Chapter 9

Setting Upper Limits on The Mixing Angle of HNL

After the selection performed on Monte Carlo (MC) samples, the analysis continues with assessing and propagating uncertainties. This is critical step to evaluate the statistical fluctuation due to the size of the MC samples. Moreover, it is to understand the uncertainties due to the simulated physics models not only of the BNB flux and SM neutrinos, but also of the SBND detector. Then, the limit setting procedure was chosen to be the likelihood-based hypothesis setting for exclusion limits, which is common statistical procedure developed by particle physicists [1]. The limit setting was performed using the beam bucket distribution with fully propagated uncertainties, which exploits the exceptional high signal-to-background ratio of bins in the edge of the distribution to achieve competitive sensitivity.

The following chapter will provide details of the analysis steps after selection to acquire the final result. Sec. 9.1 outlines the assessment of statistical as well as systematic uncertainties. Following that, Sec. 9.2 delves into the limit setting procedure. The results of the upper limits on the mixing angle of Heavy Neutral Leptons (HNLs) are then presented in Sec. 9.3. Finally, Sec. 9.4 concludes the chapter with some remarks.

9.1 Uncertainty Assessment

Since many physics measurements to be made by SBND will heavily rely on the comparison of data and MC samples, it is vitally important to understand the input physics models used to simulate the MC sample. As previously discussed in Chapter 3 describing the detector and Chapter 5 detailing the simulation framework, many different theoretical as well as data-driven models are used to predict the incoming flux from the BNB and the SM neutrino interaction cross sections. Each model has its own uncertainty and need to be propagated to the final physics result. The assessment of statistical and systematic uncertainties of signals and backgrounds are presented in this section.

9.1.1 The Reweighting Method

The impact of systematic uncertainties on a physics measurement can be assessed by simulating and reconstructing a number of different samples, referred to as *universes*, each with a physics parameter tweaked within its uncertainty range. The physics measurement is then performed in each universe in the same manner as done on the Central Value (CV) sample that does not have any parameters tweaked. The variation of the result across universes compared to the CV sample describes the uncertainty that the tweaked parameter has on the measurement. However, fully simulating and reconstructing a large MC sample for multiple systematic parameters can be computationally intensive, the *reweighting* technique is used instead by producing a weight associated with the tweaked parameter per universe. The weight can be applied to smear the physics result by an amount as expected by the tweaked parameter. Or vice versa, the smeared physics result can be unweighted to recover the original result without any uncertainties from the tweaked parameter [2].

The reweighting technique begins with transforming a physics parameter x to x' as follows

$$x' = x \cdot f(x) \tag{9.1}$$

where $f(x)$ describes some transformation functions as a function of x. If $f(x) = 1$, then x' is equal to the unweighted x. A common form of the transformation function is a Gaussian function, where the physics parameter x is thrown to x' by randomly sampling from a unit Gaussian with its mean and sigma is set as 1. Another common formalism is the Delta function, where x' can take only take the value of 0 or 1.

Then, from the probability $P(x)$ describing some physic properties parameterised by the physics parameter x, the weight w can be computed as

$$w = \frac{P(x')}{P(x)} \tag{9.2}$$

The weight w describes whether the probability $P(x')$ is more or less likely to occur given the transformed parameter x' compared to the CV parameter x. The distribution of w therefore describes the Probability Density Function (PDF) of the parameter x. A universe associated with a weight w represents an outcome sampled from the PDF. The weight example shown in Eq. 9.2 is associated with a single physics parameter non-correlated to any other parameters, however, a weight associated with multiple correlated parameters can also be computed in the same manner.

The reweighting framework provides a quick way to compute the PDFs, allowing for the assessment of the impact of systematic uncertainties without the computational expense of simulating and reconstructing the sample multiple times. A series of universes is first simulated and weights for every interaction, whether SM neutrino or HNL, are then calculated from the PDFs for each universe. The variation of the physics result across universes are used to quantify the uncertainties of the weighted physics parameters have on the physics measurement, of which the error propagation is formalised in Sec. 9.1.2.

In some cases, reweighting is not applicable such as evaluating uncertainties due to detector effects. For example, recombination, as detailed in Sec. 7.2, influences not only the charge and light yield but also the charge deposition on wires non-uniformly. Consequently, it is non-trivial to quantify analytically the downstream impacts due to the variation in recombination parameters on the charge reconstruction by Pandora or any high level analysis tools using the reconstructed charge information. A full simulation and reconstruction of a single universe using the varied recombination parameters is needed to fully assess the impact on the physics measurement. This method is commonly used for assessing detector systematic uncertainties.

9.1.2 Error Propagation Formulation

The impact of uncertainties can be evaluated by constructing the covariance matrix V of a series of observations n. The matrix describes the average deviation of the value in bins i and j from the CV away from the universe k, totalling U universes. The matrix is computed

as follows

$$V_{ij} = \frac{1}{U} \sum_{k}^{U} \left(n_i^n - n_i^{CV} \right) \left(n_j^n - n_j^{CV} \right) \tag{9.3}$$

The diagonal term of the covariance matrix represents the variance in a given bin and the error σ can be derived as

$$V_{ii} = \sigma_i^2 \tag{9.4}$$

Since there are multiple sources of error, as detailed in the forthcoming sections, the total covariance matrix is computed by summing the covariance matrix from each error source together, effectively adding the errors in quadrature as follows

$$V_{ij}^{Total} = \sum_{Sources} V_{ij}^{Source} = V_{ij}^{Stat} + V_{ij}^{Flux} + V_{ij}^{Cross\ Section} + ... \tag{9.5}$$

The total error is then computed from the total covariance matrix using Eq. 9.4. Additionally, the fractional matrix describing the relative error in each bin is computed, allowing for easy and direct comparison across bins of signal and background samples.

$$V_{ij}^{Frac} = \frac{V_{ij}}{n_i^{CV} n_j^{CV}} \tag{9.6}$$

9.1.3 Signal Uncertainty Sources

For the HNL signal, there are four primary sources of uncertainties that will effect physics measurements by SBND: (1) statistical, (2) cosmic mis-tagging, (3) flux and (4) detector. The statistical uncertainty is evaluated using the number of signal slices selected after the selection presented in Chapter 8, of which the following plots shows the distribution after the lenient selection. Meanwhile, the cosmic mis-tagging uncertainty is evaluated using the number of cosmic muon slices occurring in the same readout window of HNLs. The impact of the uncertainties due to flux modelling can be measured using the reweighting method. The detector systematic however is not reweightable and requires a full simulated and reconstructed MC sample for assessing uncertainties due to detector as previously explained in Sec. 9.1.1. At the time of writing, the SBND detector is not yet operational, and it is undetermined which detector parameters are impactful on physics measurements. Thus, detector systematics are not included in the limit setting presented here, but should be included in future iterations of this analysis.

The statistical uncertainty describes the statistical fluctuation of the MC sample. It is computed using the number of HNL signal slices remained after the selection binned to

the beam bucket distribution. The statistical uncertainty of each bin of the distribution is defined as follows

$$\sigma_{statistics} = \sqrt{\text{Number of entries in the bin}} \qquad (9.7)$$

The uncertainty is computed before the normalisation to the POT exposure of 3 years data taking. After the normalisation, the resulting statistical uncertainty is plotted in Fig. 9.1a, showing the uncertainty is minimal across the entire distribution. The fractional statistical uncertainty, as shown in the blue line in the bottom figure of Fig. 9.2, shows that it is is well-constrained under < 6%.

