

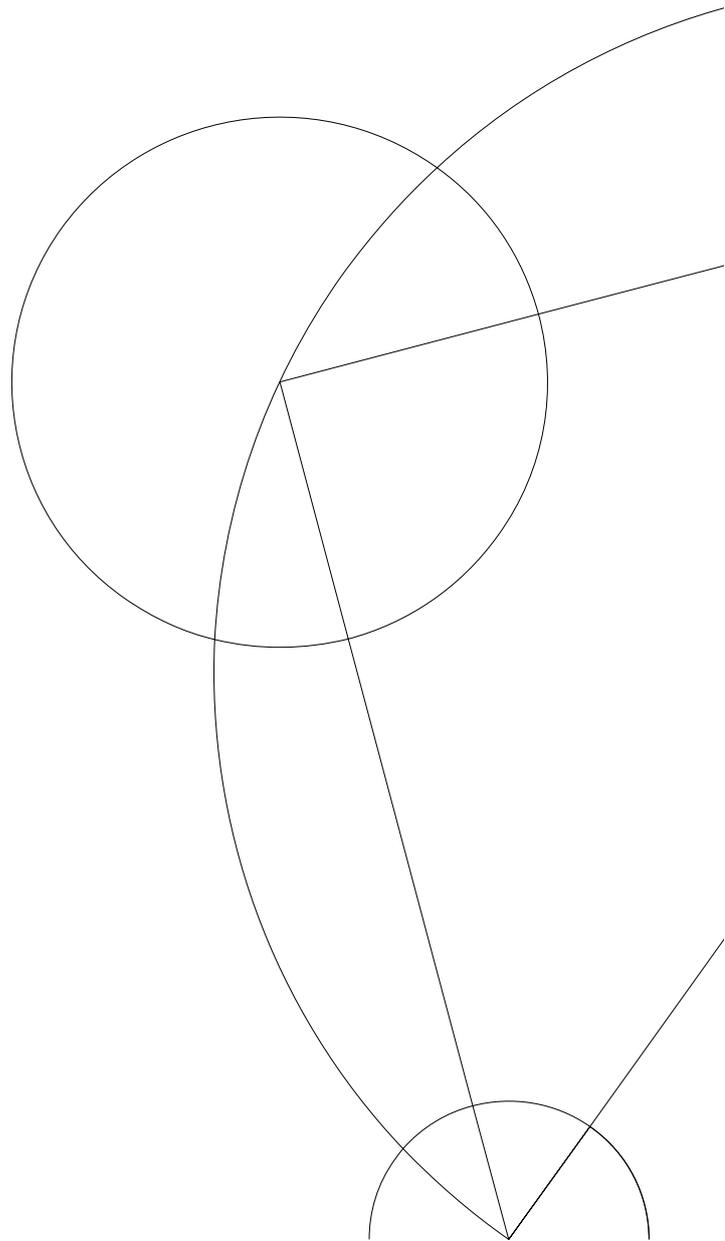


The search for right-handed neutrinos with the ATLAS experiment

Henriette Petersen
Master's Thesis in Physics
Supervisor: Assoc. Prof. Stefania Xella

Niels Bohr Institute
University of Copenhagen

Submitted:
September 4, 2017



Abstract

In this analysis it is explored what the ATLAS experiment can do to investigate the ν MSM parameters through the prompt trilepton decay channel of a right-handed neutrino. The case for mixing of a single right-handed and muon neutrino is considered here. No opposite-site sign same-flavour leptons is further required to focus on the Majorana nature of the neutrinos. This analysis uses Monte Carlo simulations for five HNL mass points (defined in terms of muon-sterile mixing angle and HNL mass) which match data sets recorded during 2015 and 2016 for a center-of-mass energy corresponding to $\sqrt{s} = 13$ TeV and an integrated luminosity of 36 fb^{-1} .

A cut-based analysis is investigated but can not be optimally exploited due to low statistics. A more intelligent approach to signal selection is therefore used in the form of a neural network. This was only done for the 50 GeV mass sample due to low statistics and the fact that this mass performed the worst in terms of significance in the cut-based optimization. The neural network approach was found to yield a significant improvement in significance and it is expected that other HNL masses will perform similarly. Furthermore, it is roughly evaluated which values of muon-sterile mixing angles can be excluded based on the given luminosity and significances obtained with the cut-based method. Good signal exclusion is obtained for lower masses but more statistics is needed for higher masses.

Acknowledgements

First of all I would like to thank my supervisor Stefania Xella for her guidance and much appreciated feedback. Furthermore, thanks is due to Fabian Thiele for his patience in answering all my technical questions. I'm also very grateful to Oleg Ruchayskiy for taking the time to answer my theoretical inquiries and last but not least, I would like to thank Alejandro Diaz and James Monk with whom I have shared very helpful discussions on the neural network approach developed in this thesis.

Contents

1	Introduction	2
2	Theory	3
2.1	The Standard Model	3
2.1.1	Chirality	4
2.1.2	CP-transformation	5
2.1.3	Beyond the Standard Model	5
2.2	Neutrino oscillations	6
2.3	Sterile neutrinos	8
2.3.1	Extended SM scenario with one sterile neutrino	9
2.3.2	Extended SM scenario with three sterile neutrinos	12
2.4	The Neutrino Minimal Standard Model	15
2.4.1	Previous experimental (target and collider) results on the N_2 and N_3 sterile neutrinos	17
3	The LHC and the ATLAS experiment	19
3.1	The LHC	19
3.2	The ATLAS experiment	21
3.2.1	ATLAS coordinate system	22
3.2.2	The magnetic system	23
3.2.3	The inner detector	23
3.2.4	Calorimeters	24
3.2.5	The muon system	25
3.2.6	Trigger system	26
4	Signal and background signatures	27
4.1	HNL signal	27
4.2	Backgrounds to the search	28
4.2.1	Drell-Yan pair production	28
4.2.2	Top quark production	28
4.2.3	Diboson production	29
4.2.4	Backgrounds arising from W + jets or multi-jets	30
4.3	Monte Carlo simulations	31
4.4	Translation of MC simulations to physical predictions	36
5	Object reconstruction and identification	38
5.1	Muon reconstruction	38
5.2	Electron reconstruction	38
5.3	Jet, b-jet and E_T^{miss} reconstruction	38
5.4	Isolation	39
5.5	Overlap removal	40
6	Truth-level analysis	42
6.1	Muon selection for 5 GeV	42
6.2	Muon selection for 50 GeV	43

7	Signal selection	44
7.1	Cut optimization	44
7.2	Isolation	52
7.2.1	Leading muon WP isolation requirements	53
7.2.2	Subleading muon and electron WP isolation requirements	54
7.3	E_T^{miss} and H_T cuts	57
7.4	Trigger selection	59
7.5	Impact of final optimal cuts on significance, signal and background	62
8	A neural network approach to signal optimization	70
8.1	Theory and functionality of neural networks	70
8.2	The neural network model	75
8.2.1	The input layer	76
8.2.2	Sample preparation	82
8.2.3	Sanity checks	84
8.2.4	Predictions	87
8.2.5	The receiver operating characteristic curve	88
8.3	Comparison of neural network approach with standard cut-based procedure	90
9	Conclusion	94
	Appendices	99
	Appendix A Kinematic distributions used in truth analysis	99
A.1	Truth-level analysis distributions for a 5 GeV HNL mass	99
A.2	Truth analysis distributions for a 50 GeV HNL mass	103
	Appendix B Invariant mass plots	107
B.1	Dilepton invariant mass plots	107
B.2	Trilepton invariant mass plots	108
	Appendix C Cut optimization of p_T for all three leptons	109
C.1	Leading muon p_T	109
C.2	Sub-leading muon p_T	112
C.3	Electron p_T	115
C.4	Sub-leading muon significance plots of individual backgrounds for 10 GeV	119
	Appendix D Distributions of the impact parameter	120
D.1	d_0/σ_{d_0} distributions for the sub-leading muon	120
D.2	d_0/σ_{d_0} distributions for the electron	123
	Appendix E Optimization of WP combinations	126
E.1	Summary of cuts used for isolation	126
E.2	WP combinations for 10 GeV	127
E.3	WP combinations for 20 GeV	128
E.4	WP combinations for 30 GeV	129

Appendix F Distributions of $E_{\mathbf{T}}^{\text{miss}}$ and $H_{\mathbf{T}}$	131
F.1 Distributions of $E_{\mathbf{T}}^{\text{miss}}$	131
F.2 Distributions of $H_{\mathbf{T}}$	134

1 Introduction

The Standard Model of particle physics (SM) [1, 2] is still an incomplete theory and fails to explain a number of issues such as neutrino oscillations, inflation, baryon asymmetry, dark matter and dark energy. To find suitable explanations for these issues one has to move beyond the SM and introduce new physics phenomena. A plethora of possible extensions exist wherein both new particles and new energy scales are suggested. The Neutrino Minimal Standard Model (ν MSM) [3] is one of these possible extensions but differs from a multitude of alternatives in its simplicity and profound implications. Three right-handed neutrinos are implemented within the SM framework without introducing any new energy scales. The additional free parameters are then used to fine tune the model to explain the problems that lie beyond the SM.

Restrictions on the ν MSM parameters have previously been set by e^+e^- colliders and fixed-target experiments. The LHC, with its high intensity collisions, represent another great place (the only one right now) where one can attempt to set further constraints on this model or potentially discover signs of the existence of right-handed neutrinos. Exploring what ATLAS can do to investigate the ν MSM parameters is the subject of this thesis. In particular, a cut-based analysis is performed to gauge the sensitivity of the ATLAS experiment in a direct search. A different approach to the cut-based analysis is also explored by using neural network classifiers to learn about kinematic topologies within the signal and background regions.

Section 2 provides an overview of the underlying theoretical framework and Section 3 introduces the ATLAS detector at the LHC. Descriptions of the experimental signatures of signal and background in addition to the Monte Carlo simulations are given in Section 4. The imposed object selection is summarized in Section 5 and Section 6 and 7 detail the chosen cut-based selection strategy. Finally, Section 8 provides a review of the alternative neural network approach. The conclusion on this thesis is given in Section 9.

2 Theory

This section provides an overview of the relevant theory and motivates the need to introduce three sterile neutrinos in the neutrino sector of the SM to explain observations that lie beyond the SM.

Section 2.1 provides an overview of the SM and a review of neutrino oscillations is given in Section 2.2. Section 2.3 provides an account of the theoretical and mathematical framework relevant to a possible extension to the SM known as the ν MSM. Finally, the details of this model and the notable parameter restrictions are given in Section 2.4.

2.1 The Standard Model

The SM summarizes the physics of fundamental elementary particles and their interactions. A visual overview of the SM can be seen in figure 1. There are two types of elementary particles: matter and force particles. The former are fermions and are characterized by their spin-value $1/2$, and the latter are bosons and have spin-value 1 .

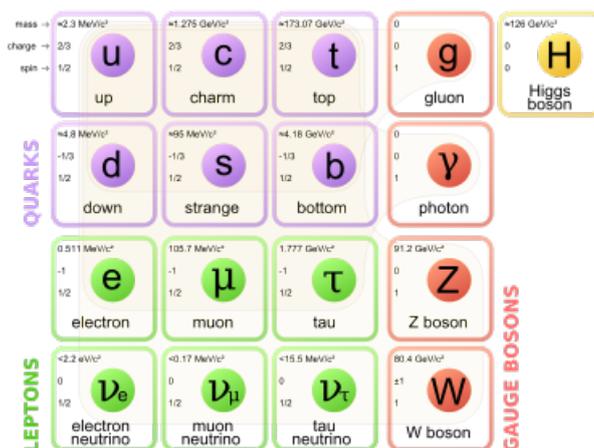


Figure 1: An overview of the SM [4].

The fermions comprise all matter in the universe and are represented by the first three columns in figure 1, where each column is known as a generation. They can be further categorized into two groups known as the leptons and the quarks which are represented by green and purple respectively. The group of quarks has twelve members with six distinct flavours called up, down, charm, strange, top and bottom. Each particle (e.g. quark) has an associated antiparticle. The group of leptons likewise has 12 members but with three distinct lepton flavours. These consist of the electron, the electron-neutrino, the muon, the muon-neutrino, the tau and the tau-neutrino in addition to their corresponding antiparticles. Leptons can be observed in isolation contrary to quarks which are confined into hadrons by the strong force. A particle with colour charge can only be observed in colour neutral combinations such as baryons (three quarks) and mesons (two quarks).

The force-carriers are represented by the red fourth column in figure 1. The term force-carrier comes from the fact that they mediate three of the four forces of nature. The photon mediates the electromagnetic interaction described by quantum electrodynamics (QED) while the strong force described by quantum chromodynamics (QCD) is carried by the gluon. The weak force is unique in the sense that it has two separate interactions and three force-carriers. The charged current is mediated by the W^+ and W^- bosons whereas the neutral current is carried by the Z boson.

A special yellow fifth column has been reserved for the Higgs boson which is set apart from the other bosons by its spin-value 0. This particle is responsible for giving mass to all other particles via the Higgs mechanism.

A coupling between a particle and a force-carrier can only take place if the particle carries the charge of the interaction. Hence, a gluon only couples to a quark or itself but never to a lepton since the lepton doesn't have a colour charge. However, quarks possess both colour and fractional electric charges and can thus couple to the three other bosons as well. The Higgs boson of course couples to all particles with a mass. A visual summary of how the particles are linked via the different interactions can be found in figure 2.

The SM can be mathematically condensed into a Lagrangian where the particles are represented as fields. This Lagrangian is specified in accordance with quantum field theory.

2.1.1 Chirality

In the following sections notations such as right-handed and left-handed will be used frequently. Therefore, a brief description is given of the meaning of these terms. Chirality is defined in terms of the γ^5 matrix which is given by

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

Particle fields which are eigenstates of this matrix are referred to as right-handed and left-handed chiral particle states, where R and L denote the two cases respectively. Any Dirac spinor can be decomposed into these fields by using right-handed and left-handed chiral projection operators defined by

$$P_R = \frac{1}{2}(1 + \gamma^5)$$

and

$$P_L = \frac{1}{2}(1 - \gamma^5)$$

respectively. Hence, for a spinor ψ one has that

$$\psi_R = P_R\psi$$

and

$$\psi_L = P_L\psi.$$

In the limit $E \gg m$ the chiral eigenstates become equivalent to the helicity eigenstates which physically means that particles have spin oriented in the direction parallel (R) or anti-parallel (L) to the direction of motion [1].

2.1.2 CP-transformation

The concept of CP-violation will be mentioned in this section and will therefore be formally introduced here. CP refers to the charge conjugation and parity transformation operators denoted by \hat{C} and \hat{P} respectively.

The effect of applying \hat{C} on a particle is to obtain the corresponding antiparticle whereas \hat{P} inverts the spatial coordinates of a particle such that

$$x' = -x, \quad y' = -y, \quad z' = -z \quad t' = t$$

represents the resulting coordinates after the transformation. Both spin and angular momentum are conserved under the action of \hat{P} , however, the momentum direction is reversed. This means that chirality will also be reversed. Both QCD and QED are conserved under \hat{C} and \hat{P} separately but weak interactions are not. However, the combined action of $\hat{C}\hat{P}$ is conserved in all three cases. In fact the SM largely conserves $\hat{C}\hat{P}$ with the exception of small CP-violating phases in the quark sector.

2.1.3 Beyond the Standard Model

The SM is a well established model but there are questions and observations that are not accounted for within the current theoretical framework. The most notable unanswered questions are listed below [3]:

- Baryogenesis: The SM fails to explain the disproportionate amount of matter and antimatter in the universe.
- Neutrino oscillations: Total lepton flavour number is conserved in the SM and hence the model fails to explain observed transitions between different neutrino flavours. Such oscillations are also not possible if the neutrinos have 0 mass which is the current SM prediction.
- Inflation: The theory of inflation explains how the early universe undergoes rapid accelerated expansion but the SM doesn't offer a mechanism for this concept.
- Dark matter and dark energy: Unknown particles constitute 80% of all matter in the universe.

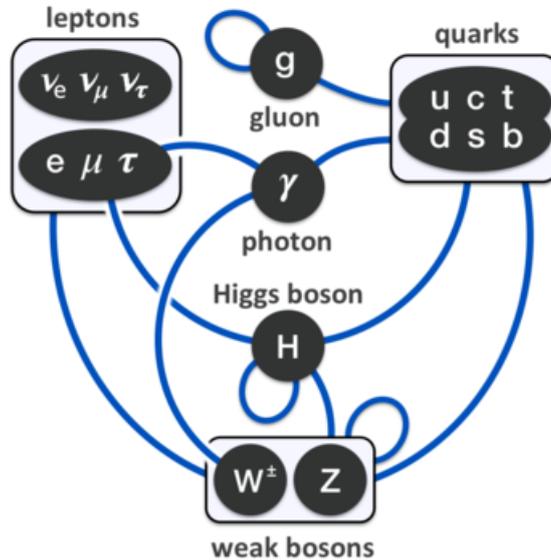


Figure 2: A summary of how matter particles interact with bosons in the SM [4].

2.2 Neutrino oscillations

In the SM neutrinos are massless particles. Moreover, a neutrino of a certain flavour is in the same family as the lepton of the corresponding flavour i.e. a tau is produced along with an anti-tau neutrino etc. Lepton number (leptons are assigned a value of 1 and the corresponding anti-leptons are denoted by -1) is totally conserved in the SM.

Due to the hydrogen burning process in the sun one should expect a large production of electron neutrinos, but in the late 90's solar neutrino experiments showed a big discrepancy between measurements of the solar flux and the theoretical value of the electron neutrino flux [1]. The SNO experiment later determined the solar flux to consist of not only electron neutrinos but also tau and muon neutrinos. This implied that the electron neutrinos produced by the sun oscillated into muon and tau neutrinos over the long distance between the sun and the earth. However, this picture was not compatible with lepton number conservation or the idea of neutrinos as massless particles since neutrino oscillations can occur only if at least 2 of the 3 neutrinos in the SM have a non-zero mass.

The three weak eigenstates given by ν_e, ν_μ and ν_τ are linear combinations of the mass eigenstates given by ν_1, ν_2 and ν_3 . The relationship between these states can be expressed by

$$\begin{pmatrix} \nu_e \\ \nu_\mu \\ \nu_\tau \end{pmatrix} = \begin{pmatrix} U_{e1} & U_{e2} & U_{e3} \\ U_{\mu1} & U_{\mu2} & U_{\mu3} \\ U_{\tau1} & U_{\tau2} & U_{\tau3} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix}.$$

Here the 3×3 matrix U is known as the PMNS matrix and describes the mixing of the states. According to [5] U can also be written as

$$U = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \\ \times \text{diag}(1, e^{i\frac{\alpha_{21}}{2}}, e^{i\frac{\alpha_{31}}{2}}),$$

where $c_{ij} = \cos \theta_{ij}$ and $s_{ij} = \sin \theta_{ij}$. The symbol θ denotes the mixing angles and δ represents the Dirac CP-violation phase while α_{21} and α_{31} represent the potential Majorana CP violation phases (the presence of the diagonal matrix term depends on whether the active neutrinos are Dirac or Majorana particles). The best-fit values of the PMNS parameters are shown in table 1 along with their 3σ bounds. These values are derived from a global fit of current data on neutrino oscillations.

parameter	best-fit	3σ
$\sin^2 \theta_{12}$	0.297	0.250-0.354
$\sin^2 \theta_{23}$	0.437	0.379-0.616
$\sin^2 \theta_{13}$	0.0214	0.0185-0.0246
δ/π	1.35	0.92-1.99

Table 1: The best-fit values of the PMNS parameters along with their 3σ bounds [5]

The oscillation probabilities are determined by the mass squared differences. Therefore, one can not put an upper limit on the individual masses by solely considering neutrino fluxes. However, experiments on the nuclear β -decay of tritium have shown that the lightest neutrino mass is $\lesssim 2$ eV [1]. Solar and atmospheric experiments have obtained the following results of the mass squared differences [3]

$$\Delta m_{\text{sun}}^2 = \Delta m_{21}^2 = m_2^2 - m_1^2 = 7.65_{-0.20}^{+0.23} \times 10^{-5} \text{ eV}^2$$

and

$$|\Delta m_{\text{atm}}^2| = |\Delta m_{32}^2| = |m_3^2 - m_2^2| = 2.40_{+0.12}^{-0.11} \times 10^{-3} \text{ eV}^2.$$

Due to the atmospheric measurement two possible hierarchies exist for the neutrino masses, namely a normal hierarchy where $m_3 > m_2$ and an inverted hierarchy where $m_3 < m_2$. This is illustrated by figure 3.

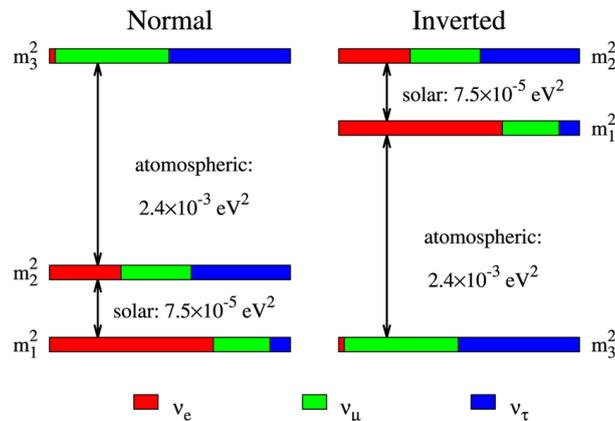


Figure 3: The two possible hierarchies of neutrino masses [6].

