_l, l \geq 2$ provide us information about primordial perturbations of the power spectrum. They are shown in Fig. 1.6.

The main characteristic features of the CMB anisotropy spectrum from Fig. 1.6 are inhomogeneities in the CMB temperature of order $\delta T/T_{\text{CMB}} \simeq 10^{-5}$, the presence of oscillations for $l \gtrsim 100$, their exponential suppression at scales $l \gtrsim 10^3$.

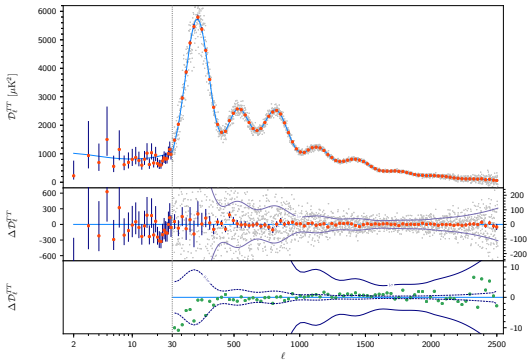

Figure 1.6: The CMB power spectrum measured by PLANCK 2018 [51]. The light blue curve corresponds to the best-fit spectrum in ΛCDM.

Qualitatively, the peaks at multipoles $l \gtrsim 100$ result from acoustic oscillations of the matter density. They originate from the competition between the pressure of photons and the gravitational pressure of matter which tries to form potential wells. There are three characteristics we can extract from the peaks:

1. Their relative height. Odd peaks correspond to compression of acoustic waves, even peaks to rarefaction. If we were to fix everything but increase the baryon density, the compression peaks (first, third, fifth, etc.) would increase in height relative to the rarefaction peaks (second, fourth, sixth, etc.). As a result, the ratio of the second to first peak amplitude tells us about the baryon density. The third peak is located at smaller scales when the Universe was more radiation-dominated. The abundance of DM alters when radiation domination stops and DM potential wells can grow. This then determines how much the sound waves can compress. The height of the third peak relative to the first or second peak thus tells us about the time of matter-radiation equality (and therefore matter density).

2. Their position as a function of angular scale. The most prominent one is the first peak and corresponds to the sound horizon r_s – the distance sound waves have traveled before recombination:
$$r_\mathrm{s} = \int c_\mathrm{s} \frac{da}{a^2 H}, \tag{1.3.17}$$
where $c_\mathrm{s} = \left[3\left(1+\frac{3\rho_\mathrm{b}}{4\rho_\gamma}\right)\right]^{-1/2}$ is the velocity of the sound waves. The corresponding angular scale is $\theta_\mathrm{s} = r_\mathrm{s}/D_A$, where $D_A = \int_{z_\mathrm{eq}}^{0} dz/H(z)$. Since T_CMB (and thus ρ_γ) has been measured, the sound speed only depends on the baryon energy density. Based on measurements of the location of this peak, one can deduce that $\Omega_\mathrm{m} + \Omega_\Lambda = 1$, i.e., the Universe is flat.

3. Their damping at smaller angular scales. Photons have a mean free path, diffuse and thus mix hot regions with cold regions. This leads to a dampening of small-scale fluctuations that goes as $\exp(-r_{\rm d}^2/\lambda^2)$, where $r_{\rm d}$ is the damping scale:

$$r_{\rm d}^2 = \int \frac{{\rm d}a}{a^3 \sigma_{\rm T} n_e H} \left(\frac{R^2 + \frac{16}{15}(1+R)}{6(1+R^2)} \right) , \tag{1.3.18}$$

where $R = 3\rho_{\rm b}/4\rho_\gamma$. Note here that $r_{\rm d} \propto H^{-1/2}$.

1.3.2.1 Impact of short-lived FIPs on CMB

Short-lived FIPs with $\tau_{\rm FIP} \ll 1$ s do not survive until the recombination and thus do not directly affect CMB. However, they may change populations of neutrinos and photons, and also affect the primordial helium abundance, which changes properties of CMB.

We may parametrize the first effect via a change in $N_{\rm eff}$, see Eq. (1.3.7). In general, it affects many parameters important for CMB: examples are sound horizon and damping scale $r_{\rm s}, r_{\rm d}$ discussed in the previous subsection, and the redshift $z_{\rm eq}$ of the radiation-matter equality. It is non-trivial to estimate the impact of the first effect on CMB, as it may be mimicked by variations of other cosmological parameters [52]. To characterize the less degenerate impact, it is possible to make a rescale of cosmological parameters which leaves $\theta_{\rm s} = r_{\rm s}/d_A, z_{\rm eq}$ invariant [53]. Under such a rescale, the remained impact of $N_{\rm eff}$ is on the damping scale, $\theta_d \propto (1 + 0.22 N_{\rm eff})^{\frac{1}{4}}$ (see also also [54]).

Even in these conditions, $N_{\rm eff}$ is still, however, degenerate with the other parameter changed by FIPs – the primordial helium abundance Y_p. The degeneracy appears since $r_{\rm d}$ depends on the amount of free electrons, which at low temperature is given by the proton abundance, $r_{\rm d} \sim n_e^{-1} \sim 1/\sqrt{1-Y_p}$. Therefore,

$$\theta_d \propto \frac{(1 + 0.22 N_{\rm eff})^{\frac{1}{4}}}{\sqrt{1-Y_p}} \tag{1.3.19}$$

The effect caused by a change in $N_{\rm eff}$ which is non-degenerate with ΛCDM parameters is the change of the CMB damping scale given by Eq. (1.3.18). There is, however, a degeneracy between $N_{\rm eff}$ and the helium abundance Y_p.

Marginalizing over the value of Y_p, the current bounds imposed by CMB on $N_{\rm eff}$ is $N_{\rm eff} = 2.89 \pm 0.62$ at 2σ [55].

1.4 Summary

In the lack of explanations of several phenomena in particle physics, the Standard model requires to be extended, probably by adding new particles. However, from current observations, we do not have a clear guideline on the choice of this extension. Therefore, it is reasonable to search for new particles in as much model-independent way as possible.

The work described in this thesis is devoted to a study of two different signatures of new physics particles: their search at future particle physics experiments, and their impact on different cosmological observables. They are complementary to each other, with the first one allowing to probe the parameter space of short particle lifetimes, and the second one constraining large lifetimes. In Chapter 2, we study different signatures with new physics particles at laboratory experiments: their decays the past experiment CHARM (Sec. 2.1) and at displaced vertices at the LHC (Sec. 2.2), and their scatterings at SND@LHC (Sec. 2.3). In Chapter 3, we study the impact of short-lived particles on cosmological observations. Namely, we consider the bounds on particles decaying hadronically on BBN in Sec. 3.1, and the effect of short-lived particles on N_{eff} in Sec. 3.2. We then apply these results to derive the constraints on HNLs in Sec. 3.3.1.

Chapter 2

Accelerator and laboratory searches[1]

In this chapter, we consider searches for FIPs at accelerator experiments. We first re-analyze the bounds from the past experiment CHARM on HNLs, demonstrating for the first time that the actual bounds are stronger by a factor of few (for the e/μ mixing) to a few orders of magnitude (for the τ mixing), in dependence on the mixing pattern, see Sec. 2.1. Next, we consider the searches for FIPs using the displaced vertices scheme at the LHC, and in particular the search with muon trackers at CMS, see Sec. 2.2. Finally, we proceed to experiments that search FIPs via their scattering, and estimate the potential of SND@LHC to search for scatterings of Light Dark Matter particles off nucleons, see Sec. 2.3.

2.1 Bounds on HNLs from CHARM experiment

In order to define the target parameter space for Intensity Frontier experiments for a given model, we need to know constraints on it coming from past experiments.

Let us look closer at constraints on HNLs. The bounds on HNLs in the GeV mass range that mix purely with electron and tau neutrino flavors as reported in [36] are shown in Fig. 2.1.

For the e mixing, below the kaon mass HNLs may be produced in decays $K \to N + e$ of copiously produced kaons, and thus are severely constrained by kaon fabrics (T2K, NA62). Being combined with the parameter space excluded by BBN, they practically rule out light HNLs. To search for heavier HNLs, we need D ($m_N < m_{D_s} \approx 1.97$ GeV), B mesons ($m_N < m_{B_c} \approx 6.3$ GeV), or W/Z bosons ($m_N < m_Z \approx 91$ GeV) in order to produce them. The amounts of these particles at experiments are much lower than amounts of kaons, and constraints are much weaker. In particular, in the mass range $m_N < m_{D_s}$ the strongest current bound comes from an old experiment CHARM, which was an SPS-based

[1] Results of this chapter are presented in papers [56–58]. The main contribution of Maksym Ovchynnikov to them are analytic estimates, simulations, and the main idea in [58].

beam dump experiment which searched for displaced decays into a di-lepton pair:

$$N \to e^+e^-, \quad N \to \mu^+\mu^-, \quad N \to \mu^\pm e^\mp \tag{2.1.1}$$

Larger masses are constrained by DELPHI experiment, which was a e^+e^- collider at energies equal to m_Z.

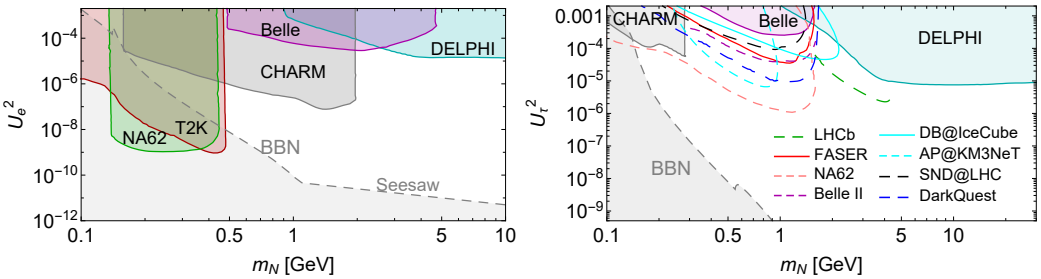

Figure 2.1: The parameter space of HNLs with the pure e (the left panel) and τ (the right panel) mixing. Constraints from the previous experiments – NA62, T2K, Belle, CHARM, DELPHI – are shown as reported in [36]. We do not show sub-dominant bounds coming from past experiments, such as NOMAD [59] and ArgoNeuT [60] for the τ mixing. For the pure τ mixing, we do not show the constraints imposed by the T2K experiment [61], since they are reported for non-zero couplings $U_{e/\mu}$ which dominate the production Constraints from the CHARM experiment are taken from the literature [62, 63], while our re-analysis for them is shown in Fig. 2.6. The light gray domain corresponds to couplings that are either excluded by BBN [37, 64] or too small to provide active neutrino masses. For the pure τ mixing, we also show sensitivities of the next generation Intensity Frontier experiments (see text for details). In cyan, we show HNL parameter space that may be probed by neutrino observatories: the solid line shows the sensitivity of IceCube to the "double bang" signature from [65], while the dashed line corresponds to the sensitivity of KM3NeT to decays of HNLs produced in the atmosphere, see text and Sec. 2.1.5 for details.

For the τ mixing, constraints at mass $m_N < m_{D_s}$ are very different, being much weaker than for the e mixing. Such HNLs cannot be constrained by kaon fabrics, as the production channel $K \to \tau + N$ is kinematically impossible. Next, constraints from CHARM are restricted by mass $m_N < 290$ MeV, with no clear reason provided. This result looks suspicious

As a result, the mass range 210 MeV $< m_N < m_D$ is reported as a poorly constrained domain, which is a reason of numerous experiments proposed to probe the unexplored parameter space: displaced decays at FASER [66, 67], Belle II [68], SND@LHC [57], DarkQuest [69], and NA62 in the dump mode [67]; prompt decays at LHCb [70, 71]; and double bang signature at IceCube, SuperKamiokande, DUNE and HyperKamiokande [65, 72].

The planned neutrino observatory KM3NeT [73] working as an atmospheric beam dump may have sensitivity to such HNLs as well. Namely, HNLs may be produced in numerous collisions of cosmic protons with atmospheric particles, then reach the detector

volume located deeply underwater in the Mediterranean Sea, and further decay into a dimuon pair inside. Such combination of decay products may be in principle distinguished from the SM events due to neutrino scatterings and penetrating atmospheric muons. We discuss this signature in more detail and estimate the sensitivity of KM3NeT to HNLs produced in the atmosphere in Sec. 2.1.5, and make the conclusions in Sec. 2.1.4.

Constraints from the CHARM experiment as reported in the literature for HNL that mix purely with e/μ and τ neutrinos are very different, with no reason provided.

In this section, we re-analyze the bounds from the CHARM experiment. We study the HNL decay channel $N \to e^+e^-\nu/\mu^+\mu^-\nu$ and show for the first time that, in addition to the constraints on the HNL's mixings with ν_e or ν_μ, the same data also implies limits on the HNLs that mix only with ν_τ and have masses in the range $290\,\mathrm{MeV} < m_N \lesssim 1.6\,\mathrm{GeV}$.

The CHARM bounds re-analysis presented in this chapter may be similarly applied for the re-analysis of bounds coming from the NOMAD experiment [59]. However, due to the smaller intensity of the proton beam at NOMAD and simultaneously similar geometric acceptance of the decay volume, the bounds imposed by NOMAD are sub-dominant, and we therefore do not make the re-analysis in this work.

2.1.1 CHARM experiment

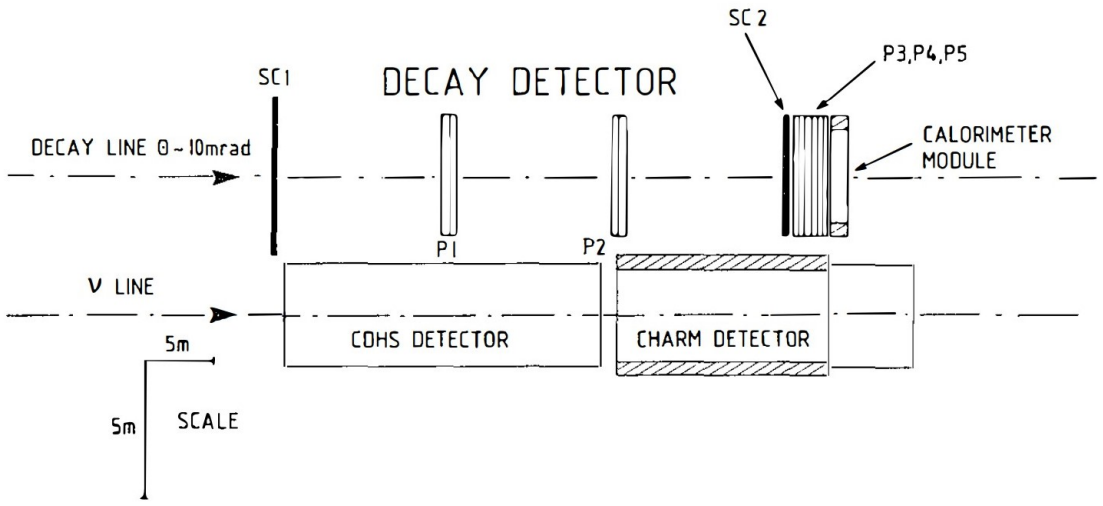

Figure 2.2: The layout of the CHARM facility, adopted from [62].

The CHARM experiment [62, 74] was a proton beam dump operating at the 400 GeV CERN SPS. The total number of exposed protons was split into $1.7 \cdot 10^{18}$ protons on a solid copper target and $0.7 \cdot 10^{18}$ on a laminated copper target with the $1/3$ effective density. Searches for decays of HNLs were performed in the $l_{\text{fid}} = 35\,\mathrm{m}$ long decay region (see Fig. 2.2) defined by the two scintillator planes SC1 and SC2, located at the distance

$l_{\text{min}} = 480\,\text{m}$ from the copper target. The decay detector covered the $3.9 \cdot 10^{-5}\,\text{sr}$ solid angle and had the transverse dimensions $3 \times 3\,\text{m}^2$, with the center displaced by $5\,\text{m}$ from the axis. The fine-grain calorimeter at CHARM was aimed to detect inelastic scattering of electrons and muons produced in hypothetical decays of HNLs [75]. The sets of tube planes P1-P5 [76] were installed to improve the reconstruction of the decay vertex and the angular resolution.

2.1.2 Bounds of CHARM on HNLs as reported in literature

As we have already discussed, in the GeV mass range, the constraints on the mixing angle U_τ^2 are orders of magnitude weaker as compared to the constraints on $U_{e/\mu}^2$ (constraints for the μ mixing are similar to the ones for the e mixing), see Fig. 2.1. Namely, for the e/μ mixing, the large values of the couplings for HNLs with masses $m_K \lesssim m_N \lesssim m_D \simeq 2\,\text{GeV}$ are excluded by the CHARM experiment [62, 74], while for the τ mixing CHARM constraints on U_τ are reported in the literature only for masses $m_N < 290\,\text{MeV}$.

The reason is the following: the original analysis [62, 74] is based on negative results for searches for decays of feebly interacting particles into one of the possible dilepton pair – $\mu e, \mu\mu, \mu e$. For HNLs, they consider only decays mediated through the charged current (CC) interaction (see Fig. 2.3, diagram (a)) that give rise to leptonic decays

$$N_\alpha \to l_\alpha \bar{l}_\beta \nu_\beta, \quad \beta = e, \mu, \tau \tag{2.1.2}$$

If only CC interactions are taken into account, the search is suitable to constrain the mixing of HNLs with ν_e and ν_μ. To search for CC mediated decays via the τ mixing (which necessarily include a τ lepton), the HNL mass should be $m_N > m_\tau \simeq m_D$ in this model. Such HNLs are mainly produced in decays of heavy B mesons, the number of which at CHARM is insufficient to provide enough events for the couplings that are not excluded (see Fig. 2.1). Therefore, HNLs that mix only with ν_τ cannot be constrained by CHARM data using only the decays via CC.

In order to constrain the τ mixing angles of the light HNLs $m_N < m_\tau$, one should include the interactions via the neutral current (NC) into the analysis, see Fig. 2.3 (diagram (b)). In this case, the dileptonic decays are

$$N_\alpha \to \nu_\alpha l_\beta \bar{l}_\beta, \tag{2.1.3}$$

and do not require the creation of a τ lepton for the pure τ mixing.

The works [63, 77, 78] have re-analyzed the CHARM constraints on HNLs by including also the neutral current processes. However, their analysis was insufficient to put the bounds on the pure τ mixing in GeV mass range. Namely, the work [63] (the results of which are used in [36]) has limited the study of the mass range by $m_N < 290\,\text{MeV}$, while [77, 78] considered the decays of HNLs via neutral currents but did not include the production of

HNLs from τ lepton (the diagrams (c) and (d) in Fig. 2.4). As a result, these works did not report any CHARM limits on the pure τ mixing.

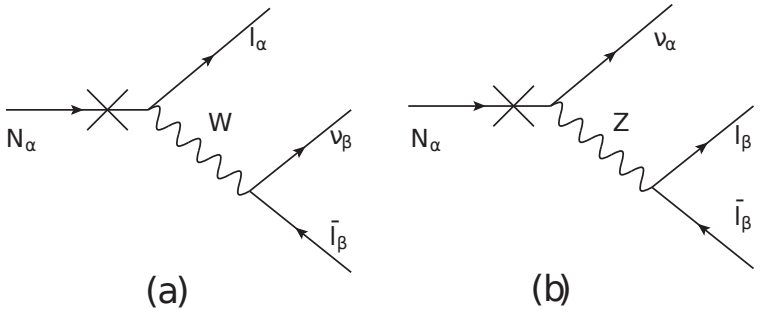

Figure 2.3: Diagrams of leptonic decays of an HNL that mixes purely with ν_α via the charged (the left diagram) and the neutral current (the right diagram).

2.1.3 Phenomenology of HNLs at CHARM

2.1.3.1 Production

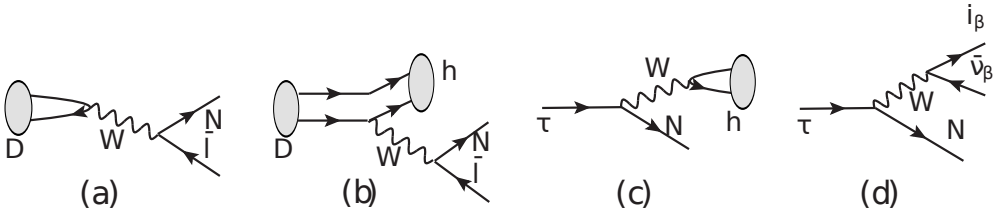

Figure 2.4: Diagrams of HNL production in leptonic and semileptonic decays of D mesons: $D_s, D^0, D^\pm$ (diagrams (a), (b)), and τ lepton, which is produced in decays of D_s meson only (diagrams (c), (d)).

At the SPS energy of $400\,\mathrm{GeV}$, HNLs with mass at the GeV scale may be produced directly either in the proton-target collisions, or in the decays of secondary particles: B, D mesons and τ leptons. The direct HNL production competes with strong interaction processes, while the production from secondary particles – with weak interactions. As a result, the latter process is dominant even taking into account small production probability of mesons [39], and the former may be completely neglected. However, similarly to the other experiment operating at SPS – NA62 in the dump mode, the CHARM experiment has no sensitivity to the HNLs produced from B mesons, implying the lower bound on the probed mass $m_N \lesssim m_{D_s} \simeq 2\,\mathrm{GeV}$.[2]

[2]To search for HNLs created in the decays of B mesons at SPS, an experiment like SHiP [20] with significantly larger beam intensity delivered to the experiment and much better geometrical acceptance would be required.

Therefore, at CHARM, HNLs may be produced only in decays of D mesons and τ leptons.

Let us define the HNL that mixes only with ν_α by N_α. Neglecting the direct production channels, the total number of N_α produced at CHARM is given by:

$$\mathcal{N}_{\text{prod}}^{(\alpha)} = 2\mathcal{N}_{c\bar{c}} \cdot \Big[\sum_{D_i} f_{c\to D_i} \text{Br}(D_i \to N_\alpha X) + \\ + f_{c\to D_s} \cdot \text{Br}(D_s \to \tau\bar{\nu}_\tau) \cdot \text{Br}(\tau \to N_\alpha X) \Big], \quad (2.1.4)$$

with $\mathcal{N}_{c\bar{c}}$ being the total number of quark-antiquark $c\bar{c}$ pairs produced at CHARM, $D_i = D^\pm$, D^0, D_s, and $f_{c\to D_i}$ the corresponding quark fragmentation fractions at SPS. The first term in the brackets describes the production from decays of D mesons (diagrams (a), (b) in Fig. 2.4) and the second – from τ leptons in the $D_s \to \tau \to N$ decay chain (diagrams (c), (d) in Fig. 2.4). $\text{Br}(D_i \to N_\alpha X)$, $\text{Br}(\tau \to N_\alpha X)$ are the branching ratios.

The amount of τ leptons is suppressed as compared to the number of D mesons, and therefore the production channel from τ is subdominant.

Indeed, the second term in Eq. (2.1.4) includes a small factor $f_{c\to D_s} \cdot \text{Br}(D_s \to \tau\bar{\nu}_\tau) \simeq 5 \cdot 10^{-3}$; for the given HNL mass, it is suppressed as compared to the first term as soon as the production from D is allowed.

The original analysis of the CHARM collaboration [62, 74] considered the mixing $\alpha = e, \mu$, for which decays from D mesons are possible for any mass in the range $m_N <$

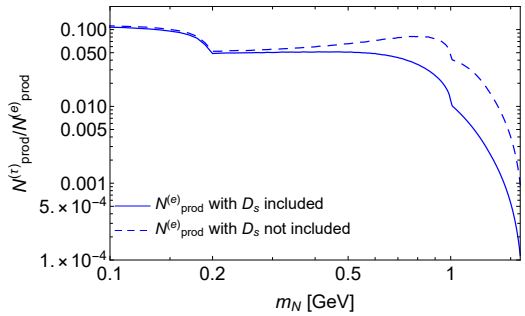

Figure 2.5: The HNL mass dependence of the ratio of the numbers of produced HNLs with pure τ and e mixing $N_{\text{prod}}^{(\tau)}/N_{\text{prod}}^{(e)}$, see Eq. (2.1.5), assuming the same values of the mixing angles $U_e^2 = U_\tau^2$ for the two models. The solid line corresponds to $N_{\text{prod}}^{(e)}$ calculated keeping the production from all D mesons D^+, D^0, D_s, while the dashed line corresponds to the estimate of $N_{\text{prod}}^{(e)} \equiv N_{\text{prod}}^{\text{CHARM}}$ calculated without the contribution of D_s, as has been done in the analysis [62] by the CHARM collaboration (see text for details).

$m_{D_s} - m_{l_\alpha} \approx 1.9\,\mathrm{GeV}$, and the production from τ decays may be completely neglected, according to the discussion above. For the τ mixing, however, the kinematic threshold of the production from D, $D_s \to \tau + N$, is $m_{D_s} - m_\tau \approx 190\,\mathrm{MeV}$, and only the second summand in Eq. (2.1.4) contributes for heavier HNLs.

Let us estimate how many HNLs with τ mixing are produced as compared to those with e mixing.

The amount of produced HNLs with the pure τ mixing ranges from 10^{-3} to 10^{-1} of those with the e mixing.

Indeed, from (2.1.4), the ratio $N^{(\tau)}_{\text{prod}}/N^{(e)}_{\text{prod}}$ is

$$\frac{N^{(\tau)}_{\text{prod}}}{N^{(e)}_{\text{prod}}} = \frac{\sum_{D_i} f_{c\to D_i}\text{Br}(D_i \to N_\tau X) + f_{c\to D_s}\text{Br}(D_s \to \tau\bar{\nu}_\tau)\text{Br}(\tau \to N_\tau X)}{\sum_{D_i} f_{c\to D_i}\text{Br}(D_i \to N_e X)}, \quad (2.1.5)$$

Assuming the same values of mixing angles $U_e^2 = U_\tau^2$ for the two models with pure e/τ mixing, the ratio $\text{Br}(\tau \to N_\tau X)/\sum f_{c\to D}\text{Br}(D \to N_e X)$ varies in the $1 - 10$ range for masses $m_N \lesssim 1.3\,\mathrm{GeV}$ and quickly drops at the kinematic threshold $m_N \approx m_\tau$ [39]. In particular, for masses $m_N \gtrsim 800\,\mathrm{MeV}$, where the dominant contribution to the HNL production with e mixing comes from D_s, we have

$$\frac{N^{(\tau)}_{\text{prod}}}{N^{(e)}_{\text{prod}}} \approx \text{Br}(D_s \to \tau\bar{\nu}_\tau) \cdot \frac{\text{Br}(\tau \to N_\tau X)}{\text{Br}(D_s \to N_e X)} < 4 \cdot 10^{-2} \quad (2.1.6)$$

The mass dependence of the ratio $N^{(\tau)}_{\text{prod}}/N^{(e)}_{\text{prod}}$ obtained from Eq. (2.1.5) is shown in Fig. 2.5.

In the original analysis [62], as well as in the re-analyses [77, 78], the production from D_s (which is the main production channel for the e and μ mixings) has not been taken into account for the e mixing.

In the mass range $m_N \gtrsim 800\,\mathrm{MeV}$, this leads to the underestimate of the number of produced HNLs, $N^{\text{CHARM}}_{\text{prod}}$, by a factor $1/6$ (see Fig. 2.4, where we show the ratio $N^{(\tau)}_{\text{prod}}/N^{\text{CHARM}}_{\text{prod}}$).

2.1.3.2 Decays and their detection

For a given number of produced HNLs, the number of detected events $N^{(\alpha)}_{\text{events}}$ for the given mixing α depends on

1. **Geometrical factors** – in order to be detected, produced HNLs need to point in the angular coverage of the CHARM decay volume, decay inside it, and their decay

products must then reach the detector and be successfully reconstructed. These factors are: geometrical acceptance ϵ_{geom}, i.e. the fraction of produced HNLs traveling in the direction of the CHARM detector; the mean HNL gamma factor γ_N; the decay acceptance ϵ_{decay}, i.e. the fraction of HNL decay products that point to the CHARM detector for HNLs that decay inside the fiducial volume.

2. The **branching ratio** $\text{Br}(N_\alpha \to l^+ l'^- \nu)$ of the channels $N_\alpha \to e^+ e^- \nu, \quad N_\alpha \to \mu^+ \mu^- \nu, \quad N_\alpha \to e^- \mu^+ \nu$ (and their charge conjugated counterparts) used for detection at CHARM [62].

The formula for $N^{(\alpha)}_{\text{events}}$ is:

$$N^{(\alpha)}_{\text{events}} = N^{(\alpha)}_{\text{prod}} \cdot \epsilon^{(\alpha)}_{\text{geom}} \cdot \sum_{l,l'=e,\mu} P^{(\alpha)}_{\text{decay}} \cdot \text{Br}(N_\alpha \to l l' \nu) \cdot \epsilon_{\text{det},ll'} \cdot \epsilon^{(\alpha)}_{\text{decay}}, \tag{2.1.7}$$

where $P^{(\alpha)}_{\text{decay}} = e^{-l_{\text{min}}/c\tau^{(\alpha)}_N \gamma^{(\alpha)}_N} - e^{-(l_{\text{min}}+l_{\text{fid}})/c\tau^{(\alpha)}_N \gamma^{(\alpha)}_N}$ is the decay probability, and $\epsilon_{\text{det},ll'}$ is the reconstruction efficiency for the given channel.

> Geometrical factors determining the sensitivity are the same for e, μ and τ mixing, while the branching ratio is smaller for the τ mixing channels, as in the former case both decays via the charged and neutral currents are relevant, while in the latter only the neutral current contribute.

Let us start by considering the lower bound of the sensitivity of the CHARM experiment, i.e. the minimal mixing angles that it may probe (the upper bound will be discussed in Sec. 2.1.4). In this regime, the decay length of the HNL $c\tau^{(\alpha)}_N \gamma^{(\alpha)}_N$ is much larger than the geometric scale of the experiment, $c\tau^{(\alpha)}_N \gamma^{(\alpha)}_N \gg l_{\text{min}} + l_{\text{fid}} \approx 515\,\text{m}$. Then $P^{(\alpha)}_{\text{decay}} \approx \frac{l_{\text{fid}}}{c\gamma^{(\alpha)}_N} \cdot \Gamma(N_\alpha)$, where $\Gamma(N_\alpha)$ is the total decay width, and it is convenient to rewrite Eq. (2.1.7) in the form

$$N^{(\alpha)}_{\text{events}} \approx N^{(\alpha)}_{\text{prod}} \times \epsilon^{(\alpha)}_{\text{geom}} \cdot \sum_{l,l'=e,\mu} \frac{l_{\text{fid}}}{c\gamma^{(\alpha)}_N} \cdot \Gamma(N_\alpha \to l l' \nu) \epsilon_{\text{det},ll'} \cdot \epsilon^{(\alpha)}_{\text{decay}}, \tag{2.1.8}$$

where $\Gamma(N_\alpha \to l^+ l'^- \nu)$ is the decay width into the dilepton pair ll'.

We will first discuss the difference in $\Gamma(N_\alpha \to l^+ l'^- \nu)$ between the cases of e and τ mixings. Decays into dileptons occur via charged and neutral current, see Fig. 2.3. For the NC mediated processes, the kinematic threshold $m_N > 2m_e \approx 1\,\text{MeV}$ is mixing-independent. In contrast, for the CC mediated process for the τ mixing this threshold is $m_N > m_\tau + m_e \approx 1.77\,\text{GeV}$, and HNLs lighter than τ lepton may decay into dileptons only via NC.

Decay widths for the processes $N_\alpha \to l^+ l'^- \nu$, assuming $m_N \gg m_l + m_{l'}$, may be

given in the unified form

$$\Gamma(N_\alpha \to l^+ l'^- \nu) = c_{ll'\nu}^{(\alpha)} \frac{G_F^2 m_N^5}{192\pi^3}, \tag{2.1.9}$$

where the coefficients $c_{ll'\nu}^{(\alpha)}$ are given in Table 2.1 [39]. For N_e, the largest decay width is $\Gamma(N_e \to \mu^+ e^- \nu_\mu)$, where only CC contributes. The width $\Gamma(N_e \to e^+ e^- \nu_e)$ is smaller:

$$\Gamma(N_e \to e^- e^+ \nu_e)/\Gamma(N_e \to e^- \mu^+ \nu_\mu) \approx 0.59, \tag{2.1.10}$$

because both NC and CC contribute in this process and interfere destructively. The smallest width is $\Gamma(N_e \to \mu^+ \mu^- \nu_e)$, with the process occurring only via NC. For N_τ, there is no process $N_\tau \to e\mu\nu$, while in the process $N_\tau \to e^+ e^- \nu_\tau$ only NC contributes, and thus the width is smaller than for N_e:

$$\Gamma(N_\tau \to e^+ e^- \nu_\tau)/\Gamma(N_e \to e^+ e^- \nu_e) \approx 0.22 \tag{2.1.11}$$

For the decay into a dimuon pair, we have $\Gamma(N_\tau \to \mu^+ \mu^- \nu_\tau) = \Gamma(N_e \to \mu^+ \mu^- \nu_e)$.

As a result, for $m_N \gg m_\mu$ the ratio of the factors $\sum_{l,l'} \Gamma(N_\alpha \to l l' \nu) \epsilon_{\text{det},ll'}$ entering Eq. (2.1.8) is given by

$$\frac{\sum_l \Gamma(N_\tau \to ll) \epsilon_{\text{det},ll}}{\sum_{l,l'} \Gamma(N_e \to ll') \epsilon_{\text{det},ll'}} \approx 0.16 \tag{2.1.12}$$

Here and below, we use the values of the efficiencies $\epsilon_{\text{det},ll'}$ as reported in [62] for the HNL mass $m_N = 1\,\text{GeV}$: $\epsilon_{\text{det},ee} = 0.6$, $\epsilon_{\text{det},e\mu} = 0.65$, $\epsilon_{\text{det},\mu\mu} = 0.75$.

In the original analysis of the sensitivity to the e mixing by the CHARM collaboration [62, 74], the Dirac nature of HNLs has been assumed (the decay widths are twice smaller), and only the CC interactions have been considered. Instead of Eq. (2.1.12), the ratio becomes

$$\frac{2\sum_l \Gamma(N_\tau \to ll) \epsilon_{\text{det},ll}}{\sum_{l,l'} \Gamma_{\text{CC}}(N_e \to ll') \epsilon_{\text{det},ll'}} \approx 0.27 \tag{2.1.13}$$

Process	$c_{ll'\nu}^{(\alpha)}$
$N_{e/\tau} \to \mu^+ \mu^- \nu_{e/\tau}$	$\frac{1}{4}(1 - 4\sin^2\theta_W + 8\sin^4\theta_W) \approx 0.13$
$N_\tau \to e^+ e^- \nu_\tau$	$\frac{1}{4}(1 - 4\sin^2\theta_W + 8\sin^4\theta_W) \approx 0.13$
$N_e \to e^- \mu^+ \nu_\mu$	1
$N_e \to e^+ e^- \nu_e$	$\frac{1}{4}(1 + 4\sin^2\theta_W + 8\sin^4\theta_W) \approx 0.59$
$N_e \to e^+ e^- \nu_e$ (CC)	1

Table 2.1: The values of $c_{ll'\nu}^{(\alpha)}$ in Eq. (2.1.9) for different decay processes. For the process $N_e \to e^+ e^- \nu_e$, we also provide the value obtained if including the charged current (CC) contribution only – the assumption used in [62].

Let us now discuss geometric factors $\epsilon_{\text{geom}}, \gamma_N, \epsilon_{\text{decay}}$. It turns out that they depend

on the mixing pattern weakly, and as a result the geometry does not influence the relative yield of events for e and τ mixing. Indeed, as was mentioned in Sec. 2.1.3.1, HNLs with τ mixing are produced in decays of τ leptons, that originate from decays of D_s. Since $m_\tau \simeq m_{D_s}$, the angle-energy distribution of τ leptons is the same as of D_s (and hence also other D mesons), whose decays produce HNLs with e mixing. The kinematics of the HNL production from D and τ is similar: two-body decays (a), (c) and three-body decays (b), (d) in Fig. 2.4 differ mainly be the replacement a neutrino or a lepton with a hadron $h = \pi, K$. However, since $m_h \ll m_{\tau,D}$, the replacement does not lead to the difference in the distribution of produced HNLs. In addition, heavy HNLs with masses $m_N \simeq 1\,\mathrm{GeV}$ share the same distribution as their mother particles, and any difference disappear. Therefore, the values ϵ_{geom}, γ_N for different mixing are the same with good precision. Next, HNL decays contain the same final states independently of the mixing, and ϵ_{decay} can also be considered the same.

To summarize, the ratio $N^{(\tau)}_{\text{events}}/N^{(e)}_{\text{events}}$ is determined only by the difference in phenomenological parameters – $N^{(\alpha)}_{\text{prod}}$ and $\Gamma(N_\alpha \to ll'\nu)$:

$$\frac{N^{(\tau)}_{\text{events}}}{N^{(e)}_{\text{events}}} \simeq \frac{N^{(\tau)}_{\text{prod}}}{N^{(e)}_{\text{prod}}} \times \frac{\sum_l \Gamma(N_\tau \to ll\nu)\epsilon_{\text{det},ll}}{\sum_{l,l'} \Gamma(N_e \to ll'\nu)\epsilon_{\text{det},ll'}} \tag{2.1.14}$$

The total number of events for the τ mixing is $10^2 - 10^4$ times smaller than for the e mixing.

To compare with the estimate of the number of events for the e mixing made by the CHARM collaboration in [62], $N^{\text{CHARM}}_{\text{events}}$, we need to take into account their assumptions on the description of HNL production and decays (see the discussion around Eqs. (2.1.5) and (2.1.13)). The resulting ratio is

$$\frac{N^{(\tau)}_{\text{events}}}{N^{\text{CHARM}}_{\text{events}}} \simeq \frac{N^{(\tau)}_{\text{prod}}}{N^{\text{CHARM}}_{\text{prod}}} \cdot \frac{2\sum_l \Gamma(N_\tau \to ll\nu)\epsilon_{\text{det},ll}}{\sum_{l,l'} \Gamma_{\text{CC}}(N_e \to ll'\nu)\epsilon_{\text{det},ll'}} \tag{2.1.15}$$

2.1.4 Results

Let us now derive the CHARM sensitivity to the τ mixing. In [62], it has been shown that the dilepton decay signature at CHARM is background free. Therefore, 90% CL sensitivity to each mixing is given by the condition

$$N^{(e,\mu,\tau)}_{\text{events}} > 2.3 \tag{2.1.16}$$

Let us define $U^2_{\text{lower,CHARM}}$ as the smallest mixing angle for which the condition (2.1.16) is satisfied for the assumptions of the original analysis of [62] (see the discussion above Eq. (2.1.15)). As the number of detected events at the lower bound $N^{(\alpha)}_{\text{events}}$ scales with the

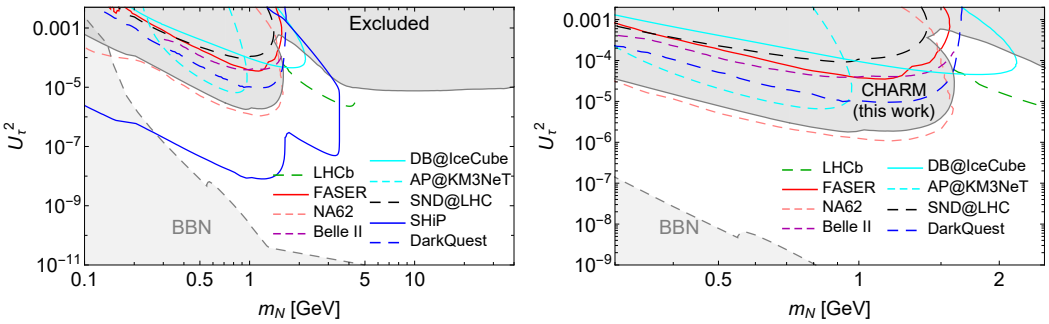

Figure 2.6: Parameter space of a single Majorana HNL that mixes with ν_τ. The excluded region is a combined reach of the DELPHI [79], T2K [61] and CHARM experiments (our result). Bounds from BBN are reproduced from [37, 64]. The sensitivity of future experiments is also shown (see text around Fig. 2.1 for details). The top panel covers the HNL mass region $m_N = 0.1 - 35\,\text{GeV}$, while the bottom panel is a zoom-in of the mass domain $m_N = \mathcal{O}(1\,\text{GeV})$.

mixing angle as $N_{\text{events}}^{(\alpha)} \propto U_\alpha^4$ (where U_α^2 comes from the production and another U_α^2 from decay probability), we can use Eqs. (2.1.15) and (2.1.5) to obtain the lower bound of the sensitivity to the τ mixing, $U^2_{\tau,\text{lower}}$, by rescaling the results reported in [62]:

$$\frac{U^4_{\tau,\text{lower}}}{U^{4\,\text{CHARM}}_{\text{lower}}} \simeq \frac{N^{\text{CHARM}}_{\text{prod}}}{N^{(\tau)}_{\text{prod}}} \cdot \frac{\sum_{l,l'} \Gamma_{\text{CC}}(N_e \to l l' \nu) \epsilon_{\text{det},ll'}}{\sum_l \Gamma(N_\tau \to l \bar{l} \nu) \epsilon_{\text{det},ll}}\bigg|_{U_e = U_\tau} . \tag{2.1.17}$$

Using the ratio $N^{\text{CHARM}}_{\text{prod}}/N^{(\tau)}_{\text{prod}}$ from Eq. (2.1.5) (see also Fig. 2.5), and the ratio of decay widths from Eq. (2.1.13), we may compare the lower bounds of the excluded regions for HNLs with e and τ mixing.

We conclude that in the mass range $m_N > 200\,\text{MeV}$ the lower bound for the τ mixing is a factor $10 - 100$ weaker than the lower bound for the e mixing reported in [62].

In the domain $m_{D_s} - m_\tau < m_N < 290\,\text{MeV}$, we validate the rescaled bound (2.1.17) by comparing it with the CHARM sensitivity to the τ mixing from [63], see Appendix 2.A.

Also, we compare our estimate for the e mixing with the CHARM sensitivity to the e mixing from [62]. In our estimates, we include neutral current interactions, the production from D_s mesons, and assume that HNLs are Majorana particles. In our estimates, we include neutral current interactions, the production from D_s mesons, and assume that HNLs are Majorana particles.

We find that for small mixing angles U_e and above $m_N \gtrsim 1\,\text{GeV}$, the bound imposed by CHARM on the e mixing may be actually improved by up to a factor $3 - 4$ as compared to [62].

At the *upper bound of the sensitivity*, the dependence of the number of events on U_α^2 is complicated and the sensitivity cannot be obtained by rescaling the results of [62]. Therefore, we independently compute the number of decay events at CHARM for HNLs with e and τ mixing and then calculate the sensitivity numerically using Eq. (2.1.16), see Appendix 2.A. In order to validate this estimate, we compare the resulting sensitivity for the τ mixing with the rescaled bound (2.1.17), and find that they are in very good agreement (Fig. 2.27).

Let us comment on errors of our estimates. We used the values of reconstruction efficiencies $\epsilon_{\text{rec},ll}$ reported in [62] for the HNL mass $m_N = 1\,\text{GeV}$. Hence, the calculation may be further refined by including HNL mass dependent reconstruction efficiencies. However, as the study [63] performed for the τ mixing and masses $m_N < 290\,\text{MeV}$ has shown similar efficiency, we do not expect any significant changes.

Our final results for the τ mixing are given in Fig. 2.6, where we show the domain excluded by previous experiments together with updated CHARM bounds, and the sensitivity of the future experiments mentioned in Sec. 2.1, together with SHiP [80]. Comparing with Fig. 2.1, we find that in the mass range $380\,\text{MeV} < m_N < 1.6\,\text{GeV}$ our results improve previously reported bounds on the mixing angle U_τ^2 by two orders of magnitude. In particular, it excludes large part of the parameter space that was suggested to be probed by the future experiments. For instance, Belle II, FASER, DarkQuest and IceCube have sensitivity only in the narrow domain above the CHARM upper bound, while NA62 may slightly push probed angles to lower values. Significant progress in testing the mixing of HNLs with ν_τ can be achieved by LHCb, which probes the complementary mass range $m_N > 2\,\text{GeV}$, and dedicated Intensity Frontier experiments, with SHiP being optimal for searches of HNLs from decays of D mesons and τ leptons.

2.1.5 Comparison with atmospheric beam dumps

Apart from the production at accelerators, HNLs with masses in GeV range may be numerously produced in decays of τ leptons, originated from the collisions of high-energy cosmic protons with the well-known spectrum [81]

$$\frac{d\Phi}{d\Omega dt dS dE_p} \approx \begin{cases} 1.7\; E_{p,\,\text{GeV}}^{-2.7}\;\text{GeV}^{-1}\text{sr}^{-1}\text{cm}^{-2}\text{s}^{-1}, & E_p < 5\cdot 10^6\,\text{GeV} \\ 174\; E_{p,\,\text{GeV}}^{-3}\;\text{GeV}^{-1}\text{sr}^{-1}\text{cm}^{-2}\text{s}^{-1}, & E_p \geqslant 5\cdot 10^6\,\text{GeV} \end{cases} \tag{2.1.18}$$

with atmospheric particles. If having significantly large lifetimes, produced HNLs may enter the detector volume of neutrino telescopes, such as IceCube and KM3NeT, located deep in ice and the Mediterranean Sea correspondingly, and decay there.

In order to probe the parameter space of HNLs, it is necessary to distinguish their decays from interactions of SM particles that are also produced in the atmosphere: neutrinos and

muons. IceCube and KM3NeT may only distinguish two event types: track-like, which corresponds to muons penetrating through the detector volume, and cascade-like, which originates from other particles such as electrons and hadrons. Scatterings of neutrinos inside the detector volume produce cascade-like (if no high-energy muons are produced) or combined cascade-like + track-like signature (if high-energy muons are produced), while penetrating atmospheric muons give rise to track-like signature.

A possible way to distinguish the SM particles events from HNLs is to look for the HNL decays into a di-muon pair, $N \to \mu\bar{\mu}\nu_\tau$. They produce a signature of two tracks originated from one point inside the detector volume, which differs from the SM events signatures.

Detectors of KM3NeT have energy and angular resolution sufficient precise for resolving the two tracks down to energies of a few tens of $10\,\mathrm{GeV}$ [73] (and much better than those at IceCube). On the other hand, characteristic energies of HNLs are $E_N \simeq 100$ GeV. Therefore, we believe that the dimuon signature may be reconstructed in the background free regime with high efficiency.[3]

2.1.5.1 Analytic estimates: comparison with CHARM

Now, let us discuss the sensitivity of KM3NeT to HNLs. We will first compare the amount of HNL decay events at CHARM and KM3NeT for the given value of the mixing angle at the lower bound of the sensitivity using simple analytic estimates. According to Eq. (2.1.8), for the ratio of decay events at these experiments we have

$$\frac{N^{(\tau)}_{\text{events,CHARM}}}{N^{(\tau)}_{\text{events,KM3NeT}}} \simeq \frac{N^{\text{CHARM}}_{c\bar{c}} \cdot \epsilon^{\text{CHARM}}_{\text{geom}} \cdot \epsilon^{\text{CHARM}}_{\text{decay}}}{N^{\text{KM3NeT}}_{c\bar{c}}} \times \frac{l^{\text{CHARM}}_{\text{fid}}}{l^{\text{KM3NeT}}_{\text{fid}}} \times \\ \times \frac{\gamma^{\text{KM3NeT}}_N}{\gamma^{\text{CHARM}}_N} \times \frac{\sum_{l=e,\mu}\Gamma(N_\tau \to ll)\epsilon_{\text{det},ll}}{\Gamma(N_\tau \to \mu\mu)} \quad (2.1.19)$$

Here, $N^{\text{CHARM}}_{c\bar{c}} \cdot \epsilon^{\text{CHARM}}_{\text{geom}} \cdot \epsilon^{\text{CHARM}}_{\text{decay}} \simeq 2 \cdot 10^{13}$ (see Fig. 2.27 is the number of $c\bar{c}$ pairs detectable fraction of HNL decay events at CHARM. $N^{\text{KM3NeT}}_{c\bar{c}}$ is the amount of $c\bar{c}$ pairs produced in the upper hemisphere propagating to KM3NeT,

$$N^{\text{KM3NeT}}_{c\bar{c}} \simeq 2\pi \times 1\ \text{km}^2 \times 5\ \text{years} \times \int \frac{d\Phi}{d\Omega dt dS dE_p} \cdot \frac{\sigma_{pp\to c\bar{c}X}}{\sigma_{pp,\text{total}}} dE_p \simeq 10^{12}, \quad (2.1.20)$$

where $\sigma_{pp\to c\bar{c}X}(E_p)$ is the energy-dependent charm production cross-section which we use from FONLL [43] and from [8], and $\sigma_{pp,\text{total}}$ is the total pp-cross-section, which we use

[3] The possible background is combinatorial and originates from pairs of oppositely charged atmospheric muons. However, it may be reduced to some extent by imposing veto on muons coming from the outer layer of the detector volume.

from [82]. The integrand in (2.1.20) is the product of two competing factors: $\frac{d\Phi}{d\Omega dt dS dE_p}$, which decreases with the proton's energy, and $\sigma_{pp\to c\bar{c}X}(E_p)$, which increases, see Fig. 2.7.

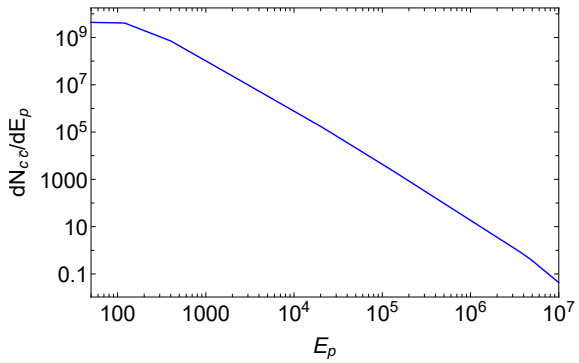

Figure 2.7: The integrand of Eq. (2.1.20).