The cosmic mis-tagging uncertainty is to account for the selected slices that are cosmic rays occurring in the same readout window as HNLs. In some cases, high energetic cosmic rays can overlap with showers from HNLs, resulting in a pile-up of charge clusters at the same location in the detector. Drifting electrons arriving at the same wires and at the same time can mislead the clustering process of Pandora during reconstruction, such that a reconstructed shower object might contain more charges from cosmic rays then HNLs. This reconstruction failure results in a number of cosmic slices remained after the selection, and therefore, is considered as the cosmic mis-tagging uncertainty. The cosmic mis-tagging uncertainty is plotted in Fig. 9.1b, however, the uncertainty is so small that it is not visible on the plot. The uncertainty can be seen in the fractional format in the bottom figure of Fig. 9.1 in the gray line, demonstrating it is almost negligible at < 1%.

The flux systematic was assessed using the reweighting method. The flux prediction and the reweighting framework for assessing the uncertainties were both developed by the MiniBooNE experiment, and furthered by the MicroBooNE experiment. The framework has now been implemented for SBND, called SBN Event Weight [3], intended for a consistent reweighting method across the experiments in the SBN program, including SBND, ICARUS and MicroBooNE. The flux systematic uncertainties are as follows

- **Proton Delivery**: The proton intensity on the target of the BNB is measured using two steroids, which have an uncertainty of 2% attributed to calibration.

- **Production**: The number of secondary mesons produced in the BNB target has an uncertainty associated with each particle type. The model predictions for $\pi^{\pm}$, $K^{\pm}$ and K_L^0 mesons are previously detailed in Sec. 3.3. The main meson parent of HNLs is K^+, of which the production is extrapolated from the global data of K^+ using the Feynman Scaling to the relevant BNB energy range, and futher constrained by the SciBooNE's direct measurement of K^+ from the BNB.

- **Hadronic Interactions**: Hadrons produced in the target may interact with the Beryllium target elastically or inelastically, effecting the kinematics of secondary mesons and consequently, tertiary daughter particles like HNLs and neutrinos. The uncertainty of interaction cross sections of $p/n + Be$ and $\pi^{\pm} + Be$ is propagated through the flux prediction. The

- **Horn Magnetic Field**: The magnetic field of the horn impacts the focusing of charged particles produced in the target. Uncertainties associating with the horn magnetic field including the current pulsed through the horn and the skin current induced on the surface of the target are computed.

Fig. 9.1c shows the combined uncertainty of all the flux systematics listed here for HNLs, where the uncertainty is consistent across every bin. The flux fractional uncertainty is plotted in the green line in the bottom figure of Fig. 9.2 that it is well-constrained < 8%.

Finally, the top figure of Fig. 9.2 shows the total uncertainty combining the statistical, cosmic mis-tagging and flux uncertainties combined. It can be seen that the uncertainty is evenly distributed across the beam bucket distribution with minimal biases in any bins. The total fractional uncertainty is shown in the purple line in the bottom figure of Fig. 9.2, showing that it is well-constrained < 10% across the full distribution.

9.1.4 Background Uncertainty Sources

For the background of SM neutrinos and cosmic muons, there are four primary sources of uncertainties: (1) statistical, (2) flux, (3) SM neutrino cross section and (4) detector. The uncertainties due to statistic and flux are computed in the same manner as the uncertainty treatment of the HNL signal. The detector systematic uncertainty is also currently not included but should be included in the future work. A new addition is the neutrino cross section uncertainty, of which the impact due to the cross section modelling is measured using the reweighting method.

The statistical uncertainty of the background was assessed using the number of SM neutrino and cosmic muon slices after the selection. The statistical uncertainty of each bin is computed using Eq. 9.7. The statistical uncertainty of background is shown in Fig. 9.3a. It is evident that the background statistics is abundant for bins at the centre of the beam bucket however, limited for bins at the edge. Therefore, the statistical uncertainty is better constrained for bins at the centre, while bins at the edge have a higher statistical fluctuation. This is visually evident in the statistical fractional uncertainty plotted in the blue line in the

(a) Statistical

(b) Cosmic Mis-tagging

(c) Flux
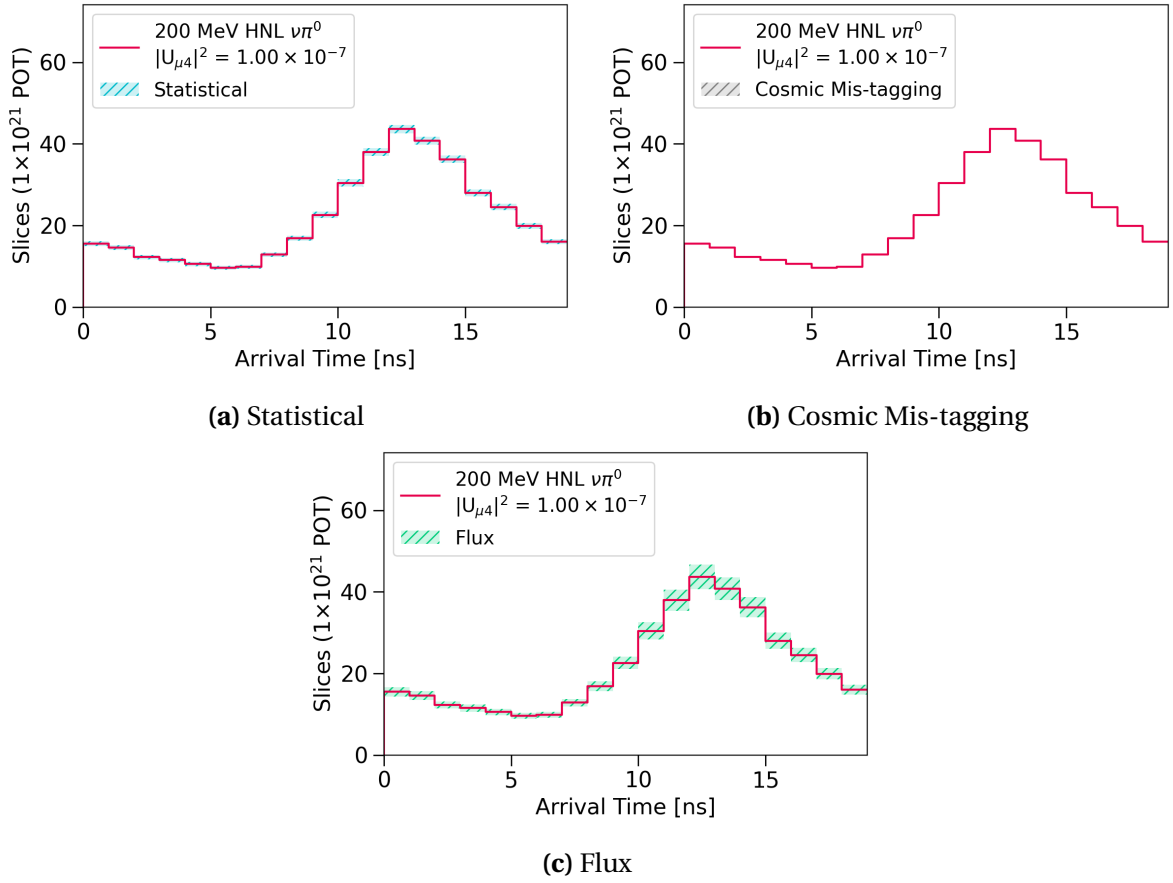
Fig. 9.1 Plot showing different sources of uncertainty (bottom) of HNLs.
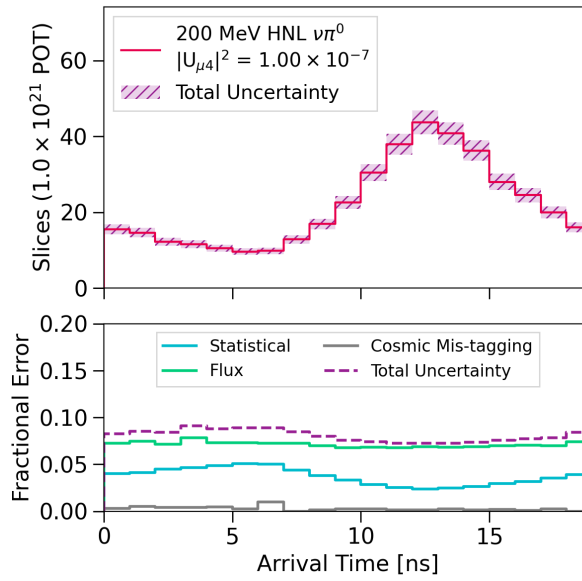
Fig. 9.2 Plot showing the total uncertainty (top) and fractional uncertainties (bottom) of HNLs.