2.3 Sterile neutrinos

The observed phenomena of neutrino oscillations, small active neutrino masses and the remaining BSM problems listed in Section 2.1 necessitate a revision of the SM. A detailed discussion of this and a proposal for an extended version of the SM can be found in Section 2.4. The discussion in this section will clarify parts of the theoretical and mathematical framework that constitutes the foundation for such an extension.

The model in Section 2.4 makes use of three right-handed neutrinos which are also referred to as sterile. The term sterile alludes to the fact that these particles have no weak, strong or electric charges and hence do not participate in the corresponding interactions described in Section 2.1. The left-handed neutrinos are referred to as active for the opposite reason.

The existence of right-handed neutrinos is not included in the SM and for this reason the Higgs mechanism cannot be applied to generate Dirac masses for the left-handed neutrinos. Suppose right-handed neutrinos are introduced. The neutrino masses are still a lot smaller than those of other fermions which would seem to indicate that the left-handed neutrino masses are generated via an alternative method to the Higgs mechanism. This has led to questions about the intrinsic character of the left-handed neutrinos and whether their nature is Dirac (particle and anti-particle fields are independent) or Majorana (a particle is its own antiparticle).

A way of explaining the experimental results in Section 2.2 is to add Majorana mass terms to the SM Lagrangian and use another method called the see-saw mechanism to explain the smallness of active neutrino masses [7]. In using this approach the SM Lagrangian, denoted by \mathcal{L}_{SM} , is extended to be of the following form [3]

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + i\bar{N}_I \partial_\mu \gamma^\mu N_I - (F_{\alpha I} \bar{L}_\alpha N_I \tilde{\phi} - \frac{M_I}{2} \bar{N}_I^c N_I + h.c.), \quad (1)$$

where N_I with $I = \{1, \dots, \mathcal{N}\}$ are the sterile neutrinos and $L_\alpha = \begin{pmatrix} \nu_\alpha \\ \alpha \end{pmatrix}$ with $\alpha = \{e, \mu, \tau\}$ are the left-handed lepton doublets. The new Yukawa couplings between active neutrinos of flavour α and sterile neutrinos of type I are given by $F_{\alpha I}$ and the Higgs doublet is denoted by ϕ . The individual terms in Eq. 1, in addition to the SM Lagrangian \mathcal{L}_{SM} , represent respectively the kinetic energy terms of the sterile neutrinos, the active and sterile couplings to the Higgs field and the Majorana mass terms of the sterile neutrinos.

The two following sections will detail how the simple case of one and the more realistic case of three sterile neutrinos are introduced into the SM and how the see-saw mechanism is applied in both scenarios. It's advantageous to discuss the case with only one active and sterile neutrino since the mathematics is simplified and the see-saw mechanism is easy to see. This in turn makes the case with three sterile neutrinos easier to understand.

2.3.1 Extended SM scenario with one sterile neutrino

The SM will now be extended with one sterile neutrino and the following discussion is based on [7]. Suppose that a right-handed particle state exists for the neutrino then the Dirac field is given by

$$\nu = \nu_{\alpha L} + \nu_{sR},$$

where $\nu_{\alpha L}$ and ν_{sR} denote the left-handed and right-handed chiral states respectively. Here $\nu_{\alpha L}$ is called the active neutrino component since it can take part in weak interactions and ν_{sR} is called the sterile neutrino component since it does not take part in any interaction. The subscript α denotes a given flavour for the active neutrino and the subscript s is used to indicate the sterile neutrino field for total clarification. The Dirac field depends on four independent components.

The Dirac mass term for ν can be written as

$$\mathcal{L}_D = -m_D \bar{\nu} \nu = -m_D (\bar{\nu}_{sR} \nu_{\alpha L} + \bar{\nu}_{\alpha L} \nu_{sR}). \quad (2)$$

In the above equation the term m_D denotes the Dirac mass and is given by $m_D = F_{\alpha s} \langle \phi \rangle$, where $F_{\alpha s}$ is the active-sterile Yukawa coupling and $\langle \phi \rangle$ is the vacuum expectation value (VEV) of the Higgs field. As stated, this term alone is not a pleasant enough way to generate the small values observed for the active neutrino masses.

To theoretically better account for the experimental results one introduces instead two Majorana neutrino fields. It should be emphasized that the neutrinos are now assumed to be Majorana particles, not Dirac particles. For one Majorana field, the Majorana condition is

$$\psi = \psi^c$$

which states that the spinor, denoted by ψ , of a massive neutral fermion must be equivalent to its charge conjugate. In other words, a particle must be its own antiparticle. To be completely general a spinor is given by

$$\psi = \psi_L + \psi_R.$$

Upon implementing the Majorana condition one obtains

$$\psi_L + \psi_R = \psi_L^c + \psi_R^c.$$

By applying the chiral projection operators to the above equation one finds that ψ_L^c is right-handed and ψ_R^c is left-handed, e.g an application of P_R on both sides gives $\psi_R = \psi_L^c$. Thus, the Majorana field considered here, ψ , only has two independent components and can be rewritten as

$$\psi = \psi_L + \psi_L^c \quad \text{or} \quad \psi = \psi_R + \psi_R^c.$$

If $\psi = \psi_L + \psi_L^c$ the Majorana mass term is given by

$$\mathcal{L}_L = -\frac{1}{2}m_L(\overline{\psi_L^c}\psi_L + \overline{\psi_L}\psi_L^c), \quad (3)$$

and if $\psi = \psi_R + \psi_R^c$ the equivalent expression is

$$\mathcal{L}_R = -\frac{1}{2}m_R(\overline{\psi_R^c}\psi_R + \overline{\psi_R}\psi_R^c), \quad (4)$$

where $m_{L,R}$ denotes the Majorana mass.

Now: provided that both the chiral left-handed and right-handed fields, i.e $\nu_{\alpha L}$ and ν_{sR} , exist AND are independent so they are two separate Majorana fields, then the possible Lagrangian mass terms are given by

$$\mathcal{L}^{D+M} = \mathcal{L}^D + \mathcal{L}_L^M + \mathcal{L}_R^M \quad (5)$$

where $\mathcal{L}_L^M = -\frac{1}{2}m_L\overline{\nu_{\alpha L}^c}\nu_{\alpha L} + \text{h.c.}$ and $\mathcal{L}_R^M = -\frac{1}{2}m_R\overline{\nu_{sR}^c}\nu_{sR} + \text{h.c.}$. The symbol ψ is replaced by ν to be specific that from here on out the Majorana discussion is not general but again refers to the neutrinos, where the labels α and s indicate the two separate Majorana fields. When written in matrix form Eq. 5 becomes

$$\mathcal{L}^{D+M} = -\frac{1}{2}(\overline{\nu_{\alpha L}^c}\overline{\nu_{sR}^c}) \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_{\alpha L} \\ \nu_{sR}^c \end{pmatrix} + \text{h.c.}, \quad (6)$$

where h.c. denotes the hermitian conjugate. From this it's clear that one must diagonalize the mass matrix to find the eigenvectors and the associated eigenvalues i.e the physical propagating neutrino fields and definite mass values. Both $\nu_{\alpha L}$ and ν_{sR}^c are left-handed and one can therefore rewrite Eq. 6 in terms of a left-handed column vector N_L such that

$$\mathcal{L}^{D+M} = \frac{1}{2}\overline{N_L^c}MN_L + \text{h.c.},$$

where

$$M = \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix}, \quad N_L = \begin{pmatrix} \nu_{\alpha L} \\ \nu_{sR}^c \end{pmatrix}.$$

The column vector N_L can be rewritten in terms of a unitary mixing matrix U and the left-handed components of the two massive neutrino fields ν_{1L} and ν_{2L} i.e.

$$N_L = Un_L \quad \text{with} \quad n_L = \begin{pmatrix} \nu_{1L} \\ \nu_{2L} \end{pmatrix}, \quad (7)$$

where U is chosen such that it diagonalizes the mass matrix M

$$U^T M U = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}.$$

One can rewrite the mixing matrix U as the product of an orthogonal matrix \mathcal{O} and a diagonal matrix of phases ρ i.e

$$U = \mathcal{O}\rho.$$

The orthogonal matrix \mathcal{O} can now be chosen such that

$$M' = \mathcal{O}^T M \mathcal{O} = \begin{pmatrix} m'_1 & 0 \\ 0 & m'_2 \end{pmatrix}$$

and suitable choices for both \mathcal{O} and ρ are

$$\mathcal{O} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \quad \text{and} \quad \rho = \begin{pmatrix} \rho_1 & 0 \\ 0 & \rho_2 \end{pmatrix},$$

where θ is the active-sterile mixing angle and ρ is chosen such that $|\rho_k^2| = 1$ with $k = \{1, 2\}$. The point of introducing phases is to ensure that m_1 and m_2 are positive. The eigenvalues $m'_{2,1}$ are found to be

$$m'_{2,1} = \frac{1}{2} [m_L + m_R \pm \sqrt{(m_L - m_R)^2 + 4m_D^2}]$$

and one can now find the definite values m_1 and m_2 from

$$U^T M U = \rho^T \mathcal{O}^T M \mathcal{O} \rho = \rho^T M' \rho = \begin{pmatrix} \rho_1^2 m'_1 & 0 \\ 0 & \rho_2^2 m'_2 \end{pmatrix} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad (8)$$

where ρ_1^2 and ρ_2^2 are set to 1 or -1 depending on the sign of m'_1 and m'_2 . Eq. 8 also gives the following relation for the mixing angle

$$\tan 2\theta = \frac{2m_D}{m_R - m_L}.$$

Thus far the approach has been completely general in allowing for a Majorana mass term for $\nu_{\alpha L}$ which technically doesn't exist due to gauge symmetries of the Standard Model. More specifically the active neutrino has $I_3 = \frac{1}{2}$, where I_3 is the third component of isospin, since it belongs to a weak isodoublet with the corresponding lepton. The left-handed Majorana mass term dictates couplings between $\nu_{\alpha L}$ and its charge conjugate i.e. $\bar{\nu}_{\alpha L}^c \nu_{\alpha L}$. This combination results in a triplet with $I_3 = 1$, but such a term is clearly not $SU(2)_L$ transformation invariant so the left-handed Majorana mass term, denoted by \mathcal{L}_L^M , in the Dirac-Majorana Lagrangian is not allowed.

One obtains the following by setting $m_L = 0$ as required to be consistent with the Standard Model

$$m_{2,1} = \frac{m_R \pm \sqrt{m_R^2 + 4m_D^2}}{2} = \frac{m_R \pm m_R \sqrt{1 + 4m_D^2/m_R^2}}{2}$$

and

$$\tan 2\theta = \frac{2m_D}{m_R}.$$

In the see-saw mechanism $|m_D| \ll m_R$. In this scenario the above expression can be approximated by

$$m_{2,1} \simeq \frac{1}{2}m_R \pm \frac{1}{2}\left(m_R + \frac{2m_D^2}{m_R}\right).$$

The active and sterile neutrino masses are then given by

$$m_1 \simeq \frac{m_D^2}{m_R} \ll |m_D| \quad \text{and} \quad m_2 \simeq m_R,$$

and the mixing angle relation becomes

$$\tan \theta \simeq \frac{m_D}{m_R} \ll 1.$$

Thus the see-saw mechanism can explain the apparent smallness of the active neutrino masses though an imbalance in the Dirac and Majorana masses. When using the fact that θ is very small in the mixing relation in Eq. 7 one obtains

$$\nu_{1L} \simeq -\nu_{\alpha L} \quad \text{and} \quad \nu_{2L} \simeq \nu_{sR}^c$$

which means that the active neutrino observable $\nu_{\alpha L}$ is practically equivalent to ν_1 and the sterile neutrino observable ν_{sR} is practically equivalent to ν_2 . Hence the light neutrino interacts with matter whereas the heavy neutrino is effectively decoupled from matter. Another important thing to note is that the Majorana mass m_R is a free parameter and is subject to theoretical tuning provided that m_D is scaled accordingly. This leads to an interesting set of restrictions on m_R that can help explain the BSM problems listed in Section 2.1. This will be elaborated upon in Section 2.4

For the final Dirac and Majorana mass term one has that

$$\mathcal{L}^{D+M} = -\frac{1}{2} \sum_{k=1,2} m_k \overline{\nu_{kL}^c} \nu_{kL} + \text{h.c.}, \quad (9)$$

where the Majorana neutrino fields are given by

$$\nu_k = \nu_{kL} + \nu_{kL}^c \quad (k = 1, 2). \quad (10)$$

This section highlights the significance of searching for sterile neutrinos as they are not only important for explaining the smallness of active neutrino masses but also provide information about the intrinsic nature of the neutrinos.

2.3.2 Extended SM scenario with three sterile neutrinos

The SM will now be extended with three sterile neutrinos and the following discussion is again based on [7]. The mathematics of adding three sterile neutrinos to the SM is very similar to the single neutrino case. The active neutrino flavour fields are now described by the left-handed chiral fields, namely ν_{eL} , $\nu_{\mu L}$, $\nu_{\tau L}$,

and the sterile neutrinos are described by the right-handed chiral fields, namely $\nu_{s_1 R}$, $\nu_{s_2 R}$ and $\nu_{s_3 R}$. The Dirac and the two Majorana Lagrangians can be written in a similar manner to Eq. 2, 3 and 4 following the same arguments. These terms are now given by

$$\begin{aligned}\mathcal{L}^D &= - \sum_{s,\beta} \overline{\nu_{sR}} M_{s\beta}^D \nu_{\beta L} + \text{h.c.}, \\ \mathcal{L}_L^M &= - \frac{1}{2} \sum_{\alpha,\beta} \overline{\nu_{\alpha L}^c} M_{\alpha\beta}^L \nu_{\beta L} + \text{h.c.} \\ \mathcal{L}_R^M &= - \frac{1}{2} \sum_{s,s'} \overline{\nu_{sR}^c} M_{\alpha\beta}^R \nu_{s'R} + \text{h.c.},\end{aligned}$$

where $\alpha, \beta = \{e, \mu, \tau\}$ and $s, s' = \{s_1, s_2, s_3\}$. The Dirac masses $M_{s\beta}^D$ are again given by $M_{s\beta}^D = F_{s\beta} \langle \phi \rangle$, where $F_{s\beta}$ is the corresponding active-sterile Yukawa coupling and $\langle \phi \rangle$ is the VEV of the Higgs field. The combined Dirac and Majorana mass term, namely \mathcal{L}^{D+M} , can once again be written as in Eq. 5 and when written in matrix form this expression is again given by Eq. 6 only now with

$$M = \begin{pmatrix} M^L & (M^D)^T \\ M^D & M^R \end{pmatrix}, \quad \nu_L = \begin{pmatrix} \nu_{eL} \\ \nu_{\mu L} \\ \nu_{\tau L} \end{pmatrix} \quad \text{and} \quad \nu_R^c = \begin{pmatrix} \nu_{s_1 R}^c \\ \nu_{s_2 R}^c \\ \nu_{s_3 R}^c \end{pmatrix}.$$

The mass coupling matrix, denoted by M , is now a 6×6 matrix with 3×3 mass matrices as elements. This means that the diagonalization of this system is more complicated than the 2×2 case and therefore the steps will be outlined explicitly.

The first step is similar to what one does in the single sterile neutrino case. The mixing relation in Eq. 7 can be generalized to

$$N_L = \mathbb{V} n_L, \quad \text{with} \quad n_L = \begin{pmatrix} \nu_{1L} \\ \vdots \\ \nu_{6L} \end{pmatrix}, \quad (11)$$

where \mathbb{V} is the unitary matrix that diagonalizes \mathcal{L}^{D+M} i.e.

$$\mathbb{V}^T M \mathbb{V} = \text{diag}(m_1, \dots, m_6).$$

Here the set (m_1, \dots, m_6) consists of distinct masses of the active and sterile neutrinos. In principle one could go about diagonalizing this system in a completely standard way which is exactly what is done for just one sterile neutrino. However, for three sterile neutrinos this is far too cumbersome. If the see-saw mechanism is invoked already at this stage one can use a different approach called block diagonalization. The see-saw mechanism for three sterile neutrinos is slightly more complicated compared to the single case. One still has to set $M^L = 0$ to avoid conflict with the SM gauge symmetries and the see-saw mechanism requires that $\det(M^D) \ll \det(M^R)$. In other words the eigenvalues of M^D are much smaller than the eigenvalues of M^R .

One can now rewrite \mathbb{V} as

$$\mathbb{V} = \mathbb{W}\mathbb{U},$$

where \mathbb{U} and \mathbb{W} are unitary matrices. Here one can think of \mathbb{W} as splitting the active and sterile masses while \mathbb{U} rotates the oscillations such that

$$M' = \mathbb{W}^T M \mathbb{W} \approx \begin{pmatrix} M_{\text{light}} & 0 \\ 0 & M_{\text{heavy}} \end{pmatrix}, \quad (12)$$

and

$$\begin{aligned} \text{diag}(m_1, \dots, m_6) &= \mathbb{U}^T M' \mathbb{U} \\ &\approx \begin{pmatrix} \text{PMNS} & 0 \\ 0 & X \end{pmatrix}^T \begin{pmatrix} M_{\text{light}} & 0 \\ 0 & M_{\text{heavy}} \end{pmatrix} \begin{pmatrix} \text{PMNS} & 0 \\ 0 & X \end{pmatrix}. \end{aligned} \quad (13)$$

An appropriate choice of \mathbb{W} is used for the approximate block diagonalization of M . This is given by

$$\mathbb{W} = 1 - \frac{1}{2} \begin{pmatrix} (M^D)^\dagger (M^R (M^R)^\dagger)^{-1} M^D & 2(M^D)^\dagger (M^R)^\dagger{}^{-1} \\ -2(M^R)^{-1} M^D & (M^R)^{-1} M^D (M^D)^\dagger (M^R)^\dagger{}^{-1} \end{pmatrix} \quad (14)$$

which leads to the following 3×3 matrices for the active and sterile neutrinos respectively

$$M_{\text{light}} \simeq -(M^D)^T (M^R)^{-1} M^D \quad \text{and} \quad M_{\text{heavy}} \simeq M^R. \quad (15)$$

From this one can see that M_{light} is suppressed by a small ratio with respect to the Dirac mass matrix. Moreover, from the off-diagonal elements in \mathbb{W} one can gauge how the definition of active-sterile mixing angles is motivated. They are defined as [9]

$$|V_{\alpha I}|^2 \equiv |(M^D (M^R)^{-1})_{\alpha I}|^2,$$

where $|V_{\alpha I}|^2$ actually refers to the off-diagonal elements of \mathbb{W} but the symbol V is used since it appears in plots referenced later on.

It should be emphasized that M_{light} and M_{heavy} are of course not diagonal yet. However, the complete diagonalization can be achieved by applying \mathbb{U} as stated. Lets start out with the regime of low-energy phenomenology: Suppose that the PMNS matrix is denoted by U then

$$U^T M_{\text{light}} U = \text{diag}(m_1, m_2, m_3),$$

where the set (m_1, m_2, m_3) consists of the three active neutrino masses. It then follows from the mixing relation in Eq. 11 that the left-handed flavour states $\nu_{\alpha L}$ can be written as linear combinations of elements in U and n_L such that

$$\nu_{\alpha L} = \sum_{k=1}^3 U_{\alpha k} \nu_{kL} \quad (\alpha = e, \mu, \tau),$$

and this can of course be immediately identified as the mixing of flavour and mass eigenstates in the model of neutrino oscillations and hence the physics discussed in Section 2.2 is reproduced using the new model. Lets now go back

to the regime of high-energy phenomenology. The mixing relation in Eq. 11 again dictates that

$$\nu_{\alpha L} = \sum_{k=1}^6 \mathbb{V}_{\alpha k} \nu_{kL} \quad \alpha = \{e, \mu, \tau\}$$

for the active neutrinos and

$$\nu_{sR}^c = \sum_{k=1}^6 \mathbb{V}_{sk} \nu_{kL} \quad s = \{s_1, s_2, s_3\}$$

for the sterile neutrinos. Since both $\nu_{\alpha L}$ and ν_{sR}^c are linear combinations of the same massive neutrino fields one can conclude that the model allows for oscillations between active and sterile neutrinos. Lastly, it should be noted that the final diagonal form of \mathcal{L}^{L+M} can once again be written on the same form as Eq. 9 and Eq. 10 with $k = 6$ instead of $k = 2$.

2.4 The Neutrino Minimal Standard Model

The ν MSM introduces three new particles with masses below the electroweak scale and the ν MSM Lagrangian is therefore given by Eq. 1 with $\mathcal{N} = 3$ i.e

$$\mathcal{L} = \mathcal{L}_{\text{SM}} + i\bar{N}_I \partial_\mu \gamma^\mu N_I - (F_{\alpha I} \bar{L}_\alpha N_I \tilde{\phi} - \frac{M_I}{2} \bar{N}_I^c N_I + h.c.),$$

where N_I and $I = \{1, 2, 3\}$. To reiterate, $F_{\alpha I}$ are the active-sterile Yukawa couplings that express the interactions of L_α and N_I with the Higgs field and are responsible for the values of all elements in M^D seen in previous equations. Figure 4 illustrates how the three new sterile neutrinos fit into the old picture of the SM given in Section 2.1.