We approximate the ratio of the mean HNL γ factors by the ratio of the mean γ factors of D mesons:

$$\gamma_N^{\text{KM3NeT}}/\gamma_{D_s}^{\text{CHARM}} \simeq \gamma_{D_s}^{\text{KM3NeT}}/\gamma_{D_s}^{\text{CHARM}} \simeq 3, \tag{2.1.21}$$

where we calculate $\gamma_{D_s}^{\text{KM3NeT}}$ using the $c\bar{c}$ distribution $\frac{d\Phi}{d\Omega dt dS dE_p} \cdot \sigma_{pp\to c\bar{c}X}$, assuming that $E_D \approx E_p/2$.

Using the fiducial lengths $l_{\text{fid}}^{\text{CHARM}} = 35\,\text{m}$ and $l_{\text{fid}}^{\text{KM3NeT}} \simeq 1\,\text{km}$, and taking into account that the last factor in Eq. (2.1.19) is $\mathcal{O}(1)$ for $m_N \gg 2m_\mu$, we finally obtain

$$\frac{N_{\text{events,CHARM}}^{(\tau)}}{N_{\text{events,KM3NeT}}^{(\tau)}} \simeq 2 \tag{2.1.22}$$

Using the analytic estimates, we conclude that even in the most optimistic case (assuming unit efficiency) the number of events at CHARM and KM3NeT are just comparable. We need more accurate estimate taking into account non-isotropic distribution of the produced HNLs.

2.1.5.2 Accurate estimate

We compute the production of D_s mesons (and hence τ leptons) using the approach from [81]. The production was found to be maximal at $O(10\,\text{km})$ height from the Earth's surface. The resulting spectrum $\frac{d\Phi_{D_s}}{dS dt dl d\cos(\theta) dE_{D_s}}$ of D_s mesons is in good agreement with Fig. 2 from [83]. The total number of D_s mesons produced in the direction of KM3NeT during the operating time 5 years was found to be $N_{D_s} \simeq 5 \cdot 10^{10}$.

Next, we use the approach from [83] in order to estimate the sensitivity of KM3NeT.

The number of decay events is

$$N_{\text{events}} \approx S_{\text{Km3NeT}} \times T \times \int \frac{d\Phi_{D_s}}{dS dt dl d\cos(\theta) dE_{D_s}} \cdot \text{Br}(D_s \to \tau \bar{\nu}_\tau) \cdot \\ \cdot \text{Br}(\tau \to N_\tau X) \cdot P_{\text{decay}}(l, E_N) d\cos(\theta) dl dE_N, \quad (2.1.23)$$

where $T = 5$ years is the operating time, $S_{\text{KM3NeT}} = 1\,\text{km}^2$ is the transverse area of KM3NeT. The decay probability is

$$P_{\text{decay}} \approx e^{-(l+l_1)/l_{\text{decay}}} - e^{-(l+l_2)/l_{\text{decay}}}, \quad (2.1.24)$$

where l is the distance from the HNL production point in atmosphere, $l_1 \approx 3\,\text{km}$ is the distance from the surface of Earth to the KM3NeT detector, while $l_2 = l_1 + 1\,\text{km}$ is the distance to the end of the KM3NeT. For simplicity, in l_{decay} we set $E_N \approx E_{D_s}/2$. In order to show the maximal reach of KM3NeT, we optimistically assume unit efficiency of the dimuon event reconstruction, and require $N_{\text{events}} > 3$ during the operating period.

The resulting sensitivity shown in Fig. 2.6 is worse than predicted by the simple estimate by a factor of few. The reason is that at masses $m_N \lesssim 500\,\text{MeV}$ there is an additional suppression from $\text{Br}(N \to \mu\mu)$, while at higher masses the scaling (2.1.8) is not valid because the lower bound is close to the upper bound.

2.2 Searches with displaced vertices at the LHC

A peculiar feature of dedicated beam experiments such as SHiP, DUNE, and MATHUSLA is that they have macroscopic distance from the collision point $l_{\text{min}} \gg 1$ m to the detector volume. On one hand, it allows to reduce background from SM particles down to controllable and even negligible level. On the other hand, such experiments cannot search for short-lived FIPs with decay lengths $c\tau\gamma \ll l_{\text{min}}$.

> Because of macroscopic distance from the FIP production point and the detector volume at Intensity Frontier experiments, there is a domain of large couplings that is neither excluded by past experiments nor may be probed by dedicated beam experiments, see Fig. 2.8.

To probe such intermediate couplings, one needs experiments that have much shorter l_{min}, and simultaneously may handle the background coming from SM events, the amount of which increases if decreasing l_{min}. Such type of experiments exists at the LHC and is called **displaced vertices scheme** (DV).

An event at a given DV experiment has to meet some selection criteria which are specific to the experiment, minimize the SM background yield and simultaneously should not diminish the detected amount of events with FIPs. The physical process involving

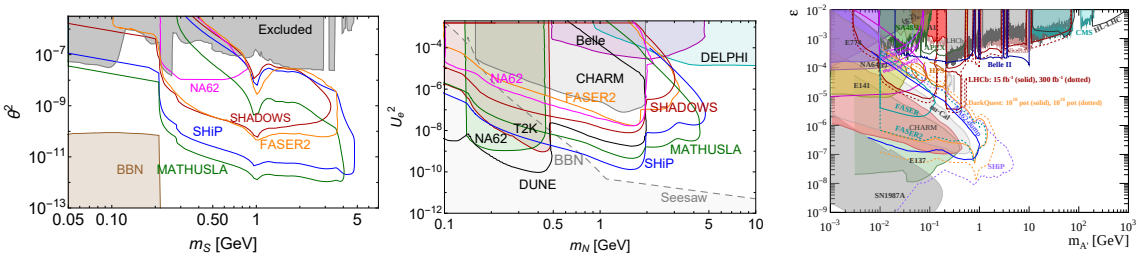

Figure 2.8: Sensitivity reach of Intensity Frontier experiment to decays of portal particles: dark scalars (1.1.1) that have zero quartic coupling $HHSS$ (the left panel), Heavy Neutral Leptons (1.1.2) that mix purely with ν_e (the middle panel), and dark photons (1.1.3) (the right panel). For the description of the lines, see Fig. 1.2. For all of the models, there is a gap between the domain that may probed by Intensity Frontier experiments and the parameter space closed by past experiments, which is due to finite distance from FIPs production point to the decay volume of Intensity Frontier experiments.

FIP consists of the production vertex (PV, in which a FIP is produced) and the displaced decay vertex, in which it decays, see Fig. 2.9; an important selection criterion is whether the requirement of the PV to be tagged in addition to the reconstruction of the DV. The tagging is done with a help of a prompt decay product such as a lepton or a jet. Currently, the PV tagging is essential to have a DV event triggering, so that it is recorded and can be analyzed offline. This type of searches is already performed at ATLAS, CMS and LHCb, see e.g. [84–89]. The second type of schemes does not require such tagging. It will be available after the phase II upgrade, during the high-luminosity LHC phase, when the possibility to use track-trigger in CMS will be introduced. This will enable a possibility to reconstruct and identify displaced tracks online [90–92], and hence will remove a need for a prompt product in the event. Estimates of the sensitivity of one of such schemes which utilizes the L1 trigger at CMS may be found e.g. in [93].

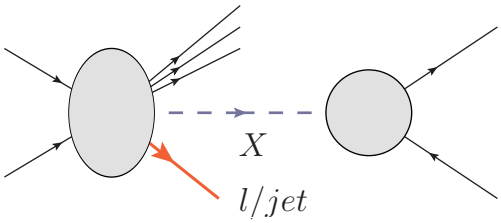

Figure 2.9: Schematic diagram of searches for FIPs with displaced vertices schemes at the LHC. A FIP X produced at the production vertex (PV) travels a macroscopic distance and then gives rise to a displaced decay vertex (DV). Displaced vertices schemes that are currently available require tagging of the PV, which is typically done with the help of a lepton or jet produced in the PV (shown in red). During Run 4, schemes that do not require such tagging will be available.

ATLAS, CMS and LHCb experiments complement each other in searches for FIPs: LHCb covers mass range $m_{\rm FIP} \lesssim m_B$, while ATLAS/CMS allow for searching in the range $m_B \lesssim m_{\rm FIP} \lesssim m_h$.

Indeed, the flux of SM particles that may produce FIPs (such as W/Z bosons, the Higgs boson, and B, D mesons) is collimated with respect to the beam axis.[4] Naively, one may expect that the best placement of a DV experiment is in the forward region that covers relatively small angles. However, decay products of these particles gain characteristic transverse momentum of order of $p_T \simeq m_{\rm mother}$, where $m_{\rm mother}$ is the mass of decaying particle. For large enough $m_{\rm mother}$ (for instance, for Higgs bosons), the characteristic angle $\theta \simeq \arctan(p_T/p_L)$ becomes much larger than the angular width of the mother particle distribution. Then, instead of the forward region, the preferable setup is the one covering large angular domain. ATLAS and CMS are located off-axis and have significant angular coverage, thus allow us to search for FIPs produced in decays of heavy particles such as the Higgs boson and W/Z bosons. LHCb, in its turn, covers much smaller angular domain but in the forward region, and thus allow to search for FIPs originating from decays of relatively light B mesons, for which the broadening is insignificant.

In this section, we discuss the potential for the search scheme at CMS that utilizes muon trackers. Advantages of this scheme is large length of the decay volume, which is essential to probe the parameter space of the LLPs with the decay lengths about 1 meter or larger, and relatively small background as compared to the schemes that search for hadronic decays. We use Heavy Neutral Leptons, Chern-Simons portal and dark scalars as three examples of long-lived particles for which the CMS muon tracker can provide essential information about their properties.

2.2.1 Displaced vertices with muon tracker at CMS

Typically, DV search schemes use inner trackers to reconstruct events. Therefore, the maximal displacement that may be reconstructed is limited by $\mathcal{O}(0.5\ \text{m})$. Alternatively, one may use muon chambers, that cover much larger distance (up to 3 m in the transverse plane at CMS).

The use of the muon chambers to reconstruct di-muon DV signatures has been explored in the past in [94–96] and recently in [97]. Ref. [98] that appeared when this work was at its final stage employed the event selection criteria that may be too optimistic with regard to the background estimates. Ref. [99] explored a potential of the CMS muon chambers *alone* to reconstruct dimuon DV. This search however, was constructed to be much more

[4] For instance, the angular distribution of B mesons and the Higgs bosons quickly drops at angles $\theta > 0.01$ rad.

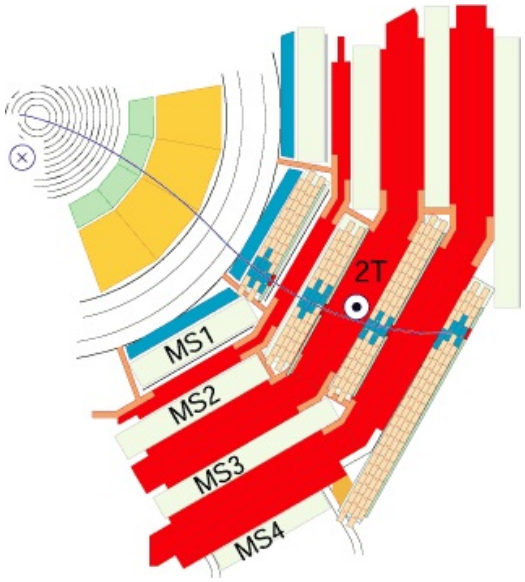

Figure 2.10: Cross-section of the CMS experiment. Layers (muon stations) of the muon detector in the plane perpendicular to the beam direction. The figure is from [103].

general, and hence could not profit from the presence of a prompt lepton in the event. This necessarily implied much more stringent cuts on p_T of either of the two muons in the muon tracker since these muons were used to record an event by a trigger, and therefore lower sensitivity.

2.2.1.1 Description of the scheme

CMS (compact muon solenoid) is a beam line azimuthally symmetric detector consisting of a solenoid generating the 3.8 T magnetic field, the inner trackers that allow to reconstruct the momentum of particles produced in the pseudorapidity range $|\eta| < 2.5$ (where $\eta = -\log[\tan(\theta/2)]$ and θ is the polar angle with respect to the anticlockwise-beam direction) and the muon trackers located outside the solenoid [100].

The muon system is located outside the solenoid and covers the range $|\eta| < 2.4$. It is a set of gaseous detectors sandwiched among the layers of the steel flux-return yoke. This allows for a muon to be detected along the track path at multiple points [101]. The magnetic field in the muon system is not uniform, and goes from 2 T in the innermost part down to almost 0 T in the outer part [102]. Schematic drawing of the muon detector is shown in Fig. 2.10.

For the LHC Run 2, new reconstruction of muons has been introduced [104], the so-called displaced standalone muon reconstruction. This reconstruction is specifically designed to address cases when muons are produced in decays far away from the production vertex. New algorithm achieves an almost 100% reconstruction efficiency for the muon production radius up to about 3 m. This is a significant gain in the efficiency compared to the reconstruction which uses also inner tracker information, but at the same time, the

momentum resolution deteriorates by about a factor of 10 and is in the range 10–60%.

The muon tracker can use two muon tracks to reconstruct a displaced vertex originating from the decay $X \to \mu\mu + \ldots$. The reconstructed DV together with the production vertex that can be tagged by prompt decays products, *e.g.* a prompt lepton, and an underlying event produced together with the X particle, is identified as a DV event. Due to the large distance between a PV and a reconstructed DV, we will call this scheme the "the long DV" scheme.

It should be noticed, that after the phase II upgrade, during the high-luminosity LHC phase, the possibility to use track-trigger in CMS will be introduced. This will enable a possibility to reconstruct and identify displaced tracks online [90–92], and hence will remove a need for a prompt lepton in the event. However, for the models discussed in this paper, current hardware configuration of the CMS allows to perform the searches with the already recorded data, as well with the data to be obtained during the Run 3 of the LHC.

At the same time, final states with a prompt, well identified, object in the event, as *e.g.* a prompt muon or electron, have much lower background rate. In this case the instrumental backgrounds and non-muon backgrounds from cosmic rays are reduced to a negligible level. The remaining cosmic-ray muon backgrounds can be suppressed by selections which do not impact signal efficiency, as described in Ref. [105]. The remaining sources of the background for the long DV scheme are processes with a presence of a prompt object (as e.g. W boson production) accompanied by decays of the SM particles into single muons, which give rise to combinatorial two-muon events, and two-muon decays of the SM particles (for example, $J/\psi, \rho, \omega$ mesons and the Z boson). The most significant displacement of such DV appears in case of two muons originating for a heavy-flavor particle decay (b or c hadrons). As we do not carry out an experimental analysis in this paper, we assume that this background is negligible if one requires the transverse position of the displaced vertex to be as far as $l_{\text{DV}} > 2$ cm from the beam collision point, since most of the SM particles decay before reaching this displacement [106]. Under this assumption we lose a part of the efficiency for LLPs with shorter lifetimes, but at the same time we provide a more robust estimate of the potential signal sensitivity. Because of the position of the muon trackers, the muon events can be reconstructed at the distances $l_{\text{DV}} < 3$ m. The muons can be reconstructed with high efficiency and low misidentification probability if each of them has the transverse momentum $p_T > 5$ GeV [107, 108].

To summarize, an event in the long DV search scheme should satisfy the following selection criteria:

- A prompt electron with $|\eta| < 2.5,\ p_T > 30$ GeV or a prompt muon with $|\eta| < 2.4,\ p_T > 25$ GeV, which are required for an event to be recorded by the single lepton triggers;
- The minimal transverse displacement of the DV from the PV is $l_{\text{min},\perp} = 2$ cm; the maximal transverse and longitudinal displacements are $l_{\text{max},\perp} = 3$ m, $l_{\text{max},l} = 7$ m;

– Two displaced muon tracks, each with $|\eta| < 2.4,\ p_T > 5$ GeV.

The requirement of a large displacement of a DV from the PV helps to significantly reduce the background from SM processes. Therefore, even in the region with the invariant mass of two muons below 5 GeV (mass of B-mesons) the SM background is considered to be negligible. The scheme is presented in Fig. 2.11.

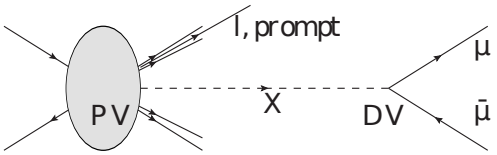

Figure 2.11: Schematic diagram of the search scheme of LLPs at CMS using the muon detectors. The production vertex (PV) is tagged by the prompt lepton l, while the displaced vertex (DV) is reconstructed by two muons produced in the decay $X \to \mu\mu$.

An event with prompt τ lepton can be tagged by its leptonic decays $\tau \to l\bar{\nu}_l\nu_\tau$, where the leptons $l = e/\mu$ satisfy the criteria for prompt leptons presented above. We do not consider the reconstruction of τ leptons by their hadronic decay products since the trigger threshold for p_T of hadronic decay products is too high for efficient reconstruction.[5] However, in the future it is wise to invest into the development of a dedicated multi-object trigger which would allow to bring down the prompt tau p_T by including additional displaced leptons in the event.

In [98] a similar search scheme was discussed, albeit with less restrictive selection criteria $l_{\text{min},\perp} = 0.5\,\text{cm}$, $l_{\text{max},\perp} = 4\,\text{m}$, and $|\eta| < 4$ for leptons.[6] A wider range of muon pseudorapidities leads to the enlargement of the selection efficiency, while a smaller displacement between a DV and the PV lifts up the upper bound of the sensitivity and hence increases the maximal mass reach. However, we caution that the background-free hypothesis for the region with smaller DV displacements adopted in [98] has not been tested. Nevertheless, to demonstrate potential improvement from considering lower displacements we provide sensitivity for two scenarios: "realistic" for the selection criteria outlined above, and "optimistic", defined according to [98].

2.2.1.2 Estimation of the number of events

The number of decay events of a new particle X that pass the selection criteria is

$$N_{\text{events}} = N_{\text{parent}} \cdot \text{Br}_{\text{prod}} \cdot P_{\text{decay}} \cdot \epsilon, \tag{2.2.1}$$

where

[5] Current trigger threshold is $p_T > 180$ GeV.

[6] After the HL-LHC upgrade the CMS will extend its pseudorapidity range to $|\eta| < 4$.

– N_{parent} is the total number of parent particles that produce a particle X at the LHC;
– Br_{prod} is the branching fraction of the production of a particle X in decays of the parent particle;
– P_{decay} is the decay probability,

$$P_{\text{decay}} = \int d\theta_X dp_X f(p_X, \theta_X) \times \left(e^{-l_{\min}/c\tau_X\gamma_X} - e^{-l_{\max}/c\tau_X\gamma_X}\right), \quad (2.2.2)$$

with τ_X being the proper lifetime of the particle X, γ_X is its γ factor, and $f(p_X, \gamma_X)$ is the distribution function of the X particle whose decay products satisfy the selection criteria;
– ϵ is the *overall efficiency* – the fraction of all decays of the X particle that occurred in the decay volume between $l_{\min}$ and $l_{\max}$, have passed the selection criteria, and were successfully reconstructed.

The efficiency is a combination of several factors:

$$\epsilon = \epsilon_{\text{sel}} \cdot \epsilon_{\text{rec}} \cdot \text{Br}_{X\to\mu\mu}, \quad (2.2.3)$$

where $\epsilon_{\text{sel}}, \epsilon_{\text{rec}}$ are the efficiencies of the selection and subsequent reconstruction of an event correspondingly, and $\text{Br}_{X\to\mu\mu}$ is the branching ratio of the decay of the X particle into two muons. Clearly, ϵ_{rec} does not depend on the nature of LLP. The reconstruction efficiency for leptons is well above 95% for muons with $p_T > 5$ GeV [101, 104, 107] and for electrons with $p_T > 30$ GeV [109]. Therefore, for simplicity the reconstruction efficiency is taken to be equal to 1 ($\epsilon_{\text{rec}} = 1$) in what follows.

We define the sensitivity curves by the condition $N_{\text{events}} \simeq 3$, corresponding to the 95% exclusion limit under the assumption of zero background. The lower boundary can be easily rescaled to other N_{events}.

The main advantage of the long DV scheme is the large length of the fiducial decay volume $l_{\max}$, which exceeds the lengths of the decay volumes of other DV search schemes at the LHC (see, e.g., [85, 110]) by $\simeq 10$ times. This has a benefit when searching for new particle with large decay lengths,

$$l_{\text{decay}} \equiv c\tau_X\gamma_X \gg l_{\max} \quad (2.2.4)$$

Indeed, in this case the decay probability (2.2.2) is in the "linear regime", $P_{\text{decay}} \approx l_{\max}/l_{\text{decay}}$, and as a result the number of events (2.2.1) is proportional to $l_{\max}$. For decay lengths that do not satisfy the condition (2.2.4) the decay probability does not depend on $l_{\max}$, and the improvement is lost (see Fig. 2.12).

In order to probe the domain (2.2.4) there must be sufficient production of the X particles, i.e.

$$N_{\text{prod}} \cdot \text{Br}_{X\to\mu\mu} > 3, \quad (2.2.5)$$

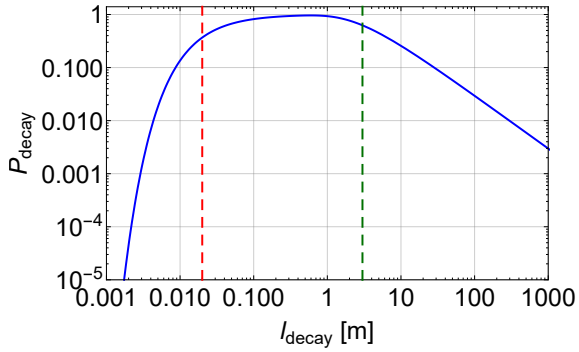

Figure 2.12: Dependence of the decay probability (2.2.2) on the decay length l_{decay}. For simplicity we assumed that all the particles travel with the same momentum and in the same direction, and set $l_{\text{max}} = 3$ m. The dashed lines denote the values $l_{\text{decay}} = l_{\text{min}}$ and $l_{\text{decay}} = l_{\text{max}}$. In the domain $l_{\text{decay}} \gg l_{\text{max}}$ the decay probability scales as $P_{\text{decay}} \simeq l_{\text{max}}/l_{\text{decay}}$, while in the domain $l_{\text{min}} \lesssim l_{\text{decay}} \lesssim l_{\text{max}}$ it behaves approximately constantly and does not depend on l_{max}.

where $N_{\text{prod}} = N_{\text{parent}} \cdot \text{Br}_{\text{prod}}$. The parameter space defined by the conditions (2.2.4), (2.2.5) is optimal for being probed by the long DV scheme. A toy example of the parameter space is given in Fig. 2.13.

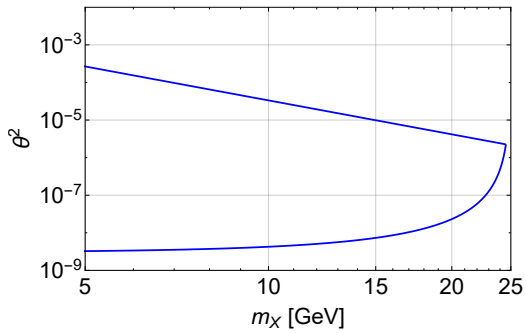

Figure 2.13: The illustration of the parameter space which is optimal for being probed with the long DV scheme, see text for details. We used a toy model with $N_{\text{prod}} = 10^9 \left[1 - m_X^2/(25\text{ GeV})^2\right]^2 \theta_X^2$, $\text{Br}_{X\to\mu\mu} = 1$ and $l_{\text{decay}} = 0.1 m_X^{-3} \theta_X^{-2}$ m.

2.2.1.3 HNLs

We start with HNLs. The main production channel of the HNLs with masses in the range $m_N \gtrsim 5$ GeV is the decay of the W bosons. We use the value $\sigma_W \approx 190$ nb for the total production cross section of the W bosons at the LHC at energies $\sqrt{s} = 13$ TeV [111]. To estimate the parameter space defined by (2.2.4) and (2.2.5), we calculated the energy spectrum and geometric acceptance ϵ_{geom} of the HNLs in the pseudorapidity range $|\eta| < 2.5$ in LO using the model *HeavyN* [112]. We found $\epsilon_{\text{geom}} \approx 0.5$ for the mass range $m_N \lesssim 20$ GeV and $E_N \approx 80$ GeV.

ϵ_{sel}	e	μ	τ
Realistic	0.16	0.17	$7 \cdot 10^{-3}$
Optimistic	0.26	0.31	$3.2 \cdot 10^{-2}$

Table 2.2: The values of the selection efficiencies for HNLs of different flavors e, μ, τ in the case of realistic and optimistic selection criteria

In Fig. 2.14, we show the parameter space for the HNLs mixing with ν_μ that can be optimally probed by the long DV scheme. We see that the domain where the long DV

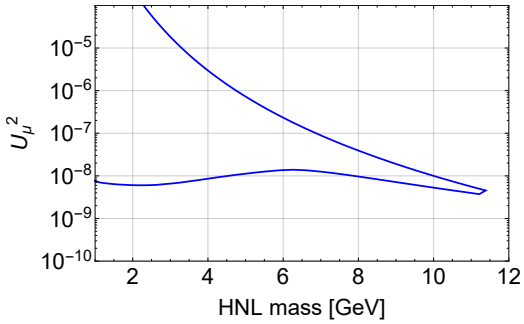

Figure 2.14: The parameters of HNLs mixing with ν_μ that satisfy criteria (2.2.4) –(2.2.5) for the LHC luminosity $\mathcal{L} = 3000$ fb^{-1}. Note, this is *not* an exclusion region, see text around equations for details.

scheme has good potential corresponds to the masses $m_N < 10$ GeV and the mixing angles $U^2 \gtrsim 10^{-9}$.

Simulations. To find the efficiency for the HNLs mixing with $\nu_{e/\mu}$, we used MadGraph5 [113] with the model *HeavyN* [112]. For simulating decays of τ lepton, we used *taudecay_UFO* model [114]. For the mixing with $\nu_{e/\mu}$ we simulated the process $p+p \to W,\ W \to l+N$, where $l = e$ for the mixing with ν_e and $l = \mu$ for the mixing with ν_μ, with subsequent decay $N \to \mu^+ + \mu^- + \nu/\bar{\nu}_l$. In the case of the mixing with ν_τ, we simulated the process $p+p \to W, W \to \tau + N$ with subsequent decays $N \to \mu^+ + \mu^- + \nu_\tau/\bar{\nu}_\tau$ and $\tau \to l + \bar{\nu}_l + \nu_\tau$, where $l = e/\mu$.

Using the selection criteria for the long DV scheme, we computed the selection efficiencies. They were found to be almost independent of the mass of the HNL in the mass range 1 GeV $< m_N <$ 20 GeV. We give their values in Table 2.2. The suppression of the efficiency for mixing with ν_τ is due mainly to the reconstruction of the prompt τ event. Indeed, the amount of the leptons produced in the decay $\tau \to l\bar{\nu}_l\nu_\tau$ and passing the p_T selection criterion for the prompt leptons is ≈ 0.1.

For the average momentum we found $p_N \approx 70$ GeV and $p_N \approx 180$ GeV for the realistic and optimistic estimates correspondingly.

Comparison with other schemes. Let us compare the sensitivity of the long DV search scheme to the HNLs with a scheme from [110, 115] that uses inner trackers at ATLAS to search for DVs events (c.f. [97]). Owing to its smaller transverse displacement

$l_{\max} = 0.3$ m we call it the "*short DV scheme*". For the estimation of the sensitivity of the short DV scheme we use parameterized efficiencies $\epsilon(m_N, U^2)$ provided by the authors of [110]. The comparison of the sensitivities is given in Fig. 2.15. We show both optimistic and realistic estimate of the sensitivity of the long DV scheme. We also show the sensitivity of the SHiP experiment from [80] that serves for an illustration of the sensitivity reach of Intensity Frontier experiments.

The long DV scheme allows to search for HNLs in the unexplored region of the parameter space that is not accessible to other Intensity Frontier experiments or other LHC searches. Its difference in the sensitivity with the short DV scheme is due to three reasons. First, for masses $m_N \lesssim 10$ GeV the decay probability for both the schemes is in the linear regime (see Sec. 2.2.1.2), and therefore the long DV scheme gets the benefit from the 10 times larger length of the decay volume $l_{\max}$. Second, for the masses 5 GeV $\lesssim m_N \lesssim 10$ GeV there is a drop of the overall efficiency for the short DV scheme. This is caused by the selection criteria on the reconstructed invariant mass of the DV, $m_{\text{DV}} > 5$ GeV, and the charged tracks, $N_{\text{trk}} \geqslant 4$, that are needed to remain in the background free region [87]. Third, because of absence of the hadronic background the long DV scheme can probe the parameter space $m_N \lesssim 5$ GeV, which is not reachable by the short DV scheme.

Nevertheless, both the schemes are complementary to each other and provide a cross-check in the mass region 5 GeV $< m_N < 15$ GeV.

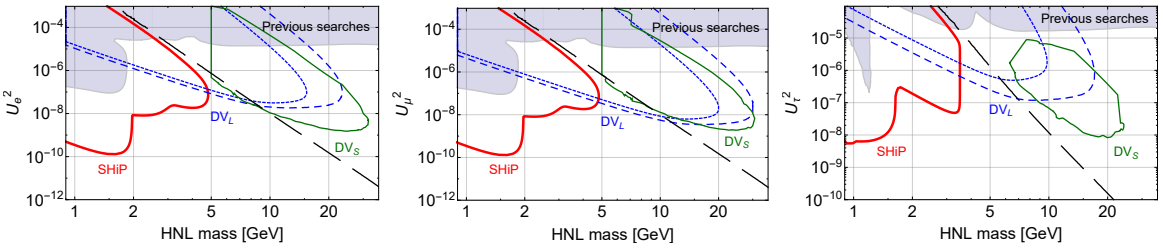

Figure 2.15: The sensitivity of the long DV (DV_L) and short DV (DV_S) search schemes to HNLs mixing with ν_e (upper panel), ν_μ (middle panel) and ν_τ (lower panel). By the blue short dashed line we denote the realistic sensitivity obtained using the selection criteria presented in this paper, while the blue dashed line corresponds to the optimistic estimate of the sensitivity using relaxed selection criteria from [98], see Sec. 2.2.1.2 for details. The sensitivity of the SHiP experiment is taken from [80]. Black long-dashed line indicates HNL parameters that correspond to $l_{\text{decay}} = 3$ m. The estimates are for the high luminosity LHC phase, $\mathcal{L} = 3000$ fb^{-1}. For the DV search schemes sensitivity we require $N_{\text{events}} \geq 3$ and assume zero background (see text for details).

2.2.1.4 Chern-Simons portal

Chern-Simons portal introduces a vector particle X interacting with pseudo-Chern-Simons current of the SM gauge bosons [82, 116]:

$$\mathcal{L}_{CS} = c_W \epsilon^{\mu\nu\lambda\rho} X_\mu W_\nu \partial_\lambda W_\rho + \\ c_\gamma \cos\theta_W \epsilon^{\mu\nu\lambda\rho} X_\mu Z_\nu \partial_\lambda \gamma_\rho + c_Z \sin\theta_W \epsilon^{\mu\nu\lambda\rho} X_\mu Z_\nu \partial_\lambda Z_\rho \quad (2.2.6)$$

We can add the interaction of the X boson with SM leptons in the form

$$\mathcal{L}_{X\mu\mu} = c_W g_{Xll} X^\nu \sum_{l=e,\mu,\tau} \bar{l}\gamma_5\gamma_\nu l, \quad (2.2.7)$$

where g_{Xll} is a dimensionless constant.[7]

Let us consider the case when $c_\gamma, c_Z \ll c_W$. Then the production of the X particle in pp collisions goes through the XWW vertex, while the decay goes through the vertex (2.2.7) down to very small couplings $g^2_{Xll} \simeq 10^{-7}$ for the X bosons as heavy as $m_X \simeq 40$ GeV, see Appendix 2.2.1.4. These vertices are parametrically independent, and for particular values of g_{Xll} it is possible to probe the parameter space in the optimal domain for the long DV scheme. The process of interest is

$$W \to X + l + \bar{\nu}_l, \quad X \to \mu^+ + \mu^- \quad (2.2.8)$$

The lepton l produced in the W decay can be triggered as a prompt lepton, while the muon pair from the decay of the X boson can be reconstructed as displaced muons, which meets the selection criteria of a DV event within the long DV scheme.

To find the selection efficiency and the energy spectrum of the W bosons, we implemented the model of the X boson (2.2.6), (2.2.7) into the MadGraph using FeynRules [117, 118]. The model is publicly available [119]. We simulated the processes $p+p \to e^+/\mu^+ + \nu_{e/\mu} + X$ (plus the charge conjugated final states) with subsequent decays $X \to \mu^+\mu^-$. We have found that the overall efficiency is $\epsilon \approx 2.3 \cdot 10^{-2}$ for m_X ranging from 1 GeV to 20 GeV. The average momentum of the X boson $p_X \approx 40$ GeV.

The sensitivity to the Chern-Simons portal is shown in Fig. 2.16. We conclude that the long DV scheme can probe masses up to $m_X \simeq 30$ GeV and couplings down to $c^2_W \simeq 10^{-9}$. We note that the probed parameter space is well below the current experimental bound on c_W, which is $c^2_W \lesssim 10^{-3}(m_X/1\text{ GeV})^2$ [82].

[7] The coupling g_{Xll} can be generated effectively by the interaction (2.2.6) or be an effect of new physics.

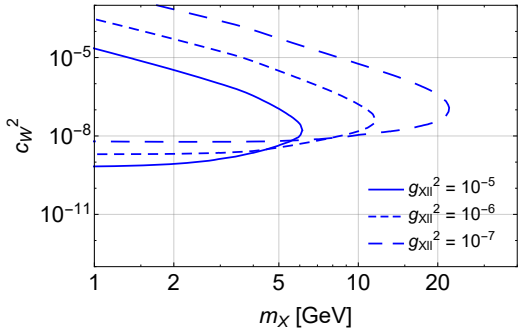

Figure 2.16: The sensitivity of the long DV scheme at the high luminosity phase to the Chern-Simons portal for different values of the coupling to muons (2.2.7). For the DV search schemes sensitivity we require $N_{\text{events}} \geq 3$ and assume zero background (see text for details).

2.2.1.5 Dark scalars with quartic coupling

In the case of dark scalars with quartic coupling (remind Eq. (1.1.1)) and masses $m_S > m_B$, the production and decay channels are independent of each other [120], similarly to the case of Chern-Simons portal.

Namely, the production occurs via the quartic coupling – it is a decay of Higgs bosons $h \to SS$, while decays are mediated by the mixing angle [38]. The branching ratio of

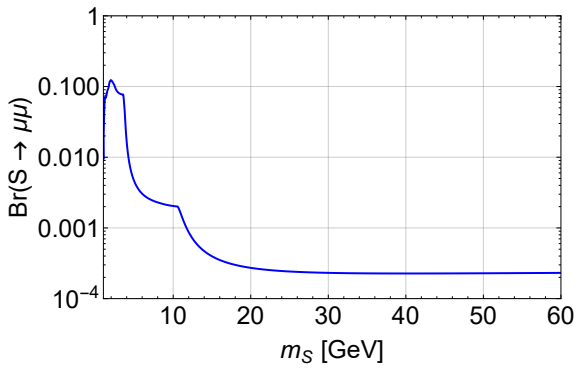

Figure 2.17: The branching ratio of decays of scalars into two muons. It does not include muons produced as secondary particles from decays of heavier decay products, whose contribution may dominate at larger masses.

decays of scalars into two muons is shown in Fig. 2.17. The interaction of scalars with fermions is similar to the interaction of Higgs bosons. Therefore, above scalar masses $m_S \simeq 2m_\pi$, it is strongly suppressed because of the Yukawa couplings. Nevertheless, despite the suppression of the signal yield by small $\text{Br}(S \to \mu\mu)$ it is still possible to search for heavy scalars using the muon trackers. In addition, as in Fig. 2.17 we do not include pair

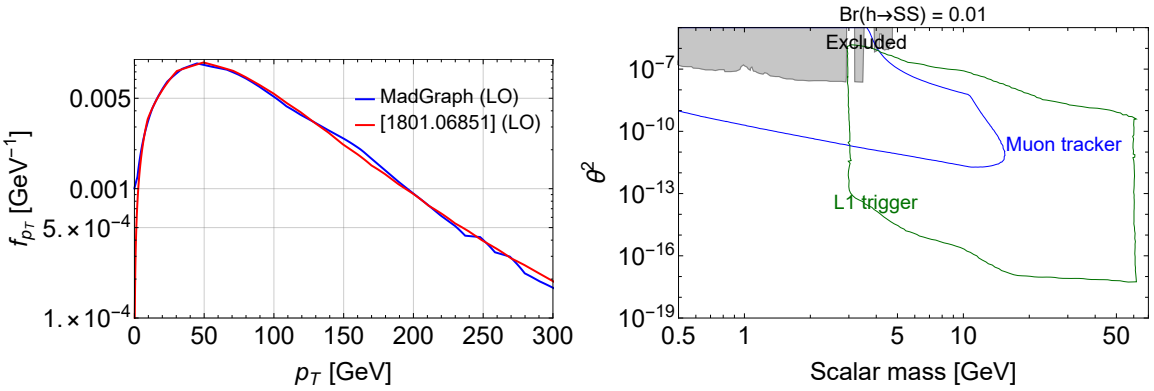

Figure 2.18: *Left panel*: The p_T spectrum of Higgs bosons produced in the process $p+p \to h+W^+,\ W^+ \to l^+ + \nu_l$ obtained in our simulations using MadGraph (blue) and in [121] for the selection criteria $|\eta_l| < 2.5$, $p_{T,l} > 15$ GeV. *Right panel*: the sensitivity of the muon tracker DV scheme to scalars, assuming the production branching ratio $\text{Br}(h \to SS) = 0.01$. In green, we also show the sensitivity of the search scheme that uses L1 trigger from [122], assuming zero background.

of muons from decays of heavier decay products of scalars, our estimate of the sensitivity is conservative.

Unlike HNLs and Chern-Simons portal mediator, the main production channel of scalars, $h \to SS$, does not include a lepton, and naively it is impossible to tag the production vertex. However, one may use leptons produced together with Higgs bosons in a process

$$p+p \to h+W/Z, \quad W/Z \to l+X, \quad h \to S+S \tag{2.2.9}$$

Indeed, h, W, Z bosons have extremely small lifetimes, and all the processes (2.2.9) occur practically at one point. The process (2.2.9) is $\simeq 50$ times less frequent than the main production channel of the Higgs boson – the gluon fusion.

Simulation. In order to simulate events of scalar decays, we have first implemented the model of a scalar into MadGraph using FeynRules. Next, we have simulated the chain (2.2.9) at LO, together with further scalar decays. NLO corrections to the spectra are known to be negligibly small [121]. To validate our simulations for the production of Higgs bosons and leptons in this process, we compare of the p_T spectrum of Higgs bosons produced in the process $p+p \to h+W^+,\ W^+ \to l^+ + \nu_l$ with the spectrum from [121], see Fig. 2.18.

Sensitivity. In Fig. 2.18, we show the sensitivity to the scalar portal assuming the branching ratio $\text{Br}(h \to SS) = 0.01$, which is well below the reach of future searches for invisible decays of Higgs bosons at LHC Run 3 and at the High-Luminosity (HL) LHC (HL-

LHC, Run 4), which are projected to have sensitivity at the level $\mathrm{Br}_{h\to\mathrm{inv}} \sim 0.05$ — 0.15 at 95% CL [123] maybe going all the way to a few *percents* [124].

We also include the sensitivity of the scheme with L1 tracker at CMS [122], which does not require a prompt lepton for triggering. Together with the fact that it requires 4 charged tracks each with $p_T > 1$ GeV, and may search for decay products other than muons, this leads to significantly better sensitivity at the lower bound. However, independent observation (or non-observation) of events at this scheme and at the scheme with muon trackers would allow to measure independently branching ratios of different decay channels of scalars, and in particular to distinguish it from other models that have similar decay modes (such as an ALP with additional trilinear coupling to Higgs bosons).

2.2.1.6 Summary

In this section, we have proposed a new method of searching for long-lived particles at LHC ("*the long DV scheme*") that utilizes the muon tracker at the CMS experiment. It uses a prompt lepton and a displaced muon pair to reconstruct a displaced vertex event. The scheme is optimal for probing the parameter space of the LLPs with the decay lengths $l_{\mathrm{decay}} \gtrsim 3$ m. We demonstrated the potential of the scheme using three exemplary models: heavy neutral lepton (HNL), Chern-Simons portal, and scalars with quartic coupling. For HNLs, we made a comparison between the long DV scheme and other planned searching schemes at ATLAS/CMS, see *e.g.* [97, 110].

Our conclusions are the following:

- For the HNLs, the long DV scheme can probe the parameter space in the mass range $m_N \lesssim 20$ GeV and down to the mixing angles $U^2 \sim 10^{-8}$ (when mixing with ν_μ).
- The long DV scheme has a unique opportunity to probe the LLPs that decay predominantly into leptons, which is demonstrated by the example of the Chern-Simons portal;
- The long DV search scheme has a sufficiently low SM background even for LLPs with the masses $m \lesssim 5$ GeV, which is unavailable for DV search schemes at the LHC that look for hadronic decay products. In the case of HNLs, this gives a possibility to probe the parts of the parameter space that have not been probed by previous experiments and are outside the reach of the planned Intensity Frontier experiments.
- Although scalars have small branching ratio of decay into muons, the scheme still has potential to search for them, and may be complementary to other search schemes, e.g. those that do not require a prompt lepton for tagging.

2.3 Searches for light dark matter at SND@LHC

Historically, the first model of dark matter was WIMP (Weakly Interacting Massive Particle) – a massive electrically neutral fermion that interacts with SM particles via weak interactions. Similarly to neutrinos, WIMPs were at thermal equilibrium in the early Universe, which has maintained by processes WIMP + WIMP $\leftrightarrow$ SM + SM.

> Because of tiny interaction strength, light WIMPs may leave thermal equilibrium at high temperatures $T \simeq m_{\text{WIMP}}$, when their number density is not Boltzmann-suppressed. Therefore, their energy density would overproduce the Universe. This happens if masses are $m_{\text{WIMP}} \lesssim 5$ GeV. The corresponding bound is known as the Lee-Weinberg bound [125].

To evade the Lee-Weinberg bound, one may assume the existence of a mediator other than W, Z bosons that mediates the WIMP annihilation with larger strength. DM candidate particles in such models may be made lighter than 5 GeV and are called light dark matter (or LDM).

> In the minimal model with an LDM χ and mediator V, there are several parameters: LDM mass, mediator's mass, mediator's coupling to SM particles g, mediator's coupling to LDM g_χ. The information about these parameters may be obtained using combined results from searches by different experiments, see Sec. 1.2 and Fig. 2.19.

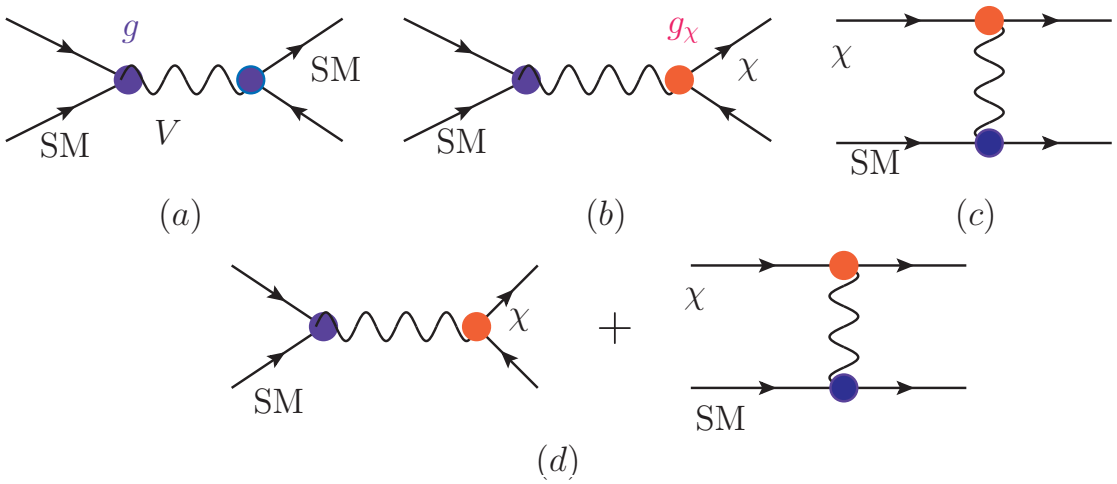

Figure 2.19: Different search schemes for LDM and mediator: visible (a) and invisible (b) decays of mediator, scatterings of LDM χ from the local DM population in our galaxy (c), and scatterings of LDM produced at accelerator experiments (d).

For instance, direct dark matter detection experiments probe scatterings of non-relativistic DM particles off nuclei/electrons, and the number of events scales as $N_{\text{events}} \propto \sigma_{\text{scatt}} = g_\chi^2 g^2/m_V^2$. Accelerator experiments that search for scatterings of LDM particles require also its production, and therefore the number of events is $N_{\text{events}} \propto N_{\text{prod}} \cdot \sigma_{\text{scatt}} \propto g^2 \cdot \text{Br}(V \to \text{LDM}) \cdot g_\chi^2 g^2 f(m_V, m_\chi)$. Experiments that search for visible decays of V constrain the combination $N_{\text{prod}} \cdot P_{\text{decay}} \propto g^2 \cdot g^2 \text{Br}(V \to \text{visible})$. Finally, at experiments that search for invisible decays of the mediator bounds, the number of events scales as $g^2 \cdot \text{Br}(V \to \text{LDM})$.

Therefore, we see that in dependence on the scaling $N_{\text{events}} \propto g^\alpha$ there are g^2 experiments and g^4 experiments. In the absence of limitations that are specific to the given search scheme, g^2 experiments are obviously more sensitive to small couplings than g^4 experiments. This is indeed the case for electron fabric experiments such as NA64, Belle/Belle II, and BaBar, which search for events with missing energy/momentum. However, the situation is different for monojet searches at hadron colliders, which suffer from backgrounds and require severe cut on the transverse momentum/missing transverse energy, $p_T \gtrsim 100$ GeV, which is unrealistic for the case of light GeV-scale mediators [126].

LDM with mediators that interact with electrons/photons is strongly constrained by "g^2 experiments": Belle, BaBar, NA64. Such bounds may be diminished if considering models in which mediators do not interact with leptons.

2.3.1 Scattering off nucleons: different signatures

Let us look closer at the LDM scattering off nucleons. This scattering may be mimicked by neutral current (NC) neutrino scattering events, and therefore, such a search is not background-free. Typically, to observe a signal over background, many LDM scattering events are required [127–131]. Under this condition, one can look for an excess of a signal over the numerous neutrino background, and in particular to distinguish events with LDM and neutrinos kinematically by comparing their reconstructed energy spectra. It would be therefore attractive to consider signatures that require less amount of events.

There are two types of LDM scattering off protons: elastic and inelastic, producing an isolated proton or hadronic showers, respectively. For light mediators with masses $m_V \lesssim 1$ GeV, elastic events yield is dominant, while for heavier mediators inelastic scatterings become the main scattering channel. For neutrinos, the dominant channel is inelastic scatterings.

Indeed, both elastic and inelastic differential cross sections depend on the mediator mass m_V as $d\sigma/d\Omega \propto (Q^2 + m_V^2)^{-2}$ due to the propagator, where Q^2 is the momentum transfer. However, the elastic cross section also includes the proton form factor that limits

the possible momentum transfer to $Q^2 \lesssim r_p^{-2} \simeq 1\,\mathrm{GeV}^2$. For large masses m_V, this leads to an additional suppression as compared to the inelastic cross section, to which all $Q^2 \lesssim m_V^2$ contribute without the suppression [131]. As a result, the ratio $\sigma_{\rm el}/\sigma_{\rm inel}$ is a decreasing function of m_V. We illustrate this feature in Fig. 2.20, considering a model of a scalar LDM that interacts with protons via a vector-like mediator. We see that in the case of light mediator $m_V \lesssim 1\,\mathrm{GeV}$, the elastic and inelastic scattering yields may be comparable, and therefore, the elastic signature is more sensitive due to the low background. However, with the increase of m_V, $\sigma_{\rm el}/\sigma_{\rm inel}$ quickly diminishes, and the inelastic signature starts to dominate.