169

bottom figure of Fig. 9.4. For bins at the centre of the beam bucket, the statistical fractional uncertainty is constrained at < 20% while, for bins the edge of the beam bucket distribution, particular the first and last 4 bins, where the statistical fractional uncertainty reaches as high as 100%.

The flux systematic of the background was measured using the reweighting method similarly as HNLs, as discussed in the previous Sec. 9.1.3. One key difference in the flux systematics between the signal and the background is the uncertainties due to the secondary meson production in the BNB. Unlike HNLs coming from only K^+, SM neutrinos can result from $\pi^{\pm}$, $K^{\pm}$ and K_L^0 as previously plotted in Fig. 3.18. Particularly, the Sanford-Wang parametrisation for modelling the π^+ production introduces biases for NC π^0 interactions with energy < 300 MeV, which is the main background contributor. Due to the selection requiring very high energetic showers, this bias is mitigated. The resulting flux uncertainty of the background is plotted in Fig. 9.3b and the flux fractional uncertainty is plotted in the green line in the bottom figure of Fig. 9.4. The flux fractional uncertainty is very well-constrained < 20% and consistent across the entire beam bucket distribution.

The SM neutrino cross section uncertainty was assessed using the SBN Event Weight, with systematic variable inputs from the GENIE generator. There are two types of weights for SM neutrino interaction cross section, weights associated with a group of correlated parameters and weights associated with a single non-correlated physics parameter. The SM neutrino systematic for a group of physics parameters are as follows

- **Charged Current Quasi-Elastic Scattering (CC-QE)**: Coefficients of the Z expansion of the axial form factor for CC-QE interactions are varied.

- **Deep Inelastic Scattering (DIS)**: Parameters and correction factors of the Bodek-Yang model, which is used for modelling DIS cross sections, are varied.

- **Neutral Current Elastic Scattering (NC-EL)**: The axial mass and the strange axial vector of the dipole form factor of NC-EL interactions are varied.

- **Neutral Current Resonant Scattering (NC-RES)**: The axial mass and the strange axial vector of the dipole form factor of NC-RES interactions are varied.

- **Charged Current Resonant Scattering (CC-RES)**: The axial mass and the strange axial vector of the dipole form factor of CC-RES interactions are varied.

- **Hadron Transport Interactions**: The mean free path, inelastic scattering, absorption, and pion production cross section are varied for both pions and nucleons.

The SM neutrino systematic uncertainties for a single non-correlated physic parameter are as follows

- **Charged Current Quasi-Elastic Scattering (CC-QE)**: The shape of vector form factor, the random phase approximation and the strength of the Coulomb corrections of CC-QE interactions are varied.

- **Meson Exchange Current Interactions (MEC)**: The decay angle and the normalisation factor of CC-MEC and NC-MEC are varied.

- **Coherent Scattering (COH)**: Normalisation factors of NC-COH and CC-COH cross section interaction are varied.

- **NonRES Backgrounds**: Non-resonant backgrounds are varied for CC and NC, ν and $\bar{\nu}$, neutron and proton, 1π and 2π for a total of 16 systematic parameters.

- **Angular Distribution**: The angular distribution of γ and π are varied.

- **Branching Fraction**: The branching fraction scale factor for resonant decays with either a single γ or π are varied.

Fig. 9.3c shows the combined uncertainty of all the SM neutrino interaction cross section systematics. Compared to statistical and flux uncertainties, the magnitude of the cross section uncertainty is significantly larger. This is due to the primary background after selection is NC π^0, of which the cross section is not well-measured. Uncertainties associated with NC-COH and NC-RES scattering are the main contributors for this channel. Moreover, a small fraction of CC ν_e interactions remains after selection and the cross section is also not well-measured. For this channel, uncertainties associated CC-QE scattering systematics contributes the most. Finally, uncertainties for modelling hadron transport interactions and NonRES background of NC1π interactions are significant contributors to the final cross section uncertainty.

The cross section fractional uncertainty is plotted in the orange line in the bottom figure of Fig. 9.4. For bins at the centre of the beam bucket distribution, the uncertainty is the lowest in the entire distribution at < 50%. The uncertainty increases towards the bins at the edge of the beam bucket, reaching more than 100% for very low statistics bins. The total fractional uncertainty combining the statistical, flux and cross section uncertainty, is plotted in the purple line. For bins at the edge of the bucket, which are the bins driving the sensitivity limits, the total uncertainty is particular high mostly due to the contribution of the statistical and cross section uncertainties.

(a) Statistical

(b) Flux

(c) SM Neutrino Cross Section
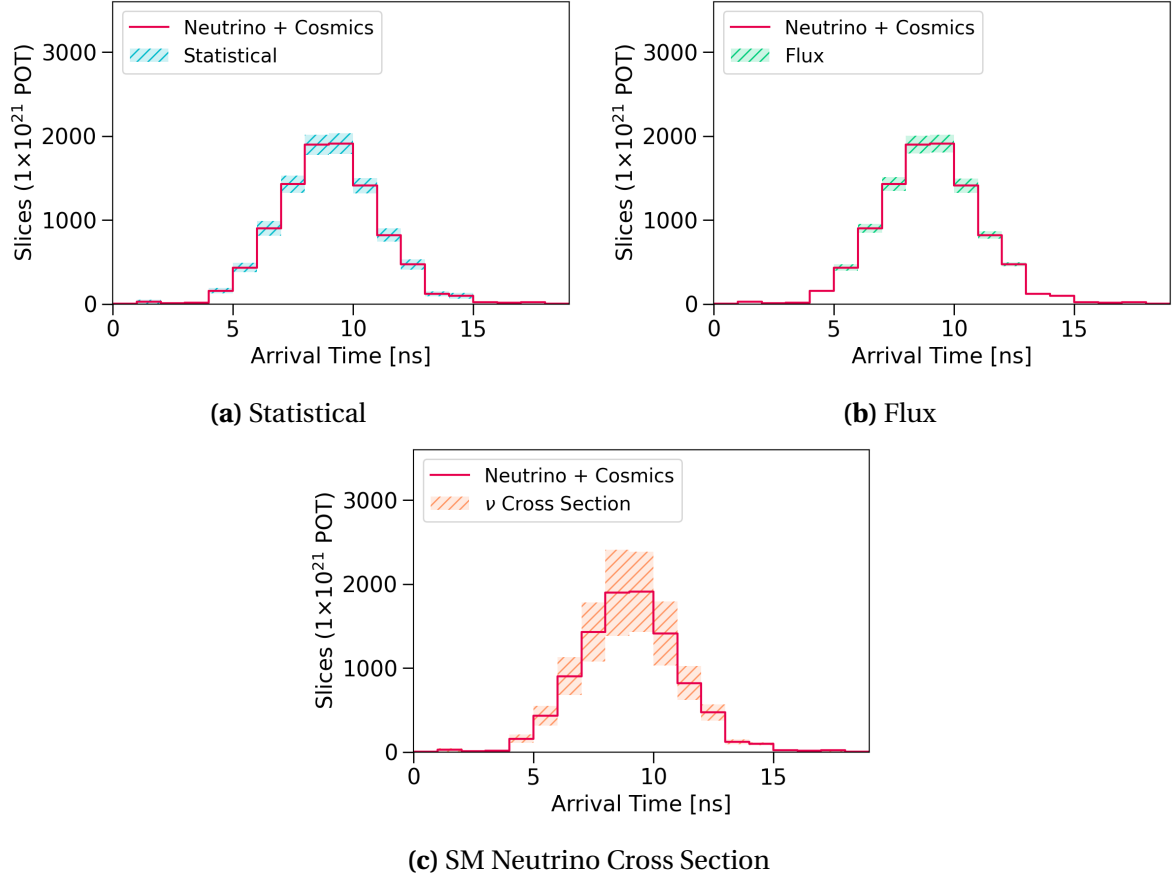
Fig. 9.3 Plot showing different sources of uncertainty (bottom) of backgrounds.
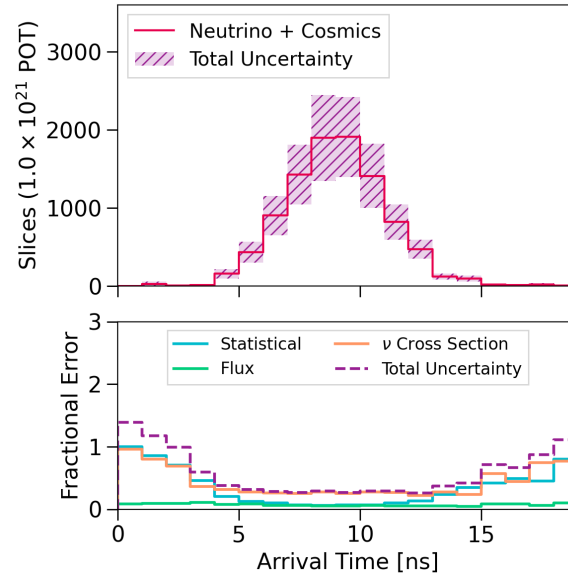
Fig. 9.4 Plot showing the total uncertainty (top) and fractional uncertainties (bottom) of backgrounds.