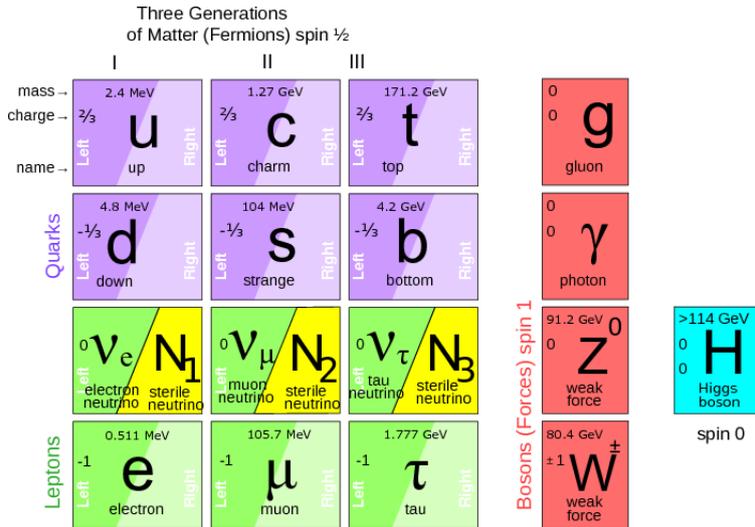


Figure 4: An overview of the Neutrino Minimal Standard Model. The neutrino sector of the Standard Model is altered to accommodate three right-handed neutrinos [8].

Here it can be seen that the three sterile neutrinos restore the symmetry between the quark and neutrino sector.

The choice to introduce three sterile neutrinos leads to 18 free parameters; 3 Majorana masses, 3 Dirac masses, 6 mixing angles between sterile and active neutrinos and 6 CP-violating phases. In the ν MSM the 18 free parameters are chosen such that the model explains all the BSM problems listed in Section 2.1. Some notable parameter restrictions will now be explained:

Restrictions from experimental results: The choice of three sterile neutrinos (no more, no less) is in part due to the experimental results of neutrino oscillations which tells us that at least 2 out of 3 neutrinos have mass. Eq. 15 in Section 2.3.2 is also valid for \mathcal{N} sterile neutrinos. This can be written on the following form under the assumption that M_{heavy} is diagonal with masses M_I :

$$(M_{\text{light}})_{\alpha\beta} = - \sum_{I=1}^{\mathcal{N}} \frac{(M_D^T)_{\alpha I} (M_D)_{I\beta}}{M_I}. \quad (16)$$

Since Δm_{sol}^2 and Δm_{atm}^2 are both non-zero the number of sterile neutrinos has to fulfil $\mathcal{N} \geq 2$. If $\mathcal{N} = 1$ then M_D is a 1×3 column matrix and it follows from simple linear algebra calculations that M_{light} only has one non-zero eigenvalue. Similarly, if $\mathcal{N} = 2$ then M_D is a 2×3 column matrix. This leads to two non-zero eigenvalues which is coherent with the experimental results.

Restrictions from dark matter: In the ν MSM the lightest sterile neutrino is treated as a candidate for dark matter (DM). This DM candidate is always denoted by N_1 and is set apart from N_2 and N_3 due to its long lifetime and *superweak* interaction with matter. In fact to be considered a candidate for DM, N_1 must have a lifetime that is greater than the age of the universe, and this condition is fulfilled since N_1 has a lifetime of at least 10^{24} s [3]. The mass of N_1 is ≥ 400 eV [3] and its mixing angle summed over all active flavours is restricted by $V_1^2 \lesssim 1.8 \times 10^{-5} \left[\frac{\text{KeV}}{M_1} \right]^5$ [3]. Moreover, N_1 is an example of decaying dark matter and its primary decay channel is to three active neutrinos, but N_1 also decays to a photon and neutrino through an extremely rare one-loop decay channel which gives rise to the constraint on the mixing angle. Furthermore, it can be shown that the maximum contribution by N_1 to the active neutrino masses is less than the error bar on Δm_{sol}^2 . Hence, 3 neutrinos are needed to explain both DM and the experimental results on the active neutrino masses.

Restrictions from Baryogenesis: In the ν MSM a total left-handed lepton asymmetry (ΔL) can be generated through CP-violating oscillations of active and sterile neutrinos. Special field configurations known as sphalerons violate conservation of baryon number (B) and lepton number (L) but conserve B-L. The action of sphalerons then transform the total left-handed lepton asymmetry given by ΔL into a total baryon asymmetry. It can be shown that the masses of M_2 and M_3 must be nearly degenerate to obtain a maximal left-handed lepton asymmetry and thereby a maximal baryon asymmetry [10]. Baryogenesis, furthermore, confines the masses M_2 and M_3 to the region $150 \text{ MeV} \lesssim M_{2,3} \lesssim 100 \text{ GeV}$ [3].

Restrictions on active-sterile Yukawa couplings: In [11] three benchmark models are constructed for the Yukawa couplings between a single sterile neutrino ($N_2 = N_3$) and each flavour α . These Yukawa couplings are denoted by $f_\alpha^2 = |F_{\alpha 2}|^2 = |F_{\alpha 3}|^2$. In each model f_α, f_β are taken to be as small as possible with respect to f_γ , where $\alpha \neq \beta \neq \gamma$. The coupling to the electron neutrino (f_e) is strongest in model I, the coupling to the muon neutrino (f_μ) is strongest in model II and the coupling to the tau neutrino (f_τ) is strongest in model III. The relative strengths in these three scenarios are found to be

$$\text{model I : } f_e^2 : f_\mu^2 : f_\tau^2 \approx 52 : 1 : 1$$

$$\text{model II : } f_e^2 : f_\mu^2 : f_\tau^2 \approx 1 : 16 : 3.8$$

$$\text{model III : } f_e^2 : f_\mu^2 : f_\tau^2 \approx 0.061 : 1 : 4.3$$

All three models are coherent with oscillation results for the active neutrinos. This analysis draws inspiration from model II. Of course one needs to investigate all three regimes, and for example colleagues in the analysis subgroup are currently looking into model I.

2.4.1 Previous experimental (target and collider) results on the N_2 and N_3 sterile neutrinos

The current experimental exclusion limits [12] in the mass and mixing angle plane for a single sterile neutrino mixing with ν_μ are shown in figure 5. The least explored parameter space is $M_N > m_b$ (m_b denotes the mass of the b quark) where the sterile neutrino is predominantly produced in the gauge boson decays $Z \rightarrow \nu N$ and $W^\pm \rightarrow l^\pm N$. When the sterile neutrino has a long decay length it will be both boosted and long-lived. The leptonic decay of such a sterile neutrino can be identified as a displaced lepton jet in which the leptons cannot be individually resolved. The exclusion limit for a displaced search is set by [12] and is represented by the blue dotted line in figure 5. When the sterile neutrino has a short decay length it is characterized by a prompt trilepton decay to three separately resolved objects and such a signature for $W^\pm \rightarrow \mu^\pm N$ is the focus of this analysis. The prompt trilepton exclusion limit is set by [12] and is shown by the brown dotted line in figure 5. Both the displaced and trilepton exclusion limits correspond to a 2σ reach with a center-of-mass energy of $\sqrt{s} = 13$ TeV and an integrated luminosity of 300 fb^{-1} . Furthermore, it should be mentioned that the BBN (bright grey) exclusion limit is set by Big Bang Nucleosynthesis while the PS191 (brown), BEBC (red), NuTeV (green) and CHARM (dark purple) exclusion limits come from fixed-target experiments. Finally, the $K \rightarrow \mu\nu$ (bright blue), Belle (olive green), DELPHI (cyan), L3 (violet), CMS (pink) and ATLAS (navy) exclusion limits all come from collider experiments. The red and purple dashed lines (not included in the legend) show the reach of the proposed SHiP and DUNE/LBNF experiments respectively. Both SHiP and DUNE/LBNF are scheduled to begin sometime after 2025. This means that the LHC is currently the only place at which one can improve upon the existing exclusion limits and this further highlights the importance of the search presented in this analysis.

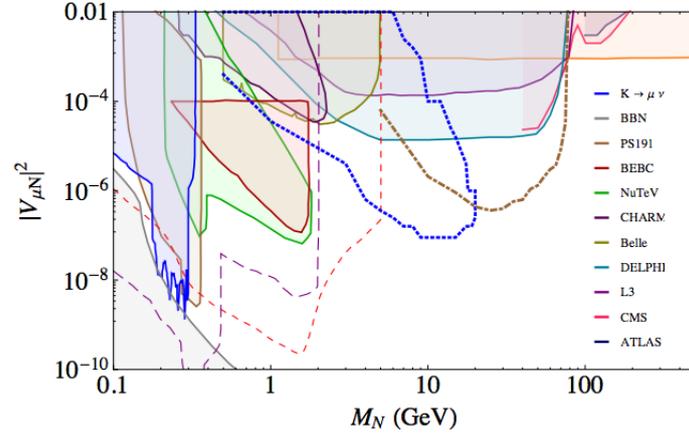


Figure 5: The current experimental exclusion limits in the mass and mixing angle plane for a single sterile neutrino mixing with ν_μ [12].

3 The LHC and the ATLAS experiment

This section gives an introduction to the ATLAS experiment at the LHC. Section 3.1 provides a brief overview of technical details concerning the LHC in general while Section 3.2 highlights key features of ATLAS.

3.1 The LHC

The Large Hadron Collider (LHC) [13] at CERN is a particle accelerator that measures approximately 27 kilometres in circumference and is located 100 meters below ground-level. The LHC stretches across both France and Switzerland and is a truly international endeavour as over 100 countries are currently involved in its experiments.

Both protons and heavy ions can be injected into the LHC but here the focus will be on the former. The protons are obtained by stripping a sample of hydrogen gas of its electrons. Before actually entering the LHC ring these protons have to undergo a series of steps to ramp up their energy.

The accelerator complex at CERN [14] has a long history that spans more than 60 years and the predecessors to the LHC are now used as its booster system as shown in figure 6.

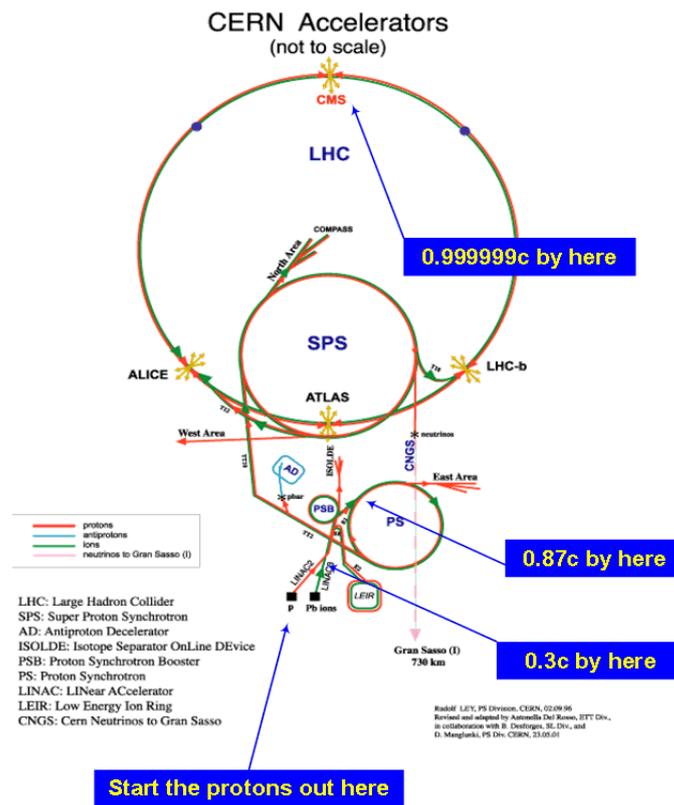


Figure 6: An overview of the Large Hadron Collider and its accelerator complex [15].

The protons' first encounter is with LINAC2 which is a linear accelerator that boosts the proton energy to 50 MeV. The proton sample is then split into 4 and passed to the Proton Synchrotron Booster (PBS); a circular accelerator consisting of four superimposed rings. Here the proton bunches reach an energy of 1.4 GeV after which they are picked up by the proton synchrotron (PS). The PBS cycle is completed one more time and 2 additional bunches are sent to the PS. The resulting 6 bunches are then organized into a train of 72 smaller bunches and accelerated to an energy of 25 GeV. These are then sent to the final stage in the accelerator complex before the LHC, namely the Super Proton Synchrotron (SPS). The SPS accelerates the protons to 450 GeV and awaits three to four fillings from the PS before sending the conjoined beam trains on their way to the LHC. The SPS fills the two LHC rings 13 times each resulting in two beams consisting of up to 2808 bunches with a 25 ns spacing. These are then ramped up to their final center-of-mass energy of 13 TeV as of Run 2. The two beams are sent in opposite directions around the beam pipe and collide at the locations of the four detectors, namely ATLAS, CMS, ALICE and LHCb. Both ATLAS and CMS are built to be general purpose detectors whereas ALICE is only built for heavy ion collisions and LHCb specializes in b-physics.

Several inelastic proton-proton collisions can occur within each bunch crossing and this is known as pile-up [16]. The average number of these collisions per bunch is denoted by $\langle\mu\rangle$ and can be calculated from the instantaneous luminosity (L), the total inelastic cross section for proton-proton collisions (σ_{incl}), the number of bunches in the LHC (N_{bunch}) and the revolution frequency of the LHC (f_{LHC}):

$$\langle\mu\rangle = \frac{L \times \sigma_{\text{incl}}}{N_{\text{bunch}} \times f_{\text{LHC}}}.$$

Pile-up is of course troublesome for any analysis and can lead to confusion when finding which particles belong to which event.

The LHC is designed for a center-of-mass energy of $\sqrt{s} = 14$ TeV, but a major technical incident shortly after the first beam injection in 2008 means that the LHC will have to reach this value gradually. The LHC has had one complete run, namely Run 1 which lasted from December 16 in 2009 until the end of 2012. Run 1 started with a center-of-mass energy of $\sqrt{s} = 2.36$ TeV in 2009 and increased to $\sqrt{s} = 7$ TeV during the years of 2010 and 2011. The total integrated luminosity corresponds to 0.047 fb^{-1} and 5.5 fb^{-1} in 2010 and 2011 respectively. In 2012 the center-of-mass energy was ramped up to $\sqrt{s} = 8$ TeV and the total integrated luminosity increased to 22.8 fb^{-1} .

At the time of writing the LHC is in the midst of Run 2 lasting from 2015 to 2018. This analysis focuses on simulations matching the data sets recorded during 2015 and 2016 for a center-of-mass energy corresponding to $\sqrt{s} = 13$ TeV and an integrated luminosity of 36 fb^{-1} .

The pile-up profiles measured by ATLAS for the individual 2015 and 2016 data sets in addition to the combined data set are shown in figure 7. The mean number of collisions per bunch crossing is $\langle u \rangle = 23.7$.

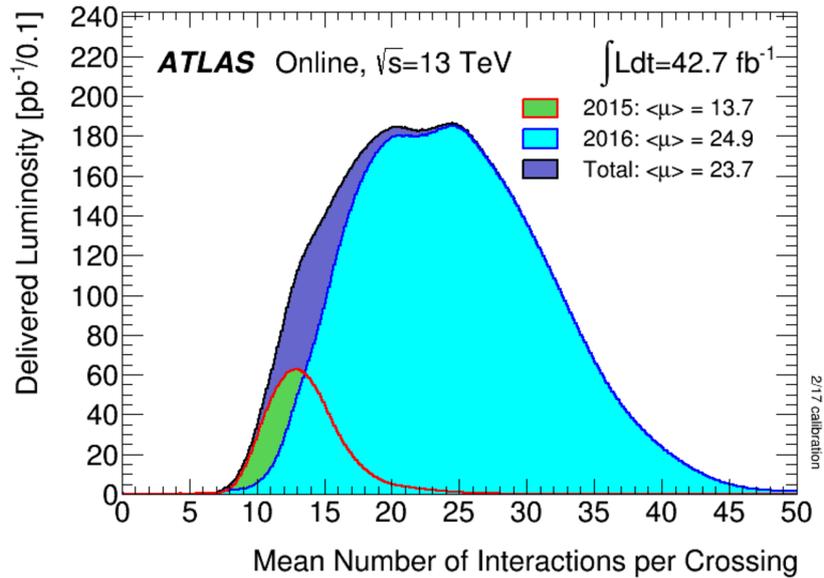


Figure 7: The luminosity-weighted distribution of the mean number of collisions per bunch crossing is shown for both the individual 2015 and 2016 data sets in addition to the two data sets combined [17].

3.2 The ATLAS experiment

ATLAS is short for A Toroidal LHC ApparatuS and the detector is a massive feat of engineering weighing 7000 tonnes at a height of 25 meters and a length of 44 meters. The ATLAS detector [13] exhibits the typical onion-like structure of a particle detector. The inner detector is situated at the core embedded in a magnetic solenoid. An electromagnetic and a hadronic calorimeter constitute the next layers and a muon spectrometer housed in a system of toroid magnets makes up the final layer. These sub-detectors are structured in three compartments. All detector components form concentric layers around the beam pipe in what is referred to as the barrel but are also mounted on disks at either side of the barrel known as end-caps. A schematic overview of the ATLAS detector is shown in figure 8 and an illustration of experimental signatures left by different particles in the ATLAS sub-detectors can be found in figure 9.

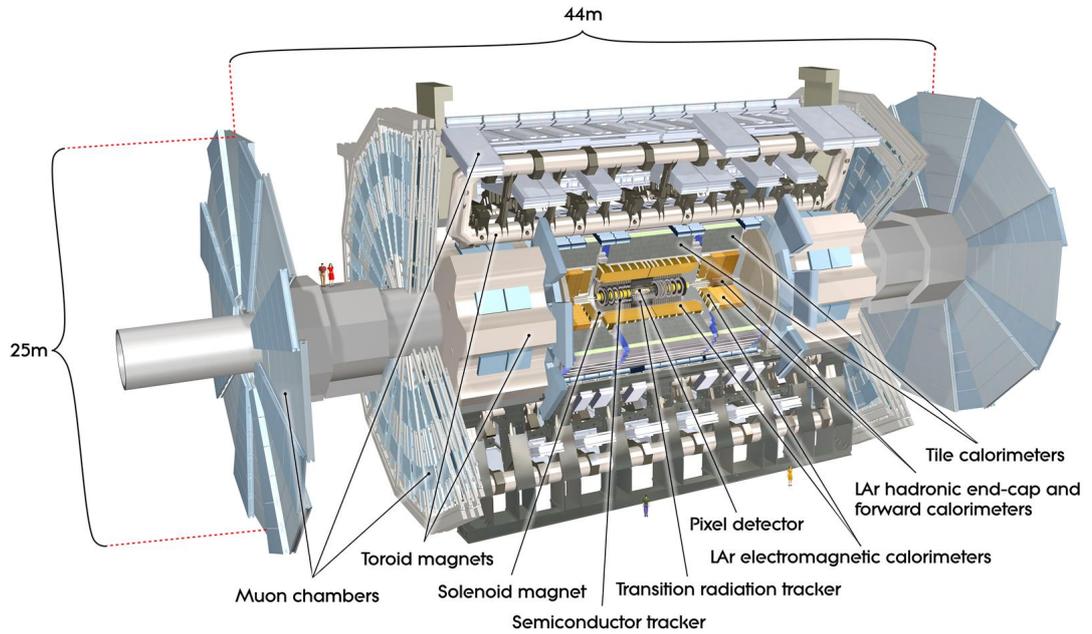


Figure 8: A schematic overview of the ATLAS detector [18].

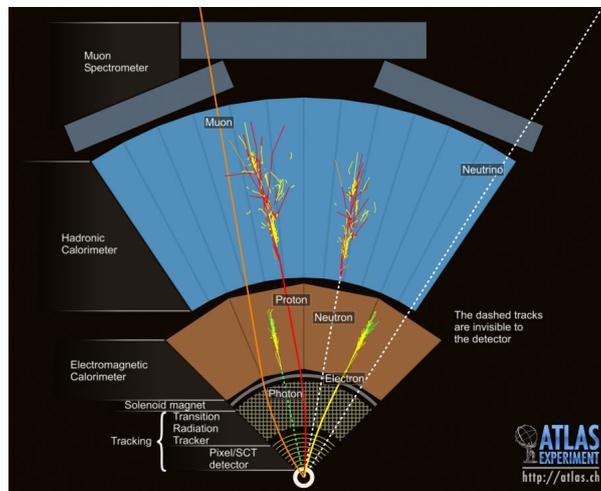


Figure 9: An overview of experimental signatures left by different particles in the ATLAS sub-detectors [18].

3.2.1 ATLAS coordinate system

The coordinate system used at ATLAS is defined with respect to the nominal interaction point of the proton-proton collisions. The z -axis lies along the beam-axis and the positive x -axis points towards the centre of the LHC ring while the positive y -axis points towards the sky.

Spherical coordinates are most often used. The polar angle represented by θ is measured from the z-axis in the xz-plane and the azimuthal angle denoted by ϕ is measured from the x-axis in the xy-plane. Pseudorapidity and rapidity are often used instead of θ and are defined by

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right),$$

and

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right)$$

respectively, where the energy of a particle is denoted by E and the z-component of the momentum is given by p_z . When the mass of a particle is negligible in comparison to its momentum the two definitions are equivalent. This is often the case for particles like electrons, muons and pions etc. However, when the mass of a particle (such as a heavy jet) is not insignificant in comparison to its momentum the rapidity definition is used.