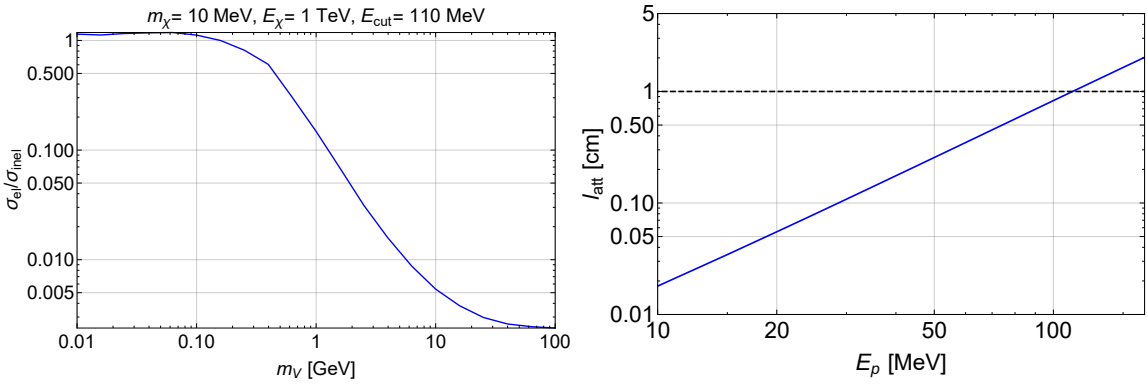

Figure 2.20: *Left panel*: the ratio $\sigma_{\rm el}/\sigma_{\rm inel}$ of elastic and inelastic scattering cross sections in the model with a vector mediator V interacting with protons and a scalar dark sector particle χ of mass $m_\chi = 10\,\mathrm{MeV}$ and energy $E_\chi = 1\,\mathrm{TeV}$. The minimal proton kinetic energy $E_{\rm cut} \gtrsim 110\,\mathrm{MeV}$ is assumed, for which protons may travel 1 cm in tungsten before being absorbed (see text for details). For the description of the elastic and deep inelastic scattering (DIS) used in the estimates, see Appendix 2.C.2. *Right panel:* proton's attenuation length ($l_{\rm att} = \int_0^{E_p} \frac{dE}{dE/dx}$, where dE/dx is the energy loss per unit length) in tungsten as a function of its kinetic energy. The value is calculated using the data from [132].

For masses $m_V \gtrsim m_p$, LDM is more likely to scatter inelastically. In this case, we need to see these events over the numerous neutrino scattering background. A good signature in this case is an increase of the ratio of neutral current and charged current events $N_{\rm NC}/N_{\rm CC}$ for neutrinos. On one hand, it is uniquely predicted within the SM. For the tungsten target, under the approximation of equal differential distributions of ν and $\bar{\nu}$, the ratio $N_{\rm NC}/N_{\rm CC}$ for deep inelastic scattering is equal to [32]

$$P = \frac{1}{2}\left[1 - 2\sin^2\theta_W + \frac{20}{9}\sin^4\theta_W - \lambda(1 - 2\sin^2\theta_W)\sin^2\theta_W\right] \approx 0.33 \qquad (2.3.1)$$

where $\lambda = 0.040$ for the tungsten target. Another advantage of the NC/CC signature is that it is free from the total neutrino flux normalization. This is crucial as it is a subject of theoretical uncertainties for some experiments, for instance, for those at the LHC that are

located in the far-forward direction [133]: currently approved FASER/FASERν [17, 33], and SND@LHC [32].[8]

For LDM particles that scatter of nucleons mostly inelastically, a good search signature at any experiment is an increase of the NC/CC ratio as compared to SM predictions.

2.3.1.1 Model example: leptophobic portal

An example of a model with LDM in which the mediator does not interact with photons and electrons is the leptophobic portal [127, 128, 131, 134, 135]:

$$\mathcal{L}_{\text{leptophob}} = -g_B V^\mu J^B_\mu + g_\chi V^\mu (\partial_\mu \chi^\dagger \chi - \chi^\dagger \partial_\mu \chi), \quad J^B_\mu = \frac{1}{3}\sum_q \bar{q}\gamma_\mu q \tag{2.3.2}$$

Here, g_χ, g_B are coupling constants of the mediator to χ and SM sector, and the sum in J^μ_μ is made over all quark flavors.

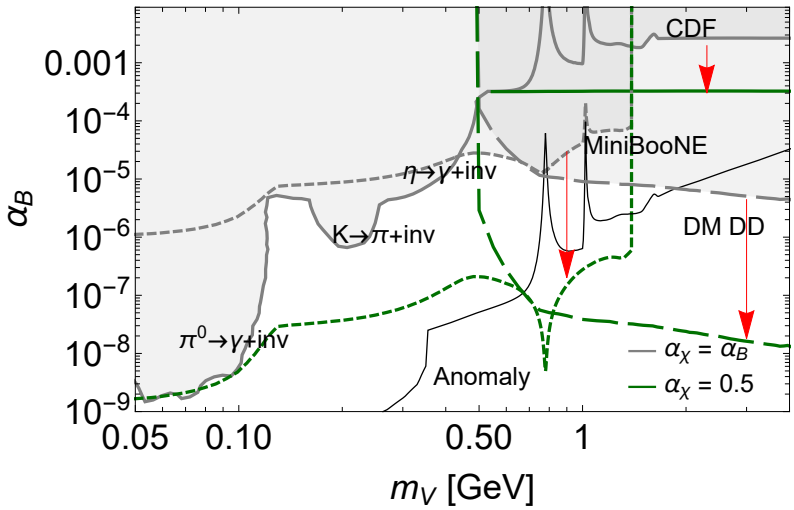

Figure 2.21: Constraints on the leptophobic portal for the case $m_\chi = m_V/3$ and two values of the coupling $\alpha_\chi = g_\chi^2/4\pi$: $\alpha_\chi = \alpha_B$ (in gray) and $\alpha_\chi = 0.5$ (in green). See text for details.

Constraints on the model (2.3.2) are summarized in Fig. 2.21. For $m_V \gtrsim 0.1\,\text{GeV}$, they come from searches for decays $\pi, K, \eta \to V\gamma$ at CB [136], E949 [137], and NA62 [138] experiments (for $m_V \lesssim 0.5\,\text{GeV}$), searches for scattering of χ particles off nucleons at MiniBooNE [139] (for $0.5\,\text{GeV} \lesssim m_V \lesssim 1.5\,\text{GeV}$), a monojet signature analysis at CDF [140] (for $m_V \gtrsim 1.5\,\text{GeV}$), and direct DM searches at CRESST III [34].

The weakness of the CDF monojet signature and the absence of direct constraints from the LHC is caused by the requirement of a large missing transverse momentum

[8] One of goals of these experiments is to study the production of mesons in the far-forward direction.

$p_T \sim 100\,\mathrm{GeV}$ for signal tagging and background suppression. Such large p_T may be provided only by large mass of a decaying particles, which is definitely not the case of light $\mathcal{O}(1\,\mathrm{GeV})$ mediators considered in this section. The bounds from MiniBooNE, being one of the strongest in the region $m_V \lesssim 1\,\mathrm{GeV}$, are much weaker at larger masses due to small center-of-mass energy of the pp collisions, $\sqrt{s} \approx 4\,\mathrm{GeV}$.

Another constraint comes from DM direct detection experiments (DD) [127] that search for scattering of DM particles off nuclei. The sensitivity of these experiments depends on DM particle mass. Indeed, it determines the maximal kinetic energy of DM (which is $T_\chi = m_\chi v_{\text{escape}}^2/2$, where $v_{\text{escape}} = 544\,\mathrm{km/s}$ is the escape velocity), and, therefore, the maximally possible nuclear recoil energy T_N. The DD experiments have finite energy threshold, being $T_N > 30.1\,\mathrm{eV}$ for CRESST-III [34] that is currently the most sensitive experiment. As a result, current constraints from DD are limited by $m_\chi \gtrsim 160\,\mathrm{MeV}$. In addition, the DD bounds may be significantly relaxed even for heavy χ particles if assume their axial-vector interaction with V instead of vector-like one (see [141]) that results in the velocity-suppressed scattering cross section.

Finally, in [142–144], it was argued that the strongest constraint may come from negative results of searches for decays

$$K \to \pi + \text{inv},\ B \to K + \text{inv},\ Z \to \gamma + \text{inv} \tag{2.3.3}$$

at LHCb. In the model of the leptophobic portal (2.3.2), the decays (2.3.3) may result from the anomalous violation of the baryon current conservation, which requires a UV completion in order to cancel the anomaly. Namely, in [142, 143], it was considered a UV completion with some heavy fermions such that the full theory is anomaly-free. At energies much lower than masses of these fermions, the effective theory contains, apart from the Lagrangian (2.3.2), pseudo-Chern-Simons (pCS) interaction operators between V and electroweak bosons W, Z, γ that result from the contribution of massive fermions to the anomalous triangle diagrams. The latter include two summands: a mass-independent, and a mass-dependent. The sum of the first terms over all fermions vanishes due to the anomaly cancellation, while the net mass-dependent part is in general non-zero (for instance, if there is a hierarchy in fermion masses). The corresponding interactions mediate the process $Z \to \gamma + X$, and generate effective flavor changing neutral current couplings bsV, sdV between quarks and the leptophobic mediator (via penguin loop diagrams) that mediate the first two processes in Eq. (2.3.3).

pCS terms generically appear in effective theories with chiral fermions. However, their contribution to the processes (2.3.3) depends on the UV completion of the model (2.3.2). For instance, one could consider a 3+n+1 dimensional model with SM physics localized on a 3+1 dimensional sub-manifold (brane) and a large mass gap for the bulk modes (see e.g. [145, 146]). The higher-dimensional theory is anomaly free by construction without adding extra fermions. The anomaly of the low-dimensional 3+1 effective theory is done by the "anomaly inflow" mechanism, non-local from 3+1 dimensional point of view. In

this case, the anomaly cancellation by massive modes does not contribute to decays.

Due to the model dependence, the status of the anomaly constraint is different from the status of the other bounds discussed above, as the latter require only the effective Lagrangian (2.3.2). Therefore, in Fig. 2.21 we just indicate the parameter space potentially constrained by processes (2.3.3) by showing its lower bound only, while for the other constraints discussed in this subsection the whole parameter space is shown in solid gray.

Choice of parameters. The parameters in the model are LDM particle and mediator masses $m_{\chi,V}$, and the couplings $\alpha_B = g_B^2/4\pi, \alpha_\chi = g_\chi^2/4\pi$.

The of the previous bounds with α_χ is the following. While the scaling of the number of events at MiniBooNE is $N_{\text{events}} \propto \alpha_B \cdot \text{Br}(V \to \chi\bar{\chi}) \cdot \alpha_B \cdot \alpha_\chi$, the number of events at the other experiments scales as $N_{\text{events}} \propto \alpha_B \cdot \text{Br}(V \to \chi\bar{\chi})$ for the collider experiments and $\alpha_B \cdot \alpha_\chi$ for DD experiments (remind the discussion in the beginning of Sec. 2.3).[9]

Further, we consider two values of α_χ. The first one is $\alpha_\chi = \alpha_B$, which is typically considered in the literature, and for which $N_{\text{events}} \propto \alpha_B^3 \cdot \text{Br}(V \to \chi\bar{\chi})$ at the accelerator direct detection experiments. The second one is $\alpha_\chi = 0.5$, for which $N_{\text{events}} \propto \alpha_B^2$.

Let us now comment on the choice of m_χ. As we have discussed previously, masses $m_\chi > 160\,\text{MeV}$ are significantly constrained by the DM direct detection experiments. Therefore, we consider two different choices: $m_\chi = m_V/3$, which is commonly used in literature and for which the DD constraint is important above $m_V = 480\,\text{MeV}$, and $m_\chi = 20\,\text{MeV}$, for which there is no bound from DD at all.

2.3.2 SND@LHC

2.3.2.1 Description of experiment

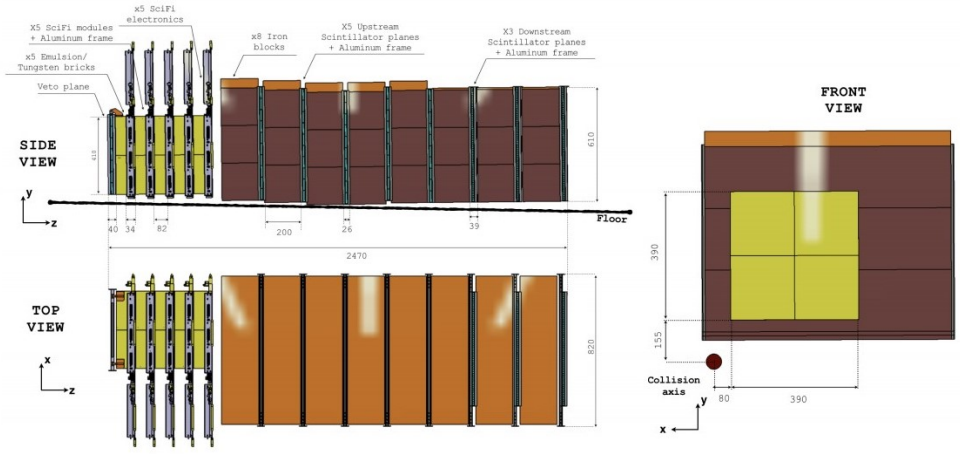

Figure 2.22: Overview of the SND@LHC detector facility: the side view (on the left) and the front view (on the right). The figure is taken from [32].

[9] For the calculation of the branching ratio $\text{Br}(V \to \chi\bar{\chi})$, see Appendix 2.C.

SND@LHC facility is planned to be installed in the TI18 tunnel at the distance of 480 m from the ATLAS interaction point along the beam collision axis. The SND@LHC detector consists of the target region followed by the muon system, see Fig. 2.22. The pseudorapidity range covered by the target will be $7.2 < \eta < 8.6$, in which ν_e, ν_τ are produced in decays of heavy mesons, with an additional component of muon neutrinos originated from decays of pions and kaons. The actual angular position of the target is $(\theta_x, \theta_y) \in [0.17, 0.98] \times [0.32, 1.14]\,\mathrm{mrad}^2$.

The target has brick structure: bricks of emulsion cloud chambers (ECC) followed by Scintillating Fibre (SciFi) plates. Each of five emulsion bricks consists of 60 emulsion films interleaved with 59 tungsten plates of 1 mm thickness, which serve as target. The total scattering length of the emulsion bricks is $29.5\,\mathrm{cm}$, which corresponds to 84 radiation lengths (X_0), and the total target length of about $40\,\mathrm{cm}$. The ECC provide micrometric accuracy that allows one to measure accurately tracks of charged particles, and reconstruct vertices of neutrino interactions (for events without muons) and any other event that deposit their energy within one emulsion brick.

SciFi predicts the neutrino interaction vertex location, connects the emulsion track with the muon track identified by the muon detector, and provides timing information for the events (with timing resolution of the order $100\,\mathrm{ps}$). In this way, it serves for disentangling the piled up events occurring in one emulsion layer. In addition, the whole facility works as a hadronic calorimeter with 9-11 interaction lengths.

An important feature of SND@LHC is high neutrino type identification efficiency. The target construction allows track detection of charged particles produced in primary interactions and subsequent decays. Muons are identified as the most penetrating charged particles, while τ-leptons – via a displaced vertex with an electron or a muon track.

There are two phases of the event reconstruction at SND@LHC [32]. The first phase uses electronic detectors: events are reconstructed based on veto, the target tracker and the muon system. The second phase adopts the emulsion target, and the event reconstruction will be available six months after the exposure. It identifies EM showers, complements the target tracker for EM energy measurements, and allows for the neutrino vertex reconstruction. The matching between these phases is required for events containing muons; it is the subject of ongoing studies.

Although the main goal of SND@LHC is to probe high-energy neutrino scatterings, it may also search for scatterings for LDM and probably even for decays of mediators.

Below, we illustrate the potential of SND@LHC to probe FIPs via decays and scatterings by estimating the sensitivity to several models. We consider two experimental setups of the detector: one that will operate during Run 3, and a possible upgrade that will work during Run 4 (see [32]). Their parameters are summarized in Table 2.3.

Setup	$\mathcal{L}$, fb^{-1}	$l_{\rm det}^{\rm scatt}$, cm	$l_{\rm det}^{\rm decay}$, cm
Setup 1	150	30	50
Setup 2	3000	75	125

Table 2.3: Experimental setups of the SND@LHC detector used in this work. The parameters are the integrated luminosity $\mathcal{L}$, the detector's length available for scatterings $l_{\rm det}^{\rm scatt}$, the detector's length available for decays $l_{\rm det}^{\rm decay}$. See text for details.

2.3.2.2 Search for scatterings of LDM

Let us discuss signal efficiency and background at SND@LHC for the *elastic signature*.

> Even a few of LDM elastic scattering events would lead to observable deviation from SM predictions at SND@LHC.

According to [32], the selection criterion for the elastic scattering off protons is a single isolated track with the momentum $p > 170\,\mathrm{MeV}$ observed in the emulsion. Studies of MC simulation containing produced particles prior to their interaction with the detector have shown that once this selection is applied to events with neutrinos, the only surviving background comes from neutrino NC resonant and deep inelastic scatterings, where only one charged track is visible, with the total number of events being 1.7. To overcome this background with 2σ CL, we require 5 elastic events.[10]

At the same time, low-momentum protons with $p \simeq 170\,\mathrm{MeV}$ that satisfy the selection criterion considered in [32] have small attenuation length[11] in tungsten, $l_{\rm att} \lesssim 1\,\mathrm{mm}$, and get absorbed in the single tungsten plate before reaching the emulsion [147]. Therefore, this requirement is not appropriate for estimate of the number of elastic χ scatterings that can be detected. Instead, for LDM signature, we require proton kinetic energy $E_p > 110\,\mathrm{MeV}$, for which protons have attenuation length $l_{\rm att} = 1\,\mathrm{cm}$, and therefore may pass through ten emulsion layers, see Fig. 2.20. We note that this criterion is conservative and may be further relaxed.

Since the requirement on the proton energy is more tight than the one assumed in the simulation in [32], the background would be lower and can go down to zero. However, precise background estimates require additional studies. Therefore, we conservatively assume the same number of background events as for the looser requirement of $p > 170\,\mathrm{MeV}$ as in [32].

Still, even with the stronger cut, such protons may not reach SciFi plates, which may lead to piled up events and potentially decrease the efficiency of the reconstruction of the signal. Namely, there may be a coincidence between the neutrino and χ scattering events,

[10] We estimate it using the relation $\sum_{n_{\rm ev}=0}^{\infty} P(n_{\rm ev}|b+s)P(n \geq n_{\rm ev}|b) < 4.5\%$, where $b = 1.7$ is the background, s is the signal, and P is the Poisson distribution.

[11] Contrary to high-energy protons that produce hadronic cascades, protons with low energies lose this energy electromagnetically and travel around $\mathcal{O}(1\,\mathrm{cm})$ in tungsten before being absorbed [52].

accumulated during six months of the SND emulsion phase operation period. However, in practice this feature does not affect the efficiency. The expected amount of neutrino events at SND@LHC during Run 3 is $\simeq 2 \cdot 10^3$ [32]. Since the DIS event may be reconstructed by the single emulsion layer, for estimating of the pile up effect we are interested in the number of neutrino events per layer, instead of their total amount. During six months of the operation period and given that the SND target consists of 60 emulsion layers, only 6 neutrino events will occur per each layer. Taking into account the micrometric accuracy of the emulsion, this amount is vanishingly small to affect the signal.

Let us comment on effects that have not been included in the simulation discussed in [32]. The first effect is a possible background from neutrino DIS events, in which soft particles get absorbed in tungsten layer before reaching the emulsion. As a result, the DIS event may mimic an elastic scattering event. However, this effect can be neglected, as high-energy neutrinos typically deposit large amount of energy in their scattering leading to many tracks in the event. Indeed, assuming the operating period corresponding to the LHC Run 3, the MC simulations in [32] have shown no events of neutrino DIS with only one track having $E > 110\,\mathrm{MeV}$ and all other tracks with energies $E \ll 100\,\mathrm{MeV}$ (such that they may be absorbed in tungsten before reaching the emulsion layer) that would be recognized as an elastic event. The second effect is a proton-to-neutron conversion, which may reduce the signal from the elastic scattering. However, the nuclear interaction length in tungsten, which is a characteristic scale of the conversion, is $\mathcal{O}(5\,\mathrm{cm})$ corresponding to about 50 emulsion layers in the target. Therefore, the proton would produce a visible track in large fraction of events, and this process will not affect the elastic event reconstruction. Finally, the third effect is possible background from radioactive isotopes that may be hypothetically present in the target. However, the decay products are typically low-energy, as energy release for most isotopes does not exceed $10\,\mathrm{MeV}$, and therefore, they get absorbed in the single tungsten layer without being detected.

For masses $m_V \gtrsim m_p$, LDM is more likely to scatter inelastically, and we consider the *NC/CC signature*.

For the NC/CC signature, the required number of events at SND@LHC is $\mathcal{O}(100)$ at 2σ confidence level.

Indeed, SND@LHC claims to measure the NC/CC ratio with the precision of 10%. Assuming $N_{\mathrm{CC}}^{\mathrm{SND@LHC}} = 1395$ and $N_{\mathrm{NC}} = 450$ as predicted by simulations for the SND@LHC setup [32], we require the yield of LDM inelastic scattering events to be $2\sqrt{450 + 45.0^2} \approx 100$ in order to reach the 2σ confidence level.

2.3.2.3 Search for decays of mediators

Although SND@LHC is constructed to probe neutrino scatterings, it may also be capable of searching for decays of FIPs, for instance scattering mediators. It is attractive to probe the parameter space simultaneously by scatterings of LDM and decays of mediators.

However, for the given coupling g, the decay length of the mediator is typically much shorter than the scattering length of LDM particles. Because of this, it is typically not possible to probe decays of mediators and scatterings of LDM within the same domain of parameter space.

Indeed, consider scatterings and decays using as an example vector (dark photons) and scalar portals. The scattering probability is

$$P_{\text{scat}} = \sigma_{\text{scat}} n_{\text{atom}} L_{\text{det}}, \qquad n_{\text{atom}} \sim (1\,\text{keV})^3 \tag{2.3.4}$$

with the scattering cross section

$$\sigma_{\text{scat}} \sim \frac{\alpha_{S\chi\chi} y_N^2 \theta^2}{m_N E_\chi}\ (\text{scalar}), \qquad \sigma_{\text{scat}} \sim \frac{\alpha_D \epsilon^2}{m_V^2}\ (\text{vector}) \tag{2.3.5}$$

In its turn, the decay probability is

$$P_{\text{dec}} = \frac{\Gamma L_{\text{det}}}{\gamma}, \qquad \Gamma \sim \theta^2 \frac{m_S^3}{v^2}\ (\text{scalar}), \qquad \Gamma \sim \epsilon^2 m_V\ (\text{vector}) \tag{2.3.6}$$

Comparing these two probabilities, one gets

$$\frac{P_{\text{scat}}}{P_{\text{dec}}} \sim \gamma \alpha_{S\chi\chi} y_N^2 \left(\frac{v}{m_S}\right)^2 \frac{n_{\text{atom}}}{m_S m_N E_\chi}\ (\text{scalar}), \qquad \frac{P_{\text{scat}}}{P_{\text{dec}}} \sim \gamma \alpha_D \frac{n_{\text{atom}}}{m_V^3}\ (\text{vector}) \tag{2.3.7}$$

As a result, for large couplings, that are required to see scatterings, the decay length is microscopic, and mediators decay before reaching the detector. It may be still possible though to probe large couplings via scatterings of LDM and smaller couplings via decays of the mediator.

There are potentially background-free signatures of mediators decays at SND@LHC. These are decays into a di-lepton pair, into a lepton and a meson, or into two mesons.

Indeed, a clear background-free signature may be decays of a FIP into a di-lepton pair, $V \to ll'/ll'\nu$, as scatterings of neutrinos produce at most one lepton.[12] For the decays of FIPs into a lepton and a meson, or into a pair of mesons there is a background that comes from the neutrino deep inelastic CC- and NC-scatterings correspondingly. However, decay products typically carry large energies $E \gtrsim 100\,\text{GeV}$ and, therefore, can be distinguished from (inelastic) neutrino scatterings with such large energy transfers as the latter typically

[12]The di-muon events may be produced by the scattering of photons in the detector. However, the photons occur in scattering of neutrinos, and apart from the di-muon pair there would be a lot of other tracks.

produce a lot of hadrons. Therefore, we believe that the mentioned background may be rejected. This question requires an additional study.

To use these signatures, it is necessary to disentangle tracks from decays of FIPs. At the SND@LHC detector, this is possible if the transverse distance between the tracks exceeds the spatial resolution, which is of the order of $1\,\mu\mathrm{m}$ for the emulsion films. The transverse distance between two tracks is determined by the flight angle that can be estimated as $\Delta\theta \simeq \frac{m_{\text{FIP}}}{E_{\text{FIP}}}$, and the distance l charged particles travel inside the target. For electrons, l is the radiation length, which in tungsten is equal to $3.5\,\mathrm{mm}$. Muons pass through the whole target without deflection, and therefore, we may conservatively restrict l to the thickness of a single SND@LHC emulsion brick $7.8\,\mathrm{cm}$. For FIPs flying in the far-forward direction, the typical energy is $E_{\text{FIP}} \simeq 1\,\mathrm{TeV}$. Thus, the corresponding masses are

$$m_{\text{FIP}} \gtrsim \min\left[E_{\text{FIP}}\,\frac{1\,\mu\mathrm{m}}{l}, 2m_{e\text{ or }\mu}\right] \simeq \begin{cases} 290\,\mathrm{MeV}, & \text{FIP} \to e\bar{e} \\ 210\,\mathrm{MeV}, & \text{FIP} \to \mu\bar{\mu} \end{cases} \tag{2.3.8}$$

If the disentanglement is not possible, instead of tracks we observe a mono-cascade. A similar signature may come from FIPs decaying into neutral pions, such as from HNLs that mix with tau flavor that decay into π^0 and a neutrino. This type of events may still be distinguished from neutrino scatterings, as the latter typically contain many tracks, and hence may be a new physics signature.

Background evaluation and event reconstruction for both charged pair and monocascade signatures are challenging tasks and require dedicated studies. Further, we will show the fixed signal events contours, assuming that all events are detected.

2.3.3 Sensitivity of SND@LHC

2.3.3.1 Leptophobic portal

Let us now estimate the sensitivity of SND@LHC to LDM scattering in the model (2.3.2). The number of scattering events may be estimated using the formula

$$N_{\text{events}} = 2\cdot N_{\chi}^{\text{SND@LHC}} \times n_{\text{detector}} \times \begin{cases} Z\cdot\sigma_{\text{scatt}}^{\text{el}}(\langle E_\chi\rangle)\cdot l_{\text{det}}^{\text{scatt}}, & \text{elastic signature} \\ A\cdot\sigma_{\text{scatt}}^{\text{inel}}(\langle E_\chi\rangle)\cdot l_{\text{det}}^{\text{scatt}}, & \text{inelastic signature} \end{cases} \tag{2.3.9}$$

Here, $N_{\chi}^{\text{SND@LHC}}$ is the number of χ particles produced in the direction of the SND@LHC detector volume (a factor of 2 stays for $\bar{\chi}$), n_{detector} is the detector's atomic number density (the tungsten material is considered), Z, A are atomic and mass numbers of the target material, and $\sigma_{\text{scatt}}^{\text{el/inel}}$ is the elastic or inelastic scattering cross section of χ particles. For simplicity, in cross section calculation we assume that all χ particles have the same energy equal to their average energy $\langle E_\chi\rangle$.

We adopt the description of the elastic scattering process from [134]. For the estimate of the cross-section for inelastic scattering, we use the calculation based on the parton model from [148], for which parton distribution functions are given by CT10nlo PDF sets from LHAPDF package [149] (see also Appendix 2.C.2).

Let us now consider the production of χ particles. The $\chi\bar{\chi}$ pairs originate from decays of V. Similarly to the dark photon case, the mediator may be produced:

1. in decays of unflavored mesons π, η,

$$\pi \to V + \gamma, \quad \eta \to V + \gamma, \tag{2.3.10}$$

2. by proton bremsstrahlung,

$$p + p \to V + X, \tag{2.3.11}$$

3. in Drell-Yan process,

$$q + \bar{q} \to V + X, \tag{2.3.12}$$

see Fig. 2.23. For the description of these channels, we mainly follow [144, 150, 151].

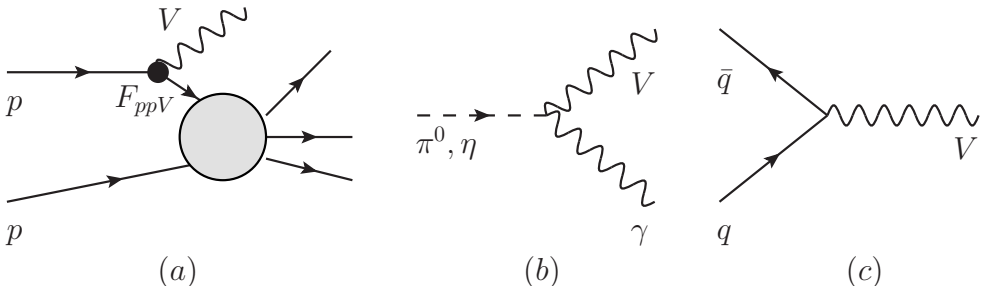

Figure 2.23: Diagrams of the production of the leptophobic mediator V: by proton bremsstrahlung (a), in decays of light unflavored mesons (b), and in Drell-Yan process (c).

For the production from mesons, we use the polar angle and energy distributions of π, η mesons generated by EPOS-LHC [152] as a part of the CRMC package [153]. The resulting spectra of V and χ particles are obtained semi-analytically using an approach presented in [120].

For obtaining the angle-energy distribution of the leptophobic mediator produced by the proton bremsstrahlung, we consider the kinematic range $p_T < 1\,\mathrm{GeV}$ and $0.1 < z < 0.9$. The corresponding production probability is affected by the mixing of V with isoscalar ω and ϕ mesons. To describe this effect, we follow the procedure described in [154] (see Appendix 2.C for details). The distribution of subsequent χ particles produced by the bremsstrahlung is obtained in a similar way as for the case of the production from mesons.

For the production in the Drell-Yan process, we use our implementation of the model (2.3.2) in MadGraph5 [113] with FeynRules [117, 118]. We then obtain the geo-

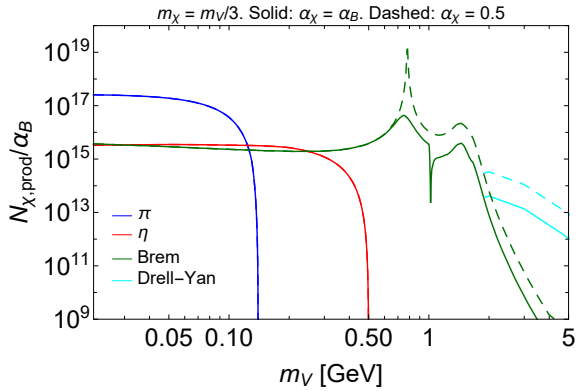

Figure 2.24: The number of χ particles produced in the direction of the SND@LHC experiment, assuming the integrated luminosity $\mathcal{L} = 150\,\mathrm{fb}^{-1}$. $m_\chi = m_V/3$ is assumed. Wiggles around V masses of $782\,\mathrm{MeV}, 1020\,\mathrm{MeV}$ and $\simeq 1.7\,\mathrm{GeV}$ are caused by the mixing of the mediator with isoscalar vector mesons ω, ϕ, and their excitations, which leads to the resonant enhancement of 1) the ppV form-factor for the production by the proton bremsstrahlung, and 2) the decay width of the leptophobic mediator V into hadrons (and hence to a suppression of $\mathrm{Br}(V \to \chi\bar{\chi})$). See text and Appendix 2.C for details.

metric acceptance and energy distribution of χ particles traveling into the direction of the SND@LHC detector by simulating the leading-order process $p + p \to V,\ V \to \chi\bar{\chi}$.

We find that the main production channel for masses $m_V \lesssim m_\eta$ is decays of mesons, for masses $m_\eta \lesssim m_V \lesssim 3\,\mathrm{GeV}$ is the proton bremsstrahlung, and, finally, for $m_V \gtrsim 3\,\mathrm{GeV}$ it is the Drell-Yan process, see Fig. 2.24.

Most of the produced χ particles have γ factors $\sim 10^3$, independently of the production channel. This means that the time-of-flight measurement is not efficient in separating signal χ particles and neutrinos.

Sensitivity. Let us now discuss the sensitivity. The parameters in the model are LDM particle and mediator masses $m_{\chi,V}$, and the couplings $\alpha_B = g_B^2/4\pi, \alpha_\chi = g_\chi^2/4\pi$.

The choice of α_χ affects the parameter space probed by SND@LHC in the following way. The number of scattering events at SND@LHC scales as

$$N_{\mathrm{events}} \propto \alpha_B \cdot \mathrm{Br}(V \to \chi\bar{\chi}) \times \alpha_B \cdot \alpha_\chi \tag{2.3.13}$$

Here, a factor $\alpha_B \cdot \mathrm{Br}(V \to \chi\bar{\chi})$ comes from the production, while a factor $\alpha_B \cdot \alpha_\chi$ – from the subsequent scattering of χ particles. The scaling of the previous bounds is somewhat different. While the scaling of events at MiniBooNE is similar, the number of events at the other experiments scales as $N_{\mathrm{events}} \propto \alpha_B \cdot \mathrm{Br}(V \to \chi\bar{\chi})$ for the collider experiments and $\alpha_B \cdot \alpha_\chi$ for DD experiments. Therefore, the dependence on α_B and α_χ is weaker.[13] Therefore, marginalizing over α_χ, the optimal choice would be $\alpha_\chi \simeq 1$, for

[13]For the calculation of the branching ratio $\mathrm{Br}(V \to \chi\bar{\chi})$, see Appendix 2.C.

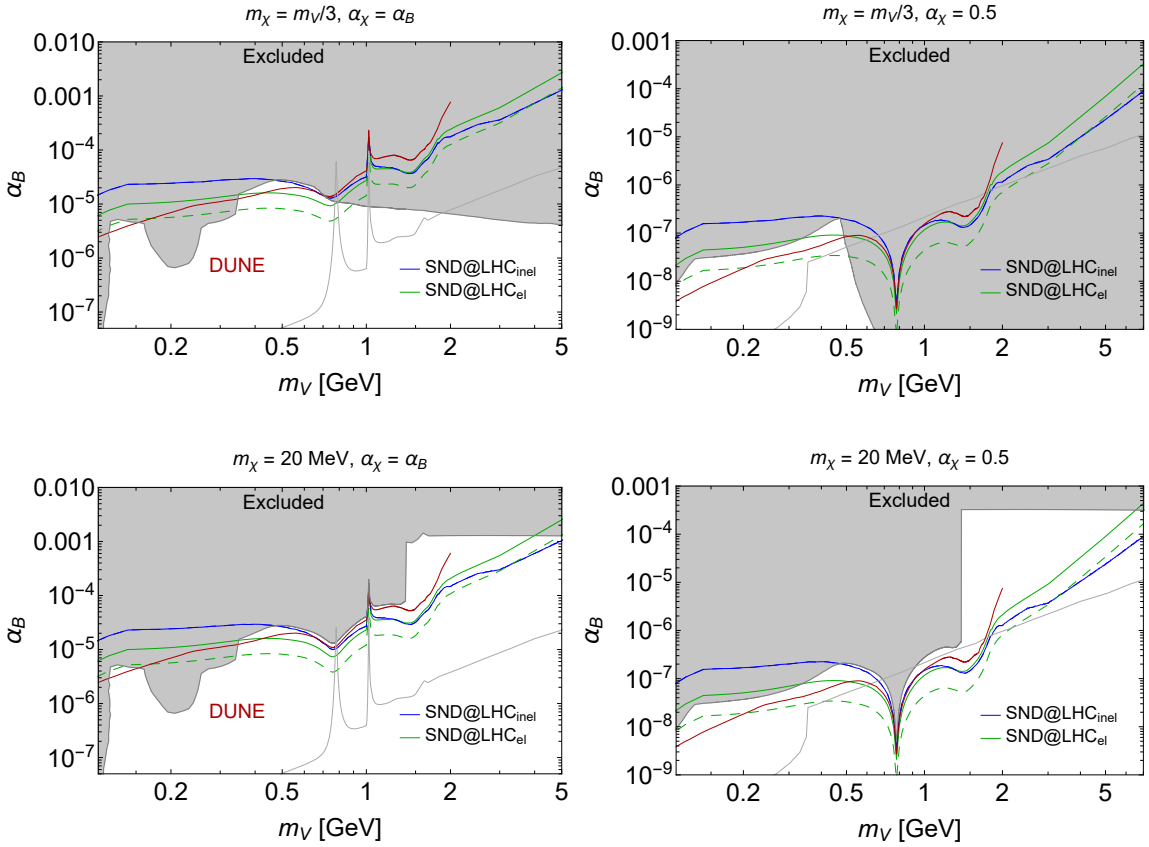

Figure 2.25: Sensitivity of the SND@LHC experiment to the leptophobic portal (2.3.2) (2σ CL). The sensitivity is shown under an assumption $m_\chi = m_V/3$ (top panel) and $m_\chi = 20\,\mathrm{MeV}$ (bottom panel), and for two different choices of the coupling of mediator to χ particles: $\alpha_\chi = \alpha_B$ (left figures), and $\alpha_\chi = 0.5$ (right figures). The considered signatures are the elastic scattering off protons (the green line) and the deep-inelastic scattering (the blue line, corresponding to 100 signal events during Run 3), see text for details. For the elastic signature, the solid line corresponds to the sensitivity during Run 3 (corresponding to 5 signal events), while the dashed line denotes the sensitivity of the upgraded setup that may operate during Run 4 (see text for details). We assume that the level of background to the elastic signature during Run 3 is 1.7 events, as reported in [32] for much weaker cut on the proton's momentum $p > 170\,\mathrm{MeV}$ that the cut used in our estimate, $p \gtrsim 500\,\mathrm{MeV}$. Therefore, the sensitivity is conservative. By the red line, we show the 100 event contour for the DUNE experiment from Ref. [130]. We rescale the previous bounds according to our description of the proton form-factor used in bremsstrahlung and $\mathrm{Br}(V \to \chi\chi)$. The thin gray line corresponds to model-dependent constraints from invisible decays (2.3.3) as derived in [144] (see text for details).

which SND@LHC would probe larger range of mediator masses.

The sensitivity of the SND@LHC experiment to the leptophobic portal for two different setups from Table 2.3 is shown in Fig. 2.25. Following the discussion in Sec. 2.3.2.2, we require $N_{\text{events}} > 5$ for the elastic signature and $N_{\text{events}} > 100$ for the inelastic signature

during Run 3.

The parameter space that may be probed by SND varies in dependence on the values of parameters α_χ, m_χ. Namely, for the choice $m_\chi = m_V/3$, SND@LHC only may probe masses $350\,\text{MeV} < m_V < 700\,\text{MeV}$. For the choice $m_\chi = 20\,\text{MeV}$, it is possible in addition to probe masses $700\,\text{MeV} < m_V \lesssim 7\,\text{GeV}$. Moreover, for the choice $\alpha_\chi = 0.5$, the probed range of the coupling α_B even competes with the model-dependent bound from the signature $B \to K +$ inv at the lower bound.

Unlike the case of the direct detection experiments, the sensitivity of SND@LHC depends only weakly on the choice of m_χ, as the production probability and the scattering cross section of χ particles is determined mainly by m_V. In this way, SND@LHC and direct DM detection experiments may probe complementary mass ranges of χ.

In the figure, we also show the sensitivity of DUNE experiment from [130]. The background estimate has not been made for this experiment. Therefore, we show the contour corresponding to 100 events.

Finally, let us discuss the improvement of the sensitivity of SND@LHC for the upgraded setup. For the curve describing elastic scattering signature, the scaling of the lower bound of the sensitivity with the integrated luminosity $\mathcal{L}$ and detector length $l_{\text{det}}^{\text{scatt}}$ is

$$\alpha_B^{\text{el}} \propto \left(N_{\chi,\text{prod}}^{\text{SND@LHC}} \cdot l_{\text{det}}^{\text{scatt}} / \sqrt{N_{\text{bg}}} \right)^{-1/n} \propto (\mathcal{L} \cdot l_{\text{det}}^{\text{scatt}})^{-1/n}, \tag{2.3.14}$$

where the scaling of the number of neutrino background events is $N_{\text{bg}} \propto \mathcal{L} \cdot l_{\text{det}}^{\text{scatt}}$, and $n = 2$ or 3, depending on the choice for α_χ. The improvement of the sensitivity for the upgraded setup reaches a factor of 2 to 3. We note, however, that due to tighter requirement on the proton energy used in our estimate, the elastic signature may be background-free, in which case the improvement increases by up to a factor 10. For the inelastic signature, the scaling depends on the improvement of the uncertainty in the NC/CC ratio measurement during the time frame of the upgraded setup. In the optimistic scenario, the uncertainty will be significantly smaller, and the lower bound will become better by the same factor as for the elastic signature.

2.3.3.2 Decays

To illustrate the potential of SND@LHC to probe decays, we estimate the sensitivity to scalar, neutrino and vector portals, which introduce correspondingly a light Higgs-like scalar, a heavy neutral lepton (HNL) and a dark photon (see, e.g., [67] for the description of the models). Decays with pairs of charged particles in the final state – muons, electrons and pions – are the main decay channels for all the portal particles, except for GeV scale HNLs N that mix with ν_τ, for which the main decay channel is a decay $N \to \pi^0 \nu$.

In order to obtain the sensitivity of SND@LHC to various decaying FIPs, we use the following estimate:

$$N_{\text{events}} = \sum_{\text{i}} N^i_{\text{prod}} \cdot \epsilon^i_{\text{geom}} \cdot P^i_{\text{decay}} \cdot \text{Br}_{\text{vis}} \tag{2.3.15}$$

Here, N^i_{prod} is the total number of FIPs of species X produced via channel i, ϵ^i_{geom} is the geometric acceptance for particle X decay products, and P^i_{decay} is the decay probability averaged over energies E_X of particles X,

$$P^i_{\text{decay}} = \int (e^{-l_{\text{min}}/c\tau_X\gamma_X} - e^{-l_{\text{max}}/c\tau_X\gamma_X}) f^i_{E_X} \mathrm{d}E_X, \tag{2.3.16}$$

with f_{E_X} being the energy E_X distribution of FIPs that fly in the decay volume, and τ_X and γ_X their lifetime and Lorentz boost factor, respectively. Finally, Br_{vis} is the branching ratio of visible decays of particle X. Details of estimates are summarized in Appendix 2.B. The sensitivities are shown in Fig. 2.26, where we show the estimate for the Run 3 setup, as well as for the possible upgrade that may operate during Run 4. For the upgraded setup, the number of events at the lower bound is higher by a factor of $l^{\text{dec}}_{\text{det,upgr}}/l^{\text{dec}}_{\text{det}} \cdot \mathcal{L}_{\text{Run 4}}/\mathcal{L}_{\text{Run 3}} = 50$.

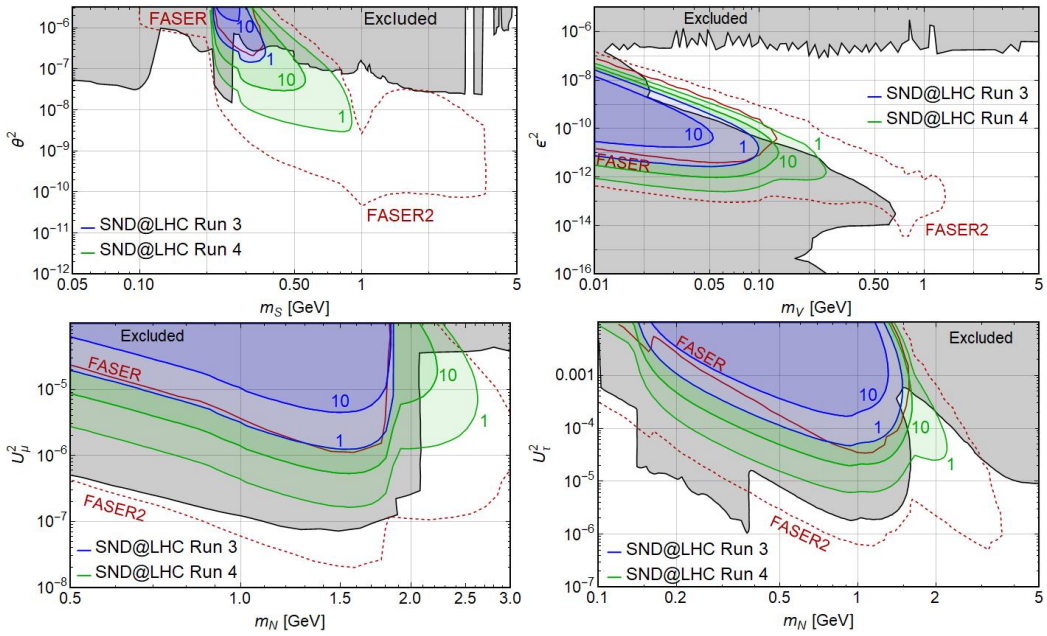

Figure 2.26: Sensitivity of SND@LHC to (top left) dark scalars, (top right) dark photons, and HNLs that mix with (bottom left) ν_μ and (bottom right) ν_τ. Blue (green) contours correspond to 1 and 10 events in the (upgraded) SND@LHC target. Sensitivities of previous experiments and of the FASER/FASER2 experiment are reproduced from [58, 67].

We conclude that for the Run 3 setup, SND@LHC may probe only a tiny parameter space for dark scalars, dark photons and HNLs that mix with τ flavor. For the upgraded setup, it may be possible to probe HNLs that mix exclusively with ν_μ in the mass range $\lesssim 2.5\,\text{GeV}$, and in $\lesssim 2.0\,\text{GeV}$ for pure mixing with ν_τ. SND@LHC may also probe dark

photons at the upper bound of the sensitivity with masses $m_V \lesssim 0.1\,\mathrm{GeV}$, and dark scalars with masses $m_S \lesssim 0.8\,\mathrm{GeV}$.

2.3.4 Comparison with FASER

There is a similarity between the facilities of SND@LHC and FASER/FASERν experiments. They are both placed in a large η region and at the same distance $l_{\text{min}} = 480$ m from the ATLAS interaction point, but in the opposite tunnels. Parameters of the experiments are summarized in Table 2.4. Below, we make a qualitative comparison of the sensitivities of the SND@LHC and FASER experiments, and then comment on the changes due to upgrades.

Detector	l_{min}, m	l_{det}, m	θ_{min}, mrad	θ_{max}, mrad	$\Omega \cdot 10^7$, sr	$\mathcal{L}, \text{fb}^{-1}$
SND@LHC	480	0.5	0.3	1.5	6.9	150
FASER		1.5	0.	0.2	1.4	
FASERν		1.	0.	0.4	2.7	
SND@LHC upgr.		1.25	0.3	1.5	6.9	3000
FASER2		5	0.	2.1	138	

Table 2.4: Parameters of SND@LHC and FASER experiments: the distance to the decay volume, the length of the decay volume, the polar coverage, covered solid angle, total integrated luminosity.

Let us summarize the main differences between SND@LHC and FASER/FASERν detectors in the reconstruction of signal. For scattering, SND@LHC competes with the FASERν detector. FASERν consists of emulsion films interleaved with tungsten plates, only providing the information of spatial position of different tracks with 30% energy reconstruction accuracy for neutrino events (see also [155]). For muons, the situation is much better, as they, being produced in FASERν, may penetrate it and enter FASER, which allows timing and momentum measurements. This option is unavailable, however, for other particles (hadrons, electrons), as they are effectively absorbed in the detector. On the contrary, SND@LHC provides timing measurements by the use of the SciFi technology and the energy reconstruction accuracy of 22% for electrons. For both experiments, timing is needed for rejecting the background induced by high-energy muons and secondary particles.[14]

In the case of decays, SND@LHC competes with FASER, and their detectors provide comparable FIP parameters reconstruction accuracy, thanks to good spatial resolution of the emulsion. Assuming that SND@LHC is a background free experiment when searching for decays, the only relevant quantity for comparing the experiments is the number of correctly identified FIP decay events.

[14] In this work, we compare the signal of new physics to the number of neutrino interactions, which was already obtained under assumption of possible background. Therefore, these key features of the detectors' concept are omitted in our analysis.

2.3.4.1 Lower and upper bound of the sensitivity

Let us now consider the differences in the number of events at these experiments. Two factors are important. First, SND@LHC is slightly off-axis, whereas FASER(ν) is placed directly on-axis. Second, SND@LHC covers $\simeq 5$ (1.25) times larger solid angle than FASER(ν).

The different placements of SND@LHC and FASER/FASERν cause two effects that directly affect the lower and upper bound of the sensitivity (we follow [9] here): *(i)* off-axis placement of SND@LHC causes smaller γ factor and thus worse potential to probe the parameter space of short-lived FIPs; *(ii)* large angular coverage leads to larger fraction of particles from heavy mesons flying to the detector.

First, particles X flying off-axis have smaller energies than those flying on-axis. This is important for probing FIPs that have small decay lengths $l_{\text{decay}} \lesssim l_{\text{min}}$. Indeed, in this regime, the decay probability is $P_{\text{decay}} \approx \exp[-l_{\text{min}}/c\tau_X\gamma_X]$. The sensitivity to such large couplings g determines the upper bound, which is very sensitive to the mean energy of X:

$$\frac{g^2_{\text{upper,SND@LHC}}}{g^2_{\text{upper,FASER}}} \sim \frac{\gamma_X^{\text{SND@LHC}}}{\gamma_X^{\text{FASER}}} \tag{2.3.17}$$

The upper bound is important for particles that may be probed by the FASER and SND@LHC experiments only in the regime of small decay lengths, including dark photons and axion-like particles (see Fig. 2.26). The ratio of the mean γ-factors of dark photons A', flying in the detector, is $\gamma_{A'}^{\text{SND@LHC}}/\gamma_{A'}^{\text{FASER}} \approx 1/3$. The resulting estimate (2.3.17) agrees with the sensitivities in Fig. 2.26.

Second, the off-axis placement may affect the geometric acceptance. Light portal particles X are often produced in decays of mesons. The angular distribution of particles X is similar to the distribution of parent mesons at angular scales larger than $\Delta\theta \simeq 2p_{X,\text{rest}}/\langle E_{\text{meson}}\rangle$, where $p_{X,\text{rest}}$ is the momentum of the daughter particle at rest frame of the decaying meson, being $\simeq m_{\text{meson}}$ if masses of all decays products are $\ll m_{\text{meson}}$. If $\Delta\theta > \theta_{\text{SND@LHC}} \simeq \mathcal{O}(1\,\text{mrad})$, the ratio of geometric acceptances ϵ_{geom} for the SND@LHC and FASER experiments scales with their solid angle coverage. Using characteristic energies $\langle E_{\text{meson}}\rangle \simeq 1\,\text{TeV}$ for mesons produced in the far-forward region, we find that this scaling is indeed the case of light particles produced in decays of D, B-mesons.