9.2   Limits Setting Procedure

Setting the sensitivity limit employs the likelihood-based hypothesis test [1], of which likelihood functions are constructed using the beam bucket distribution of signals and backgrounds. The end-to-end limit setting procedure was performed using the `pyhf` package [4, 5]. The following section will detail the hypothesis testing done under the hood of the `pyhf` package. It begins with defining the hypotheses for exclusion limits, then constructing the likelihood-based probability distribution for each hypothesis and finally, computing the modified p-value using the CL_s method. Validation of the result acquired from the *pyhf* package and the construction of likelihood functions using *pyhf* are also presented.

9.2.1   Hypothesis Definition

The setting limit procedure employs the frequentist approach by performing hypothesis testing to quantify the level of agreement between the observed data and a given hypothesis H. To exclude a signal region, the test hypothesis H_b is defined as describing only known processes, or a background-only model. Meanwhile, the null hypothesis H_{s+b} is defined as the model including both background and signal processes.

A given hypothesis H is constructed to represent the expectation value for an observable distribution, which is chosen to be the beam bucket distribution. For a series of observations n, the expectation value the ith bin can be constructed as

$$E[n_i] = \mu s_i + b_i \tag{9.8}$$

where s_i and b_i are the number of entries from the signal and background distributions for bin i. The *parameter of interest* μ determines the strength of the signal process, where $\mu = 0$ corresponds to the background-only H_b hypothesis and $\mu = 1$ corresponds to the nominal signal H_{s+b} hypothesis. It is an unconstrained parameter and is written separately from other parameters in the hypothesis. In the HNL search, varying the μ parameter is equivalent to varying the mixing angle $|U_{\mu 4}|^2$. Other parameters in the hypothesis characterising the shape of the distributions are grouped together as *nuisance parameters*, denoted as θ. Then, the null and test hypotheses can be formally written as

$$Null\ hypothesis:\ H_{s+b} = H(\mu = 1, \theta) \tag{9.9}$$

$$Test\ hypothesis:\ H_b = H(\mu = 0, \theta) \tag{9.10}$$

173

9.2.2 Likelihood-based Test Statistic

To exclude the null hypothesis H_{s+b} at some confidence levels, a test statistic is performed for a hypothesised μ against H_{s+b} and H_b. To construct a test statistic for a multi-binned histogram, likelihood-based functions are chosen. The likelihood function is the product of Poisson probabilities of all bins in the histogram [1]

$$L(\mu, \theta) = \prod_{i=1}^{N} \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)} \prod_{\theta \in \theta} c_\theta(a_\theta | \theta) \qquad (9.11)$$

The first product describes the likelihood of the histogram n with bin index i, which is constructed using the beam bucket distribution of HNLs as signals and SM neutrinos with cosmics as backgrounds. The second product is a constraint term $c_\theta(a_\theta | \theta)$ with measurement a_θ constraining the nuisance parameter θ. Individual constraint term is added for each uncertainty of the signal and background respectively, as previously discussed in Sec. 9.1.

For a hypothesised value of μ, the profile likelihood ratio can be constructed from the likelihood function as [1]

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} \qquad (9.12)$$

The numerator is the *conditional* maximised likelihood for the hypothesised value μ and $\hat{\hat{\theta}}(\mu)$ is a function of μ. The denominator is the *unconditional* maximised likelihood, where $\hat{\mu}$ and $\hat{\theta} = \hat{\hat{\theta}}(\hat{\mu})$ are the best fit parameters to the observed data.

A modification in the likelihood ratio is required for cases where the signal process only increases the mean event rate, such that $\mu \geq 0$. If the observed data results in the best fit $\hat{\mu} < 0$, then the best level of agreement between the data and the prediction is forced to occur at $\hat{\mu} = 0$. This is equivalent to the late arrival of HNLs relative to the SM neutrino beam bucket, which can only increase the event rate for bins at the edge of the beam bucket. The profile likelihood ratio is modified as [1]

$$\tilde{\lambda}(\mu) = \begin{cases} \dfrac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(0, \hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\[3mm] \dfrac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})} & \hat{\mu} \geq 0 \end{cases} \qquad (9.13)$$

where $\hat{\hat{\theta}}(0)$ is a function of $\mu = 0$.

From the modified profile likelihood ratio, the test statistic for setting upper limits is defined as [1]

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \le \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0. \\ -2\ln\frac{L(\mu,\hat{\theta}(\mu))}{L(\hat{\mu},\hat{\theta})} & 0 \le \hat{\mu} \le \mu, \\ 0 & \hat{\mu} > \mu \end{cases} \tag{9.14}$$

In the region $\hat{\mu} > \mu$, equivalent to an upward fluctuation in the data compared to the prediction, the test statistic $\tilde{q}_\mu$ is set to 0. For setting an upper limit, this fluctuation does not represent less incompatibility and therefore is not taken into the rejection region of the test. A larger value of $\tilde{q}_\mu$ corresponds to an increasing disagreement between the data and the prediction for the hypothesised μ.

The distribution $f(\tilde{q}_\mu|\mu')$ of a test statistic $\tilde{q}_\mu$ can be interpreted as a PDF. The subscript of $\tilde{q}$ refers to the value of μ being tested in the numerator of the likelihood ratio in Eq. 9.12. The second argument μ' refers to the value of μ being assumed in the denominator of Eq. 9.12. For setting an upper limit with hypotheses testing, the PDF of interest is $\tilde{q}_\mu$ under the assumption of a different strength parameter $\mu' \ne \mu$. Therefore, the corresponding PDFs for a test signal strength μ, under the assumption of the null hypothesis H_{s+b} and the test hypothesis H_b, are respectively as

$$f(\tilde{q}_\mu|s+b) = f(\tilde{q}_\mu|\mu' = 1) \tag{9.15}$$

$$f(\tilde{q}_\mu|b) = f(\tilde{q}_\mu|\mu' = 0) \tag{9.16}$$

9.2.3 The CL$_s$ Method

To quantify the level of significance of a test statistic, a p-value is computed as an integration of the $\tilde{q}_\mu$ PDF, on either side of $\hat{q} = \tilde{q}_\mu(\mu = \hat{\mu})$ corresponding to the value of the test statistic for the observed data. Typically, one might only consider the p-value resulting from the null hypothesis H_{s+b}, denoted as p_{s+b}, to exclude the null hypothesis. This represents the probability of finding data more background-like under the assumption that the signal is true, often known as the probability of getting the type I error in hypothesis testing.