The angular distance between particle objects is expressed as

$$R = \sqrt{\Delta\phi^2 + \Delta\eta^2}$$

and is used throughout for counting non overlapping particles observed in the detector.

3.2.2 The magnetic system

A system of magnets is used to obtain precision measurements of particle momenta. The Lorentz force on a particle with charge q moving at non-zero velocity \vec{v} under the influence of a magnetic field \vec{B} can be expressed as

$$\vec{F} = q \times \vec{v} \times \vec{B}$$

Here \vec{v} can be rewritten in terms of the particle momentum such that $\vec{p} = \gamma m \vec{v}$. The magnetic field will cause the particle trajectories to deflect and by relating the Lorentz force to the centripetal force one can calculate the momentum by finding the radius of curvature using the signals from the inner tracking detectors.

The inner detector is encompassed by a 2 Tesla magnetic solenoid and the toroid magnets surrounding the muon spectrometer have a strength of 0.5 Tesla both in the barrel and each of the end-caps.

3.2.3 The inner detector

The inner detector consists of a pixel detector (PD), a semiconductor tracker (SCT) and a transition radiation tracker (TRT) immersed in a magnetic solenoid. The location of the inner detector imposes additional requirements on its experimental design in order to properly reconstruct trajectories of charged particles. The presence of a strong magnetic field is necessary to bend tracks so close to the interaction region. Moreover, the high-density of tracks require a fine granularity detector.

The PD is closest to the beam-axis and therefore uses Silicon pixels instead of strips to achieve better resolution. It is made up of three pixel layers in the barrel and has three pixel disks in the two end-cap regions which results in a detector acceptance window of $|\eta| < 2.5$. The PD utilizes semiconductor technology in which a pixel sensor is made to function as a p-n junction. A reverse bias can be used to increase the depletion region and remove almost all free charge carriers. A traversing charged particle will then excite some electron-hole pairs in this region and thereby manifest itself as an electric current [19]. A staggering number of 80.4 million readout channels are used to meet the granularity requirement. An extra layer of pixels has been inserted between the existing B-layer (the innermost pixel layer) and a new thinner beam-pipe during a Run 2 update. This is referred to as the Insertable B-Layer (IBL) [20] and offers a number of advantages such as adding an additional measurement point, a redundancy which helps once pixel detector layers start to suffer from extensive radiation and increases in luminosity rate. Furthermore, a closer measurement point improves vertex reconstruction, so particles with appreciable lifetime such as b-hadrons can be better tagged.

The SCT uses the same semiconductor technology. Since the SCT covers a larger area and is further away from the interaction point than the PD, silicon micro-strips are used as opposed to pixels. Albeit cheaper the strips are not on par with the pixels in terms of resolution. To reconstruct tracks one needs to know the coordinate along a strip. This can be achieved by overlaying the strips with a small stereo angle [19]. Thus, when a particle traverses the SCT it will hit two strips at the same time effectively determining the coordinates but limiting the resolution to the cross section of the two strips. The detector acceptance provided by the SCT is again $|\eta| < 2.5$.

The TRT uses drift tubes filled with Xenon gas to detect particles. Xenon gas is chosen since it is sensitive to transition radiation emitted by high γ particles like electrons. This type of radiation will result in a much higher signal in the straws and the TRT can therefore help with particle identification of electrons in particular. An anode wire sits at the centre of each drift tube and when a charged particle enters it ionizes the Xenon gas freeing electrons in an avalanche effect that then travel towards the wire. By measuring the drift time of these electrons one can detect the coordinate of the initial particle trajectory at that point. Almost 300000 straws are used to provide precision tracking and are situated in a concentric circle around the beam-axis. A high p_T particle typically intersects 36 straws and by gathering the information from all of these hits one can reconstruct the track [19]. The resolution of the TRT is again comparably worse to that of the PD and SCT but the TRT technology is the more economical option at this volume. The coverage provided by the TRT is $|\eta| < 2.0$.

3.2.4 Calorimeters

The energies of charged and neutral particles are measured using electromagnetic and hadronic calorimeters. These are shortened to ECAL and HCAL respectively.

The ECAL and HCAL are sampling calorimeters which means that they consist of consecutive layers of absorber plates and active material. The absorber plates are very dense and thus results in a plethora of new particles as an initial particle traverses the material. The secondary particles in such a shower can then be detected in the active layers by using scintillating material.

Both the barrel and end-cap parts of the ECAL utilize a special accordion geometry to provide better particle detection [19]. A zig-zag pattern of lead absorber plates clad with copper electrodes prevent a particle shower from travelling in directions that are insensitive to picking up an electronic signal. Here liquid Argon is used as the scintillating material (LAr). The ECAL, furthermore, uses longitudinal segmentation which splits the calorimeter into three separate longitudinal compartments shaped like towers pointing at the origin. These can be used to provide a fast trigger response to a high energy particle originating from the interaction region.

The barrel HCAL is a sampling tile calorimeter [13]. Here a different calorimeter structure is used since pions constitute the majority of hadrons and are much harder to stop than electrons rendering the ECAL insufficient. Wave-length shifting fibres and photon multiplier tubes are used to register when a particle traverses plastic scintillator tiles embedded within iron absorbers. The end-cap HCAL differs from the barrel component by using a Copper-LAr sampling type.

A LAr forward calorimeter is also present (FCAL) to extend the coverage to $|\eta| < 4.9$. The FCAL consists of one ECAL with copper absorber plates and two HCAL with tungsten used instead.

3.2.5 The muon system

Muons are very penetrating particles, hence, the muon spectrometer (MS) is the outermost layer. The MS has four sub-detectors each of which utilizes a different technology, namely Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT), and Cathode Strip Chambers (CSC). This results in 4000 individual muon chambers that can detect muons with a minimum p_T of 3 GeV. The MS was optimized to be largely independent of other detector components in terms of muon reconstruction and therefore also has its own magnetic field consisting of large-scale air-core toroid magnets allowing for precision tracking.

The MDT is the main component of the MS and there are three MDT layers in the barrel and three corresponding layers in each of the end-caps. It is very similar to the TRT of the inner detector in terms of function. When a muon enters an MDT chamber it ionizes the gas and freed electrons drift towards the central anode wire. The track coordinate is again found by considering the drift time. However, in the end-cap regions the muon rate is too high to ensure proper function of the MDT and hence the multi-wire proportional chambers in the CSC are used instead. Furthermore, the MDT is too slow to provide a fast trigger response and is thus supplemented by RPC in the barrel and TGC in the end-caps. The RPC uses a parallel electrode-plate technology while TGC

again uses multi-wire proportional chambers. Although their main purpose is related to triggering they also provide measurements of the ϕ coordinate which is not obtainable from the MDT due to its orientation.

3.2.6 Trigger system

Around 1000 particles are produced every 25 ns within the detector's field of view at $|\eta| < 2.5$. This leads to an enormous demand on data storage far beyond what is feasible or sensible. Therefore, a two-level trigger system is implemented to efficiently select the events that matter. The first trigger level, called L1, looks at whether an event contains high- p_T muons, electrons, photons, jets or hadronically decaying taus. Large E_T^{miss} and total transverse energy are also considered. These snap decisions rely on coarse muon and calorimeter information and the passing events are given to the central trigger processor which implements a set trigger menu consisting of individual analysis-specific triggers. L1 lowers the event rate to approximately 100 kHz and for each event one or more regions of interest (ROI) based on significant features such as high- p_T are specified in terms of η and ϕ . This information is then used by the next level referred to as the high level trigger (HLT) to weed out even further by using the full detector information available within the ROIs. This lowers the event rate to approximately 1.5 kHz.

The triggers relevant within the context of this analysis are

HLT_mu26_ivarmedium or HLT_mu40 or HLT_mu50

and

HLT_mu22_mu8noL1.

HLT_mu26_ivarmedium and HLT_mu22_mu8noL1 require a leading muon with $p_T \geq 26$ GeV and $p_T \geq 22$ GeV respectively. The latter requires an additional muon with $p_T \geq 8$ GeV. The name ivarmedium in HLT_mu26_ivarmedium means that the muon must be isolated according to a set of criteria classified as medium. In events where either muon has $p_T \geq 40$ GeV or $p_T \geq 50$ GeV one doesn't want to use a trigger with isolation as there will often be energy deposits within the vicinity of such a lepton and one would therefore end up killing actual isolated leptons. Hence, the OR inclusion of HLT_mu40 (requires a leading muon of $p_T \geq 40$ GeV) and HLT_mu50 (requires a leading muon of $p_T \geq 50$ GeV) alongside HLT_mu26_ivarmedium.

4 Signal and background signatures

4.1 HNL signal

The experimental signature of the sterile neutrino signal such as that predicted by the ν MSM consists of two same-sign muons, one electron and missing transverse energy associated with the electron neutrino. Such a final state represents only one of the possible decay channels but a choice has been made to focus on this one alone.

The experimental signature arises from the final state of a prompt trilepton decay in which a W boson decays into a prompt muon and a sterile neutrino also referred to as a heavy neutral lepton (HNL). The HNL is created when the muon neutrino from the W decay mixes into an HNL and back into an anti-muon neutrino (due to its Majorana nature this can happen) which subsequently decays to an additional muon, an electron and an electron neutrino via an intermediate W decay. The Feynman diagram for this process is shown in figure 10. Note that the HNL signature is characterized by no opposite-sign, same-flavour (OSSF) lepton pairs and isolated leptons. Furthermore, this process clearly represents a violation of lepton flavour number conservation.

Furthermore, a trilepton and a dilepton invariant mass can be used as distinguishing features of the signal. The trilepton invariant mass reconstructed from the transverse momenta of the two muons and the electron will yield a peak just short of 80 GeV (the mass of the W boson) whereas the dilepton invariant mass reconstructed from the transverse momenta of the displaced muon and the electron will result in a peak slightly below the mass of the HNL. This can be used to discriminate against background processes. The additional requirement of no OSSF lepton pairs vetoes a significant amount of noise arising from backgrounds due to Drell-Yan pair production, $W + \text{jets}$ and $t\bar{t}$.

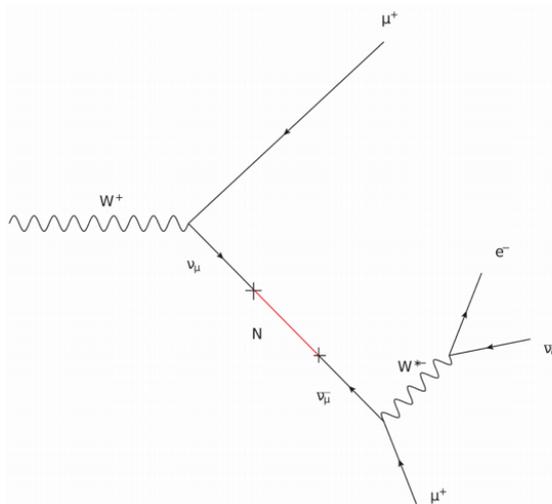


Figure 10: The experimental signature of the production and decay of an HNL exploited in the analysis described here.

4.2 Backgrounds to the search

This section provides some examples of production mechanisms associated with backgrounds to the search. In the following analysis all possible productions/decays are passed through selection to estimate the total background accurately.

4.2.1 Drell-Yan pair production

The process of Drell-Yan pair production (figure 11) can imitate the experimental signature in this analysis as illustrated by the Feynman diagram in figure 12 which shows Drell-Yan production including gluon emission. The majority of the Drell-Yan background is vetoed by the request of two leptons of the same flavour and the same charge but it is not negligible due to detector acceptance (one lepton from Z/γ is not within the range of the detector) and the presence of leptons in jets in $Z/\gamma + \text{jets}$ events having a non negligible chance of passing the event selection.

In addition there is of course $Z/\gamma \rightarrow \tau\tau$. Here if both taus decay leptonically to a muon and an electron then another muon with the same sign as the one from the tau decay can come from a jet.

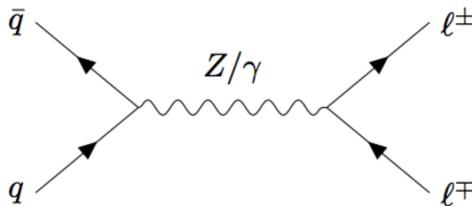


Figure 11: The Drell-Yan process.

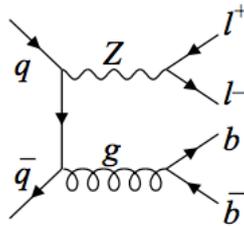


Figure 12: Drell-Yan + gluon emission [21].

4.2.2 Top quark production

Figure 13 shows how the HNL signature is predominantly imitated by $t\bar{t}$. Two leptons are produced if both W bosons decay leptonically. The remaining lepton can then arise from the associated b/c-quarks through semi-leptonic hadron decays. The majority of $t\bar{t}$ events can be vetoed due to b-tagging but since such an algorithm isn't 100% efficient a significant amount of $t\bar{t}$ events remains as background.

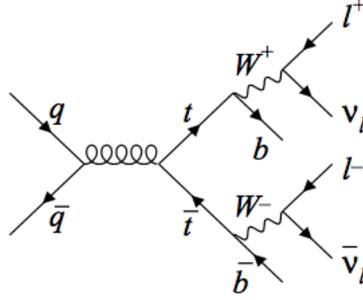


Figure 13: The $t\bar{t}$ production process [21].

Figure 14 shows how background events arise from a single top where a lepton is produced in a leptonic W decay. The two remaining leptons can again be found from semi-leptonic hadron decays of the associated b-quarks. Most single top events can be vetoed by the b-tagging algorithm but a non negligible amount will remain as background just like $t\bar{t}$.

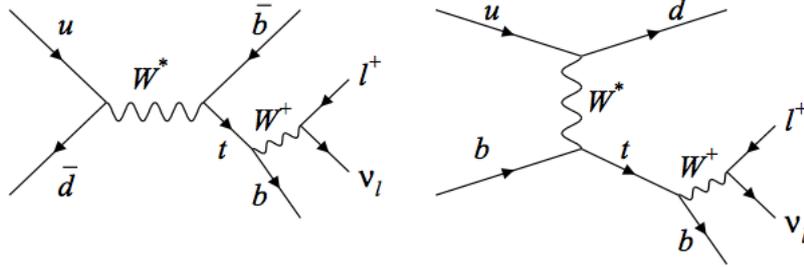


Figure 14: The s-channel (left) and t-channel (right) for the single top production process [21].

4.2.3 Diboson production

Typical diboson backgrounds arising from WW , WZ and ZZ are shown in figure 15. These backgrounds can mimic the experimental signature via hadronically and leptonically decaying bosons. The WZ background can result in a final state with all real leptons. The WW/ZZ backgrounds can have respectively one/two real lepton(s) in addition to a neutrino and two/one other lepton(s) from associated jets due to $q\bar{q}$. It should be noted that the all leptonic channel for WZ is not available in the current set of ntuples and is thus not included in the subsequent analysis.

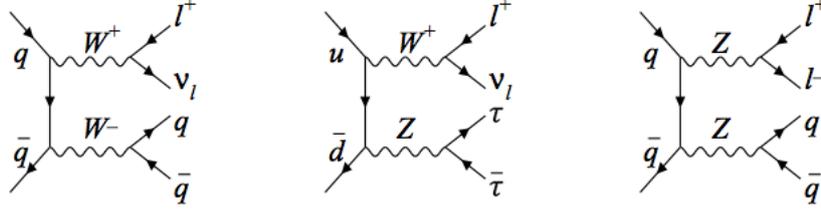


Figure 15: Diboson production [21].

4.2.4 Backgrounds arising from W + jets or multi-jets

Figure 16 shows an example of how the experimental signature is mimicked by W + jets decays. Here a W boson decays leptonically to produce a real electron or muon in addition to E_T^{miss} . The two additional leptons can then arise from b-jets or light jets produced via strong interactions.

Another possible background can be processes like that shown in figure 17. Due to the isolation and p_T requirements imposed on leptons this background can be reduced to negligible levels.

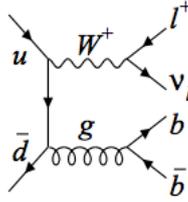


Figure 16: W + jets production [21].

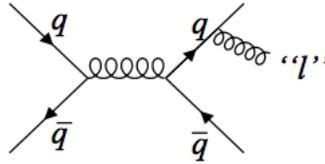


Figure 17: Multi-jet production [21].

4.3 Monte Carlo simulations

The analysis is using five signal simulation samples, namely for 5, 10, 20, 30 and 50 GeV HNL mass. Numbers for these samples are normalized in all signal predictions using the mixing angles shown in table 2. These values translate to moving just below the exclusion limit in figure 18.

HNL mass	$ V_{\mu N} ^2$
5 GeV	$3.87 \cdot 10^{-4}$
10 GeV	$1.20 \cdot 10^{-5}$
20 GeV	$3.76 \cdot 10^{-6}$
30 GeV	$4.64 \cdot 10^{-6}$
50 GeV	$3.38 \cdot 10^{-6}$

Table 2: The HNL masses and the corresponding mixing parameters used in this study.

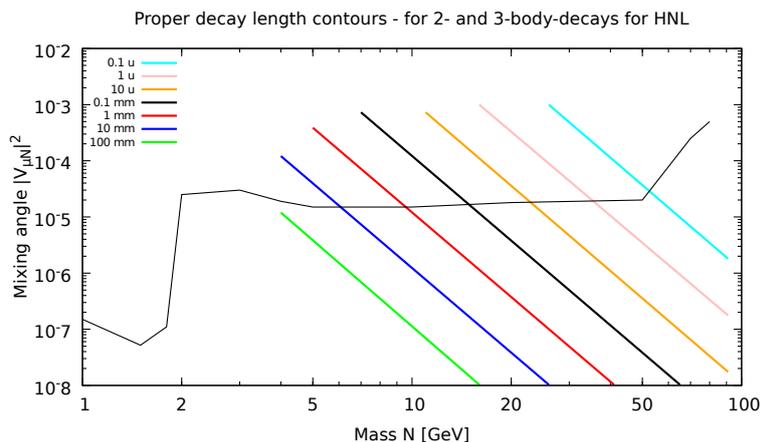


Figure 18: A sketch of the current exclusion region in the mass and mixing angle plane for a single sterile neutrino mixing with ν_μ . The coloured lines represent different proper decay lengths [22].

The background simulation samples consist of $t\bar{t}$, single top, $W + \text{jets}$, $Z/\gamma + \text{jets}$, inclusive Z/γ and dibosons. The generators Powheg [23], Pythia [24] and Sherpa [25] were employed for the Monte Carlo (MC) simulations. Both Sherpa and Pythia samples were available for Drell-Yan production so it was decided to run on both. They differ in that Sherpa uses next to leading order (NLO) diagrams to generate jets and also has high statistics for heavy and light quark flavours while Pythia is very inclusive in jets and uses jets from the parton shower process. Table 3 shows which generators are used for each sample.

The current Pythia model for MC generation uses a contact interaction approach (CI), where the HNL decays directly to a muon, an electron and an electron neutrino without an intermediate W. A model that includes an intermediate W vertex (W) would be the correct one and should be considered. Kinematic distributions of both models (CI and W) are therefore compared to

processes	MC generator
Signal	Pythia 8.2
Z/γ + jets (filter uds, b, c flavours)	Sherpa 2.2.1
inclusive Z/γ	Pythia 8.2
W + jets	Sherpa 2.2.1
$t\bar{t}$	Powheg and Pythia
single top	Powheg and Pythia
WW, ZZ and WZ	Sherpa

Table 3: The table lists which generators were used in the MC simulations.

test if there are important differences in the physical predictions of decay product kinematics.

The 5 and 50 GeV mass samples are chosen to quantify the differences in the CI and W decay modes for both low and high mass and basic phase space cuts are implemented to view the comparison in the context of this analysis. Hence $|\eta| < 2.5$ is required for all leptons. Additional lepton p_T requests are motivated by the optimal cut-based selection described in Section 7. The W and HNL muons are required to have $p_T(\mu_W) \geq 20$ GeV and $p_T(\mu_{\text{HNL}}) \geq 10$ GeV for the 5 GeV mass sample and $p_T(\mu_W) \geq 15$ GeV and $p_T(\mu_{\text{HNL}}) \geq 10$ GeV for the 50 GeV mass sample. The electron is required to have $p_T(e) \geq 5$ GeV in both cases.

The two samples were generated with 50000 events for the purpose of this study but the basic phase space cuts reduce the number of events to around one tenth and one fifth for 5 and 50 GeV respectively. Furthermore, the implementation of the basic phase space cuts result in efficiencies of 0.12 (CI) and 0.11 (W) for the 5 GeV mass and 0.22 (CI) and 0.21 (W) for the 50 GeV mass. Thus, the models perform approximately the same in terms of efficiency.