However, if the daughter particle is heavy $m_X \simeq m_{\text{meson}}$, or if the decaying meson is light (such as π, η, K), the geometric acceptance depends on the shape of the meson distribution. Experimental measurements of the meson production cross section in the region $|\eta| < 5$ [156–159] provide the following scaling:

$$\frac{\text{d}\sigma}{\text{d}p_{\text{T}}} \sim \frac{p_{\text{T}}}{(p_{\text{T}}^2 + \Lambda^2_{\text{meson}})^2}, \tag{2.3.18}$$

independently of the pseudorapidity. The values of Λ_{meson} are of order of $\Lambda_{\text{QCD}} \approx 250\,\text{MeV}$ for light mesons π, η, K, and $m_{D/B}$ for D/B-mesons. Numeric approaches (see, for instance, [43, 152, 160, 161]) predict the same behavior of $\mathrm{d}\sigma/\mathrm{d}p_{\text{T}}$ almost independently of pseudorapidity, including the far-forward region.[15] This means that the meson distribution $\mathrm{d}f/\mathrm{d}\Omega$ is flat for angles $\theta \lesssim \theta_{\text{flat}}$, where

$$\theta_{\text{flat}} \sim \frac{\langle p_{\text{T}} \rangle}{\langle E_{\text{meson}} \rangle} \sim \frac{\Lambda_{\text{meson}}}{1\,\text{GeV}}\text{mrad} \simeq \begin{cases} \mathcal{O}(1\,\text{mrad}), & B, D \\ \mathcal{O}(0.1\,\text{mrad}), & \pi, \eta \end{cases} \tag{2.3.19}$$

Using the spectra of mesons (see Appendix 2.B), we find

$$\frac{\epsilon_{\text{geom}}^{\text{SND@LHC}}}{\epsilon_{\text{geom}}^{\text{FASER}}} \simeq \begin{cases} 1, & \pi, \eta, \\ \frac{\Omega_{\text{SND@LHC}}}{\Omega_{\text{FASER}}} \approx 5, & D, B, \tau \end{cases} \tag{2.3.20}$$

2.3.4.2 Decays

Based on these findings, we can make a simple comparison of minimal couplings that may be probed by the FASER and SND@LHC experiments. Further, we will assume the most optimistic estimate for SND@LHC, according to which decays of FIPs may be clearly distinguished from backgrounds, and therefore, only 3 events are required at 95% CL.

In the regime $l_{\text{decay}} \gg l_{\text{max}}$, the number of decay events of particles that originate from mesons is

$$N_{\text{decay}} \propto \epsilon_{\text{geom}} \cdot l_{\text{det}} \cdot g^4 \cdot \text{Br}_{\text{vis}} \tag{2.3.21}$$

From this relation, we obtain

$$\frac{g^2_{\text{lower,SND@LHC}}}{g^2_{\text{lower,FASER}}} \sim \sqrt{\frac{\gamma_X^{\text{SND@LHC}}}{\gamma_X^{\text{FASER}}}} \cdot \sqrt{\frac{\text{Br}_{\text{vis}}^{\text{FASER}}}{\text{Br}_{\text{vis}}^{\text{SND@LHC}}}} \cdot \begin{cases} 1.7, & \text{particles from } \pi, \eta \\ 0.8, & \text{particles from } D,\, B \end{cases} \tag{2.3.22}$$

where we used $l_{\text{det}}^{\text{SND@LHC}} = 0.5\,\text{m}$.

Therefore, assuming that SND@LHC may reconstruct decay events in background-free regime, the reaches of SND@LHC and FASER to probe decays of particles are comparable.

Comparing the lower bounds of the numerical sensitivities of SND@LHC and FASER for dark photons and dark scalars in Fig. 2.26, we find that they agree with the estimates (2.3.22). However, for HNLs there is a disagreement as large as a factor of 3. A

[15] Some of these approaches suffer from theoretical uncertainties in far-forward direction [162]: small p_{T} and large pseudorapidity require using parton distribution functions in the domain of small Bjorken scaling variable x, which are poorly constrained. One of the goal SND@LHC and FASER may serve for checking the distributions (and in particular the property (2.3.19)) via studying the events with neutrinos produced in the meson decays.

reason for this may be different distributions of D mesons used in our analysis and in [67] (see also Appendix 2.B).

Let us now comment on the lower bounds ratio with the upgrade. With the help of the formulas (2.3.20), (2.3.22) and table 2.4, we conclude that the FASER2 experiment has much better potential:

$$\frac{g^2_{\text{lower,SND@LHC upgr}}}{g^2_{\text{lower,FASER2}}} \simeq 20 \cdot \sqrt{\frac{\gamma_X^{\text{SND@LHC}}}{\gamma_X^{\text{FASER}}}} \cdot \sqrt{\frac{\text{Br}_{\text{vis}}^{\text{FASER}}}{\text{Br}_{\text{vis}}^{\text{SND@LHC}}}} \tag{2.3.23}$$

A reason for this is mainly significantly larger angular coverage in the case of the FASER2.

2.3.4.3 Scattering

Consider now the scattering signature. For the leptophobic portal, from Eq. (2.3.9), the ratio of minimal probed couplings is (for $\alpha_\chi = \alpha_B$)

$$\frac{\alpha_{B,\text{FASER}\nu}}{\alpha_{B,\text{SND@LHC}}} \sim \left(\frac{\epsilon_{\text{geom}}^{\text{FASER}\nu}}{\epsilon_{\text{geom}}^{\text{SND@LHC}}} \frac{l_{\text{det}}^{\text{FASER}\nu}}{l_{\text{det}}^{\text{SND@LHC}}} \frac{\sigma(E_{\text{threshold}}^{\text{FASER}\nu})}{\sigma(E_{\text{threshold}}^{\text{SND@LHC}})} \sqrt{\frac{N_{\nu\,\text{bg}}^{\text{FASER}\nu}}{N_{\nu\,\text{bg}}^{\text{SND@LHC}}}} \right)^{\frac{1}{3}} \tag{2.3.24}$$

where $N_{\nu\,\text{bg}}$ is the number of neutrino background events (different for the elastic and inelastic signatures), and we assume that the detection efficiency is equal to one. The effective cross section $\sigma(E_{\text{threshold}})$ depends on momentum threshold for charged particles to be visible. The dependence of σ on $E_{\text{threshold}}$ is very important for the elastic signature, as most of the elastic scattering events are characterized by low momenta. For the inelastic signature, it is less relevant. For SND@LHC, the requirements are $E_{\text{threshold}} = 170\,\text{MeV}$ for protons and $100\,\text{MeV}$ for other charged particles. For FASERν (Ref. [163]), we have not found the information about $E_{\text{threshold}}$. Instead, we assume $E_{\text{threshold}} = 1\,\text{GeV}$ that was used for the pilot run in 2018 (see also [164], where $300\,\text{MeV}$ cut is considered for FASERν2).

For small masses $m_V \lesssim 0.5$, the mediator is mainly produced from π, η decays, as shown in Fig. 2.24. In this case, we have $\epsilon_{\text{geom}}^{\text{SND@LHC}}/\epsilon_{\text{geom}}^{\text{FASER}\nu} \approx 0.3$. A similar increase occurs for $N_{\nu\,\text{bg}}$, since neutrinos are abundantly produced in decays of pions, and therefore, we can use the same scaling for the total neutrino events, $N_{\nu\,\text{bg}} \propto \epsilon_{\text{geom}} l_{\text{det}}$. The estimate then reads:

$$\frac{\alpha_{B,\text{SND@LHC}}}{\alpha_{B,\text{FASER}\nu}} \sim 9.6^{1/6} \left(\frac{\sigma(E_{\text{cut}}^{\text{SND@LHC}})}{\sigma(E_{\text{cut}}^{\text{FASER}\nu})} \right)^{-\frac{1}{3}} \simeq 1.5 \begin{cases} 1, & \text{inelastic} \\ 0.2 - 0.9, & \text{elastic} \end{cases} \tag{2.3.25}$$

We conclude that sensitivities of SND@LHC and FASER to scatterings are also comparable.

2.3.5 Conclusions

In this section, we have demonstrated the potential of the SND@LHC experiment to probe feebly interacting particles. We have considered scattering signatures and some decay signatures as well.

Light dark matter particles coupled via mediators may be searched by looking at the scattering signature, see Sec. 2.3.2.2. These events need to be distinguished from neutrino scatterings. Because of large mass of Z and W bosons that mediate the neutrino interactions, the neutrino scattering occurs inelastically most of the times. This may be not the case for light dark matter particles interacting via a light $\mathcal{O}(1\,\mathrm{GeV})$ mediator, for which the yields of elastic and inelastic scattering events are comparable (see Fig. 2.20). Therefore, looking for an excess in the yield of elastic scattering events is suitable for probing such FIPs. For heavier mediators, FIPs scattering still may be searched via an increase in the ratio of scattering events with a lepton and those without a lepton. On one hand, this ratio may be accurately measured at SND@LHC. On the other hand, it is clearly predicted by the SM. We have illustrated the power of these two signatures by estimating the sensitivity to the scattering of light dark sector particles via the leptophobic portal, see Fig. 2.25.

SND@LHC detector may also search for decays of mediators, see Sec. 2.3.4.2. Because of good spatial resolution of the emulsion in SND@LHC, decays into two charged particles may be distinguished from the neutrino scattering events. Such decays are main decay channel in the case of heavy neutral leptons, dark scalars that mix with Higgs boson, and dark photons. It is possible to probe their parameter space at its upgraded version as described in [32], see Fig. 2.26. However, further studies of possible backgrounds are required to clarify these results.

We have also compared the potential of SND@LHC and FASER/FASERν facilities to probe new physics, see Sec. 2.3.4. Placed at the same distance but at the opposite sides of the ATLAS experiment interaction point, they are very similar. There are a few factors, however, leading to differences in the sensitivity of these facilities to new physics. First, FASER is on-axis, while SND@LHC is slightly off-axis. The off-axis placement decreases the mean momentum of particles produced in the direction of SND@LHC, which somewhat worsens its potential to probe short-lived particles with the decay lengths of the order of the distance to the detector. Second, SND@LHC covers $\simeq 5$ times larger solid angle than FASER. Because of this, depending on the FIPs production channel, a fraction of FIPs flying in the direction of FASER is smaller than that for SND@LHC. For scatterings, FASERν has higher event rate due to larger detector length and on-axis position, resulting in better sensitivity. This can be applied for the inelastic signature; for other possible signatures, e.g. elastic scattering off protons, SND@LHC might have better sensitivity thanks to its higher momentum resolution.

Appendix

2.A CHARM sensitivity based on number of decay events estimate

The number of decay events for the pure α mixing at CHARM is given by the formula

$$N_{\text{events}}^{(\alpha)} = \sum_{X=D,\tau} N_X \cdot \text{Br}(X \to N_\alpha) \times \\ \times \int dE d\theta dz \cdot f_{N_\alpha}^X(E,\theta) \frac{e^{-l(z)/c\tau_N \gamma}}{c\tau_N^{(\alpha)}\gamma} \frac{\Delta\phi(\theta,z)}{2\pi} \cdot \epsilon_{\text{decay}}(\theta,z,E) \cdot \text{Br}(N_\alpha \to l\bar{l}') \cdot \epsilon_{\text{det},ll'} \tag{2.A.1}$$

Here,

$$N_{D_i} = N_{\text{PoT}} \times \chi_{c\bar{c}} \times f_{c\to D_i}, \quad N_\tau = N_{D_s} \times \text{Br}(D_s \to \tau\bar{\nu}_\tau) \tag{2.A.2}$$

are the total numbers of D mesons ($D_i = D_s, D^+, D^0$) and τ leptons, with $N_{\text{PoT}} = 2.4 \cdot 10^{18}$ being the total number of proton-target collisions at CHARM and $\chi_{c\bar{c}} \approx 4 \cdot 10^{-3}$ the production fraction of the $c\bar{c}$ at SPS energies for a thick target [44]. $\text{Br}_{D_s\to\tau} \approx 5.43\%$ [52] and $f_{c\to D_i}$ are given from [80]. $f_{N_\alpha}^X$ is the distribution of HNLs produced in decays of X particles in polar angle and energy. $z \in (480, 515)\,\text{m}$ is the longitudinal distance, $\theta \in (3.5/515, 6.5/515)$ is the polar angle coverage of the end of the CHARM's decay volume, while $\Delta\phi(\theta)/2\pi$ is the azimuthal acceptance for HNLs decaying inside the decay volume. ϵ_{decay} is the decay acceptance – a fraction of decay products of HNLs that both point to the detector. Finally, $\epsilon_{\text{det},ll'}$ are reconstruction efficiencies for leptonic decays: $\epsilon_{ee} \approx 60\%$, $\epsilon_{\mu\mu} \approx 75\%$, and $\epsilon_{e\mu} \approx 65\%$, which we use from [62].

Computing of $f_{N_\alpha}(E,\theta)$ requires knowing the distribution of D mesons and τ leptons $f_\tau(E,\theta)$ produced at the CHARM target. We approximate f_τ by the distribution of D_s mesons, while for the distribution of D mesons we use FairShip simulations [44] for collisions of the SPS proton beam with a thick Tungsten target.[16] The distribution of HNLs $f_{N_\alpha}^X(E,\theta)$ has been obtained from $f_X(E,\theta)$ semi-analytically using the method from [120].

We have estimated ϵ_{decay} by using a toy simulation for decays of HNLs inside the decay volume into three massless particles, and requiring the momenta of the two charged leptons to point towards the end of the decay volume. The acceptances are shown in Fig.2.27.

[16] Although at CHARM the target material is different, we believe that it is still a reasonable approximation.

In order to obtain the excluded domain, we assume the absence of background and require $N_{\text{events}} > 2.3$, which corresponds to the 90% C.L.

The comparison of this estimate with the rescale from Sec. 2.1.4 and [63] is shown in Fig. 2.27. We find that the estimates are in very good agreement. We also show our estimate of the CHARM bounds on the e mixing, which differs from the bounds obtained from [62] by including the production from D_s mesons, which dominates masses $m_N \gtrsim 700\,\text{MeV}$ (see also Fig. 2.5). The resulting sensitivity at the lower bound improves by up to a factor $3-4$ for this mass region.

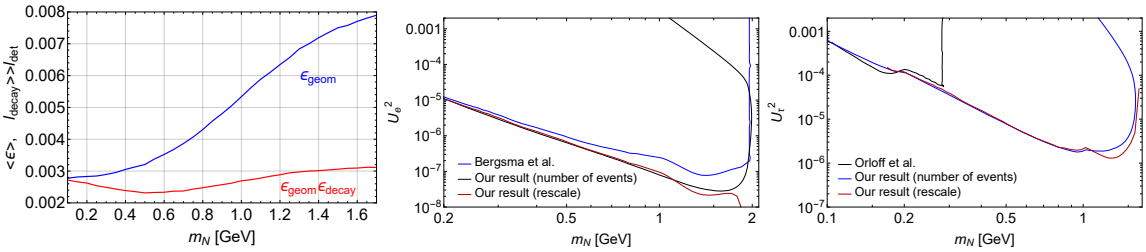

Figure 2.27: The left panel: fraction of HNLs that point towards the detector (blue line) and fraction of HNLs whose decay products point towards the detector (red line). The middle and right panels: comparison of our estimates of the constraint from the CHARM experiment on the pure e (the middle panel) and τ mixing (the right panel), with bounds reported in [62] and [63]. We show two estimates: the red line corresponds to the rescale of the bound on the e mixing from [62] (see Sec. 2.1.4 for details), while the blue line is our independent estimate based on Eq. (2.A.1).

2.B Decay events at SND@LHC

We estimate the number of decays using the following formula:

$$N_{\text{events}} = \sum_{\text{i}} N^i_{\text{prod}} \cdot \epsilon^i_{\text{geom}} \cdot P^i_{\text{decay}} \cdot \text{Br}_{\text{vis}}, \tag{2.B.1}$$

Here, N^i_{prod} is the total number of particles X produced via a channel i, ϵ_{geom} is the geometric acceptance, and P_{decay} is the decay probability averaged over energies of X,

$$P^i_{\text{decay}} = \int (e^{-l_{\text{min}}/c\tau_X\gamma_X} - e^{-l_{\text{max}}/c\tau_X\gamma_X}) f^i_{E_X} dE_X, \tag{2.B.2}$$

Here, $l_{\text{min}} = 480\,\text{m}$ is the distance to the SND@LHC detector, $l_{\text{max}} - l_{\text{min}} = l^{\text{decay}}_{\text{det}}$, f_{E_X} is the energy distribution of particles X that fly in the decay volume. Finally, Br_{vis} is the branching ratio of visible decays.

HNLs that mix with ν_μ are produced in decays of $D_s/D^{+/0}$ mesons. HNLs that mix with ν_τ are produced mainly in decays of τ-leptons, which, in their turn, originate from decays $D_s \to \tau\bar{\nu}_\tau$ [39]. We have obtained the distribution of D mesons using SIBYLL

2.3c [161, 165] as a part of the CRMC package [153]. As a cross-check, for the charm production we have compared the predictions of SIBYLL with results of the FONLL program [40, 42, 43]. We have found that the results agree well for angles $\theta > 0.8$ mrad.[17] Having the D distribution, we have obtained the distribution of τ-leptons and, subsequently, HNLs angles and momenta using the approach described in [120]. For simplicity, we approximate the angle-momentum distribution of HNLs by that of particles produced in a two-body decay $\tau \to \pi N$ (for the mixing with ν_τ) and $D_s \to \mu N$ (for the mixing with ν_μ).

Dark photons V in sub-GeV mass range are produced in decays $h = \gamma V$ of π- and η-mesons, and by proton bremsstrahlung [134]. We use the angle-energy distributions of the mesons generated by EPOS-LHC [152] as a part of the CRMC package [153], and follow [134] for the bremsstrahlung.

Dark scalars S are produced in decays $B \to X_s S$ of B-mesons, where X_s is a hadron including an s-quark, and by the proton bremsstrahlung [38]. We use FONLL in order to obtain the angle-energy distribution of B-mesons, and follow [38] for the proton bremsstrahlung.

Using the obtained distribution, we have reproduced the sensitivity of FASER to scalars and dark photons from [67]. However, we have not reproduced the sensitivity to HNLs, see Fig. 2.28. A reason may be in different distributions of D_s-mesons used in the estimates.

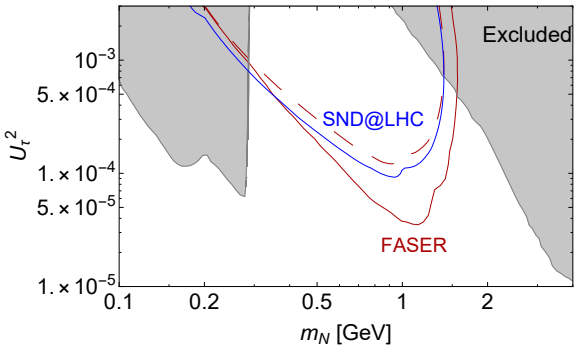

Figure 2.28: The sensitivity of FASER to HNLs that mix with ν_τ. The solid line corresponds to the contour given in Ref. [67], while the dashed line – to our estimate. For the comparison, we also show the sensitivity of SND@LHC (in blue).

2.C Leptophobic mediator: production, decays and scatterings

2.C.1 Production and decay

In order to describe interactions of V with hadrons, we follow [166] (see also [150]), in which vector mesons m play the role of gauge bosons of a "hidden" local $SU_f(3)$ symmetry in the space of pseudoscalar mesons nonet. The EM field is included as a background field

[17] For smaller angles, FONLL (both the online form and installed program) predicts zero or negative cross sections, which indicates some internal problem.

that is associated with the appropriate generator $Q = \text{diag}\left(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}\right)$, and mix with the vector mesons. The coupling of the vector mesons to the pseudoscalar mesons is fixed by the anomalous decay $\pi^0 \to \gamma\gamma$. This model is very successful in describing the EM scattering data $e^+e^- \to$ hadrons and decay widths of vector mesons. We assume that it may be also used for describing the phenomenology of the leptophobic boson.

For the lephophobic mediator, the generator is $T_V = \frac{\not{V}}{3}$, and its mixing coupling is given by

$$f_{Vm} = -2g_B g_m \text{Tr}[T_V T_m], \tag{2.C.1}$$

where T_m is a generator associated with the given meson, and $g_m/m_m^2 = 1/\sqrt{12\pi}$, as fixed by the anomaly. The mixing occurs only with isosinglet ω- and ϕ-mesons, for which

$$T_\omega = \frac{1}{2}\text{diag}(1,1,0), \quad T_\phi = \frac{1}{\sqrt{2}}\text{diag}(0,0,1) \tag{2.C.2}$$

The decay width of V may be extracted from the experimental data on the EM ratios $\sigma(e^+e^- \to \text{hadrons})/\sigma(e^+e^- \to \mu^+\mu^-)$, where the hadronic final states correspond to ϕ-like and ω-like decays. This has been made in [144], in which the data have been used for describing the decay widths up to masses $m_V \simeq 1.7\,\text{GeV}$, while for larger masses perturbative calculations were used. We use the results of this paper.

The resonant enhancement is also important when considering the production of the mediator by the proton bremsstrahlung by affecting the form-factor F_{ppV} in the ppV vertex. The baryonic form factor F_{ppV} may be related to the proton and neutron EM dipole form-factors $F_{p/n}$, which are, in its turn, related to the isoscalar form factor $F_\omega \equiv \frac{F_p+F_n}{2}$, which in the extended vector meson dominance model coincides with the ω contribution [154]:[18]

$$\langle p|J_B|p\rangle = \langle p|J_{\text{EM}}|p\rangle + \langle n|J_{\text{EM}}|p\rangle \longrightarrow F_{ppV} = 2F_\omega \tag{2.C.3}$$

Unfortunately, the experimental data on $e^+e^- \to p^+p^-$, which may be used for extracting the EM form-factors in the time-like region, is limited by the physical threshold $q^2 > 4m_p^2$. Following [154] (see also [151]), for extrapolating in the domain of lower invariant masses we use

$$F_{ppV}(q^2) = \sum_\omega f_\omega \frac{m_\omega^2}{m_\omega^2 - q^2 - i\Gamma_\omega m_\omega}, \tag{2.C.4}$$

where the sum goes over $\omega(782), \omega(1420), \omega(1680)$, $f_\omega = 2f_{NN\omega}/g_\omega$, with $f_{NN\omega}$ being the meson's coupling to the nucleon, while g_ω is the meson's coupling to photon. We use the couplings $f_{NN\omega(782)} = 17.2$ and $g_{\omega(782)} = 17.1$ [154]. The couplings to the other two resonances are unknown. However, the remaining two coefficients $f_{\omega(1420)} = -2.16 + 0.77i$ and $f_{\omega(1680)} = 1.14 - 0.57i$ in Eq. (2.C.4) may be fixed by two requirements: $F_{ppV}(0) = 1$, and $F_{ppV}(-q^2) \sim 1/q^4$. The first requirement comes from the fact that the form-factor F_{ppV}

[18] We assume no contribution of the ϕ-meson to the form-factor, since the corresponding coupling $f_{\phi NN}$ is expected to be suppressed [167, 168] (i.e., neglecting the s-quark contribution in the proton PDF).

is reduced to the baryon charge at low momenta transfer. The second requirement comes from the behavior of the proton's dipole form-factor in the space-like region predicted by the quark counting rules [169].

The behavior of the branching ratio into a $\chi\chi$ pair and the form-factor is shown in Fig. 2.29. Note that for the choice $\alpha_\chi = \alpha_B$, commonly considered in the literature, the enhancement of F_{ppV} near $m_V = m_{\omega(770)}$ and suppression of $\text{Br}(V \to \chi\chi)$ due to the ω resonances cancel each other.

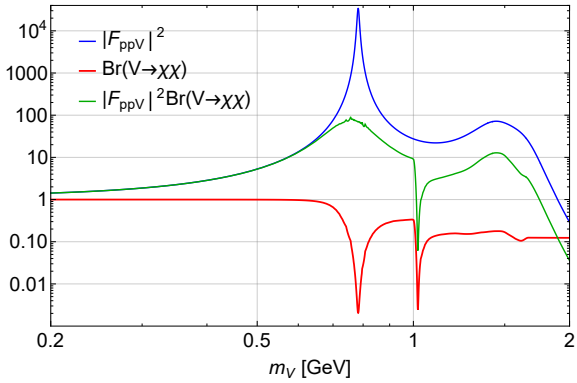

Figure 2.29: The behavior of the ppV form-factor (2.C.4) and the branching ratio for the process $V \to \chi\chi$. The coupling $\alpha_\chi = \alpha_B$ is assumed, and $m_\chi = m_V/3$.

2.C.2 Elastic and inelastic scattering cross sections

2.C.2.1 Elastic scattering

The cross section of the elastic scattering is

$$\sigma_{\text{elastic}} = \int dE_\chi f_{E_\chi} \int_{E_{N,\min}}^{E_{N,\max}(E_\chi)} dE_N \frac{d\sigma_{\chi N \to \chi N}}{dE_N}, \tag{2.C.5}$$

where $E_{N,\min}$ is the minimal recoil energy that may be detected, the maximal recoil energy of the nucleon is

$$E_{N,\max} = \frac{m_N(2E_\chi^2 + 2E_\chi m_N + m_N^2 - m_\chi^2)}{2E_\chi m_N + m_N^2 + m_\chi^2}, \tag{2.C.6}$$

$Q^2 = 2m_N(E_N - m_N)$ is the modulus of the squared momentum transfer, $Q^2 = -(p_\chi - p'_\chi)^2$. Finally, the differential cross section is

$$\frac{d\sigma_{\chi N \to \chi N}}{dE_N} = 4\pi\alpha_D^2 F_N(Q^2) \frac{m_N(2E_\chi^2 + 2E_\chi m_N + m_\chi^2) - E_N(2E_\chi m_N + m_\chi^2)}{(E_\chi^2 - m_\chi^2)(2E_N m_N - 2m_N^2 + m_V^2)^2}, \tag{2.C.7}$$

where $F_N(Q^2)$ is the elastic form-factor, which we assume to be $F_N(Q^2) = \frac{1}{(1+\frac{Q^2}{0.71\,\text{GeV}^2})}$.

2.C.2.2 Inelastic scattering

In the case of the inelastic scattering, we follow [148], which uses the parton model. Let us introduce the variables $E_V = E_\chi - E_\chi^{'}$, Q^2. The differential cross section is

$$\frac{d^2\sigma}{dE_V dQ^2} = \frac{\pi\alpha_D^2}{9m_N}\frac{1}{E_\chi^2 - m_\chi^2}\frac{1}{(m_V^2+Q^2)^2}(2p-q)^\mu(2p-q)^\nu W_{\mu\nu}\sum_q x f_q(x,Q^2), \quad (2.C.8)$$

where $f_q(x,Q^2)$ is the parton distribution function ($q = u/\bar{u}/d/\bar{d}/s/\bar{s}$), $x = \frac{Q^2}{2m_N E_V}$, $W_{\mu\nu}$ is the hadronic tensor,

$$W_{\mu\nu} = -g_{\mu\nu} + \frac{q_\mu q_\nu}{q^2} + \frac{2x}{p_N\cdot q + 2xm_N^2}\left(p_{N\mu} - \frac{p_N\cdot q}{q^2}q_\mu\right)\left(p_{N\nu} - \frac{p_N\cdot q}{q^2}q_\nu\right) \quad (2.C.9)$$

Because of the property $q^\mu W_{\mu\nu} = q^\nu W_{\mu\nu} = 0$, we have

$$(2p-q)^\mu(2p-q)^\nu W_{\mu\nu} = \frac{4E_\chi^2Q^2 - 4E_V E_\chi Q^2 - Q^4}{E_V^2+Q^2} - 4m_\chi^2 \quad (2.C.10)$$

The kinematic limits are

$$Q^2 < 2m_N E_V, \quad 2\mu^2 < Q^2 < 4(E_\chi(E_\chi - E_V) - m_\chi^2) - 2\mu^2, \quad (2.C.11)$$

$$E_{\text{min}} < E_V < \frac{2m_N(E_\chi^2 - m_\chi^2)}{2E_\chi m_N + m_N^2 + m_\chi^2}, \quad (2.C.12)$$

where E_{min} is the minimal recoil, and the function μ is

$$\mu^2 = \frac{m_\chi^2 E_V^2}{E_\chi(E_\chi - E_V) - m_\chi^2 - \sqrt{(E_\chi(E_\chi-E_V)-m_\chi^2)^2 - m_\chi^2 E_V^2}} \quad (2.C.13)$$

To get $f_q(x,Q^2)$, we use LHAPDF with CT10nlo PDF sets. We assume that they are zero if $Q < 1\,\text{GeV}$.

Chapter 3

Probes from cosmology[1]

In this chapter, we consider two earliest messengers from cosmology: BBN and CMB, and study the impact of short-lived FIPs on corresponding observables. In the case of BBN, we concentrate on short-lived particles that decay hadronically, and derive analytically the bound on lifetimes that comes from the impact of mesons on the $p \leftrightarrow n$ conversion in the primordial plasma, see Sec. 3.1. In the case of CMB, we study the impact of short-lived FIPs on the effective number of degrees of freedom, N_{eff}, and in particular show that even if decaying mostly into neutrinos they may decrease N_{eff}, see Sec. 3.2. Finally, we apply the findings to the case of a particular model – HNLs, for which we first study their cosmological population (Sec. 3.3.4), and then derive constraints from BBN (Sec. 3.3.3.1) and CMB (Sec. 3.3.4).

3.1 BBN and hadronically decaying particles

In this section, we discuss bounds from BBN on hadronically decaying particles. We will first discuss the current measurements of the primordial abundance of helium, then derive the bounds from BBN on hadronically decaying particles, and then comment until which lifetimes the bound extends.

3.1.1 Measurements of ^{4}He abundance

Over the last 6 years five works determined the primordial ^{4}He abundance from stellar measurements [171–175]. The formal statistical errors of Y_p are at the level of $1-3\%$, however, the scatter between different groups is larger, see Fig. 1.

All these works determine astrophysical Helium abundance through measurements of recombination emission lines of ^{4}He and H in the metal-poor extragalactic ionized regions, then *linearly* extrapolating the measurements to zero metallicity. Given the high precision of the results, it is important to take into account various smaller effects: including

[1] Results of this chapter are presented in papers [37, 170]. The main contribution of Maksym Ovchynnikov is analytic and numeric estimates of the BBN and CMB bounds.

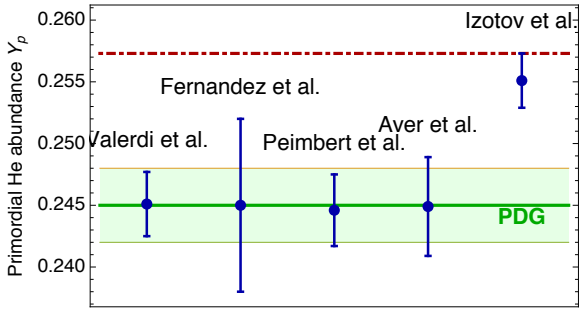

Figure 1: Measurements of Y_p of recent works [171–175]. The green shaded region is the PDG recommended value [52] (with $\pm 1\sigma$). The gray dashed line denotes the SBBN prediction $\bar{Y}_p = 0.247$ from [45]. The red dashed-dotted line is the maximal admissible value $Y_{p,\max}$ on which we base our analysis.

[4]He fluorescent emission, different ion temperatures, spatial temperature fluctuations, and others [176, 177]. Additionally, while it is true that the metallicity and Helium abundance are positively correlated, the linear extrapolation to zero-metallicity may be prone to systematic uncertainties.

The value of Y_p predicted within the framework of SBBN is $\bar{Y}_p = 0.24709 \pm 0.00019$ (see, e.g., [45]). The effect of mesons leads to an increase of Y_p as compared to the SBBN value. Therefore, in order to get a conservative upper bound we assume that the maximally allowed Y_p is given by the 1σ deviation from the maximal value predicted by [171–175], which is $Y_{p,\max} = 0.2573$. Note that this upper value significantly deviates from the PDG-recommended value [52] $Y_{p,\max} = 0.248$ at 1σ. This translates to the bound

$$\frac{\Delta Y_p}{\bar{Y}_p} < 4.35\% \tag{3.1.1}$$

3.1.2 Bound on hadronically decaying particles

Sufficiently heavy FIPs can decay into mesons $h = \pi, K$, *etc*. Charged pions drive the $p \leftrightarrow n$ conversion via processes [178]

$$\pi^- + p \to n + \pi^0/\gamma, \quad \pi^+ + n \to p + \pi^0 \tag{3.1.2}$$

The cross section of these reactions is much larger than the cross section of weak interactions driven conversion processes:

$$\frac{\langle \sigma^{\pi}_{p\leftrightarrow n} v \rangle}{\langle \sigma^{\text{Weak}}_{p\leftrightarrow n} v \rangle} \simeq \frac{1}{G_F^2 m_p^2 T^2} \sim 10^{16} \left(\frac{1\,\text{MeV}}{T}\right)^2, \tag{3.1.3}$$

Large cross section, absence of threshold and isotopic symmetry of these processes mean that if pions are present in the plasma in the amounts at least comparable with that of baryons, they drive the number densities of protons and neutrons to equal values, $n_n/n_p \simeq$

$\langle\sigma^{\pi}_{p\to n}v\rangle/\langle\sigma^{\pi}_{n\to p}v\rangle \simeq 1$.[2] The effect of kaons is qualitatively similar, but leads to a slightly different neutron-to-proton ratio (Appendix 3.B.1).

The impact of this effect on primordial ^{4}He abundance depends on how long mesons remain present in the plasma in significant amounts. Once mesons are created, they can *(i)* scatter and lose energy; *(ii)* decay; *(iii)* participate in $p \leftrightarrow n$ conversion. The corresponding rates are very different: at MeV temperatures and below, $\Gamma^h_{\rm scat} \gg \Gamma^h_{\rm decay} \gg \Gamma^h_{p\leftrightarrow n}$ (see [179]).

The instantaneous number density of mesons is an interplay between their production (via decays of FIPs) and their decays:

$$n_h^{\rm inst} = n_{\rm FIP}(T)\cdot {\rm Br}_{N\to h}\frac{\Gamma_{\rm FIP,dec}}{\Gamma_{h,\rm dec}} = n_N(T)\cdot {\rm Br}_{N\to h}\frac{\tau_h}{\tau_N}. \tag{3.1.4}$$

Here, ${\rm Br}_{{\rm FIP}\to h}$ is the branching of FIPs into mesons. $n_{\rm FIP}(T)$ is the number density of FIPs:

$$n_{\rm FIP}(T) = \left(\frac{a_{\rm dec}}{a(T)}\right)^3\cdot n^{\rm dec}_{\rm FIP}\cdot e^{-\frac{t(T)}{\tau_N}}, \tag{3.1.5}$$

where $n^{\rm dec}_{\rm FIP}$ is the FIP's number density at decoupling (i.e. when their interaction with plasma has been completely stopped), and $a(T)$ ($a_{\rm dec}$) is the scale factor at temperature T (correspondingly, at FIP decoupling).

The number of $p \leftrightarrow n$ reactions per nucleon occurring after time $t \gg \tau_{\rm FIP}$ (or below some corresponding temperature $T(t)$) is thus

$$N^h_{p\leftrightarrow n}(T) = \sum_h \int\limits_{t(T)}^{\infty} dt\; n_h^{\rm inst}(T)\cdot\langle\sigma^h_{p\leftrightarrow n}v\rangle \approx \left(\frac{a_{\rm dec}}{a}\right)^3\frac{n^{\rm dec}_{\rm FIP}}{n_B}\cdot e^{-\frac{t(T)}{\tau_{\rm FIP}}}\cdot {\rm Br}_{{\rm FIP}\to h}\cdot P_{\rm conv}, \tag{3.1.6}$$

where n_B is the baryon number density, the sum goes over meson species and $P_{\rm conv}$ is the probability for a single meson to interact with nucleons before decaying:

$$P_{\rm conv} \simeq \frac{n_B\cdot\langle\sigma^h_{p\leftrightarrow n}v\rangle}{\Gamma^h_{\rm decay}}. \tag{3.1.7}$$

At $\mathcal{O}(1\text{ MeV})$ temperatures, $P_{\rm conv}\sim 10^{-2}-10^{-1}$, see Appendix 3.B.

[2] For each of the processes (3.1.2), there are no inverse reactions. Indeed, π^0 decays very fast, whereas γs quickly lose their energy. Therefore, the conversion (3.1.2) is highly non-equilibrium, and the corresponding value of n_n/n_p is not given by the usual Boltzmann exponent.

The meson driven conversion keeps the value $n_n/n_p \simeq 1$ roughly until a temperature T_0 when the number of reactions per nucleon drops below one,

$$N^h_{p\leftrightarrow n}(T_0) \simeq 1, \tag{3.1.8}$$

and weak SBBN reactions start to relax the n/p ratio down to its SBBN value, see Fig. 2 (left panel).

However, if T_0 is close enough to the freeze-out of weak $p \leftrightarrow n$ processes, occurring roughly at $T_n \simeq 0.8$ MeV, the relaxation is not complete (Fig. 2, right panel). This leads to a positive correction $\Delta(n_n/n_p)$ as compared to the SBBN case, which translates to an increase of the ^{4}He abundance ΔY_p.

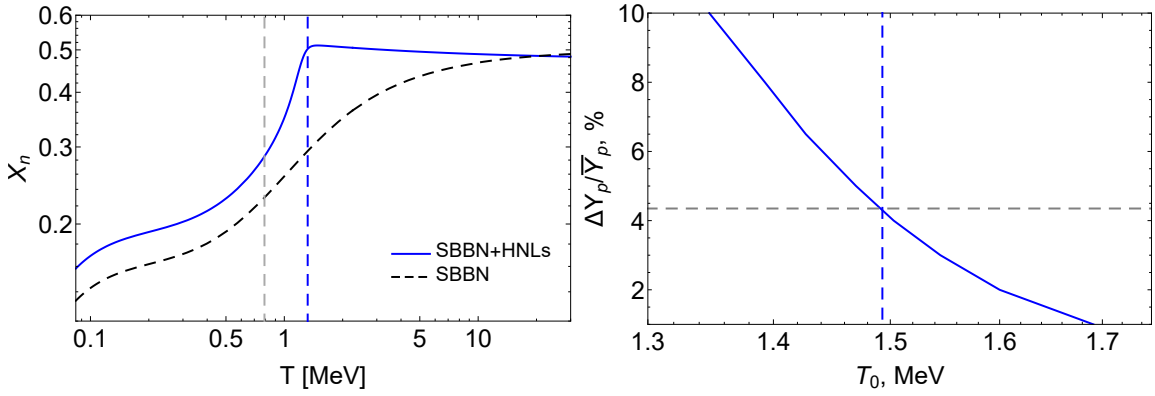

Figure 2: *Left panel*: temperature evolution of the neutron abundance $X_n = n_n/(n_n + n_p)$ in the presence of pions from decays of an HNL with mass $m_N = 400$ MeV and lifetime $\tau_N = 0.03$ s. Below $T \simeq 100$ MeV, pions drive the neutron abundance to $X_n \approx 0.5$. At temperatures $T_0 \simeq 1.3$ MeV (the blue vertical dashed line) pions disappear, and X_n starts relaxing towards its SBBN value but does not reach it. After the neutron decoupling (the gray vertical line) X_n evolves mainly due to the neutron decays. *Right panel*: a relation between the temperature T_0 (defined by Eq. (3.1.8)) and corrections to the ^{4}He abundance, as compared to the SBBN central value $\bar{Y}_p \approx 0.247$. It corresponds to the case of when only charged pions are present in plasma. The gray horizontal line corresponds to maximally allowed correction $\Delta Y_p/\bar{Y}_p = 4.35\%$ that we adopt in this work (see Appendix 3.1.1). The intersection of gray and colored lines defines the temperature $T_0^{\rm min}$.

In this way, the *upper bound* on the ^{4}He abundance $Y_{p,\rm max}$ is translated to the *lower bound* $T_0 \geq T_0^{\rm min}$. Together with the relations (3.1.6)–(3.1.8), this allow us to find an upper limit on the FIP lifetime $\tau_{\rm FIP}$:

$$\tau_{\rm FIP} \lesssim \frac{t(T_0^{\rm min})}{\ln\left[\sum_h \left(\frac{a_{\rm dec}}{a_0}\right)^3 \frac{n^{\rm dec}_{\rm FIP} P_{\rm conv} {\rm Br}_{N\to h}}{n_\gamma(T_0^{\rm min})\eta_B}\right]}. \tag{3.1.9}$$

Here, n_γ is the number density of photons, η_B is the baryon-to-photon ratio, and $t(T)$ is time-temperature relation. $t(T)$ is given by the Standard Model relation: $t(T) = \frac{M_*}{2T^2}$, with $M_* = \frac{M_{\rm Pl}}{1.66\sqrt{g_*}}$ the reduced Planck mass, where $g_*(T) \simeq 10.6$ for $T \simeq 1 - 2$ MeV.[3]

Let us rewrite the logarithmic factor in (3.1.9) as

$$\left(\frac{a_{\rm dec}}{a_0}\right)^3 \frac{n_{\rm FIP,dec}}{n_\gamma(T_0^{\rm min})} = \frac{n_{\rm FIP,dec}}{n_\gamma(T_{\rm dec})} \cdot \zeta, \tag{3.1.10}$$

where $\zeta = \left(\frac{a_{\rm dec} T_{\rm dec}}{a_0 T_0^{\rm min}}\right)^3$ is the "entropy dilution" factor. For example, if FIPs were in thermal equilibrium and decoupled while being ultrarelativistic, $T_{\rm dec} \gg m_{\rm FIP}$, we have $n_{\rm FIP,dec}/n_\gamma(T_{\rm dec}) \approx \mathcal{O}(1)$. The dilution factor ζ is a product of the SBBN value times the value induced by FIPs during their evolution:

$$\zeta = \left(\frac{a_{\rm dec}^{\rm SBBN} T_{\rm dec}}{a_0^{\rm SBBN} T_0^{\rm min}}\right)^3 \times \left(\frac{a_0^{\rm SBBN} T_0^{\rm min}}{a_0 T_0^{\rm min}}\right)^3 \equiv \zeta_{\rm SBBN} \times \zeta_{\rm FIP}, \tag{3.1.11}$$

where we used that $a_{\rm dec} \approx a_{\rm dec}^{\rm SBBN}$; this approximation is valid since at temperatures $T_{\rm dec}$ there are many SM particles, and FIPs only contribute a small fraction to the total energy density of the Universe. In SBBN at temperatures $T \gtrsim 1$ MeV, all particles are at local equilibrium, which defines the dynamics of the scale factor and hence the value of $\zeta_{\rm SBBN}$:

$$\zeta_{\rm SBBN} \approx \frac{g_*(T_0^{\rm min})}{g_*(T_{\rm dec})} \simeq \frac{1}{8}, \tag{3.1.12}$$

where we used that $a_{\rm SBBN}(T) \propto g_*^{-1/3}(T) \cdot T^{-1}$. Decays of heavy FIPs violate the thermal equilibrium at $\mathcal{O}(1\text{ MeV})$, and the scaling (3.1.12) changes. For GeV-scale particles with lifetimes $\tau_{\rm FIP} \sim 0.01\text{ s} - 0.1\text{ s}$ that were in thermal equilibrium, the factor $\zeta_{\rm FIP}$ reaches $\mathcal{O}(0.1)$, see Appendix 3.3.1 using HNLs as an example.

The simple analytic estimate leads to the model-independent bound on FIPs that decay hadronically:

$$\tau_{\rm FIP} \lesssim \frac{0.023 \left(\frac{1.5\text{ MeV}}{T_0^{\rm min}}\right)^2 \text{ s}}{1 + 0.07 \ln\left[\frac{P_{\rm conv}}{0.1} \frac{{\rm Br}_{{\rm FIP}\to h}}{0.4} \frac{2n_{\rm FIP,dec}}{3n_\gamma(T_{\rm dec})} \cdot 24 \left(\frac{a_{\rm dec} T_{\rm dec}}{a_0 T_0^{\rm min}}\right)^3\right]}. \tag{3.1.13}$$

The presence of mesons increases the ^{4}He abundance. Therefore, in order to fix $T_0^{\rm min}(m_{\rm FIP})$, we need to adopt an upper bound on the primordial ^{4}He abundance, $Y_{p,\rm max}$, that is consistent with measurements [52]. The smallest error bars come from measuring Y_p in

[3] This is indeed the case for short-lived FIPs with $\tau_{\rm FIP} \ll 0.1$ s.

low-metallicity interstellar regions and extrapolating its value to zero metallicity (pioneered in [180]). Several groups [171–175] have determined Y_p using this method, albeit with different data and assumptions. The resulting scatter between results is larger than the reported error bars. We treat this difference as an additional systematic uncertainty and adopt the maximal value $Y_{p,\text{max}} = 0.2573$ (see Appendix 3.1.1). The maximally allowed relative deviation is therefore

$$\Delta Y_p / Y_{p,\text{SBBN}} \approx 4.35\%. \tag{3.1.14}$$

To relate ΔY_p and T_0^{min}, we study how the n_n/n_p ratio is relaxed below T_0. The relaxation occurs solely via the SBBN reaction,

$$\frac{dX_n}{dt} = \Gamma^{\text{SBBN}}_{p\to n}(1 - X_n) - \Gamma^{\text{SBBN}}_{n\to p} X_n, \quad X_n = \frac{n_n}{n_n + n_p}, \tag{3.1.15}$$

albeit with the altered initial condition $X_n(T_0) = X_n^h \simeq 1/2$. ($\Gamma^{\text{SBBN}}_{p\leftrightarrow n}(t)$ are SBBN rates, see [45]). Non-SBBN value of $X_n(T_0)$ *is the dominant effect* of short-lived HNLs on Y_p. At temperatures $T \lesssim T_0$, for HNLs with lifetimes $\tau_N \lesssim 0.02$ s, all other quantities that are relevant for BBN dynamics – η_B, time-temperature relation, the nuclear reactions chain – remain the same as in SBBN, which is because most of HNLs are no longer left in the plasma at these temperatures (see also Appendix 3.B.1). As a result, a value of $X_n(T_0)$ is translated into ΔY_p via

$$\frac{\Delta Y_p}{Y_{p,\text{SBBN}}} = \frac{\Delta X_n(T_{\text{BBN}})}{X_{n,\text{SBBN}}(T_{\text{BBN}})}, \tag{3.1.16}$$

where $T_{\text{BBN}} \approx 84$ keV is the temperature of the onset of nuclear reactions in SBBN [45].

To obtain the bound (3.1.13), we considered exclusively meson-driven $p \leftrightarrow n$ processes for $T > T_0^{\text{min}}$ and only weak SBBN processes for $T < T_0^{\text{min}}$. We also solved numerically the equation (3.1.15) for the neutron abundance in the presence of both mesons-driven and SBBN $p \leftrightarrow n$ conversion rates in Appendix 3.B.1 using HNLs as an example model, and obtained results being in perfect agreement with the bound (3.3.24). We have also repeated our analysis for the case of the GeV-mass scalar that mixes with the Higgs and found an excellent agreement with [181, 182].

> We conclude that BBN may constrain hadronically decaying FIPs with lifetimes as small as $\tau_{\text{FIP}} \simeq 0.02$ s.

3.1.3 Limits of applicability of the bound

Eq. (3.1.13) defines the lower bound on FIP lifetimes that may be constrained from the meson-driven ^{4}He overproduction. Our simplified approach is limited by lifetimes for which FIPs or their decay products survive until the onset of nuclear reactions. In this case, the dynamics of nuclear reactions gets changed by

1. meson-driven $p \leftrightarrow n$ conversion and nuclear dissociation processes;
2. change of time-temperature relation by FIPs;
3. change of η_B during nuclear reaction;
4. photo-dissociation processes by high-energetic photons originating from EM decays of FIPs.

Among these effects, the effect which firstly manifests with the increase of the lifetime is the meson-driven nuclear dissociation. Indeed, a change of η_B and $t(T)$ requires a FIP to contribute to the energy density significantly, while the effect of mesons only requires the amount of meson-driven reactions to be comparable with n_B. As for the photo-dissociation, it becomes relevant only from lifetimes of order $\tau_{\text{FIP}} \gtrsim 10^4$ s, which is the time scale at which photons with energies large enough to dissociate deuterium no longer instantly disappear because of the annihilation $\gamma + \gamma_{\text{SM plasma}} \to e^+ + e^-$ (see, e.g., [182]). Let us now estimate the upper bound on the FIP lifetimes at which the simple analysis presented above is valid. The ^{4}He threshold-less dissociation processes with mesons are (see [178])

$$\pi^- + {}^4\text{He} \to T + n, \quad \pi^- + {}^4\text{He} \to D + 2n, \quad \pi^- + {}^4\text{He} \to p + 3n \tag{3.1.17}$$

To estimate the lifetimes at which the processes (3.1.17) can be neglected, we compare the number density of mesons *available for the dissociation* with the number density of ^{4}He nuclei:

$$n^h_{\text{He diss}}(T_{\text{BBN}}) \ll n_{\text{He}}(T_{\text{BBN}}), \tag{3.1.18}$$

c.f. Eq. (3.1.8). Here, $n^\pi_{\text{He diss}}$ is defined via

$$n^\pi_{\text{He diss}}(T_{\text{BBN}}) = n_{\text{FIP}} \cdot \text{Br}_{\text{FIP}\to\pi^-} \cdot P_{\text{He diss}}, \tag{3.1.19}$$

Here,

$$n_{\text{FIP}}(T) = \left(\frac{a_{\text{SBBN}}(t_{\text{dec}})}{a_{\text{SBBN}}(T)}\right)^3 n^{\text{dec}}_{\text{FIP}} \cdot \zeta_{\text{FIP}} \cdot e^{-t/\tau_{\text{FIP}}} \tag{3.1.20}$$

is the FIP's number density, and $P_{^4\text{He diss}}$ is the probability for a single meson to dissociate ^{4}He nuclei before decaying:

$$P_{\text{He diss}} = \frac{\langle \sigma^\pi_{\text{He diss}} v\rangle n_{\text{He}}}{\Gamma^\pi_{\text{decay}}} \simeq 8.3 \cdot 10^{-2} \cdot \frac{4 \cdot n_{\text{He}}}{n_B} \left(\frac{T}{1\text{ MeV}}\right)^3, \tag{3.1.21}$$

where we used the total cross-section of the dissociation processes (3.1.17), $\langle \sigma^\pi_{\text{He diss}} v\rangle \simeq 6.5 \cdot F_{\text{He}\pi^-}$ mb (a factor $F_{\text{He}\pi^-} \simeq 3.5$ accounts for the Coulomb attraction, Eq. (3.B.3)).