However, this approach does not factor in the probability of finding data under the assumption of the background-only hypothesis. In scenarios where the test statistic is not sensitive to the prediction models, the PDFs of the H_{s+b} and H_b hypotheses would greatly overlap each other. One might make the mistake of interpreting having observed data highly

contaminated with background-like events as a statement on the nominal signal hypothesis [6]. Therefore, a penalty on poor background modelling is necessary to account for how sensitive the test statistic is to the prediction model.

The CL_s method is a modified frequentist method to compute a modified p-value. This statistical method was developed specifically by particle physics experiments for the purpose of discovery as well as exclusion. This is a conservative approximation of the confidence level to account for poor background modelling and/or statistical fluctuations. Using the CL_s method, the p-values for the hypotheses H_{s+b} and H_b are defined as follows [1, 6, 7]

$$p_{s+b} = \int_{\hat{q}}^{\infty} f(\tilde{q}_\mu | s + b) d\tilde{q}_\mu \tag{9.17}$$

$$p_b = \int_{-\infty}^{\hat{q}} f(\tilde{q}_\mu | b) d\tilde{q}_\mu \tag{9.18}$$

p_{s+b} represents the probability of finding data more background-like under the assumption of the signal+background hypothesis H_{s+b}. p_b represents the probability of finding data more signal-like under the assumption of the background-only hypothesis H_b.

The modified p-value is then computed from the p-value for each hypothesis as [1, 6, 7]

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p_{s+b}}{1 - p_b} \tag{9.19}$$

Since $(1 - p_b)$ only varies between 0 and 1, the resulting CL_s is always greater or equal to p_{s+b}. Thus, the CL_s is more conservative in quantifying the confidence level of hypothesis testing. The confidence threshold is chosen to be 0.1, such that $CL_s < (1 - 0.9)$, equivalent to setting an upper limit with a confidence level of 90%.

9.2.4   Computing Test Statistic Distributions

One way to obtain the PDF of $\tilde{q}_\mu$ for a given hypothesis is by throwing toys (or pseudo-experiments) sampled from the observables n, with constrained nuisance parameters θ. Each toy result represents a potential outcome of $\tilde{q}_\mu$ and thus, the toy distribution represent the test statistic $\tilde{q}_\mu$ distribution. For a study using MC samples without having the observed data to determine $\hat{q}$, one can consider the distribution under the assumption of the background-only hypothesis H_b. In this case, $\hat{q} = \hat{q}_{expected}$ represents the *expectation* given the model prediction and is taken is the median of the H_b test statistic distribution, equivalent to assuming the observed data is the same as the background-only prediction.

Fig, 9.5 shows an example PDF of $\tilde{q}_\mu$ for a test signal strength μ acquired from 4000 toys thrown. The $\tilde{q}_\mu$ PDF under the assumption of H_{s+b} and H_b is shown in green and yellow respectively. The value $\hat{q}_{expected}$ is plotted as the dashed red line. The area under the curve $f(\tilde{q}_\mu|s+b)$ to the right of $\hat{q}_{expected}$ gives the median p_{s+b} and the area under the curve $f(\tilde{q}_\mu|b)$ to the left gives the median p_b. The CL_s value is then calculated to give the expected limit on the signal strength μ. To identify which μ value results in $\mathrm{CL}_s < 0.1$, a range of μ is considered until the desired value CL_s is achieved. The test signal strength μ plotted here gives exactly $\mathrm{CL}_s < 0.1$.
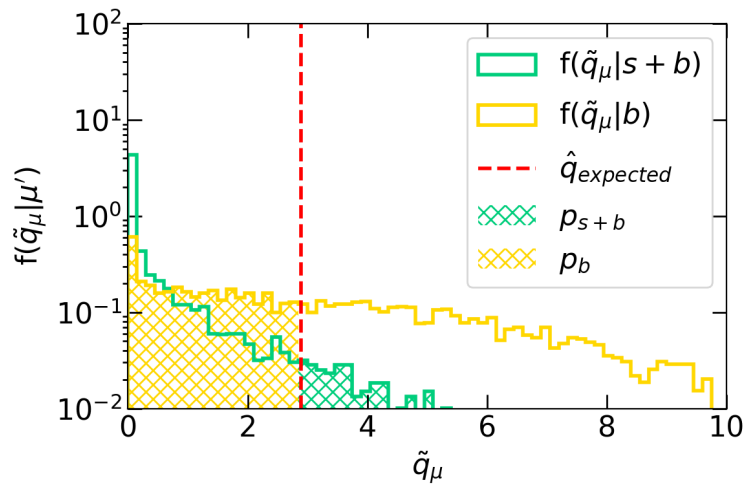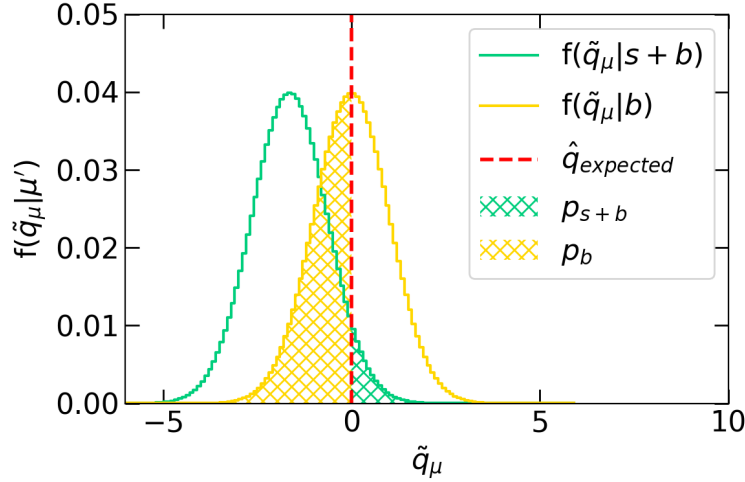

Fig. 9.5 Example of $\tilde{q}_\mu$ distributions using the toy throwing approach.

An alternative method to calculate the test statistic $\tilde{q}_\mu$ and its PDF is by asymptotic approximation. This approach assumes that $\hat{\mu}$ follows a Gaussian distribution around a mean of μ' with a standard deviation of σ. Via the `pyhf` package, PDFs are computed using the `HistFactory` module according to the formulae provided by Ref. [1]. Similar to the toy throwing approach for an MC study, $\hat{q}_{expected}$ is also taken as the median of the background-only H_b test statistic distribution. To determine $\mathrm{CL}_s < 0.1$, `pyhf` scans over a range of μ and the best value is calculated using interpolation.

Fig. 9.6 shows an example PDF of $\tilde{q}_\mu$ computed using the asymptotic approximation. The expected $\hat{q}_{expected}$ is plotted as the dashed red line. The median p-values under the H_{s+b} and H_b hypotheses are plotted as shaded green and yellow areas respectively. The test signal strength μ shown here gives $\mathrm{CL}_s < 0.1$. Standard deviations of the CL_s can also be computed by substituting μ' with $\mu' \pm N\sigma$, where $N\sigma$ is the number of standard deviations away from the mean μ'. A validation was also carried out to compare the signal strength μ

resulting from the toy throwing and asymptotic approximation. Both approaches returned μ values within 5% of each other, showing a good agreement.

Fig. 9.6 Example of $\tilde{q}_\mu$ distributions using the asymptotic approximation.