Distributions of p_T , $\Delta\eta$ and $\Delta\phi$ are shown in figure 19 and 20 for the 5 GeV HNL mass. The corresponding distributions for the 50 GeV HNL mass are shown in figure 21 and 22. The CI decay mode is shown in blue and the W decay mode is shown in red. One can see that the two models are in complete agreement (despite statistical fluctuations) when comparing distributions of the W boson, prompt muon and HNL as expected. By looking at the p_T distributions of the 5 and 50 GeV displaced muons one observes that an implementation of the CI decay mode results in a slight shift towards higher p_T , however, the overall difference between the two models can be considered negligible. No such model specific trends can be observed for $\Delta\eta$ or $\Delta\phi$.

Since there is no discernible difference in the efficiencies and spectra of the two models beyond statistical fluctuations and a slight shift in p_T there is no apparent need for re-weighting. To conclude, the CI and W decay modes are found to be comparable within the phase space relevant to the analysis and therefore the subsequent results presented in this thesis are valid.

The current analysis doesn't use a filter and consequently suffers from a crippling lack of statistics. A filter can be implemented where muons and electrons are

required to be within the detector's field of view; i.e $p_T > 5$ GeV and $|\eta| < 2.5$. An additional requirement can be made such that at least one muon always has $p_T > 15$ GeV. Thus, one avoids generating events that are thrown away at the object selection stage.

On a final note, the W decay mode tested here is now the default used in ATLAS along with the additional filter implementation which will help solve the issues with statistics.

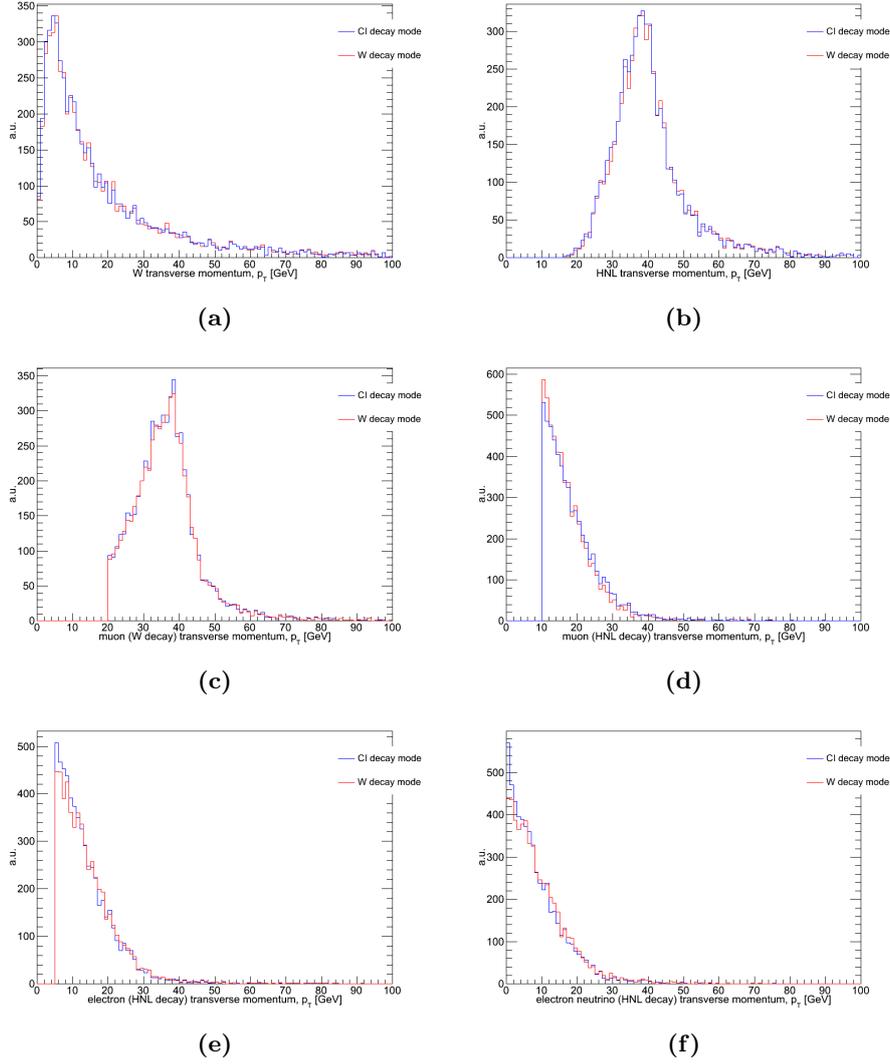


Figure 19: p_T distributions of the W boson (a), HNL (b), prompt muon (c), displaced muon (d), electron (e) and electron neutrino (f) for the 5 GeV HNL mass. Each distribution shows the comparison of the CI decay mode (blue) and W decay mode (red).

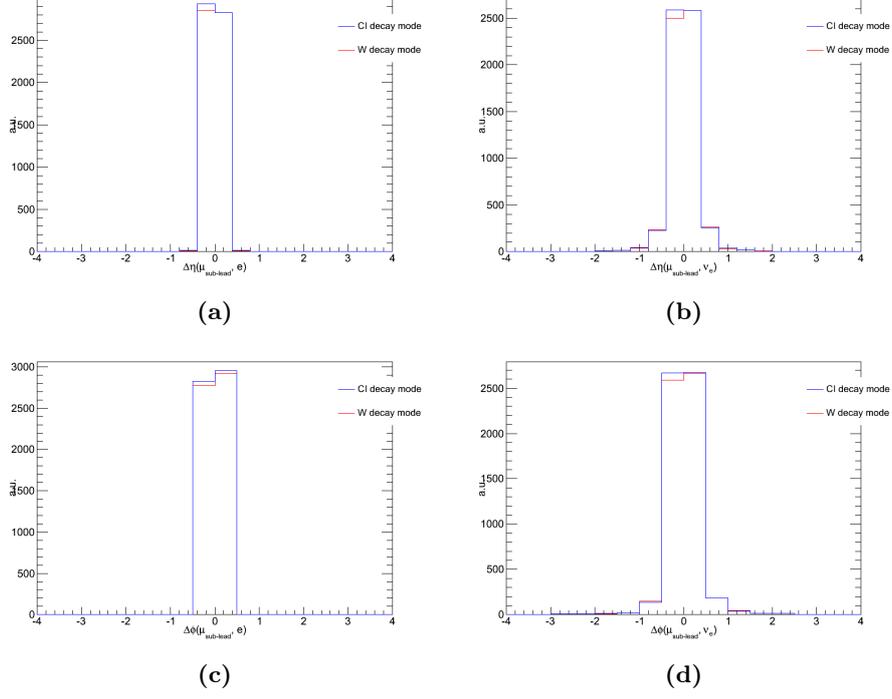


Figure 20: $\Delta\eta$ of the sub-leading muon and the electron (a), $\Delta\eta$ of the sub-leading muon and the electron neutrino (b), $\Delta\phi$ of the sub-leading muon and the electron (c) and $\Delta\phi$ of the sub-leading muon and the electron neutrino (d) for the 5 GeV HNL mass. Each distribution shows the comparison of the CI decay mode (blue) and W decay mode (red).

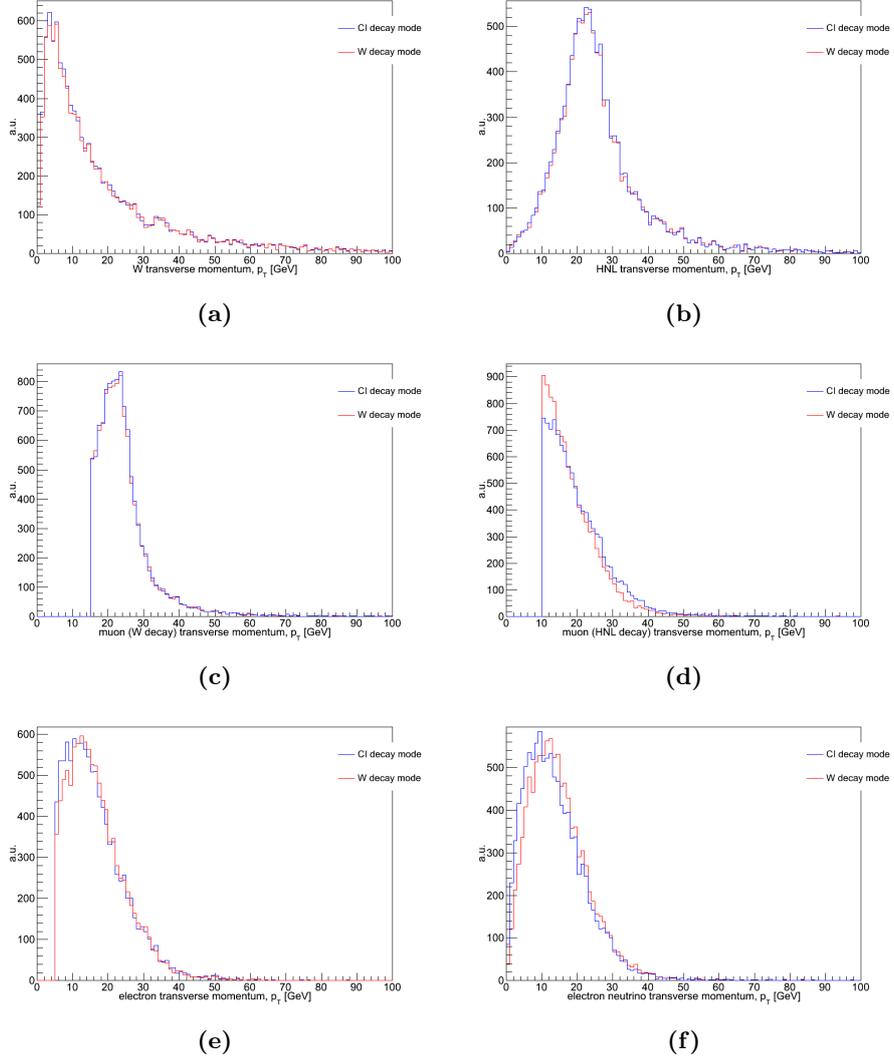


Figure 21: p_T distributions of the W boson (a), HNL (b), prompt muon (c), displaced muon (d), electron (e) and electron neutrino (f) for the 50 GeV HNL mass. Each distribution shows the comparison of the CI decay mode (blue) and W decay mode (red).

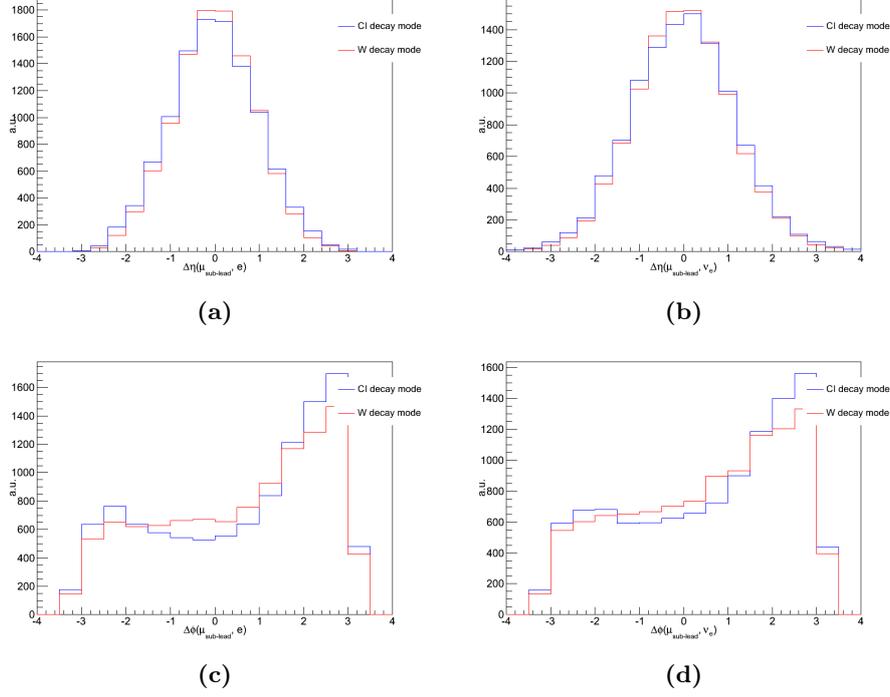


Figure 22: $\Delta\eta$ of the sub-leading muon and the electron (a), $\Delta\eta$ of the sub-leading muon and the electron neutrino (b), $\Delta\phi$ of the sub-leading muon and the electron (c) and $\Delta\phi$ of the sub-leading muon and the electron neutrino (d) for the 50 GeV HNL mass. Each distribution shows the comparison of the CI decay mode (blue) and W decay mode (red).

4.4 Translation of MC simulations to physical predictions

To go from MC simulations to the actual signal and background predictions used in this analysis requires some additional considerations. The following illustrates how the cross section is computed at each step in the signal decay process

$$\begin{aligned}
 f\bar{f} &\rightarrow W^\pm & \sigma(f\bar{f} \rightarrow W^\pm) \\
 &\rightarrow \mu^\pm \nu_\mu & \times \text{BR}(W^\pm \rightarrow \mu^\pm \nu_\mu) \\
 &\rightarrow \mu^\pm N & \times |V_{\mu N}|^2 \\
 &\rightarrow \mu^\pm \mu^\pm e^\mp \nu_e & \times \text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e),
 \end{aligned}$$

where it is used that

$$\text{BR}(W^\pm \rightarrow \mu^\pm \nu) = |V_{\mu N}|^2 \times \text{BR}(W^\pm \rightarrow \mu^\pm N).$$

It should again be emphasized that a muon neutrino mixes with a single sterile neutrino. This leads to a signal cross section of

$$\sigma_{\text{signal}} = \sigma(f\bar{f} \rightarrow W^\pm) \times \text{BR}(W^\pm \rightarrow \mu^\pm \nu_\mu) \times |V_{\mu N}|^2 \times \text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e). \quad (17)$$

In the MC Pythia simulation for the signal the branching ratios $\text{BR}(W^\pm \rightarrow \mu^\pm \nu_\mu)$ and $\text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e)$ are set to 100% to not waste processing time. Moreover, the mixing angle $|V_{\mu N}|^2$ is set to 1. To translate the Pythia simulations into actual physical predictions one has to multiply the Pythia cross section for W production by the correct values. For a signal sample with an HNL mass M_N , a proper decay length $c\tau_N$ (τ_N denotes the lifetime) and a mixing angle $|V_{\mu N}|^2$ the branching ratio $\text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e)$ can be obtained from the partial width [12] and total width given by

$$\Gamma(N \rightarrow l_\alpha^- l_\beta^+ \nu_\beta) = \frac{G_F^2 M_N^5 |V_{\alpha N}|^2}{192\pi^3} \quad (\alpha \neq \beta)$$

and

$$\Gamma = \frac{1}{\tau_N}$$

respectively, where G_F is the Fermi constant. The W production cross section $\sigma(f\bar{f} \rightarrow W^\pm) \sim 70.820$ nb is given by the Pythia generator. Furthermore, one has that $\text{BR}(W^\pm \rightarrow \mu^\pm \nu_\mu) \sim 10.1\%$ and $\text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e) \sim 26.3\%$. It should be noted that $\text{BR}(N \rightarrow \mu^\pm e^\mp \nu_e)$ is independent of $|V_{\alpha N}|^2$ (see appendix A in [11]) which means that the only mixing angle dependence of σ_{signal} is on $|V_{\mu N}|^2$.

Finally, it should be noted that each signal/background event is scaled by a Monte Carlo weight (assigned by the generator to obtain the distributions seen in data) and normalized to the appropriate signal/background cross section as observed in 36 fb^{-1} . Each event is also scaled by a pile-up weight that emulates the pile-up conditions seen in data.

5 Object reconstruction and identification

Particles traversing the detector leave signals which are then reconstructed into objects: electrons, muons, jets, taus and missing transverse energy (E_T^{miss}). This section will give an overview of the object descriptions pertaining to this analysis.

5.1 Muon reconstruction

The reconstruction of muons [26] starts by finding tracks in the inner detector (ID), in the muon spectrometer (MS) and using additional information about energy deposits in the calorimeters. There are four different types of muons: muons found only in the ID, muons found only in the MS, muons found by combining ID tracks and calorimeter measurements and muons found by combining ID tracks and MS hits together.

This analysis uses combined muons (i.e ID tracks and MS hits together). Furthermore, the identification criteria can be loose, medium or tight, where the names suggest the relative efficiency of the imposed requirements. For the purpose of this analysis loose is required. Additional requirements are imposed on the transverse momentum and pseudorapidity of the muons such that $p_T > 5$ GeV and $|\eta| < 2.5$.

5.2 Electron reconstruction

The reconstruction of electrons [26] starts by looking at ID tracks and calorimeter measurements. Like muons, the identification of electrons can be loose, medium or tight.

In this analysis loose is used for the electron and an additional requirement of 1 b-layer hit is added to follow recommendation guidelines. The transverse momentum of the electron is required to fulfil $p_T > 5$ GeV and the pseudorapidity is required to be within the calorimeter acceptance which translates to $|\eta| < 2.47$. An additional requirement is imposed on the pseudorapidity; an electron is vetoed if it falls within the region where the barrel and end-cap calorimeters meet (also known as the calorimeter crack). In terms of pseudorapidity this transition is $1.37 < |\eta| < 1.52$.

5.3 Jet, b-jet and E_T^{miss} reconstruction

The experimental signature of jets is characterized by topologically-related energy deposits in the calorimeters (topo-clusters) associated with bundles of tracks in the ID. The anti- k_t algorithm is used to perform jet reconstruction based on topo-clusters with a distance parameter of $R = 0.4$ and the jet energy is calibrated using the EM+JES scheme [27]. The transverse momentum of all jets (including b-jets) must fulfil $p_T > 20$ GeV and the pseudorapidity has to be within the acceptance range of the detector i.e $|\eta| < 2.5$. Jets with $p_T < 60$ and $|\eta| < 2.4$ are, furthermore, required to pass the default jet vertex tagger (JVT) medium working point. The JVT score is the output of a tagging algorithm used to identify and select jets coming from the hard scatter interaction through the use of tracking and vertex information [28].

Some of the jets can be classified as b-jets if they originate from a b-quark and this is known as b-tagging [29]. B-tagging algorithms generally require the jet to pass a set of quality criteria based on the long-lifetime, high mass and decay multiplicity of b hadrons (when a b-quark is produced it will become a bound state of a b-quark and lighter quarks called a b hadron). The b-tagging algorithm MV2c10 is used with a fixed working point cut (FixedCutBEff_77) which tags jets with a 77% efficiency [30].

Neutrinos are invisible but can instead be identified through E_T^{miss} . One computes the sum of the transverse momenta of identified muons, electrons, tau hadronic decays, jets and leftover energetic calorimeter clusters or tracks. The E_T^{miss} is then found by taking the negative of this sum in order to achieve a balance in the transverse plane [26].

5.4 Isolation

Isolation is defined with respect to a ΔR cone around an object trajectory. Other objects within the vicinity of the track can then be detected by looking at calorimeter energy deposits or tracks within the cone. The presence of 2 to 3 isolated leptons in the signal is one of the main features in Section 7 and it is quite effective in separating background from signal. Isolation working points (WP for short) are used in this analysis and thus a more comprehensive explanation will be given as to what these actually mean [31]. The isolation WPs are called *loose track only*, *loose*, *tight*, *gradient loose* and *gradient* each of which mean a specific cut on a discriminating variable as shown in the middle column in table 4.

For example, *loose* is defined using both $p_T^{\text{varcone30}}/p_T^\mu$ and $E_T^{\text{topocone20}}/p_T^\mu$ for muons. The first isolation variable $p_T^{\text{varcone30}}/p_T^\mu$ is the sum of p_T of the tracks that lie within a cone of $R < 0.3$ around the muon track. The second isolation variable $E_T^{\text{topocone20}}/p_T^\mu$ is the sum of E_T of the topo-clusters that lies within a cone of $R < 0.2$ around the muon track.

The requirements on the discriminating variables are summarized in the rightmost column in table 4. The same isolation WP definitions are used for both electrons and muons except $p_T^{\text{varcone30}}$ is replaced by $p_T^{\text{varcone20}}$ for electrons. For *loose* one sees that a 99% efficiency constant in η and p_T is required. This means that efficiencies are measured within 6 different p_T bins (5–6, 6–7, 7–8, 8–10, 10–12, and 12–15 GeV) for a given bin of η as seen in the left-hand plot in figure 23. The working point cuts are then set to values that generate efficiencies of 99% as specified in table 4. The right-hand plot in figure 23 shows the case for the *tight* isolation WP for which efficiencies of 96% are now required.

In this analysis the isolation WPs *gradient*, *tight* and *gradient* are used for the leading muon, the sub-leading muon and the electron objects respectively. The motivation for this is described in Section 7.

Isolation WP	Discriminating variable(s)	Definition
<i>LooseTrackOnly</i>	$p_T^{\text{varcone30}} / p_T^\mu$	99% efficiency constant in η and p_T
<i>Loose</i>	$p_T^{\text{varcone30}} / p_T^\mu, E_T^{\text{topocone20}} / p_T^\mu$	99% efficiency constant in η and p_T
<i>Tight</i>	$p_T^{\text{varcone30}} / p_T^\mu, E_T^{\text{topocone20}} / p_T^\mu$	96% efficiency constant in η and p_T
<i>Gradient</i>	$p_T^{\text{varcone30}} / p_T^\mu, E_T^{\text{topocone20}} / p_T^\mu$	$\geq 90(99)\%$ efficiency at 25 (60) GeV
<i>GradientLoose</i>	$p_T^{\text{varcone30}} / p_T^\mu, E_T^{\text{topocone20}} / p_T^\mu$	$\geq 95(99)\%$ efficiency at 25 (60) GeV
<i>FixedCutTightTrackOnly</i>	$p_T^{\text{varcone30}} / p_T^\mu$	$p_T^{\text{varcone30}} / p_T^\mu < 0.06$
<i>FixedCutLoose</i>	$p_T^{\text{varcone30}} / p_T^\mu, E_T^{\text{topocone20}} / p_T^\mu$	$p_T^{\text{varcone30}} / p_T^\mu < 0.15, E_T^{\text{topocone20}} / p_T^\mu < 0.30$

Table 4: Working point definitions for muons that also apply to electrons except $p_T^{\text{varcone30}}$ is replaced by $p_T^{\text{varcone20}}$ [31].