In our estimates, we use $T_{\text{BBN}} = 84$ keV, assuming that all free nucleons become bounded in ^{4}He nuclei at this temperature. We also do not take into account that after the dissociation of Helium the abundance of lighter elements will be also increased significantly.

Further, to make a conservative upper bound estimate, we will assume that the FIPs were in thermal equilibrium and then decoupled while being UR, such that their abundance is maximally possible. In this case, $\zeta_{\text{FIP}} \simeq 10^{-2} - 10^{-1}$ for FIPs with lifetimes $\tau_{\text{FIP}} \sim 10^2$ s, in dependence on the FIP mass.[4] Requiring $n_{\text{He}} \simeq n_B/4$ in (3.1.18), and using Eqs. (3.1.19), (3.1.21), we arrive at the upper bound on HNL lifetimes for which our analysis is applicable, $\tau_{\text{FIP}} \lesssim 50$ s.

For FIPs that decay hadronically and have lifetimes $\tau_N \lesssim 10^4$ s, the presence of mesons from their decays may lead to an increase of primordial ^{4}He abundance.

Indeed, nuclear reactions are efficient until temperatures $T \simeq$ few keV. Therefore, once mesons disappear from the plasma (and nuclear dissociation processes stop), neutrons and protons get bounded into ^{4}He. Since the meson-driven $p \leftrightarrow n$ conversion keeps n/p ratio at the level of $\mathcal{O}(1)$, the resulting abundance may be still larger than the SBBN value.

In order to derive corrections to the nuclear abundances, we need to estimate the impact of effects of long-lived FIPs on BBN, which is complicated to perform analytically. Namely, the BBN reaction chain is non-equilibrium. In addition, the impact of FIPs on η_B and $t(T)$ cannot be estimated as a perturbation, since FIP may be abundant non-relativistic particles which dominate the energy density of the Universe. Therefore, numeric solution of equations for nuclear abundances and Friedmann equations in presence of decaying FIPs is required. We do this for long-lived HNLs in Sec. 3.3.3.1.

3.2 CMB

As we have discussed in the Introduction (remind Sec. 1.3.2), the main impact of FIPs with lifetimes $\tau_{\text{FIP}} \ll t_{\text{recombination}}$ on CMB comes from their change of Y_p and N_{eff}.

The value of N_{eff} is given by

$$N_{\text{eff}} \equiv \frac{8}{7}\left(\frac{11}{4}\right)^{4/3}\left(\frac{\rho_{\text{rad}} - \rho_\gamma}{\rho_\gamma}\right) , \tag{3.2.1}$$

where ρ_{rad} and ρ_γ are the total radiation and photon energy densities respectively. We define the change in this quantity as $\Delta N_{\text{eff}} = N_{\text{eff}} - N_{\text{eff}}^{\text{SM}}$, where within the Standard Model $N_{\text{eff}}^{\text{SM}} \simeq 3.044$ [183–187]. Any deviation from the SM value is regulated by weak interactions between neutrinos and electromagnetic (EM) particles, which are efficient enough at temperatures $T \gg 1\,\text{MeV}$ to keep these species in equilibrium with each other. At lower temperatures, the interactions gradually go out of equilibrium and the energy exchange between the two sectors will stop. Decaying FIPs can affect this delicate process

[4] As the bound is sensitive logarithmically to this product, its precise value is not so important.

in different ways, depending on whether they inject most of their energy into EM particles or neutrinos.

The impact of FIPs predominantly decaying into EM particles has been extensively studied in the literature, see e.g. [64, 182, 188–191]. Such particles heat up the EM plasma and consequently decrease $N_{\rm eff}$, independently of whether the decay happens during or after neutrino decoupling.

For FIPs that mostly decay into neutrinos, we naively expect that $N_{\rm eff}$ would increase. This is indeed true for lifetimes $\tau_{\rm FIP} \gg t_\nu^{\rm dec} \sim 0.1 - 1\,{\rm s}$, where $t_\nu^{\rm dec}$ is the time of neutrino decoupling, see e.g. [192]. However, there are controversial results for the lifetimes $\tau_{\rm FIP} \sim t_\nu^{\rm dec}$.

Neutrinos are still in partial equilibrium and try to equilibrate with the injected neutrinos at such time scale. This scenario has been considered before in [193–195] that arrived at different conclusions about the impact on $N_{\rm eff}$. Namely, the work [193] studied a reheating scenario in which all the SM particles are absent before FIPs start decaying. In such a framework, all neutrinos have high energies, which means that they mainly thermalize via neutrino-EM interactions and $N_{\rm eff}$ naturally decreases. References [194, 195] considered HNLs with masses $m_N < m_\pi$ and lifetimes $\tau_N \lesssim 1\,{\rm s}$. Such HNLs are in thermal equilibrium in the early Universe, but decouple as the Universe expands and eventually decay mainly into high-energy neutrinos at MeV temperatures. These two works drew different conclusions about $N_{\rm eff}$: [194] reported $\Delta N_{\rm eff} > 0$ for the whole studied mass range, whereas [195] presented in their Fig. 3 that $\Delta N_{\rm eff} < 0$ for masses $60\,{\rm MeV} \lesssim m_N < m_\pi$ and lifetimes $\tau_N \ll 1\,{\rm s}$. The sign of $\Delta N_{\rm eff}$ is not emphasized in these two papers; [195] did not comment on the contradiction with [194] on this issue and no physical discussion of this phenomenon was provided[5].

In this section, we aim to clarify the behavior of $N_{\rm eff}$ in the presence of FIPs that decay mainly into neutrinos and have lifetimes $\tau_{\rm FIP} \sim t_\nu^{\rm dec}$. Here we will assume that a thermal bath of SM particles is already present in the primordial Universe. We will first construct a simple model in Sec. 3.2.1 that provides us with a qualitative understanding of how such particles impact $N_{\rm eff}$, the findings of which we then confirm by using the Boltzmann code `pyBBN` [64].

[5] A more recent work [196] considered *long-lived* (i.e., decaying after e^+e^- annihilation) HNLs that could decay both into EM particles and neutrinos. In this case, $N_{\rm eff}$ could both increase and decrease, as at such late times the injected energy densities from HNL decays can dominate over the SM densities of both the EM and neutrino sectors. Another recent paper [197], which appeared after our work was submitted, claims that $\Delta N_{\rm eff} \geq 0$ for all cases in which FIPs decay mostly into neutrinos. We comment on it in Appendix 3.D.

Our analysis shows that short-lived FIPs that inject most of their energy into neutrinos may decrease $N_{\rm eff}$. This is because during the equilibration process, the injected high-energy neutrinos redistribute their energy among the neutrino and EM plasma.

If the energy of the injected neutrinos is sufficiently large, the energy transfer to the EM sector occurs faster than the equilibration with the neutrino sector. This means that the EM plasma heats up more than the neutrino plasma, which eventually leads to $\Delta N_{\rm eff} < 0$. We will find that this mechanism is especially relevant for FIPs with masses larger than a few tens of MeV. We will then apply these general considerations to the well-motivated case of HNLs. Complementary details and simulation results are included in the appendix 3.2.1.2.

3.2.1 Impact of short-lived FIPs on $N_{\rm eff}$

We focus on FIPs with masses $\gg 1\,\mathrm{MeV}$ that decay when neutrinos are still in (partial) equilibrium. Such FIPs can decay into high-energy neutrinos with energies much higher than those in the primordial plasma, that then participate in interactions with thermal neutrinos and electrons/positrons.

We will find that even if most of the FIP energy is injected into neutrinos, these interactions may still cause a decrease in $N_{\rm eff}$. This feature appears since the injected high-energy neutrinos get quickly converted into electrons/positrons and drag thermal neutrinos residing in the plasma along with them. During this process, neutrino-neutrino interactions lead to the presence of residual non-thermal distortions in the distribution functions of neutrinos (neutrino spectral distortions) that keep the balance of $\nu \leftrightarrow \mathrm{EM}$ interactions shifted to the right till long after the injection (i.e., more energy is transferred from the neutrino plasma to the electromagnetic plasma than vice versa). The energy transfer from neutrinos to EM particles accumulated over time can then be sizeable enough, such that $\Delta N_{\rm eff}$ becomes negative. This effect diminishes with larger FIP lifetime, as neutrino-EM interactions go out of equilibrium and neutrinos can no longer be converted into electrons/positrons. Therefore, FIPs that decay into neutrinos after neutrino decoupling will lead to $\Delta N_{\rm eff} > 0$. In what follows, we will consider FIPs that can decay into both neutrinos and EM particles, and construct a simple model that provides a semi-analytic description of the aforementioned effect. At the end of this section, we will also highlight and further elaborate on the central role of neutrino spectral distortions in the dynamics of $N_{\rm eff}$.

3.2.1.1 Analytic considerations

We assume that the amount of injected non-equilibrium neutrinos is only a small fraction of the thermal neutrinos in the plasma. The evolution of the injected neutrinos is then mainly

governed by the following reactions:

$$\nu_{\text{non-eq}} + \nu_{\text{therm}} \rightarrow \nu_{\text{non-eq}} + \nu_{\text{non-eq}} \tag{3.2.2}$$

$$\nu_{\text{non-eq}} + \overline{\nu}_{\text{therm}} \rightarrow e^{+} + e^{-} \tag{3.2.3}$$

$$\nu_{\text{non-eq}} + e^{\pm} \rightarrow \nu_{\text{non-eq}} + e^{\pm} \, , \tag{3.2.4}$$

where 'non-eq' and 'therm' refer to neutrinos with non-equilibrium and thermal energies respectively.

Through the thermalization reactions (3.2.2)-(3.2.4), non-equilibrium neutrinos thermalize and quickly redistribute their energy among the neutrino and EM plasma.

The energy loss rate of these non-equilibrium neutrinos is higher than the interaction rates of thermal particles [49]:

$$\frac{\Gamma_{\text{non-eq}}}{\Gamma_{\text{therm}}} \sim \frac{G_F^2 T^4 E_\nu^{\text{inj}}}{G_F^2 T^5} = \frac{E_\nu^{\text{inj}}}{T} \gg 1 \, , \tag{3.2.5}$$

where E_ν^{inj} is the average energy of the injected non-equilibrium neutrinos. Note that reactions between thermal particles also exchange energy between the neutrino and EM sectors, but this energy exchange is subdominant as far as Eq. (3.2.5) holds.

The amount of energy that ends up in the EM plasma has three contributions: *1)* the direct decay of FIPs into EM particles, *2)* the energy transfer of non-equilibrium neutrinos to EM particles during thermalization and *3)* the energy transfer from thermal neutrinos to EM particles as a consequence of them being dragged by non-equilibrium neutrinos during thermalization (reactions (3.2.2) and (3.2.3)). The first process injects a fraction ξ_{EM} of the total FIP energy into the EM plasma, while the latter two increase this fraction to:

$$\xi_{\text{EM,eff}}(E_\nu^{\text{inj}}, T) = \xi_{\text{EM}} + \xi_\nu \times \epsilon(E_\nu^{\text{inj}}, T) \, , \tag{3.2.6}$$

where $\xi_\nu = 1 - \xi_{\text{EM}}$ is the energy fraction that FIPs directly inject into the neutrino sector and $\epsilon = \epsilon_{\text{non-eq}} + \epsilon_{\text{thermal}}$ is the effective fraction of ξ_ν that went to the EM plasma during the thermalization. The latter quantity can be split in a contribution from non-equilibrium neutrinos ($\epsilon_{\text{non-eq}} = E_\nu^{\text{non-eq}\rightarrow\text{EM}}/E_\nu^{\text{inj}}$) and an *effective* contribution from thermal neutrinos ($\epsilon_{\text{thermal}} = E_\nu^{\text{thermal}\rightarrow\text{EM}}/E_\nu^{\text{inj}}$).

Now, based on Eq. (3.2.6), if $\epsilon > 0.5$, then $\xi_{\text{EM,eff}} > 0.5$. This means that more than half of the FIP energy eventually ends up in the EM plasma (i.e., EM plasma heats up more than the neutrino plasma), which results in $\Delta N_{\text{eff}} < 0$ independently of the value of ξ_{EM}[6]. This simplified energy redistribution picture only holds if the non-equilibrium

[6]Note that it is not a requirement that ϵ must be larger than 0.5 in order for ΔN_{eff} to be negative. It only signifies the independence from ξ_{EM}.

neutrino energy is much larger than the average energy of thermal neutrinos. Once these two energies become similar in magnitude, backreactions cannot be neglected anymore and the evolution can only be accurately described with a system of Boltzmann equations.

Because of much faster thermalization rate of EM plasma than of neutrinos and growth of the interaction rate with neutrino energy, neutrinos may store a huge amount of their energy ϵ in EM plasma during the thermalization. We may estimate it analytically.

We can make a simple estimate of ϵ as a function of the injected neutrino energy $E_\nu^{\rm inj}$ and temperature T. We start with describing the thermalization process of a *single* injected neutrino, which causes a cascade of non-equilibrium neutrinos. Such a cascade can result after the injected neutrino participates in the processes (3.2.2)−(3.2.4). We assume that in the processes (3.2.2) and (3.2.4) each non-equilibrium neutrino in the final state carries half of the energy of the non-equilibrium neutrino in the initial state. Thus, roughly speaking, the thermalization occurs during $N_{\rm therm} \simeq \log_2(E_\nu^{\rm inj}/3.15T)$ interactions. In addition, the process (3.2.2) doubles the number of non-equilibrium neutrinos, while (3.2.3) makes neutrinos disappear and (3.2.4) leaves the number unchanged. Therefore, after the k-th step in the cascade, the average number of non-equilibrium neutrinos is given by:

$$N_\nu^{(k)} = N_\nu^{(k-1)} \left(2P_{\nu\nu\to\nu\nu} + P_{\nu e\to\nu e}\right) = N_\nu^{(0)} \left(2P_{\nu\nu\to\nu\nu} + P_{\nu e\to\nu e}\right)^k , \tag{3.2.7}$$

with $N_\nu^{(0)} = 1$, and the total non-equilibrium energy is:

$$E_\nu^{(k)} = E_\nu^{(k-1)} \left(P_{\nu\nu\to\nu\nu} + \frac{1}{2}P_{\nu e\to\nu e}\right) = E_\nu^{\rm inj} \left(P_{\nu\nu\to\nu\nu} + \frac{1}{2}P_{\nu e\to\nu e}\right)^k , \tag{3.2.8}$$

where $P_{\nu\nu\to\nu\nu}$, $P_{\nu\nu\to ee}$, and $P_{\nu e\to\nu e}$ are the average probabilities of the processes (3.2.2)-(3.2.4), respectively, and their sum equals unity. We define these probabilities as $P_i = \Gamma_i/\Gamma_\nu^{\rm tot}$, where Γ_i is the interaction rate of each process and $\Gamma_\nu^{\rm tot}$ is the total neutrino interaction rate. The relevant reactions and their corresponding matrix elements are summarized in appendix D of [64]. Assuming a Fermi-Dirac distribution for neutrinos and averaging over neutrino flavours, we find:

$$P_{\nu\nu\to\nu\nu} \approx 0.76, \quad P_{\nu\nu\to ee} \approx 0.05, \quad P_{\nu e\to\nu e} \approx 0.19\,. \tag{3.2.9}$$

Finally, the value of $\epsilon_{\text{non-eq}}$ that accounts for the energy transfer from non-equilibrium neutrinos to the EM plasma is given by:

$$\epsilon_{\text{non-eq}} = \frac{1}{E_\nu^{\rm inj}} \sum_{k=0}^{N_{\rm therm}} \left(\frac{P_{\nu e\to\nu e}}{2} + P_{\nu\nu\to ee}\right) E_\nu^{(k)} \,. \tag{3.2.10}$$

In addition to the transferred non-equilibrium energy, the non-equilibrium neutrinos catalyze the energy transfer from thermal neutrinos to the EM plasma via the processes (3.2.2) and (3.2.3). In other words, during the thermalization process non-equilibrium neutrinos drag thermal neutrinos along with them, which leads to part of the energy stored in the thermal neutrino sector to end up in the EM sector. We assume that each reaction (3.2.2) transfers an energy amount of $3.15T$ from the thermal neutrino sector to non-equilibrium neutrinos, which then via (3.2.3) ends up in the EM plasma. Moreover, each reaction (3.2.3) contributes to another energy transfer of $3.15T$ from thermal neutrinos to the EM plasma. The effective contribution coming from this transfer is therefore:

$$\epsilon_{\text{thermal}} = \frac{3.15T}{E_\nu^{\text{inj}}} N_\nu^{\text{therm}\to\text{EM}} = \frac{3.15T}{E_\nu^{\text{inj}}} P_{\nu\nu\to ee} \left(\sum_{k=0}^{N_{\text{therm}}} N_\nu^{(k)} + \left[P_{\nu\nu\to\nu\nu} + \sum_{k=1}^{N_{\text{therm}}} \left(2P_{\nu\nu\to\nu\nu}\right)^{(k)} \right] \right), \tag{3.2.11}$$

where the first term in the round brackets is the contribution from the process (3.2.3) and the terms in the square brackets are the contribution from the process (3.2.2). Note that the factor of 2 in the second sum accounts for the doubling of non-equilibrium neutrinos in the process (3.2.2). We find that $\epsilon_{\text{thermal}}$ is at least 5 times smaller than $\epsilon_{\text{non-eq}}$, which makes this a sub-dominant effect.

As the Universe expands and the temperature decreases, weak reaction rates start to compete with the Hubble rate H. The energy transfer from neutrinos to the EM plasma therefore becomes less and less efficient, and ϵ tends to zero. In order to incorporate this effect, we multiply the probabilities in (3.2.9) with a factor $\min[\Gamma_i/H, 1]$, where $\Gamma_i = \Gamma_i(E_\nu^{\text{inj}}/2^k)$ is the interaction rate of any of the processes (3.2.2)−(3.2.4). The resulting energy fraction of neutrinos that is transferred to the EM plasma $\epsilon = \epsilon_{\text{non-eq}} + \epsilon_{\text{thermal}}$ is shown in Fig. 3 for a number of injected neutrino energies E_ν^{inj}.

The analytic model tells us that ϵ can exceed 0.5 for $E_\nu^{\text{inj}} \gtrsim 60\,\text{MeV}$. This means that when FIPs decay into neutrinos with such energies at temperatures of a few MeV, the majority of the injected neutrino energy will end up in the EM plasma during the thermalization. This then leads to a decrease of N_{eff}, independently of how much energy the FIPs inject into the EM sector.

Now that we are able to estimate ϵ, we can compute the correction to N_{eff} for some benchmark FIP scenario. It is worth noting here again that ϵ only depends on the energy of the injected neutrino and the temperature at which the injection happens. This means that ϵ is an independent quantity of the FIP model considered, in contrast to ξ_{EM} and ξ_ν, which do depend on the choice of the model. As an illustrative example, we assume that $\xi_{\text{EM}} = 0$, i.e., the FIP injects all of its energy into neutrinos ($\xi_\nu = 1$). Given that in our simple model neutrinos thermalize very quickly, we assume that they have a thermal-like distribution with a temperature T_ν and follow the approach in [184, 198] to obtain the time evolution of T_ν

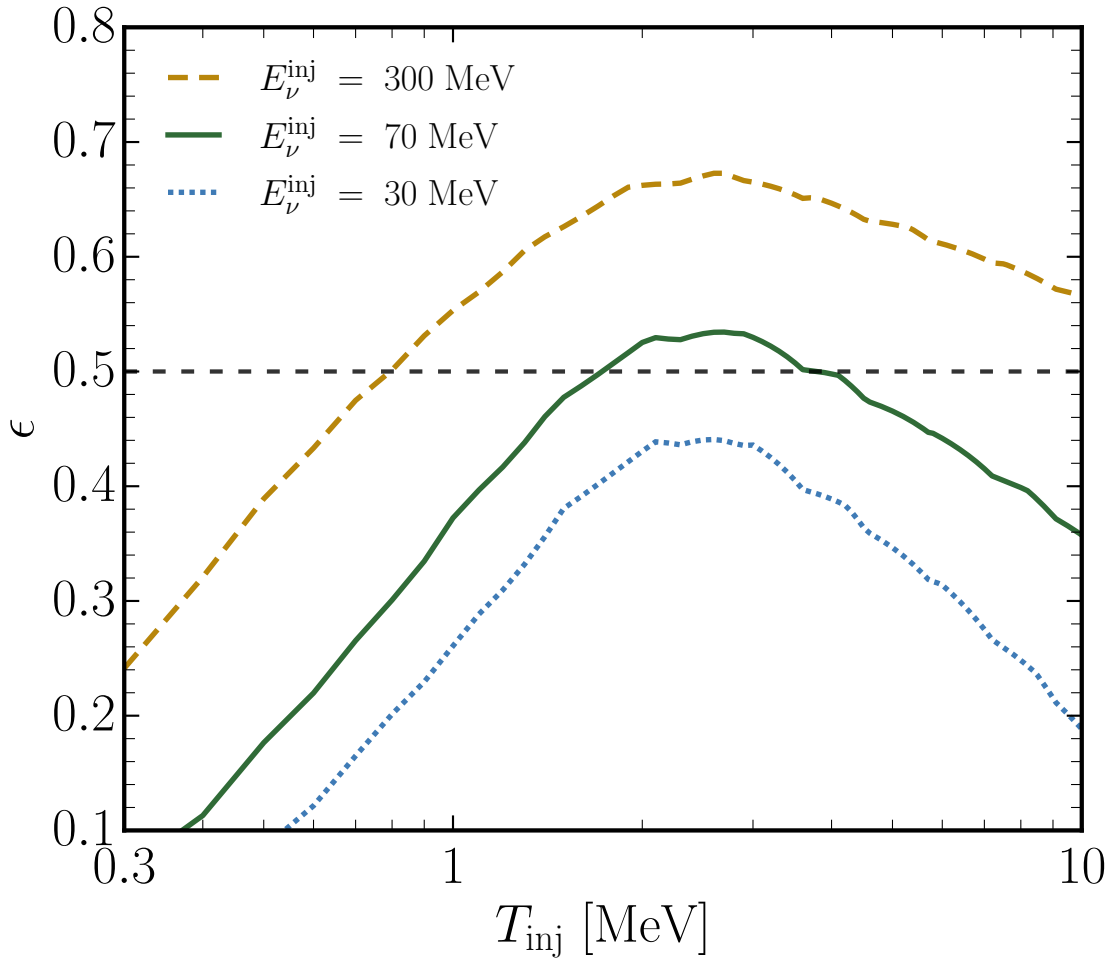

Figure 3: Estimate of the fraction of injected neutrino energy ϵ (both thermal and non-equilibrium) that gets transferred to the EM plasma during thermalization (see text for details). The three curves indicate the value of ϵ when a neutrino of energy $E_\nu^{\rm inj}$ is injected at a temperature $T_{\rm inj}$. At high temperatures of order of $T_{\rm inj} \simeq E_\nu^{\rm inj}$, the injected neutrinos are thermal-like, and hence ϵ is small. Once the temperature decreases, we enter the regime $E_\nu^{\rm inj} \gg 3.15 T_{\rm inj}$ and neutrinos transfer a significant amount of their energy to the EM plasma while thermalising. With further decrease of $T_{\rm inj}$, weak reactions go out of equilibrium and the energy transfer becomes less and less efficient, which results in a quick drop-off of ϵ.

and $T_{\rm EM}$ in the presence of decaying FIPs (see Appendix 3.C, where we provide the relevant equations). In this benchmark example, we consider a generic FIP of mass 500 MeV that can decay only into three neutrinos and show $\Delta N_{\rm eff}$ as a function of its lifetime in Fig. 5. In order to compare the accuracy of our simple model, we also include in this figure the evolution of $\Delta N_{\rm eff}$ as obtained from the publicly available Boltzmann code `pyBBN`[7] [64]. The grey band in this figure indicates the current sensitivity of $N_{\rm eff}$ by Planck, which at 2σ

[7] https://github.com/ckald/pyBBN

reads[8] $N_{\rm eff}^{\rm CMB} = 2.89 \pm 0.62$ [199, 200]. We see that $N_{\rm eff}$ can significantly decrease as a result of the thermalization of the injected neutrinos. This decrease of $N_{\rm eff}$ would only be further amplified if the FIPs were also to inject some of their energy into the EM plasma.

3.2.1.2 Effect of Residual Non-equilibrium Neutrino Distortions

The simple model described in Sec. 3.2.1 relies on the assumption that the remaining fraction $1-\epsilon$ of the injected neutrino energy is perfectly thermal. In reality, this may not be the case and the full thermalization would occur during a much larger number of interactions than $N_{\rm therm} \simeq \log_2(E_\nu^{\rm inj}/3.15T)$.

Therefore, this simple model underestimates the energy fraction that goes into the EM plasma[9]. The remaining non-equilibrium neutrinos will manifest themselves as residual non-thermal spectral distortions in the distribution function of neutrinos. These spectral distortions keep the energy exchange balance of $\nu \leftrightarrow \text{EM}$ reactions shifted to the right till long after FIP decay. As a result, more neutrino energy will be transferred to the EM plasma and $N_{\rm eff}$ can further decrease. There is a subtlety here that the remaining $1-\epsilon$ non-equilibrium neutrinos are only slightly hotter than the thermal neutrinos, and we cannot describe their thermalization as an instant process: The corresponding rate is comparable to the thermal energy exchange rate. As such, the energy transfer process is extended in time, and a proper study of this effect requires solving the Boltzmann equation for the neutrino distribution function.

To study the impact of neutrino spectral distortions on the $\nu \to \text{EM}$ energy balance shift, we consider a simple scenario where high-energy neutrinos are instantly injected into the primordial plasma. We make use of the publicly available Boltzmann code `pyBBN`[10] [64] to simulate this process and to track the evolution of the neutrino distribution functions. Within this setup, neutrinos with energy $E_\nu^{\rm inj} = 70\,\text{MeV}$ are instantly injected at $T = 3\,\text{MeV}$. They amount for a fixed percentage of the total neutrino energy density and are equally distributed over the three neutrino flavours. All Standard Model interactions as specified in [64] are included, but with neutrino oscillations turned off (without any loss of generality). In order to highlight the importance of neutrino spectral distortions, we perform this procedure a second time, but with neutrino spectral distortions turned off. In that case, the neutrino distribution function is given by a Fermi-Dirac distribution with temperature $T_{\nu_\alpha} = \left(\frac{240\rho_{\nu_\alpha}}{7\pi^2 g_{\nu_\alpha}}\right)^{1/4}$, where ρ_{ν_α} and $g_{\nu_\alpha} = 2$ are the energy density (of both neutrinos and anti-neutrinos) and number of degrees of freedom of neutrino flavour α respectively.

[8] This value is obtained from the Planck 2018 baseline TTTEEE+lowE analysis, where $N_{\rm eff}$, $Y_{\rm P}$ and the six base parameters in ΛCDM are varied.

[9] Once the energy of the non-equilibrium neutrinos is close to the average thermal energy of $3.15T$, they lose roughly $\Delta E_\nu = (E_\nu - 3.15T)/2$ of energy per scattering. Therefore, the number of scatterings required to diminish E_ν down to $3.15T$ is larger.

[10] https://github.com/ckald/pyBBN

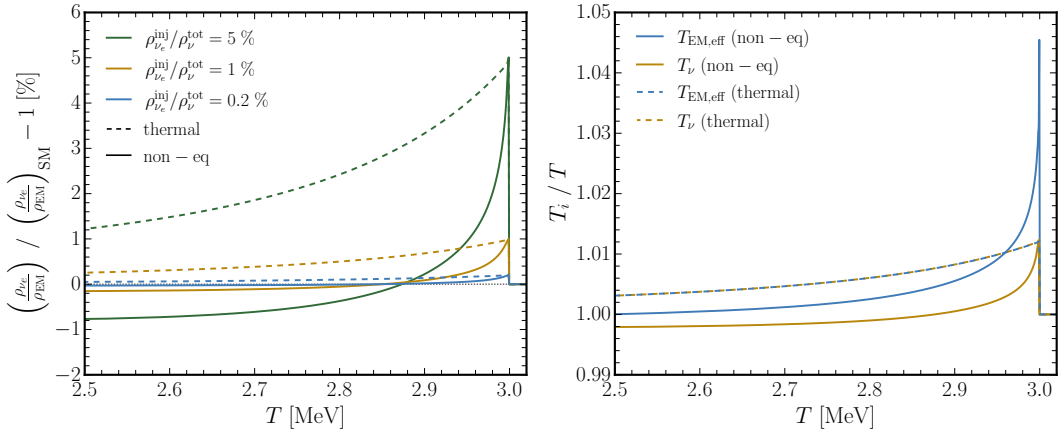

Figure 4: Evolution of the neutrino and EM plasma after the instant injection of neutrinos with energy $E_\nu^{\rm inj} = 70\,{\rm MeV}$ at $T = 3\,{\rm MeV}$. *Left panel*: The ratio of electron neutrino energy density to electromagnetic energy density, relative to the SM prediction. Three fractions of the injected energy density are considered: $\rho^{\rm inj}_{\nu_e}/\rho^{\rm tot}_{\nu} = \{0.2\%, 1\%, 5\%\}$. The solid lines are obtained by taking into account the full non-equilibrium spectrum of neutrinos, whereas the dashed lines correspond to the evolution assuming that neutrinos always have a thermal-like spectrum with temperature $T_\nu \propto \rho_\nu^{1/4}$. *Right panel*: Evolution of the neutrino temperature (dashed) and effective EM plasma temperature (solid) for which the energy transfer rate in Eq. (3.2.13) vanishes. An injected fraction of $\rho^{\rm inj}_{\nu_e}/\rho^{\rm tot}_{\nu} = 5\%$ is considered here. The solid and dashed lines indicate when non-equilibrium and thermal-like neutrino distributions are used respectively.

The evolution of the ratio $\rho_{\nu_e}/\rho_{\rm EM}$ (relative to the one in the SM) is shown in the left panel of Fig. 4 for different amounts of injected neutrino energy. In agreement with the story in Sec. 3.2.1, we observe a fast drop-off in the ratio right after the injection, which signifies the quick transfer of energy from the neutrino plasma to the EM plasma. After reaching the SM value (which naively corresponds to an equilibrium state), the ratio continues decreasing. This is the effect of the extended thermalization due to neutrino spectral distortions, as caused by the remaining fraction $1 - \epsilon$ of non-equilibrium neutrinos. Eventually, the ratio will be smaller than the SM value and $\Delta N_{\rm eff}$ becomes negative. In this plot, the dashed lines correspond to the same simulations but with a thermal-like distribution for the neutrinos. It is clear that without spectral distortions, the energy transfer from the neutrino sector to the EM sector is much less efficient.

Another way to look at this shift in the energy transfer balance from the neutrino plasma to the EM plasma is to ask the question: Which temperature $T_{\rm EM,eff}$ is the EM plasma *trying* to reach after the injection? As we will see, depending on whether neutrinos have a non-equilibrium or a thermal-like distribution, this temperature can be either larger than or equal to the neutrino temperature[11]. In the former case, it means that the EM plasma

[11] In all cases, with 'neutrino temperature' we refer to the quantity $T_\nu = \left(\frac{240\rho_\nu}{7\pi^2 g_\nu}\right)^{1/4}$, where $g_\nu = 2$ and ρ_ν is the energy density of both neutrinos and anti-neutrinos.

temperature can exceed the neutrino temperature (and thus $\Delta N_{\rm eff}$ can be negative), while in the latter case $\Delta N_{\rm eff}$ *cannot* be negative.

In more technical terms, the exchange of energy between neutrinos and EM particles is regulated by the Boltzmann collision integral $I_{\rm coll}$, which encodes all interactions between the species. For neutrinos that participate in reactions of the form $\nu + 2 \leftrightarrow 3 + 4$, the collision integral is given by [201]:

$$\begin{aligned} I_\nu =& \frac{1}{2 g_\nu E_\nu} \sum_{\text{reactions}} \int \prod_{i=2}^{4} \left(\frac{\mathrm{d}^3 p_i}{(2\pi)^3 2E_i} \right) |\mathcal{M}|^2 \times \\ & \times \left[(1-f_\nu)(1-f_2) f_3 f_4 - f_\nu f_2 (1-f_3)(1-f_4) \right] \times \\ & \times (2\pi)^4 \delta^4 (P_\nu + P_2 - P_3 - P_4) \, , \end{aligned} \tag{3.2.12}$$

where f_i and P_i are the distribution function and four momentum of species i respectively, and $|\mathcal{M}|^2$ is the unaveraged squared matrix element summed over degrees of freedom of initial and final states. The energy transfer rate between the neutrino and EM plasma can be written as:

$$\Gamma(T_{\rm EM}) = \int \frac{\mathrm{d}^3 p_\nu}{(2\pi)^3} I_{\rm coll}(T_{\rm EM}) E_\nu \, , \tag{3.2.13}$$

where we consider $I_{\rm coll}$ to be a function of the EM plasma temperature $T_{\rm EM}$. There exists a temperature $T_{\rm EM,eff}$ for which this rate is equal to 0. This corresponds to the temperature the EM plasma *tends* to during thermalization, since then the system would be in equilibrium. In the case where neutrinos would have a thermal-like spectrum with temperature T_ν, the rate vanishes when $T_{\rm EM,eff} = T_\nu$. On the other hand, when a non-equilibrium neutrino spectrum is considered, we find that $T_{\rm EM,eff} > T_\nu$ when $\Gamma = 0$. In the former case $N_{\rm eff}$ *cannot* decrease, while in the latter case the EM plasma temperature can exceed T_ν and $N_{\rm eff}$ can thus decrease. We show the evolution of $T_{\rm EM,eff}$ and T_ν as obtained from the instant neutrino injection simulations in the right panel of Fig. 4.

The conclusion here is that neutrino spectral distortions play a central role in transferring energy from the neutrino sector to the EM sector. When considering short-lived FIPs that can decay into neutrinos, the impact of these distortions on the evolution of $N_{\rm eff}$ should not be neglected.

Comparing the analytic model with the numeric simulations, we find that when using the Boltzmann equation $N_{\rm eff}$ decreases more than predicted by our semi-analytic model.

> The remaining fraction $1 - \epsilon$ of the injected neutrinos is not perfectly thermal, manifesting themselves as residual spectral distortions in the distribution function of neutrinos that further lead to a transfer of energy from the neutrino sector to the EM sector.

We see that in some cases the inclusion of this effect can make the difference between

being excluded by current data or not. We elaborate more on the effect of spectral distortions in Appendix 3.2.1.2. In short, the semi-analytic model is useful in providing a qualitative understanding of the behavior of $N_{\rm eff}$ in the presence of decaying FIPs. On the other hand, if the aim is to obtain accurate predictions for $N_{\rm eff}$ (relevant for setting bounds and forecasting), it is crucial to use the Boltzmann equation to track the evolution of the neutrino distribution functions. As such, we will use `pyBBN` in the remainder of this paper to simulate the impact of FIPs on $N_{\rm eff}$.

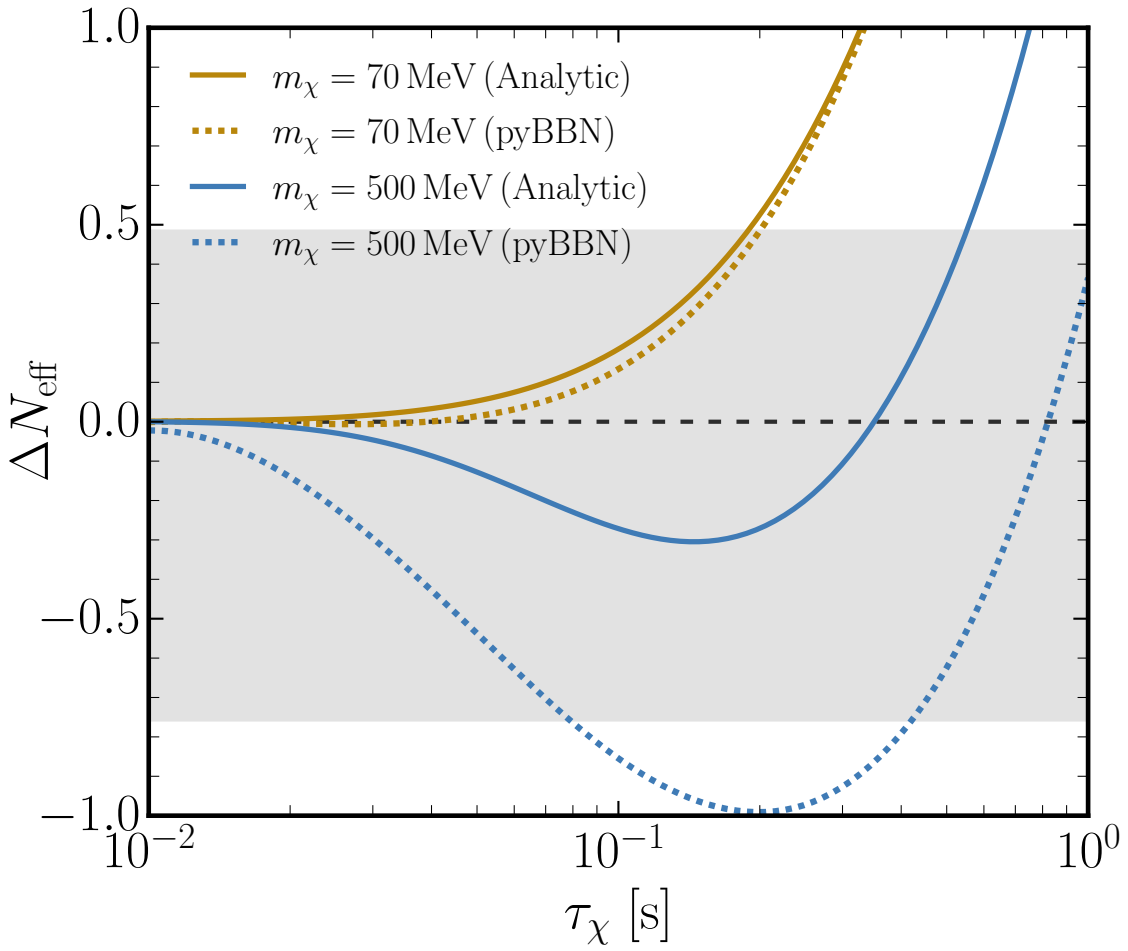

Figure 5: $\Delta N_{\rm eff}$ as a function of the lifetime of a FIP χ that can only decay into neutrinos through $\chi \to \nu_e + \nu_\mu + \overline{\nu}_\mu$. The initial FIP abundance is assumed to be $n_\chi/s = 0.01$ at $T = 1\,\mathrm{GeV}$, where s is the total entropy density of a universe consisting of photons, neutrinos and electrons/positrons. The solid lines are the result of our semi-analytic model, while the dotted lines are obtained with the Boltzmann code `pyBBN`. The grey band is the current sensitivity by Planck (see text for details). The golden curves roughly indicate the lowest FIP mass for which $N_{\rm eff}$ can decrease due to the thermalization of the injected neutrinos. The stronger decrease of the blue, dotted curve as compared to the solid curve highlights the significance of residual neutrino spectral distortions in the evolution of $N_{\rm eff}$ (see Appendix 3.2.1.2 for more details).

As a final point, we can make a rough model-independent estimate for which neutrino energies the decrease of $N_{\rm eff}$ happens. In the particular FIP scenario considered here, we find that this effect occurs for masses higher than $\sim$70 MeV (see Fig. 5). Given that in this case the neutrinos are created via 3-body decays, this would correspond to an average injected neutrino energy of roughly $E_\nu^{\rm inj} \sim m_{\rm FIP}/3 \sim 25\,{\rm MeV}$.

> As long as a FIP injects most of its energy into neutrinos around neutrino decoupling, $N_{\rm eff}$ could decrease if neutrinos with energies of at least $E_\nu^{\rm inj} \sim 25\,{\rm MeV}$ are produced.

3.2.2 Summary

In this work, we have studied how heavy, unstable FIMPs that can decay into neutrinos impact the number of relativistic species $N_{\rm eff}$ in the Early Universe. A particularly interesting effect that could occur with these particles, is when they inject most of their energy into neutrinos but still decrease $N_{\rm eff}$. This could happen if FIMPs decay when neutrinos are still in (partial) equilibrium ($\tau_{\rm FIMP} \sim \mathcal{O}(0.1)\,{\rm s}$) and is a direct consequence of the thermalization process of the injected high-energy neutrinos (see Sec. 3.2.1 for a semi-analytical treatment of this effect). Here we identify neutrino spectral distortions as the driving power behind this effect, since they lead to an efficient transfer of energy from the neutrino plasma to the electromagnetic plasma (see Figs. 5 and 4). Some of the injected neutrino energy gets quickly transferred to the EM plasma, while the remaining will stay as residual spectral distortions in the neutrino distribution functions. These spectral distortions keep the energy transfer balance of $\nu \leftrightarrow {\rm EM}$ reactions shifted to the right till long after FIMP decay. In order to accurately account for this effect, it is therefore important to solve the Boltzmann equation and track the evolution of the neutrino distribution functions. Using a thermal-like distribution for neutrinos as an approximation can lead to incorrect results, e.g., that $N_{\rm eff}$ can never decrease when FIMPs inject most of their energy into neutrinos.

From our simulations, done with the publicly available Boltzmann code `pyBBN` [64], we find that this mechanism is especially relevant for FIMPs that can decay into neutrinos with average energies $E_\nu^{\rm inj} \gtrsim 25\,{\rm MeV}$. In case such neutrinos are created via 2- or 3-body decays, this roughly corresponds to FIMP masses $m_{\rm FIMP}^{\text{2-body}} \gtrsim 50\,{\rm MeV}$ and $m_{\rm FIMP}^{\text{3-body}} \gtrsim 70\,{\rm MeV}$ respectively. This is in agreement with the results presented in [193]. As such, this effect may be relevant for many classes of FIMPs[12], including Higgs-like dark scalars [38], dark photons [202], neutralinos in supersymmetric models with broken R-parity [203], vector portals coupled to anomaly-free currents [151] and short-lived neutrinophilic scalars [204].

[12] While `pyBBN` is mainly built to simulate the cosmological history in the presence of Heavy Neutral Leptons, it can in principle be modified to include many other classes of FIMPs.

3.3 Case study: HNLs

In this section, we consider applications of the findings of previous sections to the case of HNLs. In what follows, we will consider two quasi-degenerate HNLs [205, 206], as motivated by the Neutrino Minimal Standard Model (or νMSM) [see e.g. 207–209])

HNLs alter the cosmological history through their contribution to the total energy density of the Universe and their decay into SM particles. HNLs that decay well before the decoupling of active neutrinos, i.e. at temperatures $T \gg 1$ MeV, will leave no traceable impact. On the other hand, if HNLs live long enough, they could alter several physical quantities, such as N_{eff} and the primordial abundances of light elements [192, 194, 195, 210–212]. Indeed, strong limits have been set on their mass and lifetime by considering their impact on Big Bang Nucleosynthesis and the Cosmic Microwave Background, see e.g. [37, 64, 196] for recent works on this subject.

The influence of HNLs on BBN and CMB depends on their abundance, and we will first discuss how HNLs are produced in the primordial plasma (Sec. 3.3.1). In Sec. 3.3.3.1, we derive the bounds from BBN, while in Sec. 3.3.4 we consider the impact of HNLs on CMB.

3.3.1 Thermal history of HNLs

At large temperatures, the interaction rate of HNLs with SM particles is temperature-suppressed, although the particle densities are high.

Indeed, in the plasma without lepton asymmetry at temperatures $T \gtrsim 1$ GeV the effective mixing angle is given by [213, 214]

$$U_{\text{m}}^2(T) \approx \frac{U^2}{\left[1 + 9.6 \cdot 10^{-24} \left(\frac{T}{1\text{ MeV}}\right)^6 \left(\frac{m_N}{150\text{ MeV}}\right)^{-2}\right]^2}, \tag{3.3.1}$$

see Appendix 3.A. As a result, the interaction rate of HNLs with SM particles $\Gamma_N^{\text{int}} \propto G_F^2 T^5 U_{\text{m}}^2$ is suppressed at both high and low temperatures and reaches its maximum at the temperature

$$T_{\max} \approx 12 (m_N/1\text{ GeV})^{1/3}\text{ GeV} \tag{3.3.2}$$

(see Fig. 6).

The HNLs were in thermal equilibrium if during some period $T_- < T < T_+$ the interaction rate $\Gamma_N^{\text{int}}(T)$ exceeded the Hubble expansion rate. For heavy HNLs with

masses $m_N \gtrsim 50$ MeV, this happens for mixing angles larger than

$$U^2 \gtrsim U^2_{\text{min}} \approx 3 \cdot 10^{-12} \left(\frac{1 \text{ GeV}}{m_N} \right) \tag{3.3.3}$$

Namely, using the condition $\Gamma_N^{\text{int}}(T_{\max}) = 3H(T_{\max})$, and approximating the interaction rate as $\Gamma_N^{\text{int}} \approx 10 U_m^2 G_F^2 T^5$, we find the minimal value on the mixing angle at which HNLs may enter the equilibrium, Eq. (3.3.3).

Notice that if HNLs are responsible for the generation of neutrino masses, there exists another lower bound on the mixing angle – the seesaw bound. At least one HNL with mass m_N should have mixing angle above this bound to be responsible for the generation of the atmospheric neutrino mass difference, *c.f.* [82]. The bound depends on details of the given HNL model – mixing pattern and neutrino mass hierarchy (see, e.g., [77, 208]). For simplicity, as the scale of the see-saw bound we will use the toy-model estimate

$$U^2 \gtrsim U^2_{\text{see-saw}} \simeq 5 \cdot 10^{-11} \left(\frac{1 \text{ GeV}}{m_N} \right) \tag{3.3.4}$$

The true see-saw bound may differ from the toy model estimate by within an order of magnitude.

The resulting parameter space of HNLs is shown in Fig. 6.

It is convenient to parametrize the population of HNLs in terms of the abundance, defined by

$$Y_N = \left(\frac{n_N}{s} \right)_{T=T_-}, \tag{3.3.5}$$

where n_N is the number density of HNLs and $s = g_* \frac{2\pi^2}{45} T^3$ is the entropy density.

3.3.1.1 HNLs with mixing angles below U_{min}

Let us now calculate the abundance of HNLs that never enter thermal equilibrium, i.e. of those with $U^2 \lesssim U^2_{\text{min}}$. The temperature evolution of the HNL abundance, $\mathcal{Y}_N$, may be found with the help of a simple equation

$$\frac{d\mathcal{Y}_N}{dt} = -\Gamma_{N,\text{int}} (\mathcal{Y}_N - \mathcal{Y}_{N,\text{eq}}), \tag{3.3.6}$$

where $\mathcal{Y}_{N,\text{eq}}(T)$ is the abundance of HNLs at equilibrium, Y_N is defined as the value of $\mathcal{Y}_N(T \ll T_{\max})$, and $\Gamma_{N,\text{int}}$ is the total rate of processes $A + N \to X$. At temperatures $T \gg m_N$, we may approximate the rate $\Gamma_{N,\text{int}}$ by an expression

$$\Gamma_{N,\text{int}} \approx b G_F^2 T^5 \cdot U^2_{\text{m}}(T), \tag{3.3.7}$$

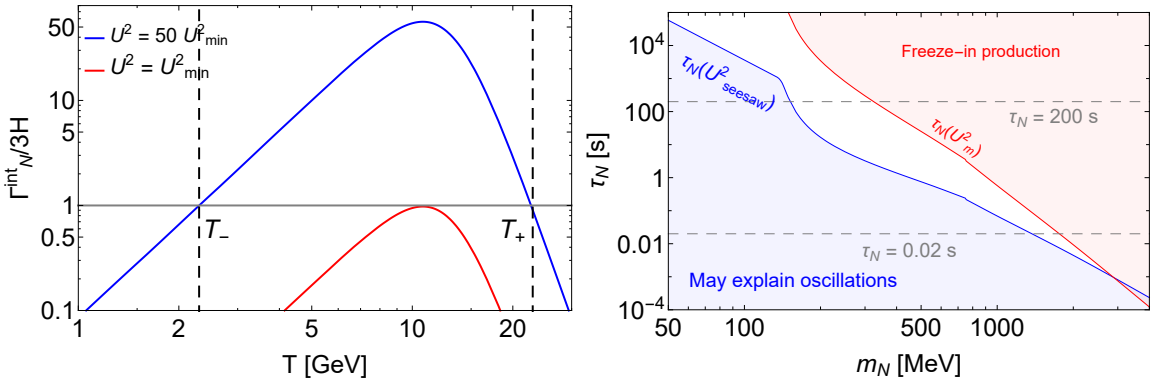

Figure 6: *Left panel*: The reaction rate of the HNL with SM particles, $\Gamma_N^{\rm int}$, compared to the Hubble rate, $H(T)$. T_+ and T_- are the temperatures at which HNLs enter and exit the thermal equilibrium. For illustration, we used HNL mass $m_N = 1$ GeV, and mixing angles $U^2 = U^2_{\rm min}$ and $U^2 = 50U^2_{\rm min}$, see Eq. (3.3.3). *Right panel*: the parameter space of HNLs that mix purely with ν_e. The blue domain roughly denotes the parameter space of HNLs that may explain neutrino oscillations, see Eq. (3.3.4). The red domain defines the parameter space for which HNLs never enter thermal equilibrium, see Eq. (3.3.3). The dashed scale $\tau_N = 0.02$ s denotes the shortest lifetime that may be constrained by BBN (the effect of the meson-driven $p \leftrightarrow n$ conversion as discussed in Sec. 3.1), while the scale $\tau_N = 200$ s defines the onset of nuclear reactions

where $b(T)$ is a factor depending on the number of SM species present in the primordial plasma. Also, if $T \gg \Lambda_{\rm QCD}$, we may use $g_* \approx 86.25$ [215], and the equilibrium abundance is $\mathcal{Y}_{N,\rm eq} \approx 0.01$ for Dirac HNLs and 0.005 for Majorana HNLs.