9.2.5 Setting The Upper Limits

The setting limit procedure is performed fully using the `pyhf` package, with the steps described in the previous section. The beam bucket distribution is input into the `pyhf` package to infer the upper limit at 90% confidence level. The example histograms for the background-only H_b and the signal+background H_{s+b} hypothesis are as following:

- H_b: The SM neutrinos and cosmic beam bucket distribution.

- H_{s+b}: The HNL beam bucket distribution on top of the background distribution.

These histograms are the ingredients for the first product in Eq. 9.11. Since the likelihood functions are products of individual bins of the histograms, one can reduce the computing time by considering only high signal-to-noise bins that contribute the most to the sensitivity when performing limit setting. As previously stated as the timing cut, the relevant high signal-to-noise bins are the first and last 4 bins of the beam bucket distribution, which lead to the same result as demonstrated later in Sec. 9.3. Appendix **??** and **??** contains the beam bucket distributions for signals and backgrounds after the lenient and stringent cut respectively, with all the uncertainties propagated for the limit setting. Additionally, appendix **??** contains the smeared truth distributions under the assumption of timing reconstruction improvement as discussed in Sec. 8.8, also used in the same limit setting.

The uncertainties of the beam bucket distribution discussed in Sec. 9.1 are the constraints on the nuisance parameters $c_\theta(a_\theta|\theta)$ in the second product Eq. 9.11. Different types of constraints can take different statistical shapes, which are called *modifiers* in the pyhf package [4, 5]. Table 9.1 summarises the uncertainties and their corresponding modifiers of the signal and background distribution.

For the signal distribution, the assumption is that the signal rate is relatively low and expected to follow a Poisson distribution. Similarly, the cosmic mis-tagging rate falls under the same assumption since it is proportional to the signal rate. Thus, the modifier for both statistical and cosmic mis-tagging uncertainties of the signal distribution follw a Poisson shape, treating each bin uncorrelated. The flux uncertainty of the signal distribution uses a Gaussian-shaped modifier instead, to enable correlation bin-wise.

For the background distribution, the statistical uncertainty needs to employ a lite version of the Beeston-Barlow method [8] for treating statistical fluctuations due to finite statistics. The modifier in this case follows a Gaussian shape for bins with high statistics and falls back to a Poisson shape for bins with low statistics, while treating each bin uncorrelated. Moreover, the flux and SM neutrino cross section uncertainty also use a Gaussian-shaped modifier, however, with correlation bin-by-bin.

	Uncertainty	Modifier	Sample Correlation	Bin Correlation
Signal	Statistical	Poisson	False	False
	Cosmic mis-tagging	Poisson	False	False
	Flux	Gaussian	False	True
Background	Statistical	Gaussian	False	False
	Flux	Gaussian	False	True
	SM Neutrino Cross Section	Gaussian	False	True

Table 9.1 Table summarising modifiers used to constrain uncertainties.

For a single mass point, pyhf performs a scanning of signal strength μ until the desired value giving the 90% CL is determined. To infer from the signal strength μ to the mixing angle $|U_{\mu4}|^2$, the proportionality of the number of signals N_{HNL} to the mixing angle is employed. The signal rate observed at the detector is proportional to the mixing angle at the HNL production as well as the mixing angle at the HNL decay, and thus, $N_{HNL} \propto |U_{\mu4}|^4$. The

upper limit of the mixing angle is then $\sqrt{\mu}$ multiplied by the input mixing angle $|U_{\mu4}|^2$. This process is repeated across all the mass points.

9.3 Results

The following section presents the sensitivity results acquired from three different beam bucket distributions. The first two distributions were acquired from using fully reconstructed MC samples mimicking data, of which one was applied with a lenient selection as shown in Fig. 9.7a and the other one was applied with a stringent selection as shown in Fig. 9.7b. The third distribution is the smeared truth as shown in Fig. 9.7c, which was constructed under the assumption of a better timing reconstruction resolving the beam bucket with a resolution of 1.73 ns. Unlike, the reconstructed distributions with full uncertainties propagated, the smeared truth distribution does not contain any uncertainties for simplification. Moreover, the first and last 4 bins of of every distribution are also shown to demonstrate the high signal-to-background ratio of these bins.

9.3.1 Comparison Across Different Beam Bucket Distributions

The sensitivity limits for Majorana HNLs are presented in Fig. 9.8, comparing results from fitting the beam bucket distribution with the lenient cut and the stringent cut. The expected limit is plotted as solid black line. The deviation band at 1 and 2σ away from the expectation, also known as the *Brazil* band, is plotted in shaded green and yellow respectively. The stringent cut results in a more competitive limit than the lenient cut. For example, examining at the same mass point of 260 MeV, the expected limit on the mixing angle $|U_{\mu4}|^2$ from the lenient cut is only ??, while the result from the stringent cut reaches ??. This is due to the stringent cut distribution having better signal-to-background ratio for bins at the edge of the beam bucket. This is evident when comparing between Fig. 9.7a and 9.7b, where the signal rate is very similar across the two distributions but the background rate is much lower in the stringent cut than the lenient cut. Particularly, the first two bins of the stringent cut distribution are background-free, thus, driving the limit significantly.