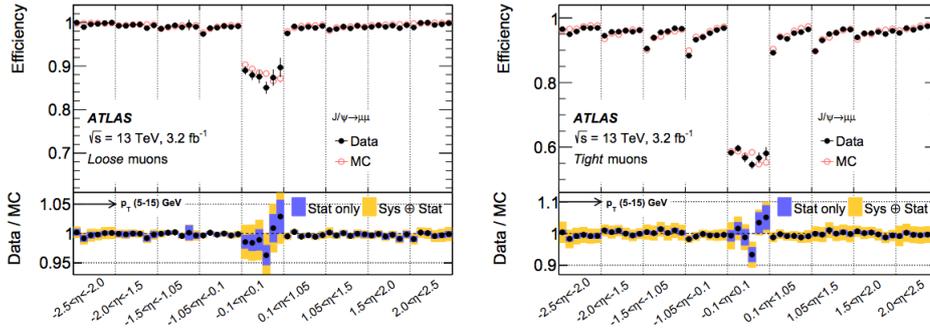


Figure 23: Distributions for loose (left) and (tight) working point definitions for muons. Efficiencies are measured within 6 different p_T bins (5–6, 6–7, 7–8, 8–10, 10–12, and 12–15 GeV) for a given bin of η . The working point cut is set to a value that generates the efficiencies specified in table 4 [31].

5.5 Overlap removal

The same particle might be reconstructed as different objects (e.g an electron can be reconstructed as an electron but also as a jet with 1 track) and therefore one needs to perform an overlap removal. This analysis removes jets based on ΔR considerations to find unique particles and not double counted objects. This is done with AssociationUtils and SUSY Tools. The overlap removal is only performed for jets, muons and electrons with respect to each other. No overlap removal is performed for muons, electrons and jets with respect to taus. The objects are removed in the following order.

1) Overlap removal for electrons and jets:

Jet overlap removal is performed first with respect to electrons. Jets with tracks

that are within $\Delta R < 0.2$ of an electron track are removed and the electron is kept. However, jets with a 20% higher p_T than the electron are not removed based on this requirement. B-jets are also not removed.

Electron overlap removal is now performed with respect to the remaining jets. If the electron is within the jet sliding cone defined by $\Delta R < 0.04 + 10 \text{ GeV}/p_T(e) < 0.4$ the electron is removed and the jet is kept.

2) Overlap removal for muons and jets:

Jet overlap removal is performed first with respect to muons. Jets with tracks that are within $\Delta R < 0.2$ of a muon track are removed. However, if a jet has three or more tracks AND either a 50% higher p_T than the muon or a jet track sum with a 30% higher p_T than the muon it is not removed based on this requirement. B-jets are also not removed.

Muon overlap removal is now performed with respect to the remaining jets. If the muon is within the jet sliding cone defined by $\Delta R < 0.04 + 10 \text{ GeV}/p_T(\mu) < 0.4$ the muon is removed and the jet is kept.

3) Overlap removal for electrons and muons:

A calorimeter tagged muon that shares the same track as an electron is removed first. The electron is removed if it shares the same track with a remaining muon.

6 Truth-level analysis

The most basic signal requirements are two muons and one electron. The current way to identify the muons from W and HNL decay is by the transverse momenta of the two muons. The prompt muon (originating from the parent W) is selected based on the assumption that it has the highest muon- p_T . Hence, it's referred to as the leading muon. On the other hand, the displaced muon (originating from the HNL) is selected by assuming that it has the second highest muon- p_T . This is then referred to as the sub-leading muon.

This assumption holds up fairly well when the HNL mass is relatively low but breaks down for high HNL masses since a high mass particle gives a significant p_T to its decay daughters. One needs to know the extent of such confusion in assignment since the mass of the HNL daughters is a possible important variable for signal discrimination.

An investigation into a better selection strategy using truth-matched particles is performed in the following section. Only the extreme mass cases are considered, namely 5 and 50 GeV.

6.1 Muon selection for 5 GeV

This section provides an investigation into the current muon selection strategy and its impact on the 5 GeV HNL mass signal. A p_T baseline motivated by the optimal cut-based selection in Section 7 is implemented to focus on the assignment problem within the relevant p_T region. The prompt muon is required to have $p_T(\mu_W) \geq 20$ GeV and the displaced muon is required to have $p_T(\mu_{\text{HNL}}) \geq 10$ GeV.

Table 5 lists the total number of displaced muons, the displaced muons found in the sub-leading muon position (sub-leading displaced muons) and the displaced muons found in the leading muon position (leading displaced muons). The latter can be inferred but is included to state the mislabelling explicitly. The leading displaced muons only constitute around 4% of the total, so as expected the assumption that the displaced muon generally carries the second highest muon- p_T holds up well in this case.

The difference in transverse momentum, pseudorapidity and azimuthal angle are plotted for the displaced/prompt muon in the leading/sub-leading position and the electron. These plots are then compared to corresponding plots for the displaced muon in the sub-leading position and the electron (see appendix A.1, figure A.2 to A.4). The reconstructed HNL mass $M(\mu, e)$ was additionally considered and similar plots were performed (see appendix A.1 figure A.1). It is observed that an $M(\mu, e)$, $\Delta\eta$ or $\Delta\phi$ check would allow one to correctly identify the prompt and displaced muons statistically. For example both sub-leading and leading displaced muons fulfil $M(\mu_{\text{HNL}}, e) < 10$ GeV while one observes $M(\mu_W, e) > 10$ GeV for the sub-leading prompt muons. Given that the misidentification is small and there is a plan to perform the signal selection using a neural network in the long term, no intermediate fix is sought for in the case of a 5 GeV HNL search.

number of displaced muons	counts
total	305
sub-leading	293
leading	12

Table 5: Mislabelling rate of muons from HNL decay in the current p_T -based selection strategy for the 5 GeV HNL mass.

6.2 Muon selection for 50 GeV

This section provides an investigation into the current muon selection strategy and its impact on the 50 GeV HNL mass signal. A p_T baseline motivated by the optimal cut-based selection in Section 7 is again used. The prompt muon is required to have $p_T(\mu_W) \geq 15$ GeV and the displaced muon is required to have $p_T(\mu_{\text{HNL}}) \geq 10$ GeV. Table 6 shows the mislabelling rate of muons from HNL decay in the current p_T -based selection strategy. The number of leading displaced muons constitutes around 30% of the total. Hence, the assumption that the displaced muon possesses the second highest muon- p_T is incorrect in approximately one third of all cases. This is clearly problematic and it's imperative to test if there is a better option to the current assignment rule.

The same type of plots described in the previous section are again used to see if one can perform a check of $M(\mu, e)$, ΔP_T , $\Delta\eta$ or $\Delta\phi$ that would allow one to correctly identify the prompt and displaced muons statistically (see appendix A.2, figure A.5 to A.8). However, there is no way to fix the swapping issue since distributions for the leading displaced and sub-leading prompt muons completely overlap with corresponding distributions for the sub-leading displaced muons. It is hoped that a neural network will be able to spot different correlations between signal and background for the lepton kinematics and thereby allow discrimination nonetheless.

number of displaced muons	counts
total	3126
sub-leading	2181
leading	945

Table 6: Mislabelling rate of muons from HNL decay in the current p_T -based selection strategy for the 50 GeV HNL mass.

7 Signal selection

In order to find a signal out of a much more abundant background, an optimal set of cuts must be found to yield the highest significance which is defined as S/\sqrt{B} for the purpose of this analysis. This is the topic of this section.

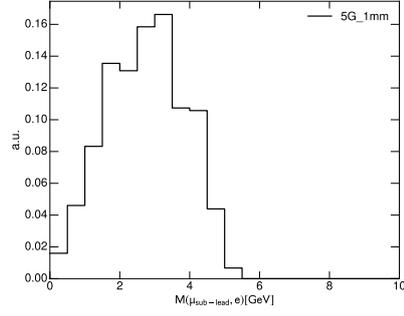
7.1 Cut optimization

Preliminary cuts for starting the cut optimization are shown in table 7. In addition to the basic object selection described in Section 5, no opposite sign, same flavour (OSSF) leptons is required. This is imposed since the analysis focuses on Majorana neutrinos.

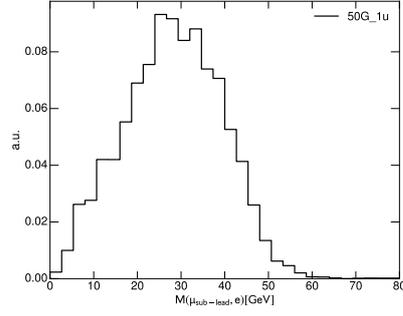
Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
$\text{Invmass}(\mu, \mu, e) < 90 \text{ GeV}$
$\text{Invmass}(\mu, \mu, e) > 40 \text{ GeV}$
$\text{Invmass}(\mu, e) < M_{\text{HNL}} \text{ GeV}$

Table 7: The pre-selection cuts are used for HNL mass point samples of 5, 10, 20, 30 and 50 GeV. All subsequent plots in this section are produced based on these cuts unless otherwise specified.

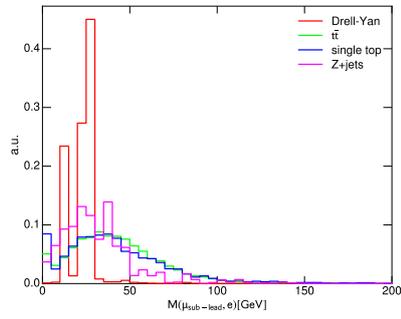
Additionally, we impose two requirements on the trilepton invariant mass, $\text{invmass}(\mu, \mu, e)$, and one requirement on the dilepton invariant mass, $\text{invmass}(\mu, e)$. The values are chosen such that the focus is on the interesting signal region where the trilepton invariant mass is consistent with the mass of the W boson and the dilepton invariant mass is consistent with the HNL mass. Figure 24 and 25 show the dilepton and trilepton invariant mass respectively for both the 5 and 50 GeV HNL mass points. Corresponding plots for the 10, 20 and 30 GeV HNL mass points can be found in appendix B. From this one can see that the chosen cuts in table 7 are justified. It should be noted that both MC samples for Drell-Yan production are shown in the figures in this section. The Pythia sample is labelled Drell-Yan and the Sherpa sample is labelled $Z + \text{jets}$. With respect to the optimization, only the Sherpa sample matters since the majority of the Pythia sample dies down after implementing the preliminary cuts.



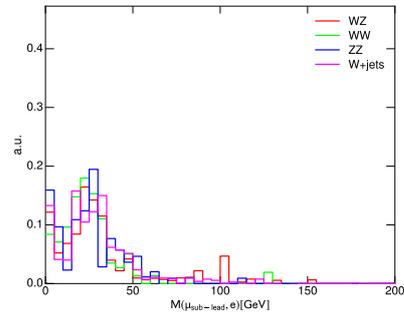
(a)



(b)



(c)



(d)

Figure 24: Dilepton invariant mass distributions for the 5 and 50 GeV HNL masses. The 5 GeV signal is shown in (a) and the 50 GeV signal is shown in (b). The backgrounds are shown in (d) and (c).

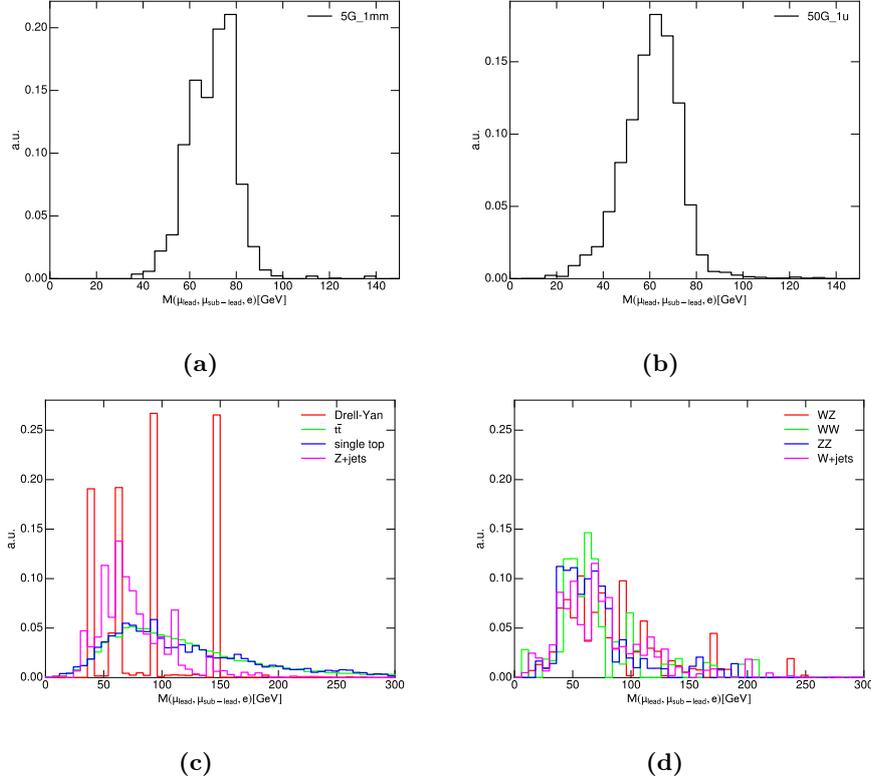


Figure 25: Trilepton invariant mass distributions for the 5 and 50 GeV HNL masses. The 5 GeV signal is shown in (a) and the 50 GeV signal is shown in (b). The backgrounds are shown in (d) and (c).

The cut optimization procedure is performed for all signal masses. However, only the extreme cases are highlighted, i.e 5 and 50 GeV. Corresponding tables and figures for the other masses can be found in the appendix. The best set of cuts is first found for a completely free choice of trigger and the best trigger is then found with respect to the optimal off-line selection.

The first cut optimization is performed on the transverse momenta of all three leptons. The leading muon was considered first, followed by the sub-leading muon and finally the electron. Thus, it should be noted that the subsequent cuts found for the sub-leading muon and the electron will be dependent on the optimal cut chosen for the leading muon. The cut procedure is shown in figure 26 and 27 for the leading muon. Figure 28 and 29 show similar plots for the sub-leading muon after implementing the optimal cut on the leading muon p_T , and figure 30 and 31 show plots for the electron after imposing the optimal cuts on both muons. Corresponding plots for the 10, 20 and 30 GeV HNL mass samples are shown in figure C.1 to C.9 in appendix C.

The cut on the leading muon p_T was placed at $p_T(\mu_{\text{lead}}) \geq 20$ GeV for the 5, 10 and 20 GeV HNL mass samples. For the 30 and 50 GeV HNL masses, the cut was placed at $p_T(\mu_{\text{lead}}) \geq 15$ GeV. The cut on the p_T of the sub-leading muon was placed at $p_T(\mu_{\text{sub-leading}}) \geq 10$ GeV for all masses. This was decided since

the sharp peak in significance at $p_T(\mu_{\text{sub-lead}}) \geq 15$ GeV was due to removing a single peak of $W + \text{jets}$. Plots showing how the significance varies with each background for an example mass of 10 GeV can be found in appendix C.4. No improvement in significance was found when varying the electron p_T for any of the HNL masses. However, a default cut of $p_T(e) \geq 5$ GeV comes from the filter used to generate the ntuples and this will be used henceforth. An additional cut was placed on the number of b-jets. The majority of the signal in all mass cases was found to have 0 b-jets while a significant number of events in the background had several b-jets. Thus keeping only events with 0 b-jets clearly results in an increase in significance while preserving most of the signal.

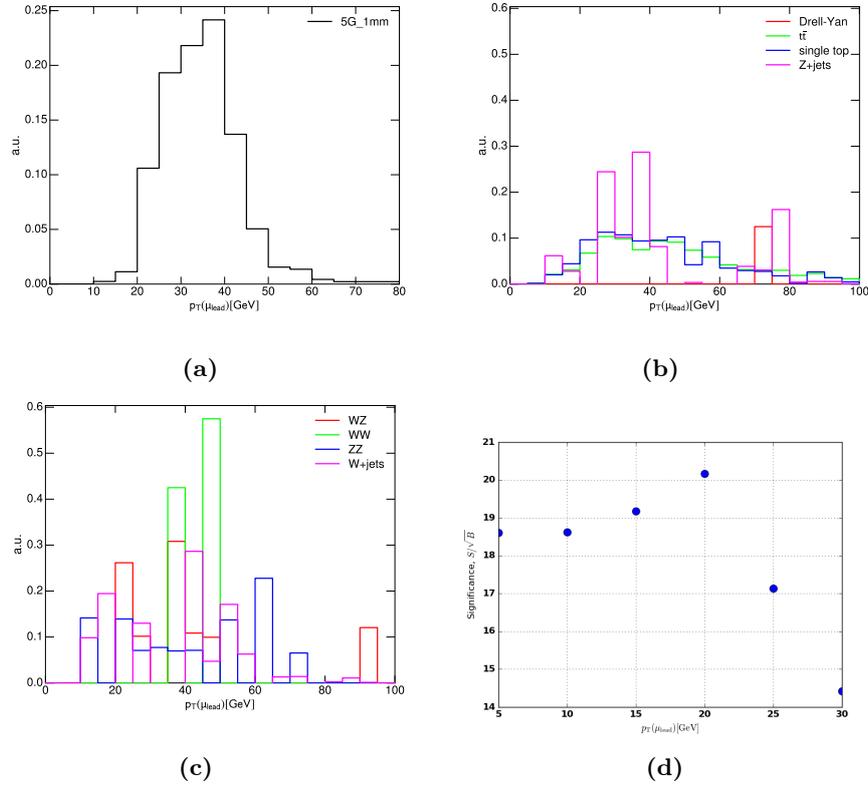


Figure 26: Choice of the leading muon p_T cut for the 5 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

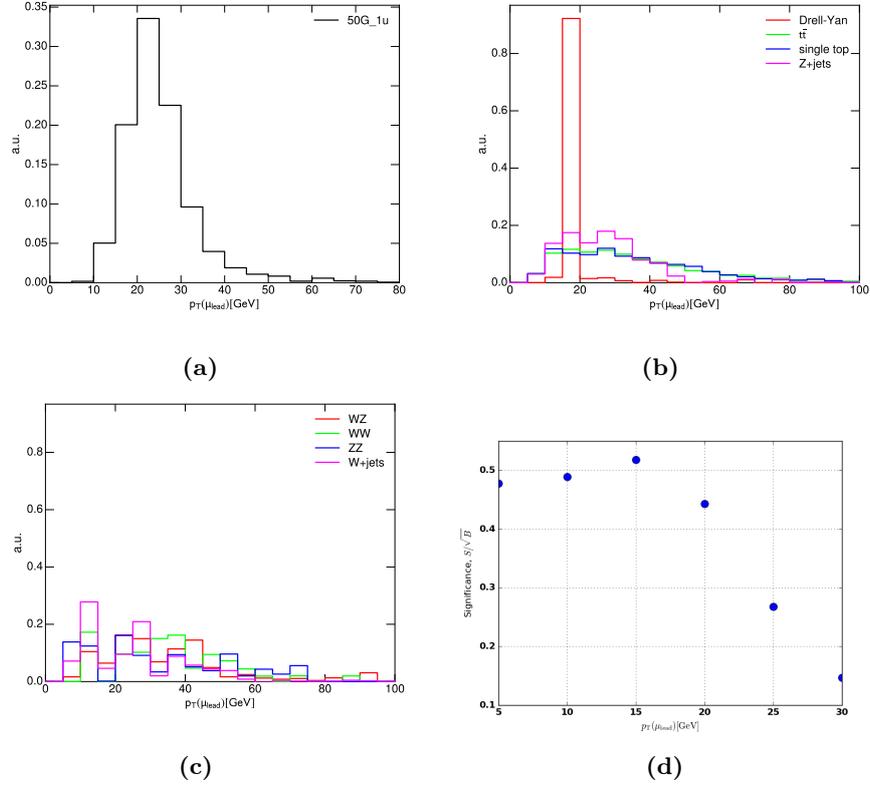


Figure 27: Choice of the leading muon p_T cut for the 50 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