Using matrix elements for processes $N + A \to B + C$ from [64], we find $b \approx 10$ for Dirac HNLs (correspondingly, $b \approx 20$ for Majorana HNLs) at $T \gtrsim 1$ GeV. The value of b for Majorana HNLs is a factor 5 larger than that is used in [196], $b \approx 3.6$ (Majorana neutrinos are considered). A reason is that [196] uses rates from [216], where temperatures below 20 MeV are considered (see Eq. (6.8) from [216]), and hence A, B, C may be $e^\pm, \nu$ only, which is a huge underestimate.

Using (3.3.1), Eq. (3.3.12) may be integrated to obtain the final abundance of HNLs Y_N:

$$\int\limits_0^{Y_N} \frac{d\mathcal{Y}_N}{\mathcal{Y}_N - \mathcal{Y}_{N,\rm eq}} = -\int\limits_0^{\infty} dT \frac{\Gamma_{N,\rm int}(E_N)}{TH(T)} \approx -2.8\cdot 10^7 \frac{b}{\sqrt{a}} \frac{m_N}{1\ \text{GeV}} U^2 \Rightarrow \tag{3.3.8}$$

$$Y_N = \mathcal{Y}_{N,\rm eq}\left(1 - e^{-6\cdot 10^{11}\frac{m_N}{1\ \text{GeV}}U^2}\right) \tag{3.3.9}$$

Using $U^2 \ll U^2_{\rm min}$, we find that the abundance of HNLs that never entered thermal equilibrium is given by

$$Y_N \approx 2.8 \cdot 10^7 \cdot \mathcal{Y}_{N,\rm eq} \cdot 6 \cdot 10^{11} \frac{m_N}{1\ \rm GeV} U^2 \approx 5.7 \cdot 10^{10} \frac{m_N}{1\ \rm GeV} U^2 \tag{3.3.10}$$

3.3.1.2 HNLs with mixing angles above $U_{\rm min}$

Let us now consider HNLs with mixing angles above $U_{\rm min}$. It is important *(i)* whether HNLs froze out while being ultra-relativistic (UR regime, $m_N \ll T_-$, no exponential Boltzmann suppression for the number density) or non-relativistic (NR regime, $m_N \gg T_-$) and *(ii)* the value of g_* at the moment of the decoupling (depending on decoupling temperature it can change rapidly - see left Fig. 7). Using $n_N = \frac{3}{4} 2 \frac{\zeta(3)}{\pi^2} T^2$ for the UR regime or $n_N \sim \left(\frac{m_N}{T_-}\right)^{3/2} e^{-m_N/T_-}$ for the NR regime, we get the abundance in these two limits:

$$Y_N \simeq \begin{cases} \dfrac{0.6}{g_*(T_-)}, & \text{UR regime} \\ \alpha(m_N, \tau_N) \left(\dfrac{m_N}{T_-}\right)^{3/2} e^{-m_N/T_-}, & \text{NR regime} \end{cases} \tag{3.3.11}$$

The coefficient $\alpha(m_N, \tau_N)$ in Eq. (3.3.11) appears since the decoupling is not an instantaneous process; in dependence on the mass and lifetime it can vary by a factor of $\mathcal{O}(10)$.

To improve these estimates, we find the abundance numerically. We assume the Boltzmann approximation for the distribution function of the plasma particles and the equilibrium shape of the energy distribution of HNLs (such that, in particular, $\langle E_N \rangle$ is $3.15T$ for $T \gg m_N$). In this case, the equation for the evolution of the abundance of HNLs has the form (see, e.g., [217])

$$\frac{d\mathcal{Y}_N}{dt} = -\Gamma_{N,\rm int}(\mathcal{Y}_N - \mathcal{Y}_{N,\rm eq}), \tag{3.3.12}$$

where $\Gamma_{N,\rm int}$ is given by

$$\Gamma_{N,\rm int} = \sum_{A,B,C} \frac{g_N g_A}{8\pi^4} \int\limits_{s_{\rm min}}^{\infty} p^2_{AN} \sqrt{s} \sigma_{N+A\to B+C} K_1\left(\frac{\sqrt{s}}{T}\right) ds \tag{3.3.13}$$

In $\Gamma_{N,\rm int}$, the threshold invariant mass is $s_{\rm min} = \min[(m_N + m_A)^2, (m_B + m_C)^2]$, and

$$p^2_{AN} = \frac{s}{4}\left(1 - \frac{(m_N - m_A)^2}{s}\right)\left(1 - \frac{(m_N + m_A)^2}{s}\right) \tag{3.3.14}$$

The values of abundances Y_N for particular lifetimes are shown in Fig. 8 (right panel).

For masses 0.2 GeV $\lesssim m_N \lesssim 2$ GeV and lifetimes above 0.001 s, HNLs decouple at temperatures $T \gg m_N$, i.e. being ultra-relativistic and above the QCD transition. Their abundance Y_N is therefore universal and almost constant, owing to the temperature dependence of g_* (left panel) for $T \gtrsim 200$ MeV. For masses $\mathcal{O}(100)$ MeV and for large lifetimes $\tau_N \gtrsim 0.1$ s, HNLs still decouple while being ultra-relativistic. With the decrease of τ_N, Y_N first grows by a factor of few (due to rapid decrease of g_*), and with further decrease it becomes strongly suppressed.

Eq. (3.3.11) means that for UR regime later decoupling (*i.e.* larger mixing angles) leads to larger HNL abundance. In Fig. 6 (the right panel), we summarize the HNL

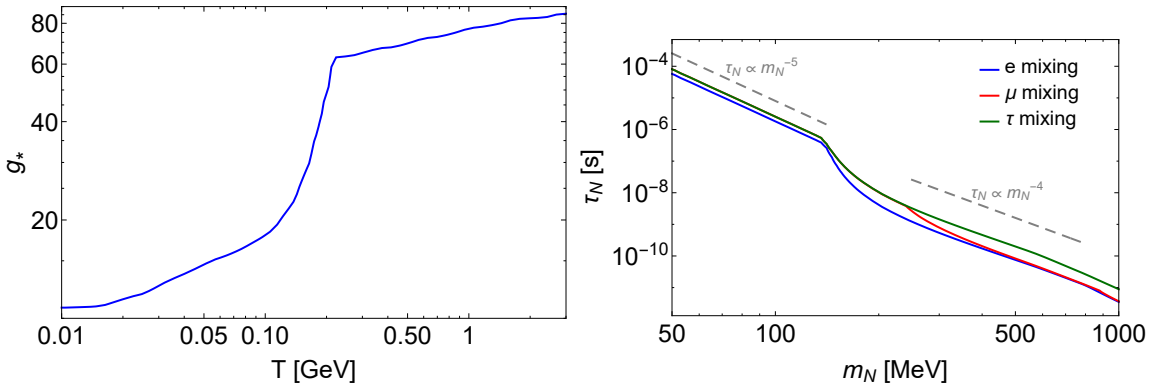

Figure 7: *Left panel*: temperature dependence of g_* in SM (reproduced from [215]). The drop around $T = 200$ MeV is caused by the entropy dilution at the QCD transition. *Right panel*: HNL lifetime as a function of mass for mixing with different flavors. The dashed gray lines show the scaling of the lifetime with mass. The lifetimes is shown for $U^2 = 1$ and scales as U^{-2}.

parameter space explored by the current study. It shows the domain in which HNLs never entered thermal equilibrium as well as the regime in which HNLs decouple while being non-relativistic. We see that these two regimes are separated by the broad parameter space for which HNLs enter thermal equilibrium and decouple while being UR. A dashed line in the middle of this region is the seesaw bound (3.3.4).

The temperature of freeze-out (T_-) is roughly defined via

$$\Gamma_N^{\rm int}(T_-) \simeq 3H(T_-), \tag{3.3.15}$$

see left panel in Fig. 6 (this equation has two solutions, the larger one defines T_+). The

factor of two estimate for T_- reads

$$T_- \simeq T_{\nu,\rm dec} \times \begin{cases} \dfrac{1}{U^{2/3}} \dfrac{1}{n_{\rm int}^{1/3}} \left(\dfrac{g_*(T_-)}{10.75}\right)^{1/6}, & \text{UR regime} \\ \dfrac{1}{U^2} \dfrac{1}{n_{\rm int}} \left(\dfrac{100\ \text{MeV}}{m_N}\right)^2 \left(\dfrac{g_*(T_-)}{10.75}\right)^{1/2} & \text{NR regime,} \end{cases} \tag{3.3.16}$$

where $T_{\nu,\rm dec} \approx 1.4$ MeV is the decoupling temperature of active neutrinos, $n_{\rm int} = \Gamma_{N,\rm int}/G_F^2 T^5$ is a factor that varies from $\simeq 2$ at $T \simeq \mathcal{O}(1\ \text{MeV})$ to $\simeq 9$ at $\mathcal{O}(1\ \text{GeV})$ temperatures. The different dependence on U^2 and on m_N in two regimes is due to the change of centre-of-mass energy ($E_{\rm cm} \sim T$ for UR and $E_{\rm cm} \approx m_N$ in the NR regimes). The values of T_- for different masses are shown in Fig. 8 (left). Instead of the mixing angles we use the lifetime $\tau_N \propto U^{-2}$ (Fig. 7, left) that is more intuitive when studying the influence on BBN.

For masses around $m_N \simeq 200$ MeV and lifetimes $\tau_N \sim 0.1$ s the HNL freeze-out occurs around the hadronization epoch. During this epoch, g_* drops by a factor ~ 3 [215] while $T_- \simeq m_N$, and therefore the abundance of HNLs can be *higher than for relativistic decoupling*.

For smaller masses, the decoupling temperature rapidly drops (see Fig. 8, left panel), which results in the Boltzmann suppression of the abundance. This effect is translated into a factor of few "kink" below the mass $m_N \simeq 200$ MeV in the final plots (Figs. 12).

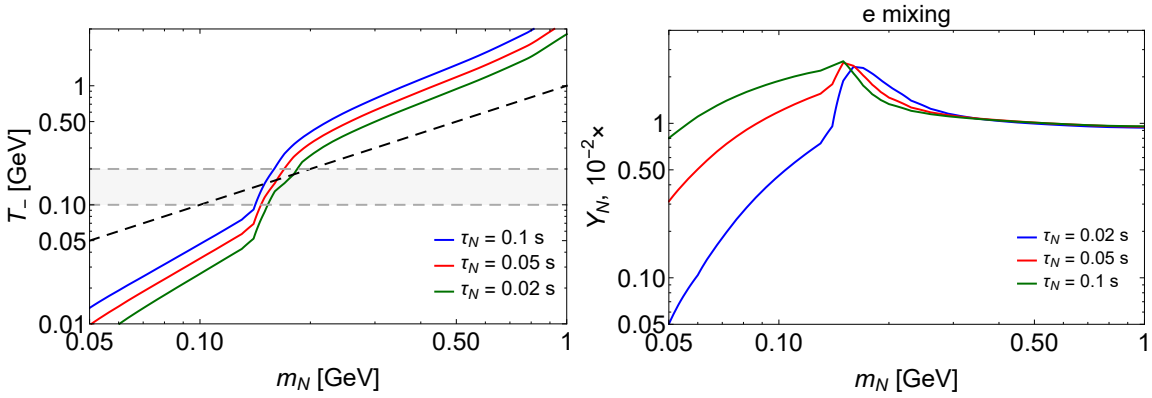

Figure 8: *Left panel*: the behavior of the decoupling temperature T_-, defined via $\Gamma_{N,\rm int}(T_-) = 3H(T_-)$, versus the HNL mass for particular lifetimes. The black dashed line defines the parameter space $T_- = m_N$, which roughly indicates the transition from relativistic to non-relativistic regime of HNL decoupling. The gray horizontal band shows a temperature when the hadronization of quarks takes place, and therefore the effective number of relativistic degrees of freedom, g_*, drops sharply (remind Fig. 7). *Right panel*: HNL abundances versus the HNL mass for particular values of the lifetime.

3.3.1.3 Evolution after decoupling

After the freeze-out, the comoving number density of HNLs changes only due to HNL decays. The physical number density thus evolves as

$$n_N(T) = n_N(T_{\text{dec}}) \cdot \left(\frac{a(T_{\text{dec}})}{a(T)}\right)^3 \cdot e^{-t/\tau_N} \tag{3.3.17}$$

Decays of HNLs inject energy into the primordial plasma. This effect changes the time-temperature relation and the scale factor evolution as compared to SBBN. The HNL decays provide additional dilution of any decoupled relics (including themselves) in comparison to the SBBN case:

$$\zeta = \left(\frac{a_{\text{SBBN}}}{a_{\text{SBBN + N}}}\right)^3 < 1, \tag{3.3.18}$$

where $a^{-1}_{\text{SBBN}}(T) \propto g_*^{1/3} T$ is the scale factor in SBBN, and the scale factors are evaluated at times $t \gg \tau_N$. To calculate ζ, we solve the Friedmann equation under an assumption that neutrinos are in perfect equilibrium and neglecting the mass of electrons:

$$\begin{aligned} H^2(t) =& \frac{1}{M_{\text{Pl}}^2}\frac{8\pi}{3}\Big[\rho_{\text{rad}} + m_N \cdot n_N(T)\Big], \\ 4\frac{\rho_{\text{rad}}}{T}\frac{dT}{dt} =& \frac{m_N n_N(t)}{\tau_N} - 4H(t)\cdot\rho_{\text{rad}}, \end{aligned} \tag{3.3.19}$$

where the number density of HNLs is given by Eq. (3.3.17). This is a reasonable assumption, since most of the HNLs with lifetimes $\tau_N \ll 0.1$ s decay much earlier than neutrinos decouple.

Effects of meson-driven conversion force us to trace the number density of HNLs even at times $t \gg \tau_N$:

$$n_N(t \gg \tau_N) = n_N(T_-) \cdot \left(\frac{a_{\text{SBBN}}(T_-)}{a_{\text{SBBN}}(T)}\right)^3 \cdot \zeta \cdot e^{-t(T)/\tau_N} \approx 0.4 Y_N \cdot g_{*,\text{SBBN}}\, T^3 \cdot \zeta \cdot e^{-t(T)/\tau_N}, \tag{3.3.20}$$

where $t(T)$ is the same as in SBBN.[13] Because of the suppression, the effect of this population on the expansion of the Universe may be neglected. However, this exponential tail still may produce mesons in amounts sufficient to change the dynamics of the n/p ratio.

The values of the HNL abundance and the dilution factor versus its mass and lifetime are given in Fig. 9.

[13] At times $t \gg \tau_N$, the time-temperature relation differs from SBBN only by the value of N_{eff}. However, the latter may change only if neutrinos are not in perfect equilibrium, and hence $t(T)$ is the same as in SBBN for lifetimes $\tau_N \ll 0.1$ s.

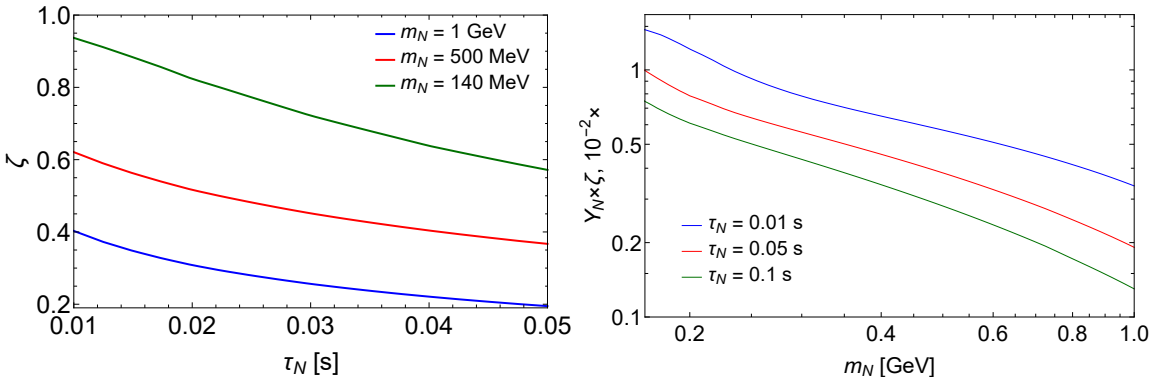

Figure 9: *Left panel*: Dilution factor (3.3.18) for short-lived HNLs mixing with ν_e. *Right panel*: HNL abundance times dilution factor as a function of mass for particular values of the lifetime. Details of the calculation of the abundances and ζ are given in [218]. Dilution factor is calculated, when most of HNLs has decayed and do not contribute to entropy density. Note, that we define abundance at the moment of decoupling, hence it does not change with decays.

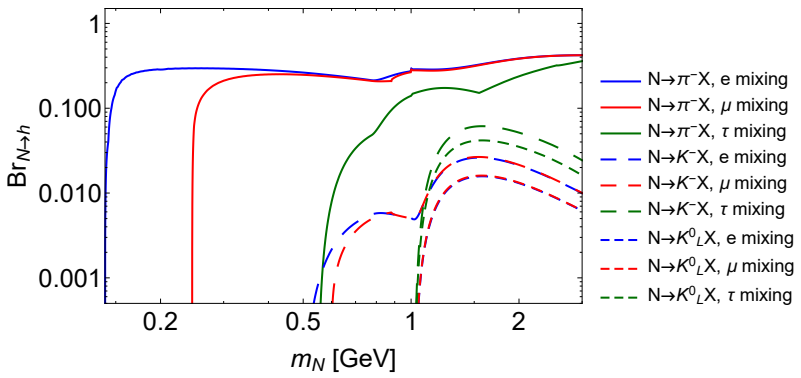

Figure 10: Branching ratios of HNL decays into mesons $h = \pi^-, K^-, K^0_L$. Secondary decays are also included (see text for details).

3.3.2 Hadronic decays of HNLs

In this work, we consider a pair of HNLs, degenerate in mass and having similar mixing angles. Two such HNLs form a single quasi-Dirac fermion [205, 219]. The abundance of a meson h produced from such HNLs is proportional to the quantity $Y_N \cdot \text{Br}_{N\to h}$. The mass dependence of $\text{Br}_{N\to h}$ for different mesons h and mixing patterns is shown in Fig. 10. We are interested only in the abundances of light mesons (pions and kaons) and for HNL masses well above pion/kaon thresholds we should account for "secondary mesons". This is discussed below, mainly following [39]. **Decays into pions.** In the case of the pure e/μ

mixings, the charged pion production threshold corresponds to $m_N = m_\pi + m_l$, where $l = e/\mu$. For τ mixing, the similar charged current-mediated channel opens up only at $m_N = m_\tau + m_\pi \simeq 1.9$ GeV. However, for all types of mixings charged pions may appear

as secondary particles in decays of neutral mesons,

$$N \to h^0 + \nu_\alpha, \quad h^0 \to \pi^\pm + X, \quad \text{where} \quad h^0 = \rho^0, \eta^0, \eta', \omega^0, \phi \tag{3.3.21}$$

Therefore, for τ mixing charged pions may appear at masses $m_N \geq m_{\eta^0}$. We use the branching ratios $\text{Br}_{\eta^0 \to \pi^\pm X} \approx 0.27$, $\text{Br}_{\rho^{0,\pm} \to \pi^\pm X} \approx 1$ [52].

Above $m_N \simeq 1$ GeV, decays of HNLs into pions cannot be approximated by single meson decays. Indeed, decays of GeV mass range HNLs are similar to decays of τ lepton [39], whereas for the latter hadronic decays are dominated by multi-pion channels [52]. We estimate the width of multi-pion decays as the difference between the total width into quarks and the width into single mesons:

$$\Gamma_{N \to n\pi} = \Gamma_{N \to \text{quarks}} - \sum_{h=\pi,K,\rho,\ldots} \Gamma_{N \to hX} \tag{3.3.22}$$

For multiplicities $\mathcal{N}$ of decays of HNLs into charged pions $N \to \pi^\pm$ (i.e., the amount of $\pi^\pm$ per multi-hadronic decay of HNLs), we will use multiplicities for multihadronic decays of τ leptons. Namely, $\mathcal{N}_{N \to \pi^+} = \mathcal{N}_{\tau^+ \to \pi^+} \approx 1.35$, $\mathcal{N}_{N \to \pi^-} = \mathcal{N}_{\tau^+ \to \pi^-} \approx 0.34$. The effective branching into π^- from multi-pion decays is

$$\text{Br}^{\text{multi-pion}}_{N \to \pi^-} = \mathcal{N}_{N \to \pi^-} \cdot \frac{\Gamma_{N \to n\pi}}{\Gamma_N}, \qquad \text{Br}^{\text{multi-pion}}_{\bar{N} \to \pi^-} = \mathcal{N}_{\bar{N} \to \pi^-} \cdot \frac{\Gamma_{N \to n\pi}}{\Gamma_N} \tag{3.3.23}$$

Since the bound on the meson driven $p \leftrightarrow n$ conversion is only logarithmically sensitive to the value of $\text{Br}_{N \to \pi^\pm}$, our results depend on these assumptions weakly.

Decays into kaons. Below $m_N = m_\phi$, charged kaons may appear only through the mixing with e/μ in the process $N \to K^- l$. This decay is Cabibbo suppressed [39] and almost two orders of magnitude smaller than into pions. Neutral kaons appear only in the final states with three or more particles (such as $N \to K^0 + \bar{K}^0 + \nu_\alpha$ and $N \to K^+ + \bar{K}^0 + \ell^-$, etc).

HNLs heavier than ϕ meson may produce both charged and neutral kaons via decays $N \to \phi\nu, \phi \to KK$. We assume that K^0 contains equal admixtures of K^0_L and K^0_S, i.e. $\text{Br}_{N \to K^0_L} = \text{Br}_{N \to K^0}/2$. We use the branching ratios $\text{Br}_{\phi \to K^-} \approx 0.5$, $\text{Br}_{\phi \to K^0_L} \approx 0.34$ [52].

3.3.3 Bounds from BBN and CMB

3.3.3.1 BBN

Let us first consider the bound from BBN. From the previous section we conclude that in dependence on the mixing patter, decays into charged pions become possible for HNLs with masses from $m_N = m_\pi + m_e$ (for the pure e mixing) to $m_N = m_\eta \approx 547$ MeV (for the pure τ mixing), see [39, 220] or Sec. 3.3.2.

HNLs with minimal lifetimes that may be constrained by BBN, $\tau_N \simeq \mathcal{O}(0.02\text{ s})$, are produced thermally, remind Fig. 6. For such HNLs, in Eq. (3.1.13) $n_{N,\text{dec}}/n_\gamma(T_{\text{dec}}) \approx 3/2$.

The dilution factor $(a_{\text{dec}}T_{\text{dec}}/a_0T_0)^3$ is of $0.1 - 0.6$ for HNL masses $m_\pi \lesssim m_N \lesssim 3$ GeV (we will use $\frac{1}{3}$ for normalization below), see Appendix 3.3.1.

Using values of $\text{Br}_{N\to h}$, P_{conv} and the scale factors ratio (which may be found in Appendix 3.B and sections 3.3.2 3.3.1), we conclude that the logarithm term in (3.1.13) is $\mathcal{O}(1)$ for HNLs in the mass range $m_N = \mathcal{O}(1\text{ GeV})$ and affects the overall bound very weakly. Therefore, the bound depends only on $T_0^{\min}$.

The maximal admissible correction (3.1.14) is reached for $T_0^{\min} = 1.50$ MeV, almost independently on HNL mass (see Fig. 2 and Appendix 3.B.1).

Plugging $T_0^{\min} = 1.50$ MeV into (3.1.13), we obtain our final limit from the analytic estimates

$$\tau_N \lesssim 0.023 \text{ s}. \tag{3.3.24}$$

Numeric calculations from Appendix 3.B.1 confirm this result, predicting constraints at the level of $0.019 - 0.021$ s.

Let us now comment on the maximal lifetimes for which our bounds are applicable. In Sec. 3.1.3, we restricted the applicability of our bounds by lifetimes $\tau_{\text{FIP}} \simeq \mathcal{O}(50\text{ s})$ – which is an estimate derived under the assumption of absence of effects of FIPs during the nuclear reaction chain. To push the bound to larger lifetimes, we solve numerically the system of equations for abundances of $d, t, ^3\text{He}, ^4\text{He}, ^7\text{Li}, ^7\text{Be}$ in presence of mesons from decaying HNLs. We incorporate the change of the time-temperature relation and the dynamics of η_B via the Friedmann equations with HNLs. We use the nuclear rates and reactions chain from [48]. Further description of our numeric approach may be found in Appendix 3.B.2. The example of the temperature behavior of the nuclear abundances in presence of HNLs is shown in Fig. 11.

Using the numeric approach, we conclude that long-lived HNLs with lifetimes $\tau_N \lesssim 10^4$ s increase nuclear abundances of all elements. The behavior of abundances with HNL lifetime for a particular mass $m_N = 200$ MeV is shown in Fig. 11.

For larger lifetimes, we need to include the effect of photo-dissociation. Therefore, using numeric approach, we have extended the domain of applicability of BBN bounds on HNLs to lifetimes $\tau_N = 10^4$ s.

3.3.3.2 Results

We demonstrated that HNLs with semi-leptonic decay channels significantly affect the primordial ^{4}He abundance, as mesons from their decays drive the $p \leftrightarrow n$ conversion rates away from their SBBN values (*c.f.* [178, 181, 182]). In order to avoid ^{4}He overproduction, mesons should disappear from the primordial plasma by $T = T_0^{\min} \simeq 1.50\,\text{MeV}$. The

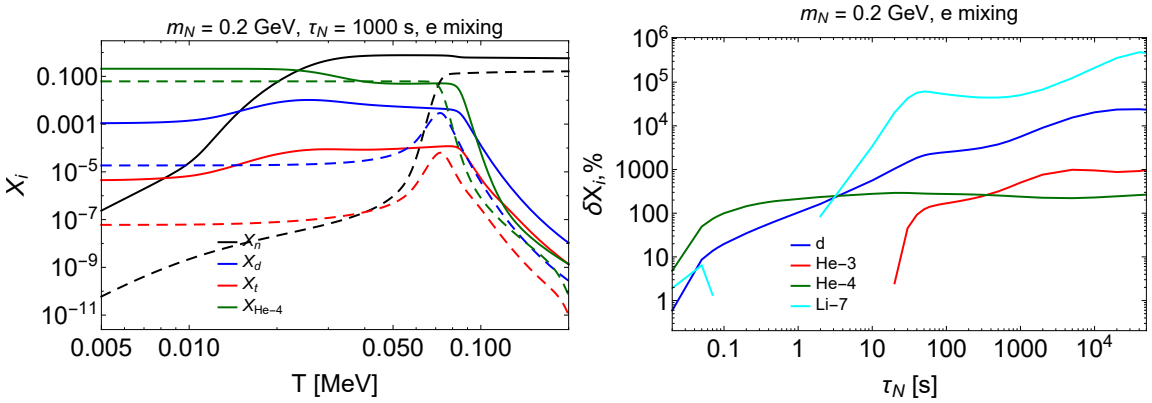

Figure 11: *Left panel*: the temperature behavior of the nuclear abundances in presence of HNLs (the solid lines), as well as their behavior in SBBN (the dashed lines). *Right panel*: the behavior of the change of nuclear abundances $\delta X_i \equiv (X_i - X_i^{(\mathrm{SBBN})})/X_i^{(\mathrm{SBBN})}$ with HNL lifetime. In both figures, an HNL with mass $m_N = 200$ MeV and pure e mixing is considered.

neutron abundance will then have enough time to relax down to its SBBN value before the onset of deuteron formation. These requirements severely constrain the parameter space of the HNLs with $0.023\ \mathrm{s} \leq \tau_N \leq 10^4\ \mathrm{s}$ for masses $m_N > 140\,\mathrm{MeV}$.

The final bounds for different mixing patterns are shown in Figs. 13 and 12. Our constraints can be generalized to other HNL models, see e.g. [221].

Confronted with the bounds from accelerator searches, we ruled out HNLs with mass below 500 MeV (for electron mixing) and 350 MeV (for muon mixing). Moreover, tighter bound means that future searches at Intensity Frontier (specifically, SHiP experiment [80]) can reach the BBN bottom line and completely rule out HNLs with the masses up to 750 MeV, which was not the case before [see e.g. 67, 222].

The comparison with the previous results [64, 194, 195] is shown in Fig. 13 (right panel). Our bound (3.3.24) is a factor of ~ 5 stronger than the previous result [194]. The recent reanalysis [64] did not take into account the effects of mesons, therefore their results are a factor $2-3$ less conservative.

The clear qualitative effect discussed in this paper not only leads to a tighter bound on HNL lifetime and provides an reachable goal for experimental searches, but also allows for an analytic description, unusual in the realm of BBN predictions driven by sophisticated numerical codes.

3.3.4 Bound from CMB

Let us now discuss the influence of HNLs on the physics at the CMB epoch. We do not consider masses higher than $m_N \simeq 1$ GeV, since there is no adequate description of HNL decay widths due to theoretical uncertainties [39], while it is crucial to know them for the calculation of N_{eff}. Indeed, this makes it complicated to compute ξ_ν (and thus ΔN_{eff}), as it depends on the branching ratios of the different multi-meson decay channels. For instance,

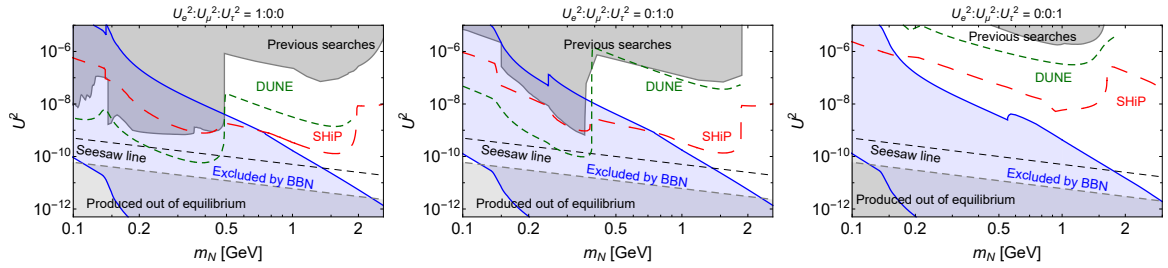

Figure 12: Bounds for HNLs mixed with a particular flavor. The blue area is excluded by our present analysis combined with [218] (for HNL masses below the charged pion production threshold). The dark gray area denotes the excluded HNL parameter space from previous searches [82], including the latest NA62 search [223]. The red and greed dashed lines show the sensitivity of several future intensity frontier experiments with the highest sensitivity in the regions of interest – SHiP [80, 224] and DUNE [225–227] (see [67]). Finally, the black dashed line denotes the seesaw bound applicable if two degenerate in mass HNLs are responsible for neutrino oscillations (as in the νMSM) [77, 82]. Our bounds are applicable up to lifetimes $\tau_N = 10^4$ s, from which EM decay products of HNLs may lead to nuclear photodissociation, see text for details.

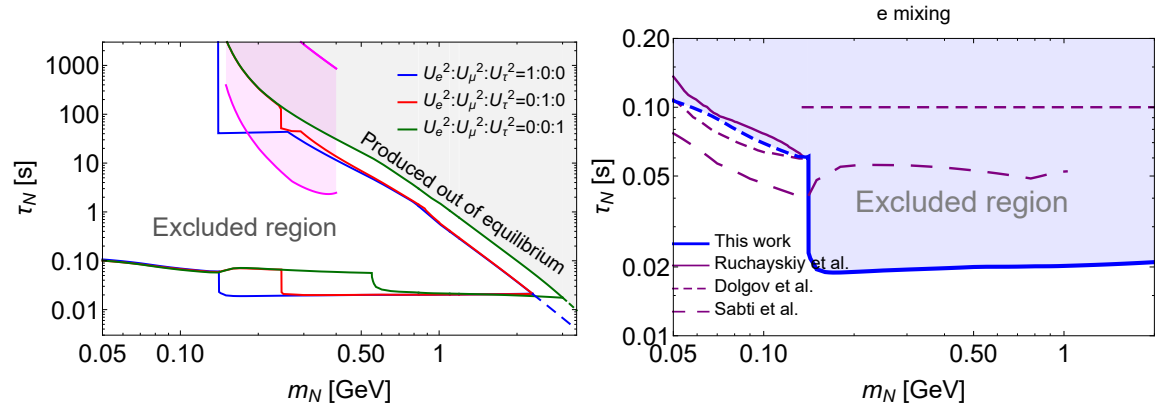

Figure 13: *Left panel*: BBN bounds on HNL lifetime for different mixing patterns. The gray region is excluded as a result of this work (for masses below pion threshold we use the results of [218]). The magenta shaded region corresponds to the domain excluded in [196]. *Right panel*: comparison of the results of this work (thick blue line) with the results of the previous works [64, 194, 195] (purple lines) assuming mixing with electron flavor only. Notice that other works have adopted different values for the maximally admissible ^{4}He abundance when deriving their bounds: $Y_{p,\max} = 0.2696$ in [194, 195] and $Y_{p,\max} = 0.253$ in [64] as compared to $Y_{p,\max} = 0.2573$ in this work (see text for details).

the decay $N \to 3\pi^0 + \nu$ injects more energy into the EM plasma and diminishes ξ_ν, while $N \to 3\pi^\pm + \ell^\mp$ may inject more energy into neutrinos and compensate for this decrease. Therefore, both such channels should be accounted for.

We make use of `pyBBN` [64] to simulate their impact on the cosmological history, in

particular on N_{eff}. We examine the region of parameter space in which HNLs inject most of their energy into neutrinos, but where ΔN_{eff} is negative, illustrating the effect described in the previous section. Finally, we derive bounds from the CMB and comment on the possible role of HNLs in alleviating the Hubble tension.

3.3.4.1 Behavior of N_{eff}

HNLs inject (eventually) all of their energy either into the neutrino or electromagnetic plasma. The fraction of the HNL energy that is injected into each of these two sectors is mass-dependent and shows a significant shift to the EM plasma once HNLs can decay into neutral pions ($\sim$135 MeV), see Fig. 16.

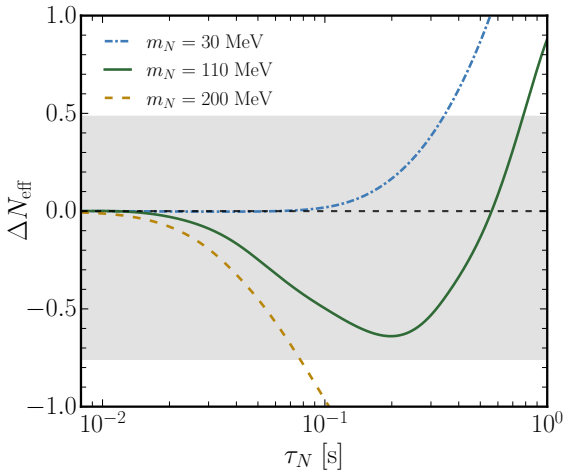

Figure 14: ΔN_{eff} as a function of HNL lifetime for a number of benchmark masses. Mixing with electron neutrinos only is considered here. The curves illustrate three cases of how HNLs can affect N_{eff}: *1)* they can decay mostly into neutrinos and simply increase N_{eff} (30 MeV curve), *2)* they can decay mostly into neutrinos and either decrease or increase N_{eff} depending on their lifetime (110 MeV curve), and *3)* they can decay mostly into EM particles and simply decrease N_{eff} (200 MeV curve). HNLs with masses $m_N \gtrsim 70$ MeV that decay mainly into neutrinos around neutrino decoupling, show an initial decrease of ΔN_{eff} as a result of the thermalization of the injected high-energy neutrinos. The grey band is the current sensitivity by Planck.

This plot shows that HNLs below the pion mass decay mainly into neutrinos and, therefore,

One would naively expect that in the mass range $m_N < m_\pi$ N_{eff} increases. However, we find that HNLs are able to decrease N_{eff} for masses already above $\sim$70 MeV, while for smaller masses an increase of N_{eff} is observed.

The origin of this sign change in ΔN_{eff} at $m_N \gtrsim 70\,\text{MeV}$ (rather than $m_N > m_\pi$ as one would guess from Fig. 16) lies in the energy transfer from the neutrino plasma to the electromagnetic plasma that is induced by the injected non-equilibrium neutrinos, as discussed earlier in Sec. 3.2.1. We run `pyBBN` simulations to examine in which region of parameter space this sign change happens[14]. We show ΔN_{eff} as a function of the HNL lifetime in Fig. 14 for a number of benchmark masses. The grey band in this figure indicates the current sensitivity by Planck. Included in this figure is an HNL of mass $110\,\text{MeV}$, which decreases N_{eff} for lifetimes below $\tau_N \lesssim 0.6\,\text{s}$ and increases N_{eff} for longer lifetimes. Such a lifetime ($\tau_N \sim 0.6\,\text{s}$) roughly corresponds to the time of neutrino decoupling, beyond which thermalization between the neutrino and EM plasma is not efficient anymore and the injected neutrinos remain in the neutrino sector. This exemplifies the ability of HNLs below the pion mass to diminish N_{eff}, even when neutrinos are on the verge of being completely decoupled. With the current sensitivity of Planck, however, this initial decrease of ΔN_{eff} for this mass falls within the error range and is thus not observable. Nevertheless, a number of upcoming and proposed CMB missions, such as the Simons Observatory [228] and CMB-S4 [229], could provide a determination of N_{eff} around the percent-level and probe this effect.

We depict the region of HNL parameter space where ΔN_{eff} changes sign in the top panel of Fig. 15. This is shown for the case of pure mixing with tau neutrinos only, as the parameter space where HNLs mix purely with electron and muon neutrinos is excluded in the lower mass range (where ΔN_{eff} can be positive) by BBN, the CMB and experimental searches [37, 64, 221]. In these latter two cases, ΔN_{eff} can only be negative in the unconstrained parameter space. This top panel shows that there is a large region of HNL parameter space, where these particles inject most of their energy into neutrinos and still decrease N_{eff}. The behavior of negative ΔN_{eff} continues for short-lived HNLs with masses $m_N > 1\,\text{GeV}$, since the neutrino energy increases with the HNL mass. On the other hand, for HNLs with lifetimes $\tau_N \gg 1\,\text{s}$, it depends on how much energy they inject into the neutrino plasma. Indeed, such HNLs decay long after neutrino decoupling, when non-equilibrium effects are not important anymore and the injected neutrinos remain in the neutrino sector. This means that the sign of ΔN_{eff} is simply determined by the value of ξ_ν. As a result, for masses where $\xi_\nu > 0.5$ (see Fig. 16) this would mean that eventually $\Delta N_{\text{eff}} > 0$ and vice versa (see Fig. 17 for an illustration).

We summarize `pyBBN` predictions for N_{eff} in the form of fitting functions for the three pure HNL mixing cases. This may provide a quick way to predict the impact of HNLs on several cosmological probes through the change in N_{eff}. They read:

[14] We note that `pyBBN` predicts a SM value for N_{eff} of 3.026, rather than 3.044. This is because the code does not include higher-order QED corrections that account for a $\Delta N_{\text{eff}} = 0.01$ increase [183–187], while the remaining is due to numerical inaccuracy. This, however, is only a minor difference and does not change any of the results presented in this work.

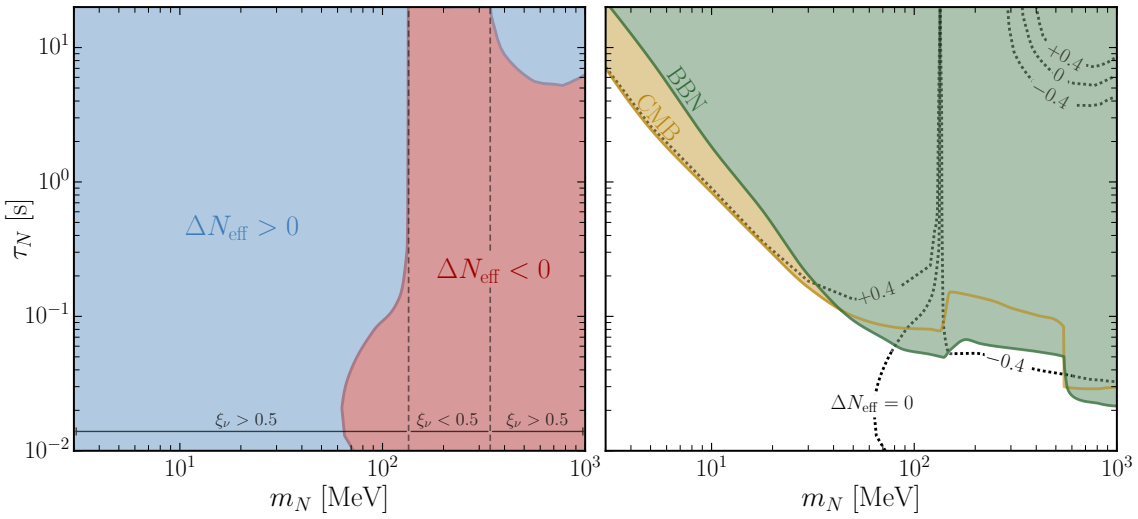

Figure 15: How HNLs change $\Delta N_{\rm eff}$ as a function of their mass and lifetime. Mixing with tau neutrinos only is considered here. *Left panel*: Regions of the HNL parameter space that predict an increase (blue) or decrease (red) of $N_{\rm eff}$ with respect to the SM value. The horizontal lines at the bottom of the plot indicate the mass ranges where HNLs inject most of their energy into neutrinos ($\xi_\nu > 0.5$) or the EM plasma ($\xi_\nu < 0.5$). In the former case, HNLs can still decrease $N_{\rm eff}$ as a result of the efficient transfer of energy from neutrinos to EM particles. *Right panel*: Regions of the HNL parameter space that are excluded by BBN abundance measurements (green) and CMB observations (yellow). The $\Delta N_{\rm eff} = \{0, \pm 0.4\}$ contours give an indication of by how much HNLs can change $N_{\rm eff}$ at the most in the unconstrained region. The BBN bound is from [64] and uses a central value for the primordial helium abundance of $Y_{\rm P} = 0.245$ [230] with an error of 4.35% (see [37] for a discussion on how this error is obtained). For masses higher than the eta-meson mass ($\sim$550 MeV), the meson effect from [37] is included in the analysis. The CMB constraint is obtained using the approach as detailed in [64] (see Sec. 3.3.5 for more details on the CMB bound). This panel also shows that there is only a relatively small unconstrained region of parameter space left that can increase $N_{\rm eff}$ and where HNLs could play a role in alleviating the Hubble tension.

$$\Delta N_{\rm eff}^{\rm Fit}\Big|_{\rm e-mixing} = -\frac{9.78\tau_N e^{5.72\tau_N}}{1+\frac{1.28\cdot 10^5}{m_N^{2.42}}} \tag{3.3.25}$$

$$\Delta N_{\rm eff}^{\rm Fit}\Big|_{\mu-\rm mixing} = -\frac{7.49\tau_N e^{12.1\tau_N}}{1+\frac{2.41\cdot 10^6}{m_N^{2.87}}} \tag{3.3.26}$$

$$\Delta N_{\rm eff}^{\rm Fit}\Big|_{\tau-\rm mixing} = -\frac{8.72\tau_N e^{13.9\tau_N}}{1+\frac{3.49\cdot 10^3}{m_N^{1.51}}}\,, \tag{3.3.27}$$

where m_N is the HNL mass in MeV and τ_N is the HNL lifetime in seconds. The

change in $N_{\rm eff}$ is with respect to the SM value of $N_{\rm eff}^{\rm SM} = 3.026$. The fitting functions are tested for masses $100\,\text{MeV} \leq m_N \leq 1\,\text{GeV}$ and lifetimes $0.02\,\text{s} \leq \tau_N \leq 0.05\,\text{s}$, and have a maximum deviation from the simulated data of roughly $\sim 3\%$.

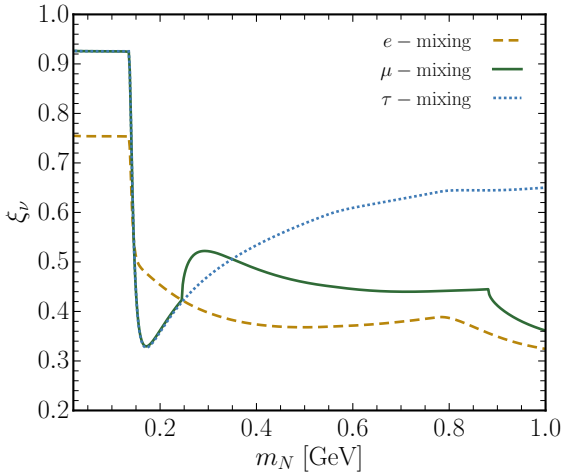

Figure 16: The fraction of HNL mass that is injected into the neutrino plasma. Contributions to this fraction from unstable HNL decay products (mesons and muons) are included and we assume that the kinetic energy of all created charged particles goes into the EM plasma. For $m_N \gtrsim 135\,\text{MeV}$, HNLs can decay into neutral pions, which in their turn decay into two photons. This causes the sudden decrease of ξ_ν around that mass. At higher masses, ξ_ν keeps increasing in the case of τ-mixing, which is due to the absence of HNL decays into charged mesons (such decays are possible in the other two mixing cases).

3.3.5 Bounds from CMB

The CMB anisotropies are mainly sensitive to $N_{\rm eff}$ through its impact on the damping tail [54, 230–232]. For example, a larger number of relativistic degrees of freedom causes a stronger suppression of the power spectrum at high multipoles, as temperature anisotropies below the scale of the photon diffusion length are more damped by the increased expansion rate. This effect is, however, degenerate if the primordial helium abundance $Y_{\rm P}$ is also considered as a free parameter [54]. $Y_{\rm P}$ is related to the number density of free electrons[15], $n_e \propto (1 - Y_{\rm P})$, which in its turn enters in the CMB damping scale. A larger $Y_{\rm P}$ leads to a lower electron density, a larger electron-photon interaction rate, a larger photon diffusion length and thus a stronger damping.

We extend the CMB constraint on HNLs for masses up to $1\,\text{GeV}$ using the same approach as detailed in [64, 233] and show the result in the bottom panel of Fig. 15. Also included in this panel are the contours where $\Delta N_{\rm eff} = \pm 0.4$, which give an indication of by how much HNLs can change $N_{\rm eff}$ at the most, given the current constraints imposed

[15]This relation between n_e and $Y_{\rm P}$ is obtained by imposing charge neutrality on the primordial plasma. Therefore, $Y_{\rm P}$ is allowed to change even if the total baryon density is fixed.

by BBN and the CMB. The CMB bound is only stronger than the BBN bound in the lower mass range, as this is where N_{eff} strongly increases. HNLs with short lifetimes and masses around $\mathcal{O}(10)\,\text{MeV}$ decouple while being non-relativistic and thus have a suppressed number density. They can therefore survive beyond the decoupling of SM weak reactions, without significantly affecting the primordial abundances. However, since the HNL energy density here falls off as $(\text{scale factor})^{-3}$, the HNLs could eventually dominate the total energy density of the Universe. As can be seen in Fig. 16, HNLs in this lower mass range inject most of their energy into neutrinos, which remains in the neutrino sector after neutrino decoupling. The result is then a significant increase in N_{eff}, which can be constrained with the CMB. On the other hand, for masses higher than $\sim$70 MeV, N_{eff} starts decreasing. This decrease is relatively small in magnitude, especially in the region that is not constrained by BBN, where $N_{\text{eff}} - N_{\text{eff}}^{\text{CMB}} \lesssim 0.4$. In addition, the error in the determination of Y_{P} by the CMB is larger than the one by BBN [199, 233]. These two properties make the CMB a weaker probe of HNLs in the higher mass range.

Currently, CMB is a weaker probe of HNLs than BBN in the mass range $m_N \gtrsim 40$ MeV. However, as mentioned before, future CMB experiments could improve upon this result.