However, the magnitude of the Brazil band of the stringent result is larger than the lenient result, indicating that the stringent result has larger uncertainties. This is likely due to the limited statistics of the background MC samples after the stringent selection. As can be observed from bins at the edge of the beam bucket distribution shown in Fig. 9.7b, the lack of background statistics might be insufficient describing the underlying distribution,
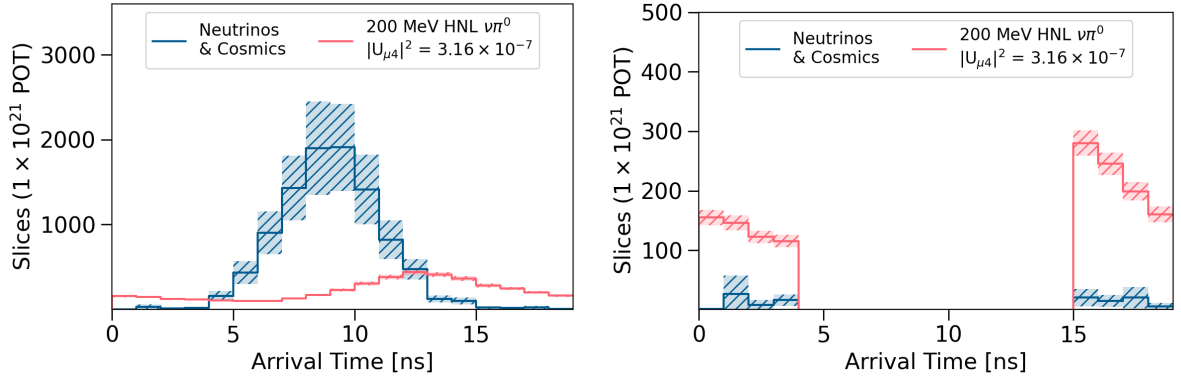
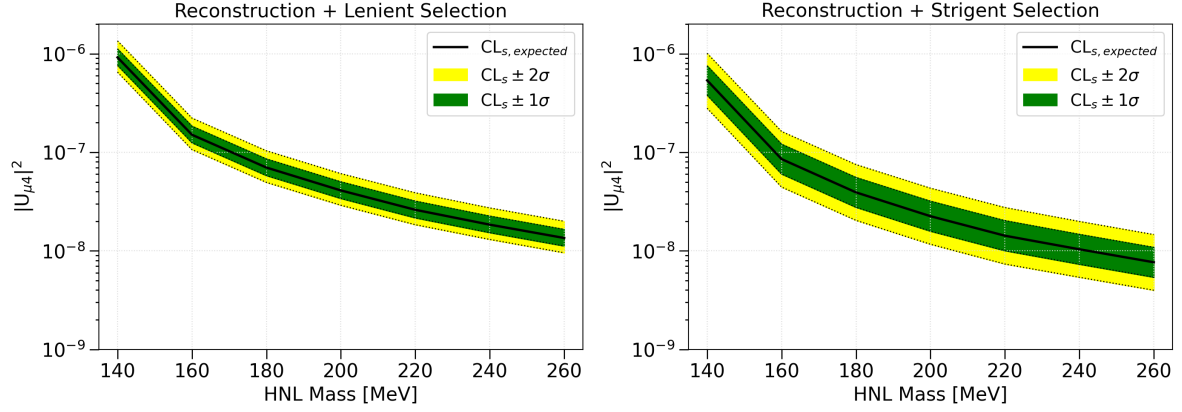**(a)** The Lenient Distribution
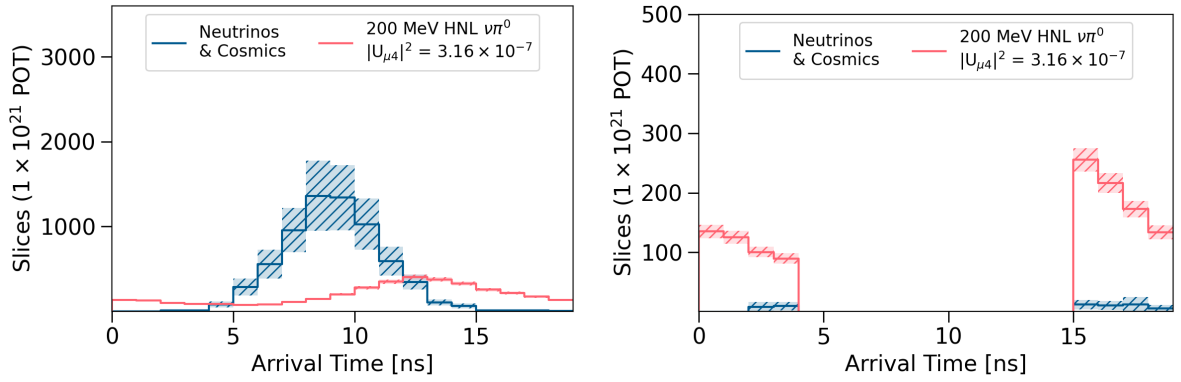
(b) The Stringent Distribution
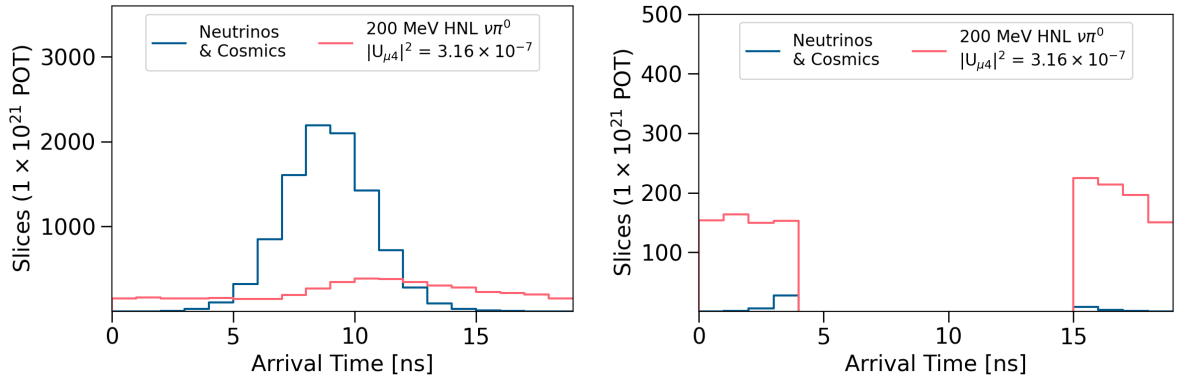

(c) The Smeared Truth Distribution

Fig. 9.7 Plots showing the beam bucket distributions used for the limit setting: the lenient cut (top), the stringent cut (middle) and the smeared truth (bottom).

which can lead to large statistical fluctuations. Although the stringent distribution leads to a more competitive result, a future iteration of this selection is recommended to perform on larger statistics MC samples so that the statistical uncertainty can be better constrained.

Fig. 9.8 Plots showing the expected upper limits and the Brazil bands on the mixing angle $|U_{\mu 4}|^2$ for Majorana HNLs, for the lenient (left) and stringent distribution (right).

Moreover, Fig. 9.9 shows the comparison between limit setting using the entire beam bucket distribution and only the first and last 4 bins, which were previously referred to as the timing cut in Chapter 8. The sensitivity limits are the same for both cases. The result demonstrates that the first and last 4 bins are the highest signal-to-background ratio bins, and therefore are the main contributor to the limits. This provides a useful insight for future iterations of this analysis indicating the region on the beam bucket distribution should
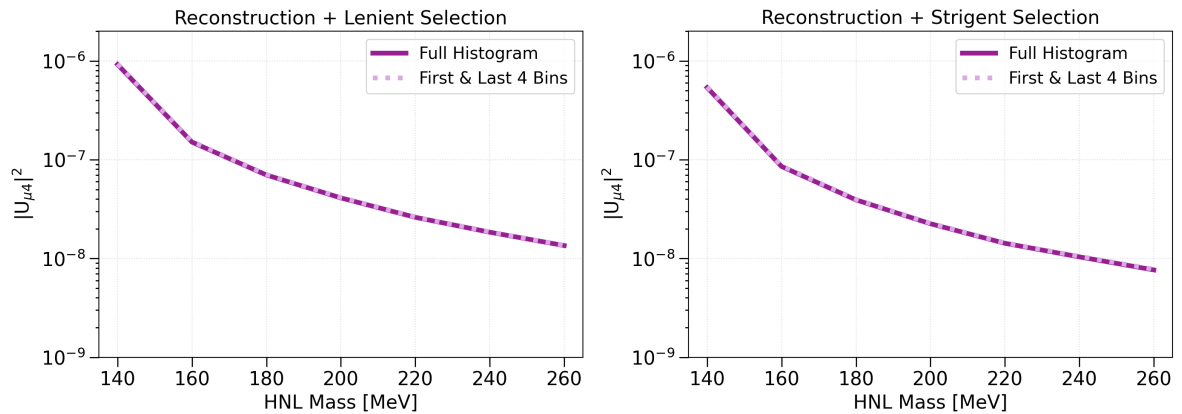
Fig. 9.9 Plots showing the upper limits on the mixing angle $|U_{\mu 4}|^2$ for Majorana HNLs, comparing results from limit setting using the entire histogram or only the edge bins of the lenient (left) and stringent distribution (right).

be focused on when performing selection. A recommendation is to focus on optimising the signal-to-background only in the edge region. Another recommendation is to develop different selections for different regions of the bucket, capitalising on their different signal-to-background ratios.