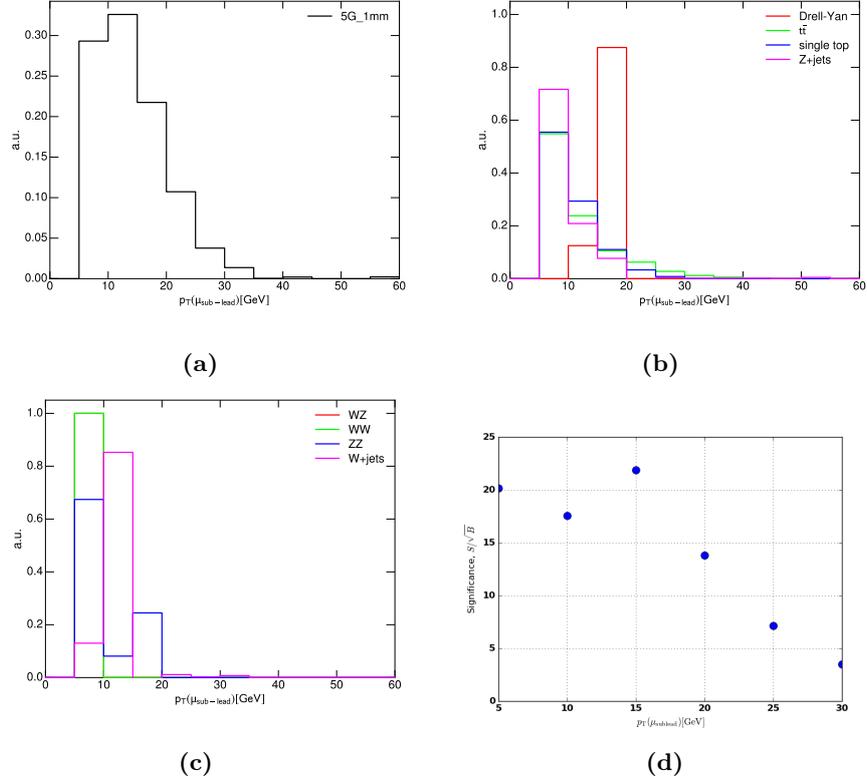


Figure 28: Choice of the sub-leading muon p_T cut for the 5 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

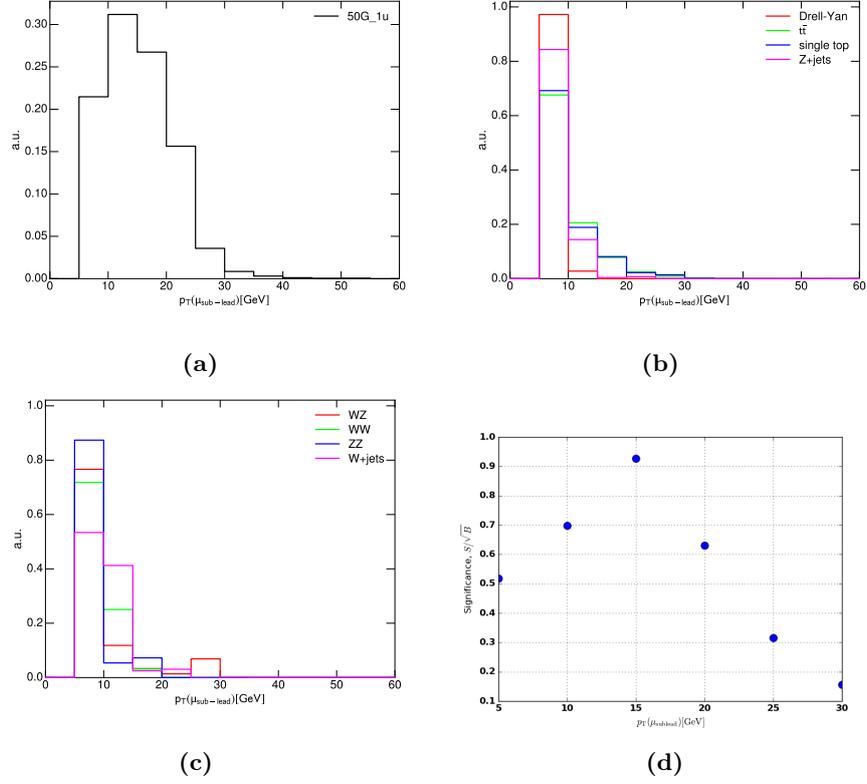


Figure 29: Choice of the sub-leading muon p_T cut for the 50 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

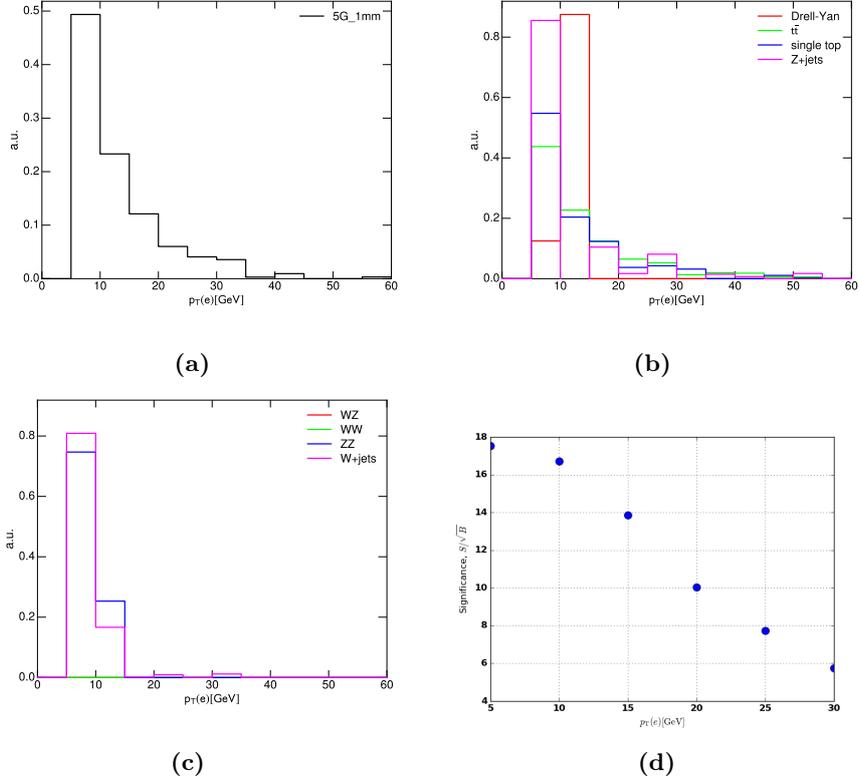


Figure 30: Choice of the electron p_T cut for the 5 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

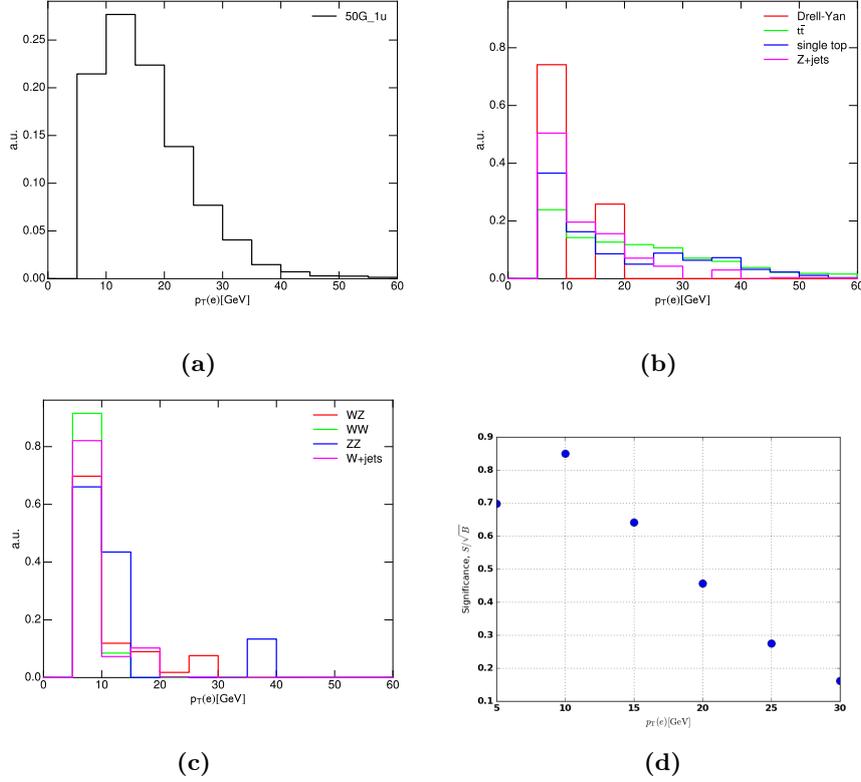


Figure 31: Choice of the electron p_T cut for the 50 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

7.2 Isolation

All leptons in HNL signal events are isolated in general, however one or more leptons in $W + \text{jets}$, $Z + \text{jets}$ and $t\bar{t}$ are not. Furthermore, the best way to estimate these types of backgrounds is to use a data-driven approach, where one identifies a suitable variable or property of the leptons to invert. This feature can then be used in a matrix method for determining these reducible backgrounds. Isolation variables for the leptons can be potential candidates to invert in such an approach. For these reasons an attempt is made to identify the best isolation cuts on the leptons.

In this analysis non-isolated and non-prompt leptons contribute mainly via the sub-leading muon and the electron. This motivates at least an isolation cut on one of them if not both. The leading muon is isolated in signal and background because it mostly stems from the decay process of a W or Z, therefore applying isolation on the leading muon is mostly helpful in reducing the multi+jet background to negligible levels without affecting the signal.

Distributions were plotted showing signal and background impact parameters divided by the corresponding errors, denoted by d_0/σ_{d_0} , for the sub-leading muon and the electron (see appendix D). It was evident from the overlap be-

tween signal and background in both cases that this variable was not suitable to cut upon. Thus only isolation will be considered for the sub-leading muon and the electron.

It should be noted that the sub-leading muon and the electron are much less isolated for a 5 GeV HNL mass than for a 50 GeV HNL mass since the decay daughters will be more boosted in the former case. Nonetheless, isolation was found to improve signal versus background discrimination for both low and high masses.

The two following sections investigate the impact of pre-defined isolation working points (WP) on the different signal samples and backgrounds.

7.2.1 Leading muon WP isolation requirements

The following section shows the significance (table 8), signal events (table 9) and background events (table 10) when applying different isolation WPs on the leading muon for the 5, 10, 20, 30 and 50 GeV HNL masses. All the isolation WPs are found to perform very similarly in terms of both significance, signal and background for all masses. As the leading muon is isolated from both the sub-leading muon and the electron when considering the actual physical decay process it is completely justified to apply gradient isolation.

Cut	Significance				
	5 GeV	10 GeV	20 GeV	30 GeV	50 GeV
No isolation	19.95 ± 1.36	0.79 ± 0.04	0.62 ± 0.02	0.74 ± 0.02	0.78 ± 0.02
Loose Track Only	20.44 ± 1.39	0.81 ± 0.05	0.62 ± 0.02	0.75 ± 0.02	0.79 ± 0.02
Loose	20.57 ± 1.41	0.81 ± 0.05	0.64 ± 0.02	0.75 ± 0.02	0.8 ± 0.02
Tight	19.86 ± 1.38	0.78 ± 0.04	0.62 ± 0.02	0.73 ± 0.02	0.78 ± 0.02
Gradient Loose	19.89 ± 1.38	0.78 ± 0.05	0.62 ± 0.02	0.75 ± 0.02	0.79 ± 0.02
Gradient	19.61 ± 1.38	0.76 ± 0.04	0.61 ± 0.02	0.73 ± 0.02	0.76 ± 0.02

Table 8: Significances after implementing different WP isolation requirements on the leading muon for the signal samples. The luminosity corresponds to 36 fb^{-1} .

Cut	Signal				
	5 GeV	10 GeV	20 GeV	30 GeV	50 GeV
No isolation	329.97 ± 19.51	14.09 ± 0.68	15.29 ± 0.41	26.42 ± 0.62	31.91 ± 0.56
Loose Track Only	328.92 ± 19.48	13.99 ± 0.67	15.22 ± 0.41	26.32 ± 0.62	31.67 ± 0.56
Loose	324.44 ± 19.35	13.8 ± 0.67	15.06 ± 0.4	26.17 ± 0.62	31.58 ± 0.56
Tight	313.09 ± 18.91	13.27 ± 0.65	14.59 ± 0.4	25.29 ± 0.61	30.48 ± 0.55
Gradient Loose	313.77 ± 18.96	13.3 ± 0.66	14.64 ± 0.4	25.33 ± 0.61	30.51 ± 0.55
Gradient	305.95 ± 18.72	12.82 ± 0.64	14.25 ± 0.39	24.67 ± 0.6	29.49 ± 0.54

Table 9: Remaining signal events after implementing different WP isolation requirements on the leading muon for the signal samples. The luminosity corresponds to 36 fb^{-1} .

Cut	Background				
	5 GeV	10 GeV	20 GeV	30 GeV	50 GeV
No isolation	273.67 ± 18.42	316.42 ± 18.41	616.85 ± 28.82	1274.12 ± 47.73	1672.24 ± 53.67
Loose Track Only	258.93 ± 17.33	301.56 ± 17.48	598.02 ± 28.14	1247.47 ± 47.54	1598.31 ± 52.9
Loose	248.67 ± 16.79	291.0 ± 16.98	557.33 ± 26.27	1203.33 ± 46.15	1552.62 ± 51.65
Tight	248.45 ± 17.06	289.98 ± 17.21	551.93 ± 26.5	1196.93 ± 46.91	1536.12 ± 52.32
Gradient Loose	248.84 ± 16.99	287.61 ± 16.96	549.17 ± 26.19	1150.57 ± 45.43	1481.2 ± 50.95
Gradient	243.46 ± 16.78	282.21 ± 16.77	541.98 ± 26.16	1137.07 ± 45.49	1496.37 ± 52.61

Table 10: Remaining background events after implementing different WP isolation requirements on the leading muon for the signal samples. The luminosity corresponds to 36 fb^{-1} .

7.2.2 Subleading muon and electron WP isolation requirements

The following section similarly investigates the best combination of the five WPs for the sub-leading muon and the electron in terms of significance, signal and background. The extreme mass cases are highlighted in this section but corresponding tables can be found for 10, 20 and 30 GeV in appendix E. Gradient isolation for the leading muon is implemented on top of the existing cuts. A summary of the cut lists is provided in table 11 and 12 for 5 and 50 GeV respectively.

The *tight-gradient* combination for the sub-leading muon and the electron is chosen for both 5 and 50 GeV as motivated by table 13 to 18. The *tight-gradient* combination was also chosen for 10, 20 and 30 GeV based on the tables shown in appendix E. The final choice reflects the decision to strive for uniformity across the different mass selections.

Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
$\text{Invmass}(\mu, \mu, e) < 90 \text{ GeV}$
$\text{Invmass}(\mu, \mu, e) > 40 \text{ GeV}$
$\text{Invmass}(\mu, e) < 5 \text{ GeV}$
$p_T(\mu_{\text{lead}}) \geq 20 \text{ GeV}$
$p_T(\mu_{\text{sub-lead}}) \geq 10 \text{ GeV}$
$p_T(e) \geq 5 \text{ GeV}$
Number of b jets is 0
Gradient isolation for leading muon

Table 11: A summary of cuts used for testing isolation WPs on the sub-leading muon and the electron. The cuts refer to a study on an HNL signal sample with a mass of 5 GeV.

Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
Invmass (μ, μ, e) < 90 GeV
Invmass (μ, μ, e) > 40 GeV
Invmass (μ, e) < 50 GeV
$p_T(\mu_{\text{lead}}) \geq 15$ GeV
$p_T(\mu_{\text{sub-lead}}) \geq 10$ GeV
$p_T(e) \geq 5$ GeV
Number of b jets is 0
Gradient isolation for leading muon

Table 12: A summary of cuts used for testing isolation WPs on the sub-leading muon and the electron. The cuts refer to a study on an HNL signal sample with a mass of 50 GeV.

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	13.43 ± 1.32	18.57 ± 2.83	19.11 ± 3.07	19.32 ± 3.18	22.96 ± 3.59
Loose	13.12 ± 1.33	17.96 ± 2.87	17.62 ± 2.86	17.78 ± 2.96	20.76 ± 3.46
Tight	25.19 ± 3.79	25.43 ± 6.3	25.73 ± 6.6	26.41 ± 7.19	60.1 ± 14.0
Gradient Loose	17.95 ± 4.65	15.31 ± 7.33	15.31 ± 7.33	15.31 ± 7.33	50.27 ± 12.33
Gradient	17.64 ± 7.12	14.37 ± 7.29	14.37 ± 7.29	14.37 ± 7.29	80.05 ± 31.81

Table 13: The significance given by S/\sqrt{B} after testing different working point combinations of the sub-leading muon and the electron for the 5 GeV HNL mass sample. The luminosity corresponds to 36 fb⁻¹.

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	174.93 ± 14.01	127.04 ± 11.74	123.69 ± 11.58	122.63 ± 11.54	116.24 ± 11.24
Loose	168.32 ± 13.74	120.44 ± 11.43	117.08 ± 11.27	116.02 ± 11.22	109.63 ± 10.91
Tight	158.37 ± 13.39	113.93 ± 11.12	112.94 ± 11.07	111.88 ± 11.02	108.67 ± 10.87
Gradient Loose	99.61 ± 11.14	59.83 ± 8.38	59.83 ± 8.38	59.83 ± 8.38	56.63 ± 8.17
Gradient	75.37 ± 9.21	54.48 ± 8.03	54.48 ± 8.03	54.48 ± 8.03	52.37 ± 7.9

Table 14: The number of signal events after testing different working point combinations of the sub-leading muon and the electron for the 5 GeV HNL mass sample. The luminosity corresponds to 36 fb⁻¹.

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	169.78 ± 19.46	46.79 ± 11.35	41.91 ± 10.92	40.3 ± 10.89	25.63 ± 6.29
Loose	164.55 ± 19.64	44.99 ± 11.59	44.17 ± 11.58	42.56 ± 11.55	27.89 ± 7.47
Tight	39.53 ± 9.82	20.08 ± 9.15	19.27 ± 9.14	17.94 ± 9.1	3.27 ± 1.38
Gradient Loose	30.81 ± 14.41	15.26 ± 13.96	15.26 ± 13.96	15.26 ± 13.96	1.27 ± 0.5
Gradient	18.25 ± 14.05	14.38 ± 13.96	14.38 ± 13.96	14.38 ± 13.96	0.43 ± 0.31

Table 15: The number of background events after testing different working point combinations of the sub-leading muon and the electron for the 5 GeV HNL mass sample. The luminosity corresponds to 36 fb⁻¹.

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	0.8 ± 0.02	0.93 ± 0.03	0.94 ± 0.03	0.93 ± 0.03	0.92 ± 0.03
Loose	0.83 ± 0.02	0.95 ± 0.03	0.97 ± 0.03	0.96 ± 0.03	0.94 ± 0.03
Tight	0.91 ± 0.03	0.97 ± 0.03	0.99 ± 0.04	0.98 ± 0.04	0.97 ± 0.04
Gradient Loose	0.97 ± 0.03	1.04 ± 0.04	1.06 ± 0.04	1.05 ± 0.04	1.04 ± 0.04
Gradient	1.01 ± 0.04	1.09 ± 0.05	1.11 ± 0.05	1.1 ± 0.05	1.09 ± 0.06

Table 16: The significance given by S/\sqrt{B} after testing different working point combinations of the sub-leading muon and the electron for the 50 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	28.98 ± 0.53	27.47 ± 0.52	27.1 ± 0.52	26.68 ± 0.51	26.09 ± 0.51
Loose	28.94 ± 0.53	27.43 ± 0.52	27.06 ± 0.52	26.65 ± 0.51	26.05 ± 0.51
Tight	28.32 ± 0.53	26.81 ± 0.51	26.45 ± 0.51	26.07 ± 0.51	25.49 ± 0.5
Gradient Loose	28.05 ± 0.52	26.54 ± 0.51	26.19 ± 0.51	25.82 ± 0.5	25.25 ± 0.5
Gradient	26.64 ± 0.51	25.15 ± 0.5	24.84 ± 0.49	24.48 ± 0.49	23.96 ± 0.49

Table 17: The number of signal events after testing different working point combinations of the sub-leading muon and the electron for the 50 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	1296.14 ± 56.17	876.52 ± 46.58	827.37 ± 44.5	817.92 ± 44.82	801.9 ± 47.0
Loose	1211.92 ± 53.26	830.63 ± 44.89	785.7 ± 42.78	776.36 ± 43.1	761.26 ± 45.26
Tight	978.08 ± 49.01	762.69 ± 45.32	718.69 ± 43.15	709.65 ± 43.52	691.21 ± 45.62
Gradient Loose	840.27 ± 50.1	652.07 ± 44.91	613.32 ± 42.53	607.38 ± 42.95	592.3 ± 44.49
Gradient	696.84 ± 52.68	530.83 ± 45.98	499.3 ± 43.57	494.27 ± 44.15	483.14 ± 45.35

Table 18: The number of background events after testing different working point combinations of the sub-leading muon and the electron for the 50 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

7.3 E_T^{miss} and H_T cuts

The results in this section are based on the cuts in table 7, the three optimal lepton p_T cuts and the requirement of no b jets. The full optimal set of cuts simply removes too much of the signal and especially of the background to sensibly place cuts for met and htsum. This approach is justified since E_T^{miss} and H_T are not correlated with isolation.