3.3.6 Implications for the Hubble Parameter

An increase or decrease of N_{eff} subsequently also changes the Hubble parameter. As such, HNLs could play a role in alleviating the longstanding tension between local determinations of the current day Hubble rate H_0 and the one as inferred from the CMB[16] [235, 236]. The usual approach involves increasing N_{eff}, while keeping the angular scale of the sound horizon $\theta_{\text{s}} = r_{\text{s}}/D_{\text{A}}$ fixed, see e.g. [237–239]. Here, r_{s} is the comoving sound horizon and D_{A} is the comoving angular diameter distance to the surface of last scattering. Both of these quantities depend on the Hubble parameter:

$$r_{\text{s}}(z_*) = \int_{z_*}^{\infty} \frac{c_s(z)\mathrm{d}z}{H(z)} \tag{3.3.28}$$

$$D_{\text{A}}(z_*) = \int_{0}^{z_*} \frac{\mathrm{d}z}{H(z)}\,, \tag{3.3.29}$$

where z_* is the redshift of the last-scattering surface and $c_{\text{s}}(z)$ is the speed of sound of the baryon-photon fluid in the early Universe. The Hubble rate in Eq. (3.3.28) depends mainly on the radiation (photons *and* neutrinos) and matter energy densities, while the one in Eq. (3.3.29) is the late-time Hubble rate and depends mostly on the dark energy and

[16] This question has been considered before in [234]. Importantly, this study used the results of [194], where the assumption was made that any change in the primordial helium abundance is due to ΔN_{eff}. In contrast, here we find that neutrino spectral distortions are the driving power behind ΔN_{eff} for short-lived HNLs. As a consequence, the results presented in our work and in [234] are rather different.

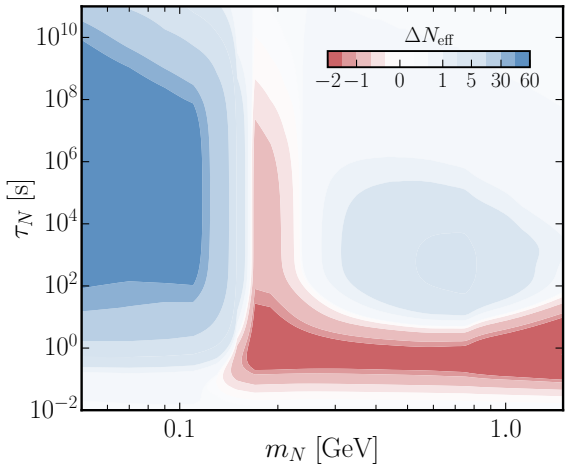

Figure 17: Semi-analytic estimate of $\Delta N_{\rm eff}$ as a function of HNL mass and lifetime in the case of pure tau mixing. This plot is obtained using the method described in Sec. 3.2.1 (and is therefore only accurate up to a factor $3-4$ for short lifetimes, when neutrinos are still in partial equilibrium). Nevertheless, it allows for a qualitative understanding of the behavior of $\Delta N_{\rm eff}$ at lifetimes larger than considered in the main analysis (Fig. 15). Importantly, for lifetimes well beyond the time of neutrino decoupling ($\mathcal{O}(1)$ s), non-equilibrium effects are absent and the sign of $\Delta N_{\rm eff}$ is thus completely determined by the fraction of HNL energy ξ_ν that is injected into the neutrino plasma, see Fig. 16. We see that HNLs with low masses and long lifetimes can still considerably affect $N_{\rm eff}$, while in the higher mass range $\Delta N_{\rm eff}$ tends to 0. This is because low-mass HNLs are more abundantly produced in this region of parameter space [196], where their mixing angles are relatively large.

matter energy densities. This means that increasing $N_{\rm eff}$ only results in a larger early-time Hubble rate and a smaller $r_{\rm s}$. In order to keep $\theta_{\rm s}$ fixed, the comoving angular diameter distance must satisfy $D_{\rm A} = r_{\rm s}/\theta_{\rm s}$, which then also decreases if $r_{\rm s}$ decreases. Looking at Eq. (3.3.29), such a decrease can be accomplished by increasing the dark energy density ω_Λ, or equivalently, H_0 (as $\Omega_\Lambda = 1 - \Omega_{\rm m}$). Since local measurements find a larger value of H_0 than the one inferred from the CMB within the Standard Model, this approach provides a way to reduce the Hubble tension.

This method, however, does not take into account the increased Silk damping induced by a larger $N_{\rm eff}$ [54, 230–232]. Therefore, a price must be paid when alleviating the Hubble tension in this way: An increase of $N_{\rm eff}$ leads to a larger disagreement with the CMB itself. Given our CMB constraint in Fig. 15, we see that HNLs can increase $N_{\rm eff}$ by at most $\Delta N_{\rm eff} \approx 0.4$. This gives us an indication of the extent to which unconstrained HNLs could increase H_0 and ameliorate the Hubble tension. We estimate the corresponding H_0 by running `Monte Python` [240, 241] with the Planck 2018 baseline TTTEEE+lowE analysis. Fixing the primordial helium abundance to[17] $Y_{\rm P} = 0.25$, we obtain[18] $H_0 =$

[17] This is approximately the value of $Y_{\rm P}$ along the $\Delta N_{\rm eff} = +0.4$ curve on the left in the bottom panel of Fig. 15.

[18] All errors in H_0 reported here are at 68% CL.

$70.5 \pm 0.7\,\mathrm{km\,s^{-1}Mpc^{-1}}$. This value can be compared to the one as obtained from, e.g., a distance ladder approach, which gives $H_0^{\mathrm{local}} = 73.0 \pm 1.4\,\mathrm{km\,s^{-1}Mpc^{-1}}$ [242]. Given the Hubble rate obtained within ΛCDM ($H_0 = 67.3 \pm 0.6\,\mathrm{km\,s^{-1}Mpc^{-1}}$ [199]), we see that HNLs which are not excluded by BBN, the CMB and terrestrial experiments can moderately alleviate the Hubble tension.

Appendix

3.A Thermal dependence of mixing angle of HNLs

3.A.1 Neutrino self-energy

Consider hot dense plasma with 4-velocity $u^\mu, u^2 = 1$, temperature $T \gg m_e$ and zero lepton asymmetry. Neutrinos in this plasma may interact elastically with electron-positron pairs:

$$\nu + e^\pm \to \nu + e^\pm \tag{3.A.1}$$

If the 4-momentum of the neutrino does not change in these processes, they contribute to the self-energy of neutrinos Σ, see Fig. 18. In explicit form, the self-energy is

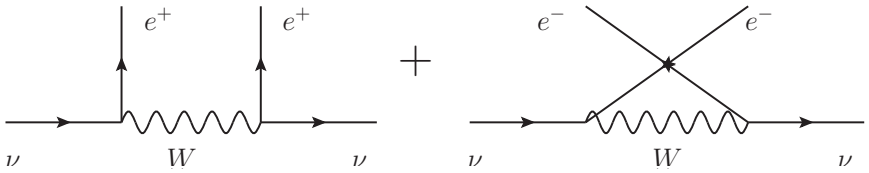

Figure 18: Diagrams of the contribution of the processes (3.A.1) to the self-energy of neutrinos.

$$\Sigma \sim 2 \int \frac{d^3\mathbf{k}}{(2\pi)^3} f_{\text{FD}}(k) \Sigma_k, \tag{3.A.2}$$

where

$$f_{\text{FD}}(k) = \frac{1}{\exp{(k \cdot u)} + 1} \tag{3.A.3}$$

is the distribution function of electrons and positrons (with u being the 4-velocity of the plasma), while Σ_k is (σ denotes the equality up to sign)

$$\Sigma_k \sim \frac{1}{2k} \frac{\sqrt{2} G_F}{2} \Big[\gamma^\mu (1-\gamma_5) u(k) D_{\mu\nu}(p-k) \bar{u}(k) \gamma^\nu (1-\gamma_5) - \\ - \gamma^\mu (1-\gamma_5) v(k) D_{\mu\nu}(p+k) \bar{v}(k) \gamma^\nu (1-\gamma_5) \Big], \tag{3.A.4}$$

with $D_{\mu\nu}$ being the W boson propagator

$$D_{\mu\nu}(p \pm k) = -\frac{g_{\mu\nu} - \frac{(p\pm k)_\mu (p\pm k)_\nu}{m_W^2}}{(p \pm k)^2 - m_W^2} \tag{3.A.5}$$

The minus sign in (3.A.4) appears because of the Pauli principle – the processes (3.A.1) differ only by exchanged in- and out- charged fermions lines.

Let us simplify estimates:

1. The self-energy (3.A.4) vanishes in the leading order on $G_F E^2$ (i.e., approximating the propagator by $-g_{\mu\nu}/m_W^{-2}$). We need to keep the next order corrections:

$$D_{\mu\nu}(p \pm k) = \frac{g_{\mu\nu}}{m_W^2} - \frac{1}{m_W^4}\left[g_{\mu\nu}(p \pm k)^2 + (p \pm k)_\mu (p \pm k)_\nu\right] + \dots \tag{3.A.6}$$

   The $(p \pm k)_\mu (p \pm k)_\nu / m_W^2$ terms in the numerator either are $\mathcal{O}(m_e/m_W)$ suppressed (this can be shown by acting them on electron-positron spinors $u(k)/v(k)$), or cancel when plugging in Eq. (3.A.4), so only the term $g_{\mu\nu}(p \pm k)^2/m_W^4$ matters. Averaging over incoming electron/positron polarizations, $v\bar{v}, u\bar{u} \to \not{k}/2$, we get

$$\begin{aligned}\Sigma_k &\sim \frac{\sqrt{2}G_F}{8km_W^2}[(p+k)^2 - (p-k)^2]\gamma^\mu(1-\gamma_5)\not{k}\gamma_\mu(1-\gamma_5) = \\ &= \frac{\sqrt{2}G_F(p\cdot k)}{2km_W^2}\gamma^\mu(1-\gamma_5)\not{k}\gamma_\mu(1-\gamma_5) \sim \frac{2\sqrt{2}G_F(p\cdot k)}{km_W^2}\not{k}(1-\gamma_5),\end{aligned} \tag{3.A.7}$$

   where in the last step we used the identity $\gamma^\mu(1-\gamma_5)\not{k}\gamma^\mu(1-\gamma_5) = -4\not{k}(1-\gamma_5)$.

2. Let us integrate Σ_k over the momenta of $e^\pm$. We have

$$\frac{1}{n_e}\int \frac{d^3\mathbf{k}}{(2\pi)^3}(p\cdot k)\frac{\not{k}}{k} f_{\text{FD}}(k) = \frac{p^\alpha \gamma^\beta}{n_e}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{k_\alpha k_\beta}{k} f_{\text{FD}}(k) = p^\alpha\gamma^\beta(Ag_{\alpha\beta} + Bu_\alpha u_\beta), \tag{3.A.8}$$

   where $n_e = 3\zeta(3)T^3/4\pi^2$ is the number density of electrons and positrons, the coefficients A, B can be obtained considering the integral at rest frame of the plasma ($u^\alpha = \delta^\alpha_0$):

$$B = -4A = \frac{4\langle E_e\rangle}{3} \tag{3.A.9}$$

   Therefore, we get

$$\Sigma = \frac{16\sqrt{2}\langle E_e\rangle n_e G_F}{3m_W^2}\left((p\cdot u)\not{u} - \frac{1}{4}\not{p}\right)(1-\gamma_5) \tag{3.A.10}$$

Restoring the overall sign and making a completely similar calculation that involves the

contribution of neutrinos themselves to the self-energy, we get

$$\boxed{\Sigma(p) = -\frac{16\sqrt{2}\langle E_e\rangle n_e G_F}{3m_W^2}\left[1+\frac{m_W^2}{2m_Z^2}\right]\left((p\cdot u)\not{u} - \frac{1}{4}\not{p}\right)(1-\gamma_5)} \tag{3.A.11}$$

This expression agrees with [243] (Eq. (12)) at the rest frame of the plasma, $u^\mu = (1,0,0,0)$ (see also Eq. (20) in [213]).[19]

At rest frame of the plasma, neglecting the neutrino mass, using the relation $\langle E_e\rangle = 7\pi^4 T/180\zeta(3)$, for the correction to the neutrino energy we have

$$\Delta E_\nu(p) = \frac{1}{2p}\text{Tr}[\Sigma u(p)\bar{u}(p)] = -\frac{14\sqrt{2}\pi^2 G_F T^4 p}{45m_W^2}\left(1+\frac{m_W^2}{2m_Z^2}\right) \tag{3.A.12}$$

This expression agrees with [213, 243].

The self-energy modifies the neutrino propagator: after the resummation we get

$$D_\nu(p) = \frac{1}{\not{p}}\sum_{n=0}^{\infty}\left(-\Sigma\frac{1}{\not{p}}\right)^n = \frac{1}{\not{p} + A\not{c}(1-\gamma_5)} \tag{3.A.13}$$

3.A.2 Derivation of $U_{\mathbf{m}}^2$

Now, consider the general matrix element $\mathcal{M}$ of the interaction of an HNL N with SM particles. It couples to the neutrino via the term $\mathcal{L}_{\text{mixing}} = m_N\theta\bar{N}\nu$, where m_N is the mass of the HNL and $\theta \ll 1$ is the mixing angle. Therefore, $\mathcal{M}$ takes the form

$$\mathcal{M} = \theta m_N \bar{N}(p) D_\nu(p)\gamma^\mu(1-\gamma_5)\ldots, \tag{3.A.14}$$

where $\ldots$ is the interaction dependent part and $\gamma^\mu(1-\gamma_5)$ comes from the neutrino vertex. Let us use the series representation of the neutrino propagator (3.A.13). With the help of the identity

$$\left[\not{c}(1-\gamma_5)\frac{\not{p}}{p^2}\right]^n\gamma^\mu(1-\gamma_5) = \left[2\not{c}\frac{\not{p}}{p^2}\right]^n\gamma^\mu(1-\gamma_5) \tag{3.A.15}$$

we get

$$\mathcal{M} = \theta m_N\bar{N}(p)\frac{1}{\not{p}+2A\not{c}}\gamma^\mu(1-\gamma_5)\cdots = \theta m_N\bar{N}(p)\frac{\not{p}+2A\not{c}}{p^2+4A(p\cdot c)+4A^2c^2}\gamma^\mu(1-\gamma_5)\ldots, \tag{3.A.16}$$

where in the last equality we multiplied the numerator and denominator of the propagator by $\not{p}+2A\not{c}$. Finally, using the dispersion relation for HNLs ($p^2 = m_N^2$), the Dirac equation $\bar{N}\not{p} = m_N\bar{N}$ and neglecting the $4A^2$ term in comparison to $4A$ (valid everywhere for

[19] The expression (3.A.11) is larger from Eq. (12) in [243] by a factor of two, but this is most likely due to a misprint, as Eqs. (21) from [213] and (13) from [243] require twice larger value.

$T \ll m_W$), we obtain

$$\boxed{\mathcal{M} = \theta_M \bar{N}(p) \left[1 + 2A \left(\frac{(p\cdot u)}{m_N}\not{u} - \frac{1}{4}\right)\right] \gamma^\mu (1-\gamma_5) \dots}, \tag{3.A.17}$$

where we introduced the effective mixing angle θ_M:

$$\theta_M = \frac{\theta}{1 + 4A\left(\frac{(p\cdot u)^2}{m_N^2} - \frac{1}{4}\right)} \approx \frac{\theta}{1 + 2.2\cdot 10^{-8}\left(\frac{T}{1\,\text{GeV}}\right)^4 \left(\gamma_N^2 - \frac{1}{4}\right)}, \tag{3.A.18}$$

with $\gamma_N = E_N/m_N$. *For practical purposes, the second term in the numerator,*

$$\hat{X} = 1 + 2A\left(\frac{(p\cdot u)}{m_N}\not{u} - \frac{1}{4}\right), \tag{3.A.19}$$

may be neglected, see a discussion below.

In the UR limit $\gamma_N \approx p_N/m_N \approx 3.15T/m_N \gg 1$, and we get

$$\boxed{\theta_M(T) \approx \frac{\theta}{1 + a\left(\frac{T}{1\,\text{GeV}}\right)^6 \left(\frac{1\,\text{GeV}}{m_N}\right)^2}}, \quad a = 2.2\cdot 10^{-7} \tag{3.A.20}$$

The value of a fully coincides with that from the literature (see, e.g., [196]).

Role of the numerator. Let us now consider the operator in the square brackets from Eq. (3.A.17):

$$\hat{X} = 1 + 2A\left(\frac{(p\cdot u)}{m_N}\not{u} - \frac{1}{4}\right) \tag{3.A.21}$$

Using $(p\cdot u)\not{u} \simeq E_N$ in the plasma rest frame and approximating $E_N \approx 3.15T$, we find that the second term of the operator cannot be neglected ($\gtrsim 1$) for temperatures

$$T \gtrsim 31\left(\frac{m_N}{1\,\text{GeV}}\right)^{1/5} \tag{3.A.22}$$

In contrast, the second term in the denominator of (3.A.20) becomes non-negligible at

$$T \gtrsim 12.9\left(\frac{m_N}{1\,\text{GeV}}\right)^{1/3} \tag{3.A.23}$$

So there is a temperature domain in which the θ_M is affected by the plasma effects, whereas the numerator can be neglected. However, the numerator prevents the matrix element from huge suppression at large temperatures: the asymptotics of the suppression of the product $\hat{X}\cdot\theta_M$ is m_N/E_N.

3.B Changes in $p \leftrightarrow n$ rates due to the presence of mesons

In this section, we provide details on our estimate of the effect of mesons on BBN.

Pions. The threshold-less processes with charged pions are

$$\pi^- + p \to n + \pi^0/\gamma, \quad \pi^+ + n \to p + \pi^0. \tag{3.B.1}$$

The cross-sections at threshold are [179]

$$\langle \sigma^{\pi^-}_{p\to n} v\rangle \approx 4.3 \cdot 10^{-23} F_c^\pi(T)\ \mathrm{m}^3/s, \quad \frac{\langle \sigma^{\pi^-}_{p\to n} v\rangle}{\langle \sigma^{\pi^+}_{n\to p} v\rangle} \approx 0.9\ F_c^\pi(T), \tag{3.B.2}$$

where F_c^h is the Sommerfeld enhancement of the cross-section due to presence of two oppositely charged particles in the in-state:

$$F_c^h = \frac{x}{1+e^{-x}}, \quad \text{where} \quad x = \frac{2\pi\alpha_{\text{EM}}}{v_e}, \tag{3.B.3}$$

where $v_e \approx \sqrt{\frac{T}{m_h}} + \sqrt{\frac{T}{m_p}}$ is the relative velocity between a nucleon and a meson. F_c is of order of one at $T \simeq 1$ MeV.

Kaons. The threshold-less $n \leftrightarrow p$ conversions driven by kaons are

$$\begin{aligned}
K^- + p &\to \Sigma^{\pm/0}/\Lambda + \pi^{\mp/0}/\pi^0 \to n + 2\pi,\\
K^- + n &\to \Sigma^{-/0}/\Lambda + \pi^{0/-}/\pi^- \to n + 2\pi,\\
\bar{K}^0_L + p &\to \Sigma^{0/+}/\Lambda + \pi^{+/0}/\pi^+ \to n + 2\pi,\\
\bar{K}^0_L + n &\to \Sigma^{\pm/0}/\Lambda + \pi^{\mp/0}/\pi^0 \to p + 2\pi,
\end{aligned} \tag{3.B.4}$$

where Λ, Σ are the lightest strange hadronic resonances [178].

Their effect is similar to the one of pions, but with small differences: *(i)* cross-sections of above reactions are higher than the cross-sections of (3.1.2)[20], *(ii)* there is no isotopic symmetry - K^+ mesons do not contribute to $p \leftrightarrow n$ conversion, since there are no threshold-less processes $n + K^+ \to p + X$. Indeed, the process $n + K^+ \to p + K^0$ has the threshold $Q \approx 2.8$ MeV, while the threshold-less processes going through s-quark resonances, similar to (3.B.4), would require resonances with negative strangeness and positive baryon number, that do not exist, *(iii)* neutral kaons do not lose the energy before decaying (however, we follow [178] and approximate the cross-sections by threshold values).

The threshold cross-sections are

$$\langle \sigma^{K^-}_{p\to n} v\rangle \approx 9.6 \cdot 10^{-22} F_c^K(T)\ \mathrm{m}^3/s, \quad \frac{\langle \sigma^{K^-}_{p\to n} v\rangle}{\langle \sigma^{K^-}_{n\to p} v\rangle} \approx 2.46\ F_c^K(T), \tag{3.B.5}$$

[20]The reason is that these reactions have higher available phase space and go through hadronic resonances.

$$\langle\sigma^{K^0}_{p\to n}v\rangle \approx 1.95\cdot 10^{-22}\ \mathrm{m}^3/\mathrm{s}, \quad \frac{\langle\sigma^{K^0_L}_{p\to n}v\rangle}{\langle\sigma^{K^0_L}_{n\to p}v\rangle} \approx 0.41. \tag{3.B.6}$$

Conversion probabilities. A probability for a meson h to convert $p \leftrightarrow n$ before decaying is given by

$$P^h_{\text{conv}} \approx \frac{\langle\sigma^h_{p\leftrightarrow n}v\rangle n_B}{\Gamma^h_{\text{decay}}}, \tag{3.B.7}$$

where Γ^h_{decay} is the decay width and n_B is the baryon number density. The decay widths of mesons are [52]

$$\Gamma^{\pi^\pm}_{\text{decay}} \approx 3.8\cdot 10^7\ \mathrm{s}^{-1}, \quad \Gamma^{K^-}_{\text{decay}} \approx 8.3\cdot 10^7\ \mathrm{s}^{-1}, \quad \Gamma^{K^0_L}_{\text{decay}} \approx 2\cdot 10^7\ \mathrm{s}^{-1} \tag{3.B.8}$$

Using (3.B.2), (3.B.5), (3.B.8), for the $p \to n$ conversion probabilities we obtain

$$\begin{aligned} &P^{\pi^-}_{\text{conv}}(T) \approx 2.5\cdot 10^{-2}\left(\frac{T}{1\ \text{MeV}}\right)^3, \quad P^{K^-}_{\text{conv}}(T) \approx 2.8\cdot 10^{-1}\left(\frac{T}{1\ \text{MeV}}\right)^3, \\ &P^{K^0_L}_{\text{conv}}(T) \approx 1.6\cdot 10^{-1}\left(\frac{T}{1\ \text{MeV}}\right)^3 \end{aligned} \tag{3.B.9}$$

The largeness of the probabilities is caused by the fact that the decay of mesons proceeds through weak interactions, while the $p \leftrightarrow n$ conversion is mediated by strong interactions. In particular, at $T \gtrsim 2$ MeV kaons participate in the conversion faster than they decay.

3.B.1 Numeric study

To verify the analytic estimate (3.3.24), we numerically solve equation for the neutron abundance X_n, where we include both weak conversion $p \leftrightarrow n$ processes and the meson driven processes (3.B.1)-(3.B.4). The system of equations has the form

$$\begin{cases} \frac{X_n}{dt} = \left(\frac{dX_n}{dt}\right)_{\text{SM}} + \left(\frac{dX_n}{dt}\right)_{\pi} + \left(\frac{dX_n}{dt}\right)_{K^-} + \left(\frac{dX_n}{dt}\right)_{K^0_L}, \\ \frac{dn_{\pi^-}}{dt} = n_N\frac{\text{Br}_{N\to\pi^-}}{\tau_N} - \Gamma^{\pi^-}_{\text{decay}}n_{\pi^-} - \langle\sigma^{\pi^-}_{p\to n}v\rangle(1-X_n)n_Bn_{\pi^-}, \\ \frac{dn_{\pi^+}}{dt} = n_N\frac{\text{Br}_{N\to\pi^+}}{\tau_N} - \Gamma^{\pi^+}_{\text{decay}}n_{\pi^+} - \langle\sigma^{\pi^+}_{n\to p}v\rangle X_nn_Bn_{\pi^+}, \\ \frac{dn_{K^-}}{dt} = n_N\frac{\text{Br}_{N\to K^-}}{\tau_N} - \Gamma^{K^-}_{\text{decay}}n_{K^-} - \langle\sigma^{K^-}_{p\to n}v\rangle(1-X_n)n_Bn_{K^-} - \langle\sigma^{K^-}_{n\to p}v\rangle X_nn_Bn_{K^-}, \\ \frac{dn_{K^0_L}}{dt} = n_N\frac{\text{Br}_{N\to K^0_L}}{\tau_N} - \Gamma^{K^0_L}_{\text{decay}}n_{K^0_L} - \langle\sigma^{K^0_L}_{p\to n}v\rangle(1-X_n)n_Bn_{K^0_L} - \langle\sigma^{K^0_L}_{n\to p}v\rangle X_nn_Bn_{K^0_L} \end{cases} \tag{3.B.10}$$

Here the quantities

$$\left(\frac{dX_n}{dt}\right)_\pi = (1-X_n)n_{\pi^-}\langle\sigma^{\pi^-}_{p\to n}v\rangle - X_n n_{\pi^+}\langle\sigma^{\pi^+}_{n\to p}v\rangle,$$

and (3.B.11)

$$\left(\frac{dX_n}{dt}\right)_K = (1-X_n)n_K\langle\sigma^{K}_{p\to n}v\rangle - X_n n_K\langle\sigma^{K}_{n\to p}v\rangle$$

are the rates of change of X_n due to different mesons ($K = K^-/K^0_L$); n_B is the baryon number density $n_B = \eta_B n_\gamma$. In equations for the number density of mesons n_h, the first term comes from HNLs, the second due to decays of mesons and the last term is due to $p \leftrightarrow n$ conversion. The time-temperature relation and the scale factor dynamics are provided by the solution of Eq. (3.3.19), and the HNL number density may be obtained using Eq. (3.3.17).

During times $t_{\rm eq} \simeq (\Gamma^h_{\rm decay})^{-1} \sim 10^{-8}$ s, which are small in comparison to any other time scale in the system, the solution for n_h reaches the dynamical equilibrium:

$$n_{\pi^-} = \frac{n_N \cdot {\rm Br}_{N\to\pi^-}}{\tau_N(\Gamma^{\pi^-}_{\rm decay} + \langle\sigma^{\pi^-}_{p\to n}v\rangle(1-X_n)n_B)}, \quad n_{\pi^+} = \frac{n_N \cdot {\rm Br}_{N\to\pi^+}}{\tau_N(\Gamma^{\pi^+}_{\rm decay} + \langle\sigma^{\pi^+}_{n\to p}v\rangle(1-X_n)n_B)}, \tag{3.B.12}$$

$$n_K = \frac{n_N \cdot {\rm Br}_{N\to K}}{\tau_N(\Gamma^{K}_{\rm decay} + \langle\sigma^{K}_{p\to n}v\rangle(1-X_n)n_B + \langle\sigma^{K}_{n\to p}v\rangle X_n n_B)}, \tag{3.B.13}$$

where $K = K^-/K^0_L$.

Therefore, we solve a single equation

$$\frac{X_n}{dt} = \left(\frac{dX_n}{dt}\right)_{\rm SM} + \left(\frac{dX_n}{dt}\right)_\pi + \left(\frac{dX_n}{dt}\right)_{K^-} + \left(\frac{dX_n}{dt}\right)_{K^0_L}. \tag{3.B.14}$$

where we use meson number densities given by Eqs. (3.B.12) and (3.B.13) in the meson-driven conversion rates (3.B.11). The results are shown in Fig. 19. Our main result is the right panel of Fig. 19 – it shows that the value $T_0^{\rm min} \simeq 1.50\,{\rm MeV}$ and that its variation as a function of the HNL mass is within $\pm 1\%$.

With the help of Eqs. (3.B.2), (3.B.5), we obtain the value of the neutron abundance driven solely by a given meson h. As long as $T \gtrsim T_0$ (see Eq. (3.1.9) and left panel of Fig. 19), the weak interaction processes may be completely neglected, and the resulting X_n are given by

$$X_n^{\pi^\pm} = \frac{\langle\sigma^{\pi^-}_{p\to n}v\rangle \cdot n_{\pi^-}}{\langle\sigma^{\pi^-}_{p\to n}v\rangle \cdot n_{\pi^-} + \langle\sigma^{\pi^+}_{n\to p}v\rangle \cdot n_{\pi^+}} \approx \frac{0.9F_c^\pi(T)}{1+0.9F_c^\pi(T)},$$

$$X_n^{K^-} \approx \frac{2.46F_c^K}{2.46F_c^K+1}, \quad X_n^{K^0_L} \approx 0.32 \tag{3.B.15}$$

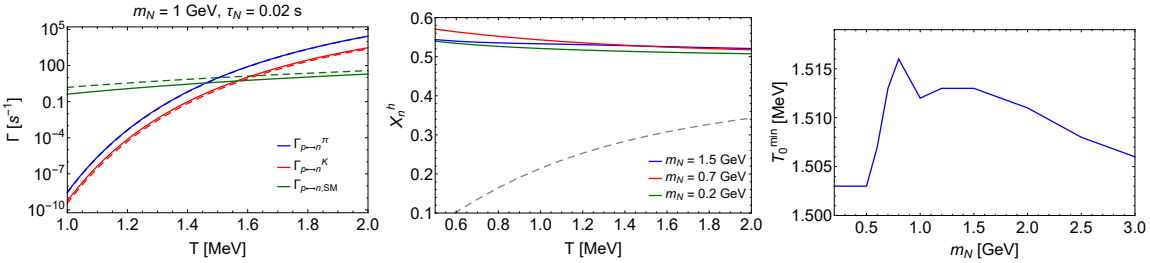

Figure 19: *Left panel*: the behavior of the $p \to n$ (solid lines) and $n \to p$ (dashed lines) conversion rates in the case of pion and kaon driven conversions and SBBN. We consider HNLs mixing with e flavor, mass $m_N = 1$ GeV and lifetime $\tau_N = 0.02$ s as an example. *Middle panel*: the temperature dependence of the neutron abundance X_n assuming that its evolution is completely dominated by the meson driven $p \leftrightarrow n$ conversions. We consider HNLs mixing with e flavor and different masses: $m_N = 200$ MeV (only pions are present), $m_N = 700$ MeV (pions and charged kaons are present), $m_N = 1.5$ GeV (pions, charged and neutral kaons are present). The dashed gray line denotes the value of the neutron abundance at equilibrium in SBBN. *Right panel*: the HNL mass dependence of the temperature $T_0^{\min}$.

The values of $X_n^{\pi^-/K^-}$ grow with the decrease of the temperature due to the growth of the Coulomb factor F_c, which enhances the rate of the $p \to n$ process.

The quantities (3.B.15) provide us the qualitative estimate of the value of X_n in presence of different mesons, Fig. 19. Below the kaon production threshold, $X_n^h = X_n^{\pi^\pm}$. At larger masses, in order to find X_n^h we need to set the whole right hand-side of Eq. (3.B.10) to zero. Below the K_L^0 production threshold (which occurs at $m_N = m_\phi$), the value of X_n^h grows, since charged kaons tend X_n to higher values than $X_n^{\pi^-}$. Above the neutral kaon production threshold, the ratio $\mathrm{Br}_{N\to K^-}/\mathrm{Br}_{N\to\pi^-}$ increases (Fig. 10) and X_n^h grows further. However, kaons K_L^0, that are present in small amounts, somewhat diminish this growth.

The value of $X_n^h(m_N)$ provide us the mass dependence of $T_0^{\min}(m_N)$, which is the smallest temperature allowed by observations (c.f. Fig. 2). We show it in Fig. 19 (right panel).

Let us now comment on the approximations of this approach. If HNLs disappear from the plasma before neutrinos froze out, the evolution of the neutron abundance and subsequent nuclear reactions proceed exactly as in SBBN case (albeit with modified initial value of X_n at $T = T_0^{\min}$).

Indeed, the onset of nuclear reactions is determined by the dynamical balance between reactions of deuterium synthesis and dissociation. This balance depends on the value of η_B. The latter gets diluted by the factor ζ due to decays of HNLs, see Section 3.3.1. However, we fix η_B at the beginning of nuclear reactions to be the same as measured by CMB. This of course means that η_B has been ζ^{-1} times higher before decays of HNLs, but no observables can probe the value of η_B in this epoch.

Another ingredient that affects dynamics of nuclear reactions is the time-temperature relation, traditionally encoded in the value of N_{eff}. If HNLs have $\tau_N \simeq 0.02$ s, neutrinos

are in equilibrium during the decay of the most of HNLs, and therefore they do not change neither N_{eff} nor weak $p \leftrightarrow n$ conversion, see detailed analysis in [170]. As a result, the evolution of primordial plasma below T_0^{min} is governed by the SBBN equations, and our prediction of the upper bound on the allowed HNL lifetimes is conservative. HNLs with larger lifetimes do change N_{eff} and rates. However, the net effect of this impact is an increase of the ^{4}He abundance [64, 170], and therefore the predictions of our approach in the increase of the ^{4}He abundance, which does not include changes in these quantities, are conservative.

3.B.2 Numeric approach for long-lived HNLs

The total system of equations for HNLs, mesons, SM plasma and nuclei reads

$$\begin{cases} n_N = \left(\frac{a_{N,\text{dec}}}{a}\right)^3 n_{N,\text{dec}} \cdot e^{-t/\tau_N}, \\ \dot{a}(t) = a(t) \cdot H(t), \\ \frac{dT_{\text{EM}}}{dt} + HT_{\text{EM}} = \Gamma_{\text{EM}\leftrightarrow\nu} \frac{\rho_{\text{EM}}}{d\rho_{\text{EM}}/dT_{\text{EM}}} + \frac{\rho_N \epsilon_{\text{EM}}}{\tau_N}, \\ \frac{dT_\nu}{dt} + HT_\nu = -\Gamma_{\text{EM}\leftrightarrow\nu} \frac{\rho_\nu}{d\rho_\nu/dT_\nu} + \frac{\rho_N \epsilon_\nu}{\tau_N}, \\ \dot{X}_i = \sum_{j,k} N_i \left(\Gamma_{j\to ki} \prod_j \frac{Y_j^{N_j}}{N_j!} - \Gamma_{ki\to j} \prod_k \frac{Y_k^{N_k}}{N_k!} \right) \end{cases} \tag{3.B.16}$$

Here, $X_i \equiv n_i/n_B$, N_i denotes the stoichiometric coefficient, $j \to kl$ is the shortland notation for

$$j_1 + \cdots + j_p \to i + k_1 + \cdots + k_q, \tag{3.B.17}$$

and $\prod_k \frac{Y_k^{N_k}}{N_k!}$ is the shortland notation for

$$\prod_k \frac{Y_k^{N_k}}{N_k!} \equiv \frac{Y_{k_1}^{N_{k_1}} \dots Y_{k_q}^{N_{k_q}}}{N_{k_1}! \dots N_{k_q}!} \tag{3.B.18}$$

$\Gamma_{j\to ik}$ are the reaction rates of SBBN reactions governed the evolution of $d, t,^3$He,^{4}He,^{7}Be,^{7}Li, as well as weak $p \leftrightarrow n$ rates from from [48], and meson-driven dissociation rates, which we use from [182].[21] The number density of mesons evolve due to Eqs. (3.B.10), where in addition to $p \leftrightarrow n$ rates there are now also nuclear dissociation rates. Our results for nuclear abundances in SBBN are in perfect agreement with predictions from [48].

We neglect the change of weak SM rates, since in presence of long-lived HNLs with $\tau_N \gg 1$ s they do not change at temperatures $T \simeq \mathcal{O}(1\,\text{MeV})$ at which weak interaction processes are important.

[21] We of course do not include the inverse reactions in which mesons occur, since these reactions are endotermic and practically impossible.

3.C Temperature Evolution Equations

Here we provide the relevant equations for the time evolution of the neutrino and photon temperatures in the presence of decaying FIPs. Assuming a Fermi-Dirac distribution for neutrinos, the equations read [184, 198]:

$$\frac{dT_\nu}{dt} + 4HT_\nu = \frac{(1-\xi_{\text{EM,eff}})\frac{\rho_{\text{FIP}}}{\tau_{\text{FIP}}} + \Gamma_{\nu\leftrightarrow\text{EM}}(T_\nu, T_{\text{EM}})}{d\rho_\nu/dT_\nu} \tag{3.C.1}$$

$$\frac{dT_{\text{EM}}}{dt} + \frac{(4H\rho_\gamma + 3H(\rho_e + p_e))}{d\rho_e/dT + d\rho_\gamma/dT} = \frac{\xi_{\text{EM,eff}}\frac{\rho_{\text{FIP}}}{\tau_{\text{FIP}}} - \Gamma_{\nu\leftrightarrow\text{EM}}(T_\nu, T_{\text{EM}})}{d\rho_e/dT + d\rho_\gamma/dT} \tag{3.C.2}$$

$$\frac{d\rho_{\text{FIP}}}{dt} + 3H\rho_{\text{FIP}} = -\frac{\rho_{\text{FIP}}}{\tau_{\text{FIP}}}\,, \tag{3.C.3}$$

where $\xi_{\text{EM,eff}}$ is given in Eq. (3.2.6), ρ_i is the energy density of particle i, τ_{FIP} is the FIP lifetime and $\Gamma_{\nu\leftrightarrow\text{EM}}(T_\nu, T_{\text{EM}}) = \left(\Gamma_{\nu_e\leftrightarrow\text{EM}} + 2\Gamma_{\nu_\mu\leftrightarrow\text{EM}}\right)/3$ is the energy density exchange rate averaged over neutrino flavours, given by Eqs. (2.12a) and (2.12b) in [198].

3.D Comment on "Massive sterile neutrinos in the early universe: From thermal decoupling to cosmological constraints" by Mastrototaro et al.

After our work was submitted, the paper [197] appeared that studies the impact of HNLs with masses $m_N < m_\pi$ on N_{eff}. The authors of this work used numerical simulations in order to obtain N_{eff} and disagree with our conclusion that N_{eff} can decrease even if most of the HNL energy is injected into neutrinos. They have presented an analytic argument in their Appendix C which aims to demonstrate that our conclusion on N_{eff} is wrong. They start with a toy model in Eq. (C.1) that describes the evolution of the distribution function of neutrinos f_ν:

$$x\partial_x f_\nu(E_\nu, x) = \frac{1}{H}\left[S(x, E_\nu) + \varsigma^2 G_F^2 T^4 E_\nu (f_{\text{eq}} - f_\nu)\right]\,, \tag{3.D.1}$$

where $x = ma$ (with a the scale factor and $m = 1\,\text{MeV}$), H is the Hubble rate, ς is a constant and $S(x) > 0$ is the source term from decays of HNLs. The second term in the brackets describes the interactions between neutrinos and EM particles, where f_{eq} is the equilibrium distribution function resulting from the interaction dynamics of neutrinos and EM particles in the presence of HNLs.

Their argument as to why $N_{\rm eff}$ cannot decrease goes as follows: as far as the source injecting rate $S(x, E_\nu)$ and the collision rate $G_F^2 T^4 E_\nu$ are much higher than the Hubble rate, the solution of Eq. (3.D.1) may be given in terms of the quasi-static solution:

$$f_\nu \approx f_{\rm eq} + \frac{S}{G_F^2 T^4 E_\nu} \,. \tag{3.D.2}$$

In the limiting case $S \ll G_F^2 T^4 E_\nu$, the solution is just $f_\nu = f_{\rm eq}$, while in the opposite case $f_\nu \gg f_{\rm eq}$. The authors conclude that in any case $f_\nu \geq f_{\rm eq}$ and thus $\Delta N_{\rm eff} \geq 0$. However, while this argument may be applicable at very early times when neutrinos are in perfect equilibrium, it is no longer valid at temperatures $T = \mathcal{O}(1\ \text{MeV})$, when they start to decouple. During the decoupling process, the dynamics of the equilibration between neutrinos and EM particles, i.e., the energy transfer between the two sectors, becomes very important and is not captured by Eq. (3.D.1).

We reiterate our argument as to why $N_{\rm eff}$ can decrease when FIPs inject most of their energy into neutrinos, but now from the point of view of the neutrino distribution function (see also the right panel of Fig. 4 and the surrounding text for a similar discussion). Before the decay of the FIP, the neutrino distribution function is the same as the equilibrium distribution, $f_\nu = f_{\rm eq}$. Right after the decay of the FIP, the neutrino distribution at high energies becomes $f_\nu > f_{\rm eq}$, while at low energies it is still $f_\nu = f_{\rm eq}$. During the thermalisation, high-energy neutrinos interact with both low-energy neutrinos and EM particles. In this process, the temperature of the equilibrium distribution function $f_{\rm eq}$ increases. Now, neutrinos in the high-energy tail of f_ν interact efficiently, see Eq. (3.2.5), and $f_\nu \longrightarrow f_{\rm eq}$ for such neutrinos. But at low energies, neutrinos do not interact efficiently anymore to catch up with the increase of $f_{\rm eq}$, which eventually leads to $f_\nu < f_{\rm eq}$ in this energy range. Given that these low-energy neutrinos contribute the most to $N_{\rm eff}$, it means that $\Delta N_{\rm eff}$ can become negative.

Bibliography

[1] ALEPH, DELPHI, L3, OPAL, SLD, LEP ELECTROWEAK WORKING GROUP, SLD ELECTROWEAK GROUP, SLD HEAVY FLAVOUR GROUP collaboration, *Precision electroweak measurements on the Z resonance*, *Phys. Rept.* **427** (2006) 257 [hep-ex/0509008].

[2] S. Bifani, S. Descotes-Genon, A. Romero Vidal and M.-H. Schune, *Review of Lepton Universality tests in B decays*, *J. Phys. G* **46** (2019) 023001 [1809.06229].

[3] ATLAS collaboration, *Precision measurement and interpretation of inclusive W^+ , W^- and Z/γ^* production cross sections with the ATLAS detector*, *Eur. Phys. J. C* **77** (2017) 367 [1612.03016].

[4] ATLAS collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys. Lett. B* **716** (2012) 1 [1207.7214].

[5] CMS collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Phys. Lett. B* **716** (2012) 30 [1207.7235].

[6] ATLAS collaboration, *Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton–proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment*, .

[7] G. Krnjaic, *Probing Light Thermal Dark-Matter With a Higgs Portal Mediator*, *Phys. Rev. D* **94** (2016) 073009 [1512.04119].

[8] F. Bezrukov and D. Gorbunov, *Light inflaton Hunter's Guide*, *JHEP* **05** (2010) 010 [0912.0390].

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

[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].

[11] ATLAS collaboration, *Search for displaced vertices of oppositely charged leptons from decays of long-lived particles in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Phys. Lett. B* **801** (2020) 135114 [1907.10037].

[12] CMS collaboration, *Search for long-lived particles decaying to jets with displaced vertices in proton-proton collisions at $\sqrt{s} = 13$ TeV*, [2104.13474].

[13] ATLAS collaboration, *Search for long-lived, massive particles in events with a displaced vertex and a muon with large impact parameter in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Rev. D* **102** (2020) 032006 [2003.11956].

[14] CMS collaboration, *Search for heavy neutral leptons in events with three charged leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. Lett.* **120** (2018) 221801 [1802.02965].

[15] CMS collaboration, *Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JHEP* **01** (2019) 122 [1806.10905].

[16] LHCb collaboration, *Search for $A' \to \mu^+\mu^-$ Decays*, *Phys. Rev. Lett.* **124** (2020) 041801 [1910.06926].

[17] FASER collaboration, *Technical Proposal for FASER: ForwArd Search ExpeRiment at the LHC*, [1812.09139].

[18] MATHUSLA collaboration, *Explore the lifetime frontier with MATHUSLA*, *JINST* **15** (2020) C06026 [1901.04040].

[19] G. Aielli et al., *Expression of interest for the CODEX-b detector*, *Eur. Phys. J. C* **80** (2020) 1177 [1911.00481].

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

[21] DUNE collaboration, *Long-baseline neutrino oscillation physics potential of the DUNE experiment*, *Eur. Phys. J. C* **80** (2020) 978 [2006.16043].

[22] A. Berlin, S. Gori, P. Schuster and N. Toro, *Dark Sectors at the Fermilab SeaQuest Experiment*, *Phys. Rev. D* **98** (2018) 035011 [1804.00661].

[23] LHCb collaboration, *Test of lepton universality in beauty-quark decays*, [2103.11769].

[24] BaBar collaboration, *The BaBar detector*, *Nucl. Instrum. Meth. A* **479** (2002) 1 [hep-ex/0105044].

[25] Belle-II collaboration, *Belle II Technical Design Report*, [1011.0352].

[26] NA62 collaboration, *The Beam and detector of the NA62 experiment at CERN*, *JINST* **12** (2017) P05025 [1703.08501].

[27] S. Andreas et al., *Proposal for an Experiment to Search for Light Dark Matter at the SPS*, [1312.3309].

[28] S. N. Gninenko, *Search for MeV dark photons in a light-shining-through-walls experiment at CERN*, *Phys. Rev. D* **89** (2014) 075008 [1308.6521].

[29] ATLAS collaboration, *Search for new phenomena in events with an energetic jet and missing transverse momentum in pp collisions at $\sqrt{s}$ =13 TeV with the ATLAS detector*, *Phys. Rev. D* **103** (2021) 112006 [2102.10874].

[30] ATLAS collaboration, *Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector*, *JHEP* **01** (2018) 126 [1711.03301].

[31] CMS collaboration, *Searches for invisible decays of the Higgs boson in pp collisions at* $\sqrt{s}$ *= 7, 8, and 13 TeV*, *JHEP* **02** (2017) 135 [1610.09218].

[32] SND@LHC collaboration, *SND@LHC - Scattering and Neutrino Detector at the LHC*, Scientific Committee Paper CERN-LHCC-2021-003; LHCC-P-016, https://cds.cern.ch/record/2750060, Jan, 2021.

[33] FASER collaboration, *Technical Proposal: FASERnu*, [2001.03073].

[34] CRESST collaboration, *First results from the CRESST-III low-mass dark matter program*, *Phys. Rev. D* **100** (2019) 102002 [1904.00498].

[35] DARKSIDE collaboration, *Low-Mass Dark Matter Search with the DarkSide-50 Experiment*, *Phys. Rev. Lett.* **121** (2018) 081307 [1802.06994].

[36] P. Agrawal et al., *Feebly-Interacting Particles:FIPs 2020 Workshop Report*, [2102.12143].

[37] A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy and V. Syvolap, *Improved BBN constraints on Heavy Neutral Leptons*, [2008.00749].

[38] I. Boiarska, K. Bondarenko, A. Boyarsky, V. Gorkavenko, M. Ovchynnikov and A. Sokolenko, *Phenomenology of GeV-scale scalar portal*, *JHEP* **11** (2019) 162 [1904.10447].

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

[40] M. Cacciari, M. Greco and P. Nason, *The P(T) spectrum in heavy flavor hadroproduction*, *JHEP* **05** (1998) 007 [hep-ph/9803400].

[41] M. Cacciari, S. Frixione and P. Nason, *The p(T) spectrum in heavy flavor photoproduction*, *JHEP* **03** (2001) 006 [hep-ph/0102134].

[42] M. Cacciari, S. Frixione, N. Houdeau, M. L. Mangano, P. Nason and G. Ridolfi, *Theoretical predictions for charm and bottom production at the LHC*, *JHEP* **10** (2012) 137 [1205.6344].

[43] M. Cacciari, M. L. Mangano and P. Nason, *Gluon PDF constraints from the ratio of forward heavy-quark production at the LHC at* $\sqrt{S} = 7$ *and 13 TeV*, *Eur. Phys. J. C* **75** (2015) 610 [1507.06197].

[44] SHIP collaboration, H. Dijkstra and T. Ruf, *Heavy Flavour Cascade Production in a Beam Dump*, Dec, 2015.

[45] C. Pitrou, A. Coc, J.-P. Uzan and E. Vangioni, *Precision big bang nucleosynthesis with improved Helium-4 predictions*, *Phys. Rept.* **754** (2018) 1 [1801.08023].

[46] T. M. Bania, R. T. Rood and D. S. Balser, *The cosmological density of baryons from observations of 3He+ in the Milky Way*, *Nature* **415** (2002) 54.

[47] L. Sbordone et al., *The metal-poor end of the Spite plateau. 1: Stellar parameters, metallicities and lithium abundances*, *Astron. Astrophys.* **522** (2010) A26 [1003.4510].

[48] C. Pitrou, A. Coc, J.-P. Uzan and E. Vangioni, *A new tension in the cosmological model from primordial deuterium?*, *Mon. Not. Roy. Astron. Soc.* **502** (2021) 2474 [2011.11320].

[49] A. D. Dolgov, *Neutrinos in cosmology*, *Phys. Rept.* **370** (2002) 333 [hep-ph/0202122].

[50] D. J. Fixsen, *The Temperature of the Cosmic Microwave Background*, *Astrophys. J.* **707** (2009) 916 [0911.1955].

[51] PLANCK collaboration, *Planck 2018 results. V. CMB power spectra and likelihoods*, *Astron. Astrophys.* **641** (2020) A5 [1907.12875].

[52] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *Phys. Rev. D* **98** (2018) 030001.

[53] J. Lesgourgues, G. Mangano, G. Miele and S. Pastor, *Neutrino Cosmology*. Cambridge University Press, 2, 2013.

[54] Z. Hou, R. Keisler, L. Knox, M. Millea and C. Reichardt, *How Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail*, *Phys. Rev. D* **87** (2013) 083008 [1104.2333].

[55] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[56] K. Bondarenko, A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy and L. Shchutska, *Probing new physics with displaced vertices: muon tracker at CMS*, *Phys. Rev. D* **100** (2019) 075015 [1903.11918].

[57] A. Boyarsky, O. Mikulenko, M. Ovchynnikov and L. Shchutska, *Searches for new physics at SND@LHC*, [2104.09688].

[58] I. Boiarska, A. Boyarsky, O. Mikulenko and M. Ovchynnikov, *Blast from the past: constraints from the CHARM experiment on Heavy Neutral Leptons with tau mixing*, [2107.14685].

[59] NOMAD collaboration, *Search for heavy neutrinos mixing with tau neutrinos*, *Phys. Lett. B* **506** (2001) 27 [hep-ex/0101041].

[60] ARGONEUT collaboration, *New Constraints on Tau-Coupled Heavy Neutral Leptons with Masses mN=280–970 MeV*, *Phys. Rev. Lett.* **127** (2021) 121801 [2106.13684].