Finally, Fig. 9.10 shows the sensitivity limits from all three beam bucket distributions. The result from the smeared truth distribution is significantly more competitive than the results using reconstructed distributions. As detailed in Sec. 8.8, the smeared truth distribution was produced to answer the hypothetical question exploring the impact on sensitivity limits due to timing resolution improvement. The background beam bucket distribution of the smeared truth follows a Gaussian with a sigma of 1.73 ns. In comparison, the reconstructed distribution after the lenient or stringent cut follows a Gaussian with a sigma of 2.0 ns. As evident on in Fig. 9.7c, particular for bins at the edge region, the background rate is lower while the signal rate is much higher in the smeared truth distribution than the reconstructed distribution. The resulting signal-to-background ratio is therefore higher for the edge bins, resulting in higher signal-to-noise ratio bins, driving the sensitivity limits more aggressively. This shows a positive outlook as an area for improvement for SBND in order to achieve more competitive results. Particularly, this can be done with a more sophisticated timing reconstruction that can resolve the beam bucket with higher resolution < 2 ns.
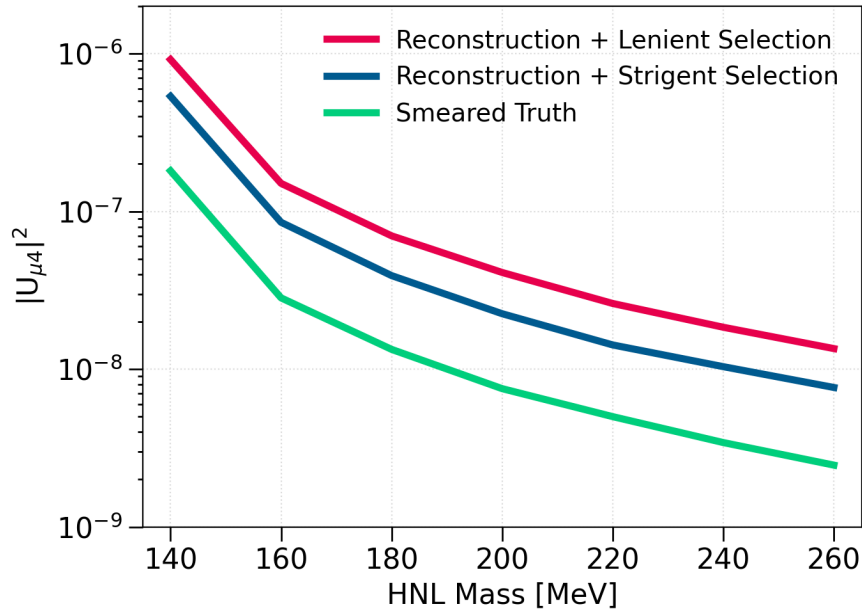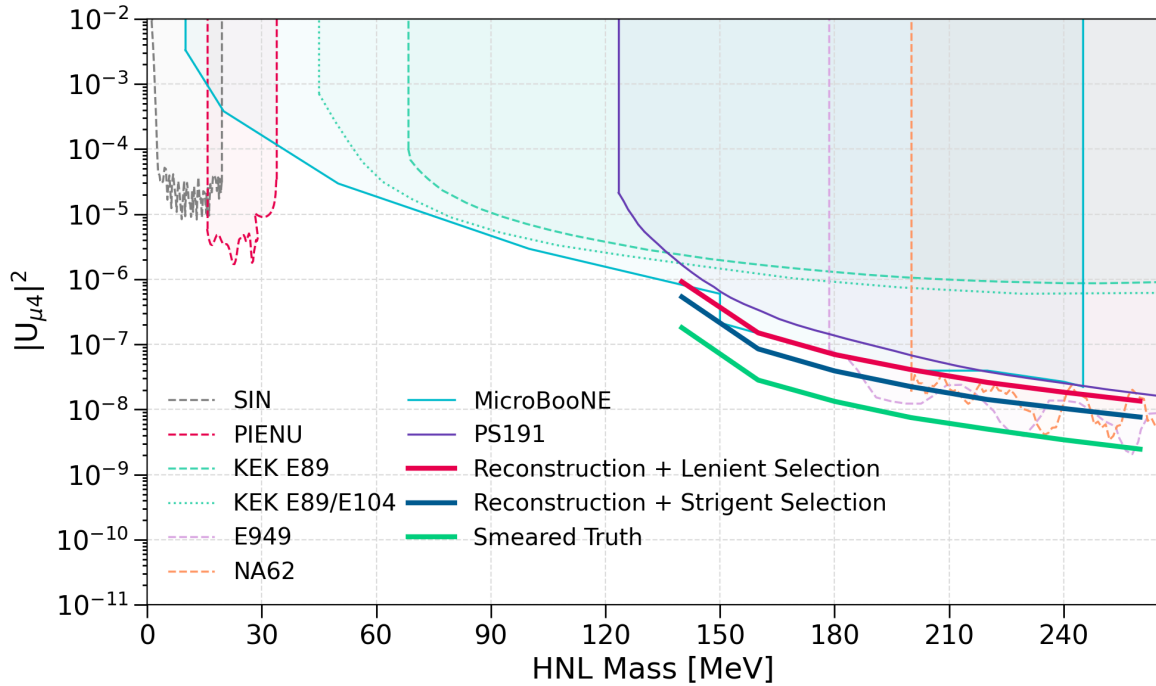

Fig. 9.10 Plots showing the upper limits on the mixing angle $|U_{\mu 4}|^2$ for Majorana HNLs, comparing results from the lenient and stringent reconstructed distributions and the smeared truth distribution.

9.3.2   Comparison With Other Experiments

Fig. 9.11 shows the expected limits of SBND on the mixing angle $|U_{\mu 4}|^2$ for Majorana HNLs, comparing against existing experimental results that were previously discussed in Sec. 1.4. For the limits acquired from the lenient and stringent distributions, as shown by the solid red and blue line respectively, they are comparable to existing limits. The lenient limit is almost the same as the limit achieved by the MicroBooNE experiment, as shown by the light blue line. On the other hand, the stringent limit is slightly more competitive than MicroBooNe, excluding new region of the mixing angle in the mass range < 200 MeV. However, in the mass range > 200 MeV, the phase space is already excluded by the E949 and NA62 experiment, as shown by the dashed pink and orang line. These two limits feature the first benchmark of SBND in this phase space of HNLs given the current detector and reconstruction capabilities.

The limit acquired from the smeared truth distribution is the most competitive out of the three results presented here. The result is projected to exclude a new region beyond the existing limits from the MicroBooNE, E949 and NA62 experiment. This demonstrates the potential of the physics capability of SBND, under the assumption of an exposure of

Fig. 9.11 Sensitivity contour of the upper limits on the mixing angle $|U_{\mu 4}|^2$ for Majorana HNLs within the mass range of $0 < m_N < 265$ MeV, comparing the expected result from SBND against other existing experimental results.

1×10^{21} POT and an exceptional timing reconstruction with a resolution of 1.73 ns. The POT exposure presented here is equivalent to 3 years of physics data taking, which allows for a lot of time and opportunities to work on improving the timing reconstruction of SBND to achieve the target resolution.

9.4 Concluding Remarks

Three limits on the mixing angle $|U_{\mu4}|^2$ of HNLs are presented in this chapter to demonstrate the range of the physics capability of SBND. The limits from the lenient and stringent beam bucket distribution demonstrate the *out of the box* result of SBND, which is currently comparable to existing experimental limits. The limit from the smeared truth distribution leads is the most competitive that can explore a new region not yet explored by other experiments. This limit also demonstrates the potential that SBND can achieve if the timing reconstruction is improved with a better timing resolution. Although ambitious, this goal is within reach in the next couple of years given the ongoing work towards a nanosecond timing resolution at SBND, one of which was presented in Chapter 4 and many more to be expected in the near future.

References

[1] G. Cowan et al., "Asymptotic formulae for likelihood-based tests of new physics", Eur. Phys. J. C **71**, [Erratum: Eur.Phys.J.C 73, 2501 (2013)], 1554 (2011).

[2] G. Cowan, *Statistical data analysis*, Oxford science publications (Clarendon Press, 1998).

[3] K. Duffy et al., "SBN Event Weighting module", (Visited 26/06/2024).

[4] L. Heinrich et al., *pyhf: v0.7.6*, version 0.7.6, https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6.

[5] L. Heinrich et al., "Pyhf: pure-python implementation of histfactory statistical models", Journal of Open Source Software **6**, 2823 (2021).

[6] A. L. Read, "Presentation of search results: The CL_s technique", J. Phys. G **28**, edited by M. R. Whalley et al., 2693–2704 (2002).

[7] T. Junk, "Confidence level computation for combining searches with small statistics", Nucl. Instrum. Meth. A **434**, 435–443 (1999).

[8] R. J. Barlow et al., "Fitting using finite Monte Carlo samples", Comput. Phys. Commun. **77**, 219–228 (1993).