[61] T2K collaboration, *Search for heavy neutrinos with the T2K near detector ND280*, *Phys. Rev. D* **100** (2019) 052006 [1902.07598].

[62] CHARM collaboration, *A Search for Decays of Heavy Neutrinos in the Mass Range 0.5-GeV to 2.8-GeV*, *Phys. Lett. B* **166** (1986) 473.

[63] J. Orloff, A. N. Rozanov and C. Santoni, *Limits on the mixing of tau neutrino to heavy neutrinos*, *Phys. Lett. B* **550** (2002) 8 [hep-ph/0208075].

[64] N. Sabti, A. Magalich and A. Filimonova, *An Extended Analysis of Heavy Neutral Leptons during Big Bang Nucleosynthesis*, *JCAP* **11** (2020) 056 [2006.07387].

[65] P. Coloma, P. A. N. Machado, I. Martinez-Soler and I. M. Shoemaker, *Double-Cascade Events from New Physics in Icecube*, *Phys. Rev. Lett.* **119** (2017) 201804 [1707.08573].

[66] FASER collaboration, *FASER's physics reach for long-lived particles*, *Phys. Rev. D* **99** (2019) 095011 [1811.12522].

[67] J. Beacham et al., *Physics Beyond Colliders at CERN: Beyond the Standard Model Working Group Report*, *J. Phys. G* **47** (2020) 010501 [1901.09966].

[68] C. O. Dib, J. C. Helo, M. Nayak, N. A. Neill, A. Soffer and J. Zamora-Saa, *Searching for a sterile neutrino that mixes predominantly with ν_τ at B factories*, *Phys. Rev. D* **101** (2020) 093003 [1908.09719].

[69] B. Batell, J. A. Evans, S. Gori and M. Rai, *Dark Scalars and Heavy Neutral Leptons at DarkQuest*, *JHEP* **05** (2021) 049 [2008.08108].

[70] G. Cvetič and C. S. Kim, *Sensitivity bounds on heavy neutrino mixing $|U_{\mu N}|^2$ and $|U_{\tau N}|^2$ from LHCb upgrade*, *Phys. Rev. D* **100** (2019) 015014 [1904.12858].

[71] E. J. Chun, A. Das, S. Mandal, M. Mitra and N. Sinha, *Sensitivity of Lepton Number Violating Meson Decays in Different Experiments*, *Phys. Rev. D* **100** (2019) 095022 [1908.09562].

[72] M. Atkinson, P. Coloma, I. Martinez-Soler, N. Rocco and I. M. Shoemaker, *Heavy Neutrino searches through Double-Bang Events at Super-Kamiokande, DUNE, and Hyper-Kamiokande*, [2105.09357].

[73] KM3NET collaboration, *Letter of intent for KM3NeT 2.0*, *J. Phys. G* **43** (2016) 084001 [1601.07459].

[74] CHARM collaboration, *A Search for Decays of Heavy Neutrinos*, *Phys. Lett. B* **128** (1983) 361.

[75] CERN-HAMBURG-AMSTERDAM-ROME-MOSCOW collaboration, *A Detector for Neutral Current Interactions of High-energy Neutrinos*, *Nucl. Instrum. Meth.* **178** (1980) 27.

[76] A. Diddens, M. Jonker, J. Panman, F. Udo, J. Allaby, U. Amaldi et al., *Performance of a large system of proportional drift tubes for a fine-grain calorimeter*, *Nuclear Instruments and Methods* **176** (1980) 189.

[77] O. Ruchayskiy and A. Ivashko, *Experimental bounds on sterile neutrino mixing angles*, *JHEP* **06** (2012) 100 [1112.3319].

[78] M. Chrzaszcz, M. Drewes, T. E. Gonzalo, J. Harz, S. Krishnamurthy and C. Weniger, *A frequentist analysis of three right-handed neutrinos with GAMBIT*, *Eur. Phys. J. C* **80** (2020) 569 [1908.02302].

[79] DELPHI collaboration, *Search for neutral heavy leptons produced in Z decays*, *Z. Phys. C* **74** (1997) 57.

[80] SHIP collaboration, *Sensitivity of the SHiP experiment to Heavy Neutral Leptons*, *JHEP* **04** (2019) 077 [1811.00930].

[81] P. Gondolo, G. Ingelman and M. Thunman, *Charm production and high-energy atmospheric muon and neutrino fluxes*, *Astropart. Phys.* **5** (1996) 309 [hep-ph/9505417].

[82] 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 [1504.04855].

[83] C. Argüelles, P. Coloma, P. Hernández and V. Muñoz, *Searches for Atmospheric Long-Lived Particles*, *JHEP* **02** (2020) 190 [1910.12839].

[84] ATLAS collaboration, *Search for exotic decays of the Higgs boson into long-lived particles in pp collisions at $\sqrt{s} = 13$ TeV using displaced vertices in the ATLAS inner detector*, [2107.06092].

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

[86] ATLAS collaboration, *Search for long-lived particles produced in pp collisions at $\sqrt{s} = 13$ TeV that decay into displaced hadronic jets in the ATLAS muon spectrometer*, *Phys. Rev. D* **99** (2019) 052005 [1811.07370].

[87] ATLAS collaboration, *Search for long-lived, massive particles in events with displaced vertices and missing transverse momentum in $\sqrt{s}$ = 13 TeV pp collisions with the ATLAS detector*, *Phys. Rev. D* **97** (2018) 052012 [1710.04901].

[88] CMS collaboration, *Search for long-lived particles with displaced vertices in multijet events in proton-proton collisions at $\sqrt{s}$ =13 TeV*, *Phys. Rev. D* **98** (2018) 092011 [1808.03078].

[89] CMS collaboration, *Search for new long-lived particles at $\sqrt{s} = 13$ TeV*, *Phys. Lett. B* **780** (2018) 432 [1711.09120].

[90] CMS collaboration, *Search sensitivity for dark photons decaying to displaced muons with CMS at the high-luminosity LHC*, CMS Physics Analysis Summary CMS-PAS-FTR-18-002, 2018.

[91] CMS collaboration, *First Level Track Jet Trigger for Displaced Jets at High Luminosity LHC*, CMS Physics Analysis Summary CMS-PAS-FTR-18-018, 2018.

[92] ATLAS, CMS collaboration, *Report on the Physics at the HL-LHC and Perspectives for the HE-LHC*, in *HL/HE-LHC Physics Workshop: final jamboree Geneva, CERN, March 1, 2019*, 2019, [1902.10229].

[93] Y. Gershtein, *CMS Hardware Track Trigger: New Opportunities for Long-Lived Particle Searches at the HL-LHC*, *Phys. Rev. D* **96** (2017) 035027 [1705.04321].

[94] S. Bobrovskyi, W. Buchmuller, J. Hajer and J. Schmidt, *Quasi-stable neutralinos at the LHC*, *JHEP* **09** (2011) 119 [1107.0926].

[95] S. Bobrovskyi, J. Hajer and S. Rydbeck, *Long-lived higgsinos as probes of gravitino dark matter at the LHC*, *JHEP* **02** (2013) 133 [1211.5584].

[96] E. Izaguirre and B. Shuve, *Multilepton and Lepton Jet Probes of Sub-Weak-Scale Right-Handed Neutrinos*, *Phys. Rev.* **D91** (2015) 093010 [1504.02470].

[97] I. Boiarska, K. Bondarenko, A. Boyarsky, S. Eijima, M. Ovchynnikov, O. Ruchayskiy et al., *Probing baryon asymmetry of the Universe at LHC and SHiP*, [1902.04535].

[98] M. Drewes and J. Hajer, *Heavy Neutrinos in displaced vertex searches at the LHC and HL-LHC*, *JHEP* **02** (2020) 070 [1903.06100].

[99] CMS COLLABORATION collaboration, *Search for long-lived particles that decay into final states containing two muons, reconstructed using only the CMS muon chambers*, Tech. Rep. CMS-PAS-EXO-14-012, CERN, Geneva, 2015.

[100] CMS collaboration, *The CMS Experiment at the CERN LHC*, *JINST* **3** (2008) S08004.

[101] CMS collaboration, *Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV*, *JINST* **13** (2018) P06015 [1804.04528].

[102] CMS collaboration, *The CMS muon project*, Technical Design Report CERN-LHCC-97-032, CMS-TDR-3, 1997.

[103] CMS collaboration, "Muon detectors." http://cms.web.cern.ch/news/muon-detectors.

[104] CMS collaboration, *Muon Reconstruction and Identification Improvements for Run-2 and First Results with 2015 Run Data*, CMS Detector Performance Summary CMS-DP-2015-015, 2015.

[105] ATLAS collaboration, *Search for heavy neutral leptons in decays of W bosons produced in 13 TeV pp collisions using prompt and displaced signatures with the ATLAS detector*, *JHEP* **10** (2019) 265 [1905.09787].

[106] CMS collaboration, *Measurement of b hadron lifetimes in pp collisions at $\sqrt{s} = 8$ TeV*, *Eur. Phys. J. C* **78** (2018) 457 [1710.08949].

[107] CMS collaboration, *Muon ID performance: low-p_T muon efficiencies*, CMS Detector Performance Summary CMS-DP-2014-020, 2014.

[108] CMS collaboration, *Muon Identification performance: hadron mis-Id measurements and RPC Muon selections*, CMS Detector Performance Summary CMS-DP-2014-018, 2014.

[109] CMS collaboration, *Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV*, *JINST* **10** (2015) P06005 [1502.02701].

[110] G. Cottin, J. C. Helo and M. Hirsch, *Displaced vertices as probes of sterile neutrino mixing at the LHC*, *Phys. Rev. D* **98** (2018) 035012 [1806.05191].

[111] ATLAS collaboration, *Measurement of $W^{\pm}$ and Z-boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, *Phys. Lett. B* **759** (2016) 601 [1603.09222].

[112] C. Degrande, O. Mattelaer, R. Ruiz and J. Turner, *Fully-Automated Precision Predictions for Heavy Neutrino Production Mechanisms at Hadron Colliders*, *Phys. Rev.* **D94** (2016) 053002 [1602.06957].

[113] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079 [1405.0301].

[114] K. Hagiwara, T. Li, K. Mawatari and J. Nakamura, *TauDecay: a library to simulate polarized tau decays via FeynRules and MadGraph5*, *Eur. Phys. J. C* **73** (2013) 2489 [1212.6247].

[115] G. Cottin, J. C. Helo and M. Hirsch, *Searches for light sterile neutrinos with multitrack displaced vertices*, *Phys. Rev.* **D97** (2018) 055025 [1801.02734].

[116] I. Antoniadis, A. Boyarsky, S. Espahbodi, O. Ruchayskiy and J. D. Wells, *Anomaly driven signatures of new invisible physics at the Large Hadron Collider*, *Nucl. Phys. B* **824** (2010) 296 [0901.0639].

[117] A. Alloul, N. D. Christensen, C. Degrande, C. Duhr and B. Fuks, *FeynRules 2.0 - A complete toolbox for tree-level phenomenology*, *Comput. Phys. Commun.* **185** (2014) 2250 [1310.1921].

[118] N. D. Christensen and C. Duhr, *FeynRules - Feynman rules made easy*, *Comput. Phys. Commun.* **180** (2009) 1614 [0806.4194].

[119] M. Ovchynnikov, "Chern-Simons portal."

[120] I. Boiarska, K. Bondarenko, A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy and A. Sokolenko, *Light scalar production from Higgs bosons and FASER 2*, *JHEP* **05** (2020) 049 [1908.04635].

[121] P. Obul, S. Dulat, T.-J. Hou, A. Tursun and N. Yalkun, *Next-to-leading order QCD and electroweak corrections to Higgs-strahlung processes at the LHC*, *Chin. Phys. C* **42** (2018) 093105 [1801.06851].

[122] Y. Gershtein, S. Knapen and D. Redigolo, *Probing naturally light singlets with a displaced vertex trigger*, [2012.07864].

[123] P. Bechtle, S. Heinemeyer, O. Stal, T. Stefaniak and G. Weiglein, *Probing the Standard Model with Higgs signal rates from the Tevatron, the LHC and a future ILC*, *JHEP* **11** (2014) 039 [1403.1582].

[124] J. de Blas et al., *Higgs Boson Studies at Future Particle Colliders*, *JHEP* **01** (2020) 139 [1905.03764].

[125] B. W. Lee and S. Weinberg, *Cosmological Lower Bound on Heavy Neutrino Masses*, *Phys. Rev. Lett.* **39** (1977) 165.

[126] A. Berlin and F. Kling, *Inelastic Dark Matter at the LHC Lifetime Frontier: ATLAS, CMS, LHCb, CODEX-b, FASER, and MATHUSLA*, *Phys. Rev. D* **99** (2019) 015021 [1810.01879].

[127] B. Batell, P. deNiverville, D. McKeen, M. Pospelov and A. Ritz, *Leptophobic Dark Matter at Neutrino Factories*, *Phys. Rev. D* **90** (2014) 115014 [1405.7049].

[128] P. Coloma, B. A. Dobrescu, C. Frugiuele and R. Harnik, *Dark matter beams at LBNF*, *JHEP* **04** (2016) 047 [1512.03852].

[129] B. Batell, A. Freitas, A. Ismail and D. Mckeen, *Probing Light Dark Matter with a Hadrophilic Scalar Mediator*, *Phys. Rev. D* **100** (2019) 095020 [1812.05103].

[130] S. Naaz, J. Singh and R. Singh, *DUNE prospect for leptophobic dark matter*, [2004.10996].

[131] B. A. Dobrescu and C. Frugiuele, *GeV-Scale Dark Matter: Production at the Main Injector*, *JHEP* **02** (2015) 019 [1410.1566].

[132] National institute of standards and technology, "Stopping power and range tables for protons in various materials."

[133] F. Kling, *Forward Neutrino Fluxes at the LHC*, [2105.08270].

[134] P. deNiverville, C.-Y. Chen, M. Pospelov and A. Ritz, *Light dark matter in neutrino beams: production modelling and scattering signatures at MiniBooNE, T2K and SHiP*, *Phys. Rev. D* **95** (2017) 035006 [1609.01770].

[135] L. Buonocore, C. Frugiuele, F. Maltoni, O. Mattelaer and F. Tramontano, *Event generation for beam dump experiments*, *JHEP* **05** (2019) 028 [1812.06771].

[136] CRYSTAL BARREL collaboration, *Search for a new light gauge boson in decays of pi0 and eta*, *Phys. Lett. B* **333** (1994) 271.

[137] BNL-E949 collaboration, *Study of the decay $K^+ \to \pi^+ \nu\bar{\nu}$ in the momentum region $140 < P_\pi < 199$ MeV/c*, *Phys. Rev. D* **79** (2009) 092004 [0903.0030].

[138] NA62 collaboration, *Search for production of an invisible dark photon in π^0 decays*, *JHEP* **05** (2019) 182 [1903.08767].

[139] MINIBOONE DM collaboration, *Dark Matter Search in Nucleon, Pion, and Electron Channels from a Proton Beam Dump with MiniBooNE*, *Phys. Rev. D* **98** (2018) 112004 [1807.06137].

[140] CDF collaboration, *A Search for dark matter in events with one jet and missing transverse energy in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, *Phys. Rev. Lett.* **108** (2012) 211804 [1203.0742].

[141] S. Caron, J. A. Casas, J. Quilis and R. Ruiz de Austri, *Anomaly-free Dark Matter with Harmless Direct Detection Constraints*, *JHEP* **12** (2018) 126 [1807.07921].

[142] J. A. Dror, R. Lasenby and M. Pospelov, *Dark forces coupled to nonconserved currents*, *Phys. Rev. D* **96** (2017) 075036 [1707.01503].

[143] J. A. Dror, R. Lasenby and M. Pospelov, *New constraints on light vectors coupled to anomalous currents*, *Phys. Rev. Lett.* **119** (2017) 141803 [1705.06726].

[144] P. Ilten, Y. Soreq, M. Williams and W. Xue, *Serendipity in dark photon searches*, *JHEP* **06** (2018) 004 [1801.04847].

[145] A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, *Observational manifestations of anomaly inflow*, *Phys. Rev. D* **72** (2005) 085011 [hep-th/0507098].

[146] A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, *Anomalies as a signature of extra dimensions*, *Phys. Lett. B* **626** (2005) 184 [hep-ph/0507195].

[147] R. Rayaprolu, S. Möller, C. Linsmeier and S. Spellerberg, *Simulation of neutron irradiation damage in tungsten using higher energy protons*, *Nuclear Materials and Energy* **9** (2016) 29.

[148] D. E. Soper, M. Spannowsky, C. J. Wallace and T. M. P. Tait, *Scattering of Dark Particles with Light Mediators*, *Phys. Rev. D* **90** (2014) 115005 [1407.2623].

[149] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht et al., *LHAPDF6: parton density access in the LHC precision era*, *Eur. Phys. J. C* **75** (2015) 132 [1412.7420].

[150] S. Tulin, *New weakly-coupled forces hidden in low-energy QCD*, *Phys. Rev. D* **89** (2014) 114008 [1404.4370].

[151] F. Kling, *Probing light gauge bosons in tau neutrino experiments*, *Phys. Rev. D* **102** (2020) 015007 [2005.03594].

[152] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko and K. Werner, *EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider*, *Phys. Rev. C* **92** (2015) 034906 [1306.0121].

[153] C. Baus, T. Pierog, and R. Ulrich, "Cosmic Ray Monte Carlo (CRMC)."

[154] A. Faessler, M. I. Krivoruchenko and B. V. Martemyanov, *Once more on electromagnetic form factors of nucleons in extended vector meson dominance model*, *Phys. Rev. C* **82** (2010) 038201 [0910.5589].

[155] A. Ismail, R. Mammen Abraham and F. Kling, *Neutral current neutrino interactions at FASERν*, *Phys. Rev. D* **103** (2021) 056014 [2012.10500].

[156] CMS collaboration, *Measurement of charged pion, kaon, and proton production in proton-proton collisions at $\sqrt{s} = 13$ TeV*, *Phys. Rev. D* **96** (2017) 112003 [1706.10194].

[157] ATLAS collaboration, *Measurement of $D^{*\pm}$, $D^{\pm}$ and $D_s^{\pm}$ meson production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, *Nucl. Phys. B* **907** (2016) 717 [1512.02913].

[158] LHCb collaboration, *Measurement of B meson production cross-sections in proton-proton collisions at $\sqrt{s}$ = 7 TeV*, *JHEP* **08** (2013) 117 [1306.3663].

[159] CMS collaboration, *Measurement of the total and differential inclusive B^+ hadron cross sections in pp collisions at $\sqrt{s}$ = 13 TeV*, *Phys. Lett. B* **771** (2017) 435 [1609.00873].

[160] P. Nason, S. Dawson and R. K. Ellis, *The One Particle Inclusive Differential Cross-Section for Heavy Quark Production in Hadronic Collisions*, *Nucl. Phys. B* **327** (1989) 49.

[161] F. Riehn, H. P. Dembinski, R. Engel, A. Fedynitch, T. K. Gaisser and T. Stanev, *The hadronic interaction model SIBYLL 2.3c and Feynman scaling*, *PoS* **ICRC2017** (2018) 301 [1709.07227].

[162] W. Bai, M. Diwan, M. V. Garzelli, Y. S. Jeong and M. H. Reno, *Far-forward neutrinos at the Large Hadron Collider*, *JHEP* **06** (2020) 032 [2002.03012].

[163] FASER collaboration, *Detecting and Studying High-Energy Collider Neutrinos with FASER at the LHC*, *Eur. Phys. J. C* **80** (2020) 61 [1908.02310].

[164] B. Batell, J. L. Feng and S. Trojanowski, *Detecting Dark Matter with Far-Forward Emulsion and Liquid Argon Detectors at the LHC*, *Phys. Rev. D* **103** (2021) 075023 [2101.10338].

[165] A. Fedynitch, F. Riehn, R. Engel, T. K. Gaisser and T. Stanev, *Hadronic interaction model sibyll 2.3c and inclusive lepton fluxes*, *Phys. Rev. D* **100** (2019) 103018 [1806.04140].

[166] T. Fujiwara, T. Kugo, H. Terao, S. Uehara and K. Yamawaki, *Nonabelian Anomaly and Vector Mesons as Dynamical Gauge Bosons of Hidden Local Symmetries*, *Prog. Theor. Phys.* **73** (1985) 926.

[167] H. Kuwabara and T. Hatsuda, *phi meson in nuclear matter*, *Prog. Theor. Phys.* **94** (1995) 1163 [nucl-th/9507017].

[168] K. Nakayama, J. W. Durso, J. Haidenbauer, C. Hanhart and J. Speth, *phi meson production in proton proton collisions*, *Phys. Rev. C* **60** (1999) 055209 [nucl-th/9904040].

[169] A. I. Vainshtein and V. I. Zakharov, *Remarks on Electromagnetic Form-Factors of Hadrons in the Quark Model*, *Phys. Lett. B* **72** (1978) 368.

[170] A. Boyarsky, M. Ovchynnikov, N. Sabti and V. Syvolap, *When feebly interacting massive particles decay into neutrinos: The Neff story*, *Phys. Rev. D* **104** (2021) 035006 [2103.09831].

[171] Y. I. Izotov, T. X. Thuan and N. G. Guseva, *A new determination of the primordial He abundance using the He i λ10830 A emission line: cosmological implications*, *Mon. Not. Roy. Astron. Soc.* **445** (2014) 778 [1408.6953].

[172] E. Aver, K. A. Olive and E. D. Skillman, *The effects of He I λ10830 on helium abundance determinations*, *JCAP* **1507** (2015) 011 [1503.08146].

[173] A. Peimbert, M. Peimbert and V. Luridiana, *The primordial helium abundance and the number of neutrino families*, *Rev. Mex. Astron. Astrofis.* **52** (2016) 419 [1608.02062].

[174] V. Fernández, E. Terlevich, A. I. Díaz, R. Terlevich and F. F. Rosales-Ortega, *Primordial helium abundance determination using sulphur as metallicity tracer*, *Monthly Notices of the Royal Astronomical Society* **478** (2018) 5301–5319.

[175] M. Valerdi, A. Peimbert, M. Peimbert and A. Sixtos, *Determination of the Primordial Helium Abundance Based on NGC 346, an H ii Region of the Small Magellanic Cloud*, *Astrophys. J.* **876** (2019) 98 [1904.01594].

[176] Y. I. Izotov, F. H. Chaffee, C. B. Foltz, R. F. Green, N. G. Guseva and T. X. Thuan, *Helium abundance in the most metal-deficient blue compact galaxies: I zw 18 and sbs 0335-052*, *Astrophys. J.* **527** (1999) 757 [astro-ph/9907228].

[177] K. A. Olive and E. D. Skillman, *A Realistic determination of the error on the primordial helium abundance: Steps toward non-parametric nebular helium abundances*, *Astrophys. J.* **617** (2004) 29 [astro-ph/0405588].

[178] M. Pospelov and J. Pradler, *Metastable GeV-scale particles as a solution to the cosmological lithium problem*, *Phys. Rev. D* **82** (2010) 103514 [1006.4172].

[179] K. Kohri, *Primordial nucleosynthesis and hadronic decay of a massive particle with a relatively short lifetime*, *Phys. Rev. D* **64** (2001) 043515 [astro-ph/0103411].

[180] Y. I. Izotov, G. Stasinska and N. G. Guseva, *Primordial 4He abundance: a determination based on the largest sample of HII regions with a methodology tested on model HII regions*, *Astron. Astrophys.* **558** (2013) A57 [1308.2100].

[181] A. Fradette and M. Pospelov, *BBN for the LHC: constraints on lifetimes of the Higgs portal scalars*, *Phys. Rev. D* **96** (2017) 075033 [1706.01920].

[182] A. Fradette, M. Pospelov, J. Pradler and A. Ritz, *Cosmological beam dump: constraints on dark scalars mixed with the Higgs boson*, *Phys. Rev. D* **99** (2019) 075004 [1812.07585].

[183] J. J. Bennett, G. Buldgen, M. Drewes and Y. Y. Wong, *Towards a precision calculation of the effective number of neutrinos $N_{\rm eff}$ in the Standard Model I: The QED equation of state*, *JCAP* **03** (2020) 003 [1911.04504].

[184] M. Escudero Abenza, *Precision early universe thermodynamics made simple: $N_{\rm eff}$ and neutrino decoupling in the Standard Model and beyond*, *JCAP* **05** (2020) 048 [2001.04466].

[185] J. J. Bennett, G. Buldgen, P. F. de Salas, M. Drewes, S. Gariazzo, S. Pastor et al., *Towards a precision calculation of $N_{\rm eff}$ in the Standard Model II: Neutrino decoupling in the presence of flavour oscillations and finite-temperature QED*, [2012.02726].

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

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

[188] M. Kawasaki, K. Kohri and N. Sugiyama, *Cosmological constraints on late time entropy production*, *Phys. Rev. Lett.* **82** (1999) 4168 [astro-ph/9811437].

[189] M. Kawasaki, K. Kohri and N. Sugiyama, *MeV scale reheating temperature and thermalization of neutrino background*, *Phys. Rev.* **D62** (2000) 023506 [astro-ph/0002127].

[190] M. Kawasaki, K. Nakayama and M. Senami, *Cosmological implications of supersymmetric axion models*, *JCAP* **03** (2008) 009 [0711.3083].

[191] T. Hasegawa, N. Hiroshima, K. Kohri, R. S. L. Hansen, T. Tram and S. Hannestad, *MeV-scale reheating temperature and thermalization of oscillating neutrinos by radiative and hadronic decays of massive particles*, *JCAP* **12** (2019) 012 [1908.10189].

[192] G. M. Fuller, C. T. Kishimoto and A. Kusenko, *Heavy sterile neutrinos, entropy and relativistic energy production, and the relic neutrino background*, [1110.6479].

[193] S. Hannestad, *What is the lowest possible reheating temperature?*, *Phys. Rev.* **D70** (2004) 043506 [astro-ph/0403291].

[194] A. D. Dolgov, S. H. Hansen, G. Raffelt and D. V. Semikoz, *Heavy sterile neutrinos: Bounds from big bang nucleosynthesis and SN1987A*, *Nucl. Phys.* **B590** (2000) 562 [hep-ph/0008138].

[195] O. Ruchayskiy and A. Ivashko, *Restrictions on the lifetime of sterile neutrinos from primordial nucleosynthesis*, *JCAP* **1210** (2012) 014 [1202.2841].

[196] G. B. Gelmini, M. Kawasaki, A. Kusenko, K. Murai and V. Takhistov, *Big Bang Nucleosynthesis constraints on sterile neutrino and lepton asymmetry of the Universe*, *JCAP* **09** (2020) 051 [2005.06721].

[197] L. Mastrototaro, P. D. Serpico, A. Mirizzi and N. Saviano, *Massive sterile neutrinos in the early universe: From thermal decoupling to cosmological constraints*, [2104.11752].

[198] M. Escudero, *Neutrino decoupling beyond the Standard Model: CMB constraints on the Dark Matter mass with a fast and precise $N_{\rm eff}$ evaluation*, *JCAP* **02** (2019) 007 [1812.05605].

[199] PLANCK collaboration, *Planck 2018 results. VI. Cosmological parameters*, *Astron. Astrophys.* **641** (2020) A6 [1807.06209].

[200] PLANCK collaboration, *Planck 2018 results. V. CMB power spectra and likelihoods*, *Astron. Astrophys.* **641** (2020) A5 [1907.12875].

[201] E. W. Kolb and M. S. Turner, *The Early Universe*, vol. 69. Front.Phys. 69 (1990) 1-547, 1990.

[202] M. Fabbrichesi, E. Gabrielli and G. Lanfranchi, *The Dark Photon*, [2005.01515].

[203] S. Dumitru, B. A. Ovrut and A. Purves, *The R-parity Violating Decays of Charginos and Neutralinos in the B-L MSSM*, *JHEP* **02** (2019) 124 [1810.11035].

[204] K. J. Kelly and Y. Zhang, *Mononeutrino at DUNE: New Signals from Neutrinophilic Thermal Dark Matter*, *Phys. Rev. D* **99** (2019) 055034 [1901.01259].

[205] M. Shaposhnikov, *A Possible symmetry of the nuMSM*, *Nucl. Phys.* **B763** (2007) 49 [hep-ph/0605047].

[206] M. B. Gavela, T. Hambye, D. Hernandez and P. Hernandez, *Minimal Flavour Seesaw Models*, *JHEP* **09** (2009) 038 [0906.1461].

[207] A. Boyarsky, O. Ruchayskiy and M. Shaposhnikov, *The Role of sterile neutrinos in cosmology and astrophysics*, *Ann. Rev. Nucl. Part. Sci.* **59** (2009) 191 [0901.0011].

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

[209] J. Klarić, M. Shaposhnikov and I. Timiryasov, *Uniting Low-Scale Leptogenesis Mechanisms*, *Phys. Rev. Lett.* **127** (2021) 111802 [2008.13771].

[210] A. D. Dolgov, S. H. Hansen, G. Raffelt and D. V. Semikoz, *Cosmological and astrophysical bounds on a heavy sterile neutrino and the KARMEN anomaly*, *Nucl. Phys.* **B580** (2000) 331 [hep-ph/0002223].

[211] A. D. Dolgov and F. L. Villante, *BBN bounds on active sterile neutrino mixing*, *Nucl. Phys.* **B679** (2004) 261 [hep-ph/0308083].

[212] A. C. Vincent, E. F. Martinez, P. Hernández, M. Lattanzi and O. Mena, *Revisiting cosmological bounds on sterile neutrinos*, *JCAP* **04** (2015) 006 [1408.1956].

[213] D. Notzold and G. Raffelt, *Neutrino Dispersion at Finite Temperature and Density*, *Nucl. Phys. B* **307** (1988) 924.

[214] A. D. Dolgov and S. H. Hansen, *Massive sterile neutrinos as warm dark matter*, *Astropart. Phys.* **16** (2002) 339 [hep-ph/0009083].

[215] L. Husdal, *On Effective Degrees of Freedom in the Early Universe*, *Galaxies* **4** (2016) 78 [1609.04979].

[216] K. Abazajian, G. M. Fuller and M. Patel, *Sterile neutrino hot, warm, and cold dark matter*, *Phys. Rev. D* **64** (2001) 023501 [astro-ph/0101524].

[217] J. Edsjo and P. Gondolo, *Neutralino relic density including coannihilations*, *Phys. Rev. D* **56** (1997) 1879 [hep-ph/9704361].

[218] A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy and V. Syvolap, *Constraints on MeV-GeV scale feebly interacting particles from primordial nucleosynthesis*, *to appear* (2020) .

[219] J. Kersten and A. Y. Smirnov, *Right-Handed Neutrinos at CERN LHC and the Mechanism of Neutrino Mass Generation*, *Phys. Rev.* **D76** (2007) 073005 [0705.3221].

[220] D. Gorbunov and M. Shaposhnikov, *How to find neutral leptons of the νMSM?*, *JHEP* **10** (2007) 015 [0705.1729].

[221] K. Bondarenko, A. Boyarsky, J. Klaric, O. Mikulenko, O. Ruchayskiy, V. Syvolap et al., *An allowed window for heavy neutral leptons below the kaon mass*, [2101.09255].

[222] R. K. Ellis et al., *Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020*, [1910.11775].

[223] NA62 collaboration, *Search for heavy neutral lepton production in K^+ decays to positrons*, *Phys. Lett. B* **807** (2020) 135599 [2005.09575].

[224] D. Gorbunov, I. Krasnov, Y. Kudenko and S. Suvorov, *Heavy Neutral Leptons from kaon decays in the SHiP experiment*, *Phys. Lett. B* **810** (2020) 135817 [2004.07974].

[225] P. Ballett, T. Boschi and S. Pascoli, *Heavy Neutral Leptons from low-scale seesaws at the DUNE Near Detector*, *JHEP* **03** (2020) 111 [1905.00284].

[226] DUNE collaboration, *Deep Underground Neutrino Experiment (DUNE), Far Detector Technical Design Report, Volume II: DUNE Physics*, [2002.03005].

[227] P. Coloma, E. Fernández-Martínez, M. González-López, J. Hernández-García and Z. Pavlovic, *GeV-scale neutrinos: interactions with mesons and DUNE sensitivity*, *Eur. Phys. J. C* **81** (2021) 78 [2007.03701].

[228] SIMONS OBSERVATORY collaboration, *The Simons Observatory: Science goals and forecasts*, *JCAP* **1902** (2019) 056 [1808.07445].

[229] K. Abazajian et al., *CMB-S4 Science Case, Reference Design, and Project Plan*, [1907.04473].

[230] PARTICLE DATA GROUP collaboration, *Review of Particle Physics*, *PTEP* **2020** (2020) 083C01.

[231] W. Hu and M. J. White, *The Damping tail of CMB anisotropies*, *Astrophys. J.* **479** (1997) 568 [astro-ph/9609079].

[232] S. Bashinsky and U. Seljak, *Neutrino perturbations in CMB anisotropy and matter clustering*, *Phys. Rev. D* **69** (2004) 083002 [astro-ph/0310198].

[233] N. Sabti, J. Alvey, M. Escudero, M. Fairbairn and D. Blas, *Refined Bounds on MeV-scale Thermal Dark Sectors from BBN and the CMB*, *JCAP* **01** (2020) 004 [1910.01649].

[234] G. B. Gelmini, A. Kusenko and V. Takhistov, *Possible Hints of Sterile Neutrinos in Recent Measurements of the Hubble Parameter*, [1906.10136].

[235] L. Verde, T. Treu and A. G. Riess, *Tensions between the Early and the Late Universe*, *Nature Astron.* **3** (2019) 891 [1907.10625].

[236] G. Efstathiou, *A Lockdown Perspective on the Hubble Tension (with comments from the SH0ES team)*, [2007.10716].

[237] L. Knox and M. Millea, *Hubble constant hunter's guide*, *Phys. Rev. D* **101** (2020) 043533 [1908.03663].

[238] S. Vagnozzi, *New physics in light of the H_0 tension: An alternative view*, *Phys. Rev. D* **102** (2020) 023518 [1907.07569].

[239] E. Di Valentino, O. Mena, S. Pan, L. Visinelli, W. Yang, A. Melchiorri et al., *In the Realm of the Hubble tension − a Review of Solutions*, [2103.01183].

[240] B. Audren, J. Lesgourgues, K. Benabed and S. Prunet, *Conservative Constraints on Early Cosmology: an illustration of the Monte Python cosmological parameter inference code*, *JCAP* **1302** (2013) 001 [1210.7183].

[241] T. Brinckmann and J. Lesgourgues, *MontePython 3: boosted MCMC sampler and other features*, [1804.07261].

[242] A. G. Riess, S. Casertano, W. Yuan, J. B. Bowers, L. Macri, J. C. Zinn et al., *Cosmic Distances Calibrated to 1% Precision with Gaia EDR3 Parallaxes and Hubble Space Telescope Photometry of 75 Milky Way Cepheids Confirm Tension with LambdaCDM*, [2012.08534].

[243] P. Elmfors, D. Grasso and G. Raffelt, *Neutrino dispersion in magnetized media and spin oscillations in the early universe*, *Nucl. Phys. B* **479** (1996) 3 [hep-ph/9605250].

Samenvatting

In de 20e eeuw hebben pogingen om een aantal verschijnselen in de deeltjesfysica te verklaren, samen met twee wetenschappelijke revoluties (de speciale relativiteitstheorie en de kwantummechanica), geresulteerd in de ontwikkeling van het Standaardmodel van de deeltjesfysica (SM). De huidige status van het SM is dubbel. Aan de ene kant is het SM extreem succesvol in het zeer nauwkeurig beschrijven van talloze deeltjesfysica-experimenten bij versnellers. Aan de andere kant slaagt het SM er niet in om verschillende bevestigde fenomenen te verklaren die bekend staan als Buiten het Standaardmodel (BSM) verschijnselen: donkere materie in het heelal, neutrino-oscillaties en de oorsprong van de materie-antimaterie asymmetrie in het heelal. Al deze fenomenen suggereren dat het SM moet worden uitgebreid.

Helaas zeggen deze BSM-fenomenen niets over de precieze manier waarop het SM moet worden uitgebreid: veel verschillende extensies kunnen de BSM-fenomenen evengoed verklaren. Om aanwijzingen te vinden waar te gaan zoeken, moeten wij observationele tekens bestuderen die kunnen worden veroorzaakt door verschillende uitbreidingen van het SM. Dit proefschrift is toegewijd aan het bestuderen van twee zulke zoekopdrachten: in deeltjes experimenten en in kosmologische waarnemingen. De twee benaderingen zijn complementair: deeltjes versnellers onderzoeken deeltjes met een korte levensduur en kosmologische waarnemingen kunnen ons iets vertellen over langlevende deeltjes.

Bij deeltjesversnellers kunnen de nieuwe deeltjes worden geproduceerd en vervolgens worden onderzocht, met name door hun verval of verstrooiing. Om het bereik in de parameter-ruimte voor toekomstige experimenten te kunnen definiëren, is het belangrijk om resultaten te bestuderen die voortkwamen uit eerdere experimenten. Voor dit doeleinde, hebben wij de resultaten van het oude experiment CHARM opnieuw geanalyseerd, in de context van een model met zware neutrale leptonen (HNL's). Dit is een uitbreiding waar nieuwe zware fermionen worden toegevoegd aan het SM die mogelijk direct verantwoordelijk zijn voor alle BSM-verschijnselen. Wij hebben geconstateerd dat eerdere gepubliceerde resultaten voor HNL's met CHARM met twee ordes van grootte werden onderschat (afhankelijk van de eigenschappen van HNL's). Vervolgens hebben wij gekeken naar zogenaamde 'displaced vertices' in de Large Hadron Collider (LHC), waar wij hun potentieel hebben bestudeerd om verval van verschillende deeltjes met een zwakke interactie (FIP's) te onderzoeken. Vanwege de korte afstand tussen het FIP-productiepunt en het vervalvolume, van de orde $l_{\text{min}} = \mathcal{O}(\text{mm})$, kan zo een zoekactie de kloof in de parameterruimte van FIP's die zijn uitgesloten door eerdere experimenten dichten en het

domein dat kan worden onderzocht met toekomstige experimenten, die grotere $l_{\text{min}} \gtrsim 1$ m hebben. In het bijzonder, hebben wij voor het eerst het potentieel van de muon-tracker bij de CMS bestudeerd om naar FIP's te zoeken, met als voorbeelden HNL's, Higgs-achtige scalars en Chern-Simons-portalen. Vervolgens hebben wij FIP-verstrooiing overwogen en onderzocht hoe verstrooiing van FIP's met hadronen kan worden onderscheiden van SM-neutrinoverstrooiing, afhankelijk van de massa van een deeltje dat de verstrooiing bemiddelt. Met deze signalen hebben wij het bereik geschat van SND@LHC (een recent goedgekeurd experiment bij de LHC) voor het model van lichte donkere materie dat via het leptofobe portaal een interactie ondergaat met SM-deeltjes.

Kosmologie bepaalt doorgaans de ondergrens voor de koppelingen van FIP's. De belangrijkste waarnemingen voor FIP's met een korte levensduur zijn BBN en CMB, die mogelijk gevoelig zijn voor FIP's met een levensduur van slechts aantal $\times\, 10^{-2}$ s. Wij hebben ontdekt dat lichte, langlevende mesonen zoals $\pi^{\pm}$ en $K^{\pm/0}$ protonen efficiënt omzetten in neutronen. Dit kan leiden tot een significante toename van de hoeveelheid ^{4}He als dergelijke mesonen aanwezig zijn in het primordiaal plasma met temperaturen onder 1.5 MeV. Met behulp van analytische technieken hebben wij met dit effect een modelonafhankelijke limiet aan FIP-levensduren afgeleid. Een belangrijke parameter die verandert door FIP's met een korte levensduur en die de CMB kan beïnvloeden, is het effectieve aantal aan vrijheidsgraden N_{eff}. Bij het bestuderen van de impact van zware FIP's met $m_{\text{FIP}} \gg T$ op N_{eff}, hebben wij ontdekt dat FIP's met een levensduur van $\tau_{\text{FIP}} \lesssim 0.1$ s een afname van N_{eff} kunnen veroorzaken, zelfs als ze grotendeels vervallen in neutrino's. Dit is te wijten aan het feit dat neutrino's, die gecreëerd zijn door het verval van FIP's, het grootste deel van hun energie opslaan in de EM-plasma tijdens hun thermalisatie.

Ten slotte hebben wij deze modelonafhankelijke resultaten toegepast in het geval van HNL's. Wij hebben hun kosmologische productie bestudeerd en rekening gehouden met de parameterruimte van HNL's die ofwel nooit in thermisch evenwicht zijn gekomen, of erin zijn gekomen en vervolgens zijn ontkoppeld. Vervolgens hebben wij limieten afgeleid van BBN van ongeveer $\tau_N \lesssim 0.02$ s. Dit is een factor vijf sterker dan wat eerder in de literatuur is vermeld en wat verkregen is met de CMB. Ten slotte, hebben wij geconcludeerd dat HNL's met massa's $m_N \gtrsim 70$ MeV N_{eff} kunnen verminderen.

Summary

In the 20th century, attempts to explain the variety of particle physics phenomena, together with two scientific revolutions – special relativity and quantum mechanics, have resulted in the development of the Standard Model of particle physics (SM). The current status of SM is dual. Being extremely successful in describing very precisely countless particle physics experiments at accelerators, the SM fails to explain several well-established phenomena that are known as Beyond the Standard Model (BSM) phenomena: dark matter in the Universe, neutrino oscillations, and the origin of matter-antimatter asymmetry of the Universe. This suggests that the SM has to be extended.

Unfortunately, the BSM phenomena do not tell us about the precise way in which SM has to be extended, and a lot of different extensions may equally explain the BSM phenomena. In order to find hints where to go we have to study observational signatures that may be caused by various extensions of the SM. This thesis is devoted to studying two kinds of such searches: accelerator experiments, and cosmological observations. The two approaches are complementary: accelerator experiments probe short-lived particles and cosmological signatures constrain long-lived particles.

At accelerator experiments, the new particles may be produced and then searched in particular by their decays or scatterings. In order to define the target parameter space for upcoming experiments, it is important to know constraints coming from past experiments. For these purposes, I have re-analyzed bounds coming from the past experiment CHARM to a model of Heavy Neutral Leptons (HNLs), which is an extension that adds massive fermions that may be directly responsible for the resolution of all BSM phenomena. I have found that previously reported bounds on HNLs from CHARM are underestimated by two orders of magnitude, in dependence on properties of HNLs. Next, I have considered searches for displaced vertices at the Large Hadron Collider (LHC) and studied their potential to probe decays of different feebly interacting particles (FIPs). Due to the short distance between the FIP production point and the decay volume, of the order of $l_{\text{min}} = \mathcal{O}(\text{mm})$, such kind of search is able to close the gap between the parameter space of FIPs that are ruled out by past experiments and the domain which may be probed by dedicated experiments, which have larger $l_{\text{min}} \gtrsim 1$ m. In particular, I have demonstrated for the first time the potential of the muon tracker at CMS to search for FIPs, using as examples HNLs, Higgs-like scalar and Chern-Simons portals. Next, I have considered FIP scattering, and studied how scatterings of FIPs off hadrons may be distinguished from

SM neutrino scatterings, dependent on the mass of a particle that mediates the scattering. Considering these signatures, I have estimated the sensitivity of SND@LHC, a recently approved experiment at the LHC, to the model of light dark matter interacting with SM particles via the leptophobic portal.

Cosmological signatures typically provide the lower bound on the values of FIP's couplings. The most important observations for short-lived FIPs are BBN and CMB, which may be sensitive to FIPs with lifetimes as small as few $\times\, 10^{-2}$ s. I have found that light long-lived mesons such as $\pi^{\pm}$ and $K^{\pm/0}$ efficiently convert protons into neutrons, which may lead to a significant increase of the ^{4}He abundance if such mesons are present in the primordial plasma at temperatures below 1.5 MeV. Using analytic considerations, I have derived a model-independent bound on FIP lifetimes from BBN from this effect. An important parameter that is changed by short-lived FIPs and may affect CMB is the effective number of degrees of freedom N_{eff}. Studying the impact of heavy FIPs with $m_{\text{FIP}} \gg T$ on N_{eff}, I have found that FIPs with lifetimes $\tau_{\text{FIP}} \lesssim 0.1$ s decrease N_{eff} even if decaying mostly into neutrinos, which is due to the fact that neutrinos from decays of FIPs store most of their energy in EM plasma during their thermalization.

Finally, I have applied these model-independent results to the case of HNLs. I have studied their cosmological production, considering the parameter space of HNLs that either never entered thermal equilibrium, or entered it and then decoupled. Then, I have derived bounds from BBN at the level $\tau_N \lesssim 0.02$ s, which are a factor of five stronger than those reported previously in the literature, and from CMB, finding in particular that HNLs with masses $m_N \gtrsim 70$ MeV may decrease N_{eff}.

List of publications

[1] A. Boyarsky, M. Ovchynnikov, V. Syvolap, N. Sabti, "Upper bound on neutrino mass from cosmological observations in presence of Heavy Neutral Leptons", in preparation

[2] I. Boiarska, A. Boyarsky, O. Mikulenko, M. Ovchynnikov, "Constraints from the CHARM experiment on Heavy Neutral Leptons with tau mixing", submitted to Phys.Rev.D, arXiv:2107.14685

[3] A. Boyarsky, O. Mikulenko, M. Ovchynnikov, L. Shchutska, "Searches for new physics at SND@LHC", submitted to JHEP, arXiv:2104.09688

[4] A. Boyarsky, M. Ovchynnikov, N. Sabti, V. Syvolap, "When feebly interacting massive particles decay into neutrinos: The $N_{\rm eff}$ story", Phys.Rev.D 104 (2021) 3, 035006, arXiv:2103.09831

[5] A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy, V. Syvolap, "Improved BBN constraints on Heavy Neutral Leptons", Phys.Rev.D 104 (2021) 2, 023517, arXiv:2008.00749

[6] I. Boiarska, K. Bondarenko, A. Boyarsky, M. Ovchynnikov, O. Ruchayskiy *et al.*, "Light scalar production from Higgs bosons and FASER 2", JHEP 05 (2020) 049, arXiv:1908.04635

[7] I. Boiarska, K. Bondarenko, A. Boyarsky, V. Gorkavenko, M. Ovchynnikov *et al.*, "Phenomenology of GeV-scale scalar portal ", JHEP 11 (2019) 162, arXiv:1904.10447

[8] K. Bondarenko, A. Boyarsky, M. Ovchynnikov, Oleg Ruchayskiy, L. Shchutska, "Probing new physics with displaced vertices: muon tracker at CMS ", Phys.Rev.D 100 (2019) 7, 075015, arXiv:1903.11918

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

[10] I. Boiarska, K. Bondarenko, A. Boyarsky, S. Eijima, M. Ovchynnikov, O. Ruchayskiy, I. Timiryasov, "Probing baryon asymmetry of the Universe at LHC and SHiP", arXiv:1902.04535

[11] *SHiP Collaboration*, C. Ahdida *et al.*, "Sensitivity of the SHiP experiment to dark photons decaying to a pair of charged particles", Eur.Phys.J.C 81 (2021) 5, 451, arXiv:2011.05115

[12] *SHiP Collaboration*, C. Ahdida *et al.*, "Sensitivity of the SHiP experiment to light dark matter", JHEP 04 (2021) 199, arXiv:2010.11057

[13] *SHiP collaboration*, C. Ahdida *et al.*, "SND@LHC", rep. num. CERN-LHCC-2020-002, LHCC-I-035, arXiv:2002.08722

[14] *SHiP collaboration*, C. Ahdida *et al.*, "The experimental facility for the Search for Hidden Particles at the CERN SPS", JINST 14 (2019) 03, P03025, arXiv:1810.06880

[15] *SHiP collaboration*, C. Ahdida *et al.*, "Sensitivity of the SHiP experiment to Heavy Neutral Leptons", JHEP 04 (2019) 077, arXiv:1811.00930. **Contributing author.**

Curriculum vitæ

I was born in Kyiv, Ukraine, on the 16th of April 1994. I have received my primary education at the school #304.

After finishing the school, I entered the Physics Department of the Taras Shevchenko National University of Kyiv in 2011, where I have got the bachelor's degree in 2015 and the master's degree in 2017. My master thesis was titled "Generation of primordial magnetic fields by axion-like particle dark matter". After graduation, I started my PhD at the Leiden Institute of Physics under the supervision of Dr. Alexey Boyarsky.

While being a PhD student, I was a teaching assistant for the courses "Effective field theory", "Origin and structure of the Standard model", and "Particle Physics in the Early Universe". I have become a member of the SHiP and SND@LHC collaborations, and studied the sensitivity of these experiments to different models of new physics. I have visited a number of schools and conferences in the Netherlands, Germany, Switzerland.

Acknowledgements

I want to thank my supervisor Alexey Boyarsky for guidance during the years of my bachelor's and master's studies and graduate school.

I am thankful for my second promotor Ana Achúcarro and my collaborators, Oleg Ruchayskiy, Lesya Shchutska, Kyrylo Bondarenko, Alex Mikulenko, Nashwan Sabti, Vsevolod Syvolap for kindness, gaining experience, and spending many hours in discussions. I would especially like to mention Kyrylo Bondarenko, who was very kind to help me start my research in searching for new physics at accelerators.

The Lorentz Institute has been a hospitable environment all these years.

I would also like to thank Nashwan Sabti for help with the Dutch translation. I am grateful the secretary of our institute, Fran, for the help in all sorts of practical things, and all the colleagues at the Lorentz Institute for providing the great atmosphere.

Finally, I would like to thank my family: my wife Iryna, who supported me both in my scientific research and in the life behind the science, my newborn son Matvii, who gave me an opportunity to keep working on my thesis. I am grateful to my parents, Oksana and Yurii, for bringing me up.