For E_T^{miss} most of the signal is below 60 GeV in all HNL mass cases, whereas the various backgrounds are scattered across 200 GeV which can be seen for the 5 and 50 GeV HNL masses in figure 32 (corresponding E_T^{miss} plots for the additional HNL masses can be found in appendix F.1). Therefore, a cut at $E_T^{\text{miss}} < 60$ was chosen for all masses. However, it should be noted that the final cut on E_T^{miss} might be revised after a data driven method is implemented if a tighter cut should prove necessary to increase statistics in validation regions.

H_T is defined as the scalar sum of the p_T of jets found in the event. A cut on H_T is motivated in the same way as E_T^{miss} . By cutting at $H_T < 120$ GeV one barely touches the signal while still throwing away some background which is illustrated by the 5 and 50 GeV HNL masses in figure 33 (corresponding H_T plots for the additional HNL masses can be found in appendix F.2). This choice results in a high significance and, therefore, seems like a sensible cut. Similar to E_T^{miss} , it should be noted that the final cut on H_T might also be revised if necessary.

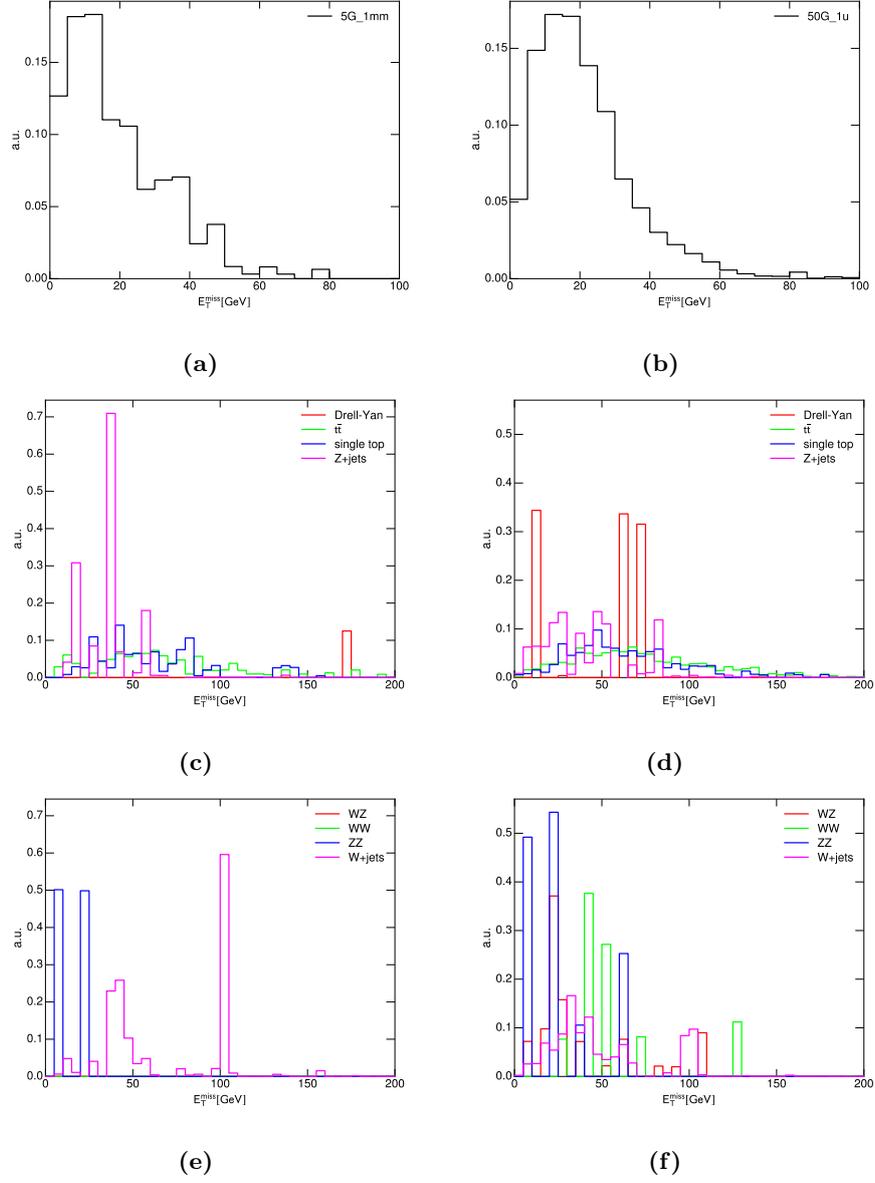


Figure 32: E_T^{miss} distributions for the 5 GeV HNL mass (first column) and 50 GeV HNL mass (second column). The signal is shown in figure (a) and (b), the first set of backgrounds is shown in figure (c) and (d) and the second set of backgrounds is shown in figure (e) and (f).

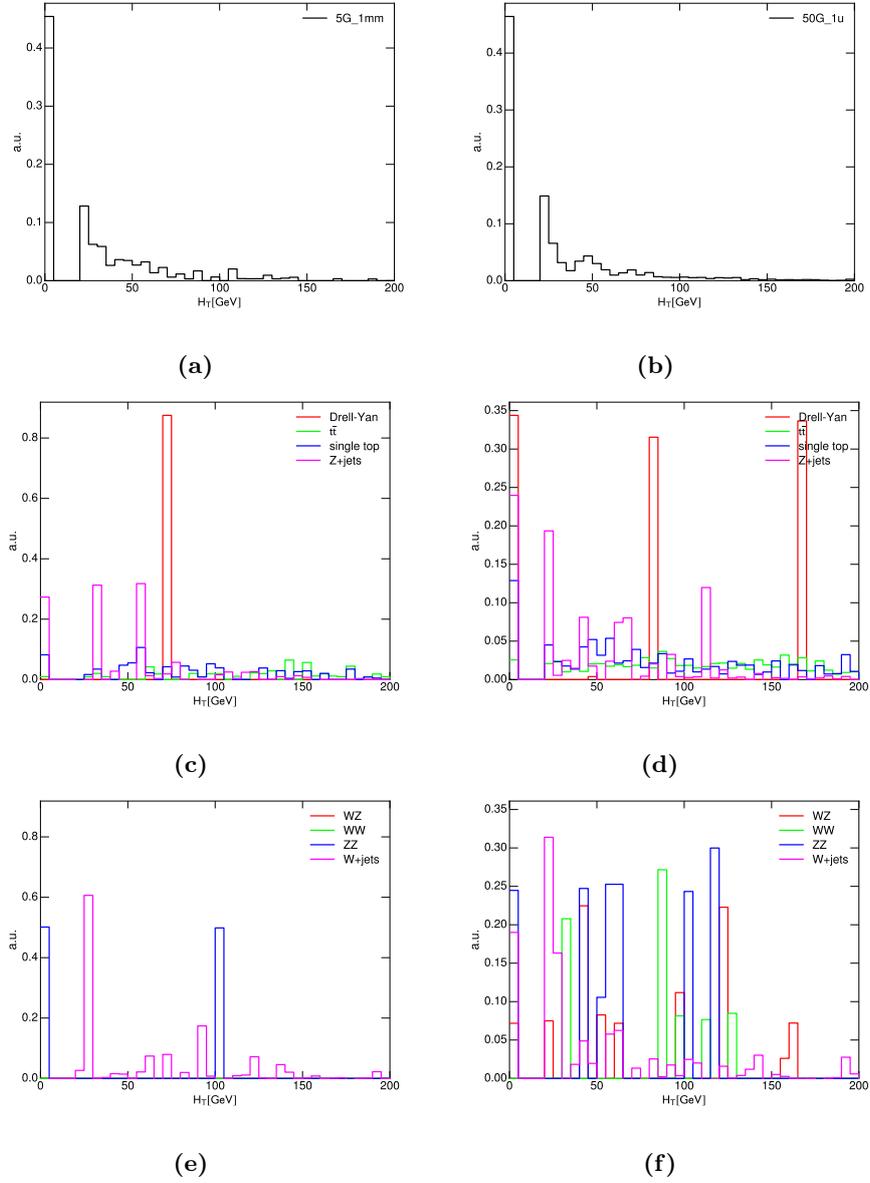


Figure 33: H_T distributions for the 5 GeV HNL mass (first column) and 50 GeV HNL mass (second column). The signal is shown in figure (a) and (b), the first set of backgrounds is shown in figure (c) and (d) and the second set of backgrounds is shown in figure (e) and (f).

7.4 Trigger selection

Once an optimal off-line selection is established the triggers in the data that were active in 2015 and 2016 are checked and the most suitable ones are chosen. The choice of trigger is tested using all off-line cuts. Two triggers were examined as previously stated; namely the high level triggers `HLT_mu26_ivarmedium`

and `HLT_mu22_mu8noL1`. The decision to test these two triggers is based on a few key features; they were both active for long periods of time throughout the collection of the 2015 and 2016 data sets and they are both looser than the chosen isolation working points. Moreover, a combination of both triggers was tried to see if they select the same events or if the combination could increase the signal statistics further. As previously stated, when testing the `HLT_mu26_ivarmedium` trigger it is necessary to include the high level P_T triggers `HLT_mu40` and `HLT_mu50` in OR.

`HLT_mu22_mu8noL1`, `HLT_mu26_ivarmedium` and the two high P_T triggers technically performed the best in all mass cases but the gain in signal was too small compared to that of the stand-alone asymmetric trigger to warrant the use of the four-trigger combination. The discrepancy between `HLT_mu22_mu8noL1` and `HLT_mu26_ivarmedium` was the most prominent for 5 GeV otherwise they performed similarly as seen in table 19. Thus, in this analysis the asymmetric trigger combination `HLT_mu22_mu8noL1` will be used henceforth in all mass cases.

Cut	Signal				
	5 GeV	10 GeV	20 GeV	30 GeV	50 GeV
No trigger requirements	100.39 ± 10.47	8.21 ± 0.51	10.26 ± 0.33	18.4 ± 0.52	24.24 ± 0.49
HLT_mu22_mu8noL1	90.4 ± 9.92	7.46 ± 0.48	9.09 ± 0.31	15.03 ± 0.47	14.85 ± 0.38
HLT_mu26_ivarmedium or HLT_mu40 or HLT_mu50	78.38 ± 9.05	6.83 ± 0.46	8.28 ± 0.3	13.15 ± 0.44	9.12 ± 0.3
HLT_mu22_mu8noL1 or HLT_mu26_ivarmedium or HLT_mu40 or HLT_mu50	91.51 ± 9.98	7.53 ± 0.49	9.26 ± 0.32	15.21 ± 0.47	14.96 ± 0.38

Table 19: Trigger impact on signal for the 5, 10, 20, 30 and 50 GeV HNL masses.

7.5 Impact of final optimal cuts on significance, signal and background

The final set of optimal cuts is summarized in table 20. The two lepton p_T cuts are adjusted to account for the trigger choice and consequently only the dilepton invariant mass cut differs for each mass selection. The final selection was optimized also with respect to both medium and tight ID for all leptons but showed a general decrease in significance in each case. Thus, it is justified to keep the loose ID requirement as is. Furthermore, one should note that the objects need to be trigger matched but this is not possible with the current set of ntuples.

Table 21 shows an overview of how the significance develops with each cut for each signal mass hypothesis. The highest significance is attained for 5 GeV and is progressively lower for the other mass cases. Table 22 to 26 show the cut-flow impact on the signal and individual backgrounds for all masses. It should be noted that the Pythia Drell-Yan sample is left out in these tables. As statistics decreases and becomes sufficiently low the $W + \text{jets}$ and $Z + \text{jets}$ backgrounds occasionally become negative and then positive again. This seemingly strange behaviour is actually a feature of the generator Sherpa 2.2.1 which assigns negative weights to account for interference of NLO order processes with real emissions when calculating the cross section.

To conclude, the selection is not optimal for high masses. Moreover, the HNL reconstructed mass is not as powerful against typical SM backgrounds as hoped and something needs to be done. Clearly objects in the background processes have different features and correlations among them with respect to objects in the signal process. Therefore, a decision was made to attempt a more intelligent selection using a neural network which - provided enough statistics is generated - is expected to perform much better than a cut-based analysis. The study described in the next section is therefore very important in preparation for the best strategy and the best result for a public statement on this search at the end of Run 2.

Cut number	Cut
1	2 muons (loose and combined) and 1 electron (loose)
2	$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
3	$q_{\mu,\text{lead}} = -q_e$
4	HLT_mu22_mu8noL1
5	$\text{Invmass}(\mu, \mu, e) < 90 \text{ GeV}$
6	$\text{Invmass}(\mu, \mu, e) > 40 \text{ GeV}$
7	$\text{Invmass}(\mu, e) < M_{\text{HNL}} \text{ GeV}$
8	$p_{\text{T}}(\mu_{\text{lead}}) \geq 23 \text{ GeV}$
9	$p_{\text{T}}(\mu_{\text{sub-lead}}) \geq 10 \text{ GeV}$
10	$p_{\text{T}}(e) \geq 5 \text{ GeV}$
11	Number of b jets is 0
12	Gradient isolation for μ_{lead}
13	Tight isolation for $\mu_{\text{sub-lead}}$
14	Gradient isolation for e
15	$E_{\text{T}}^{\text{miss}} < 60$
16	$H_{\text{T}} < 120$

Table 20: A summary of the final optimal cuts. The cuts are the same for all masses and differ only in the dilepton invariant mass cut. In the following tables the numbers listed correspond to the different cuts as shown.

Cut	Significance				
	5 GeV	10 GeV	20 GeV	30 GeV	50 GeV
1	0.84 ± 0.04	0.05 ± 0.0	0.04 ± 0.0	0.07 ± 0.0	0.07 ± 0.0
2	3.48 ± 0.16	0.19 ± 0.01	0.17 ± 0.0	0.3 ± 0.01	0.31 ± 0.0
3	3.89 ± 0.18	0.21 ± 0.01	0.2 ± 0.0	0.35 ± 0.01	0.35 ± 0.01
4	4.69 ± 0.26	0.23 ± 0.01	0.22 ± 0.01	0.36 ± 0.01	0.3 ± 0.01
5	6.81 ± 0.38	0.34 ± 0.01	0.32 ± 0.01	0.52 ± 0.01	0.44 ± 0.01
6	7.26 ± 0.41	0.36 ± 0.02	0.34 ± 0.01	0.56 ± 0.01	0.45 ± 0.01
7	17.74 ± 1.08	0.72 ± 0.03	0.48 ± 0.01	0.61 ± 0.02	0.46 ± 0.01
8	16.81 ± 1.04	0.71 ± 0.03	0.47 ± 0.01	0.6 ± 0.02	0.39 ± 0.01
9	16.55 ± 1.13	0.68 ± 0.04	0.5 ± 0.02	0.59 ± 0.02	0.43 ± 0.01
10	16.55 ± 1.13	0.68 ± 0.04	0.5 ± 0.02	0.59 ± 0.02	0.43 ± 0.01
11	18.92 ± 1.37	0.76 ± 0.05	0.55 ± 0.02	0.65 ± 0.02	0.48 ± 0.01
12	18.64 ± 1.39	0.73 ± 0.05	0.55 ± 0.02	0.66 ± 0.02	0.47 ± 0.02
13	23.75 ± 2.94	1.18 ± 0.09	0.68 ± 0.03	0.73 ± 0.03	0.53 ± 0.02
14	49.48 ± 11.7	2.48 ± 0.34	0.96 ± 0.07	0.89 ± 0.05	0.6 ± 0.03
15	58.12 ± 17.95	2.7 ± 0.42	0.99 ± 0.09	0.91 ± 0.06	0.65 ± 0.03
16	79.94 ± 35.13	2.33 ± 0.58	1.17 ± 0.13	0.94 ± 0.08	0.65 ± 0.04

Table 21: Impact of final selection on significance for the 5, 10, 20, 30 and 50 GeV HNL masses.

Cut	signal 5G	Sample										total background
		$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$Z \rightarrow \tau\tau$	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$	$t\bar{t}$	single top	dibosons		
1	550.4 ± 25.15	366479.47 ± 557.75	871.48 ± 19.84	6328.66 ± 53.49	7768.64 ± 106.1	13682.19 ± 140.96	1834.98 ± 51.55	25847.69 ± 92.71	2639.57 ± 25.69	680.68 ± 5.43	426133.36 ± 597.83	
2	544.98 ± 24.95	614.25 ± 16.45	397.07 ± 18.37	1758.8 ± 26.84	3861.25 ± 98.43	7039.21 ± 104.78	853.28 ± 37.06	9072.78 ± 54.88	927.91 ± 15.15	39.64 ± 2.37	24564.2 ± 163.15	
3	521.86 ± 24.33	353.66 ± 11.68	194.8 ± 12.65	1631.02 ± 25.8	1800.33 ± 62.6	4217.17 ± 76.91	583.54 ± 31.65	8350.6 ± 52.53	818.05 ± 14.31	27.95 ± 2.03	17977.12 ± 121.51	
4	384.56 ± 20.95	219.76 ± 9.28	0.6 ± 0.21	162.28 ± 5.57	6.96 ± 1.69	2438.34 ± 62.05	129.41 ± 13.49	3450.39 ± 33.75	303.79 ± 8.6	10.63 ± 1.31	6722.18 ± 73.27	
5	378.77 ± 20.82	186.18 ± 10.28	0.49 ± 0.25	105.21 ± 4.47	3.61 ± 1.28	1540.77 ± 52.97	58.11 ± 7.38	1087.02 ± 19.07	108.14 ± 5.04	6.97 ± 1.12	3096.5 ± 58.12	
6	377.92 ± 20.8	144.59 ± 8.75	0.49 ± 0.25	102.47 ± 4.46	1.37 ± 0.61	1298.99 ± 48.31	57.42 ± 8.2	997.91 ± 18.3	97.3 ± 4.81	6.43 ± 1.09	2706.98 ± 53.46	
7	352.21 ± 20.17	29.96 ± 3.35	0.0 ± 0.0	1.88 ± 0.36	0.0 ± 0.0	208.9 ± 14.64	-2.82 ± 0.73	135.15 ± 6.66	19.79 ± 1.96	0.64 ± 0.21	394.13 ± 16.56	
8	331.48 ± 19.3	30.92 ± 3.5	0.0 ± 0.0	1.72 ± 0.34	0.0 ± 0.0	202.93 ± 14.76	-2.82 ± 0.73	130.76 ± 6.55	18.89 ± 1.91	0.61 ± 0.21	388.64 ± 16.66	
9	291.26 ± 18.2	16.89 ± 2.46	0.0 ± 0.0	-1.37 ± 0.33	0.0 ± 0.0	175.56 ± 14.78	-14.77 ± 4.92	87.64 ± 5.28	13.51 ± 1.62	0.05 ± 0.04	309.81 ± 16.72	
10	291.26 ± 18.2	16.89 ± 2.46	0.0 ± 0.0	-1.37 ± 0.33	0.0 ± 0.0	175.56 ± 14.78	-14.77 ± 4.92	87.64 ± 5.28	13.51 ± 1.62	0.05 ± 0.04	309.81 ± 16.72	
11	290.19 ± 18.17	16.48 ± 2.57	0.0 ± 0.0	-1.86 ± 0.52	0.0 ± 0.0	170.24 ± 16.16	-14.77 ± 4.92	25.99 ± 2.79	5.88 ± 0.97	0.03 ± 0.03	235.26 ± 17.35	
12	269.22 ± 17.41	14.5 ± 2.42	0.0 ± 0.0	-1.86 ± 0.52	0.0 ± 0.0	147.65 ± 14.34	-14.77 ± 4.92	23.94 ± 2.68	5.77 ± 0.99	0.03 ± 0.03	208.52 ± 15.63	
13	163.86 ± 13.61	2.08 ± 0.6	0.0 ± 0.0	-2.64 ± 1.18	0.0 ± 0.0	17.79 ± 2.93	-15.91 ± 7.95	7.2 ± 1.47	1.96 ± 0.59	0.0 ± 0.0	47.6 ± 8.72	
14	89.48 ± 9.65	0.18 ± 0.18	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.29 ± 0.12	-1.78 ± 1.26	0.58 ± 0.41	0.45 ± 0.32	0.0 ± 0.0	3.27 ± 1.38	
15	86.8 ± 9.53	0.18 ± 0.18	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.06 ± 0.03	-1.78 ± 1.26	0.0 ± 0.0	0.22 ± 0.22	0.0 ± 0.0	2.23 ± 1.29	
16	82.31 ± 9.26	0.18 ± 0.18	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.88 ± 0.88	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	1.06 ± 0.9	

Table 22: Impact of final selection on signal, individual major backgrounds and total background for the 5 GeV HNL mass.

Cut	signal 10G	Sample										total background
		$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$Z \rightarrow \tau\tau$	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$	$t\bar{t}$	single top	dibosons		
1	29.85 ± 1.01	366479.47 ± 557.75	871.48 ± 19.84	6328.66 ± 53.49	7768.64 ± 106.1	13682.19 ± 140.96	1834.98 ± 51.55	25847.69 ± 92.71	2639.57 ± 25.69	680.68 ± 5.43	426133.36 ± 597.83	
2	29.2 ± 0.99	614.25 ± 16.45	397.07 ± 18.37	1758.8 ± 26.84	3861.25 ± 98.43	7039.21 ± 104.78	853.28 ± 37.06	9072.78 ± 54.88	927.91 ± 15.15	39.64 ± 2.37	24564.2 ± 163.15	
3	28.67 ± 0.98	353.66 ± 11.68	194.8 ± 12.65	1631.02 ± 25.8	1800.39 ± 62.6	4217.17 ± 76.91	583.54 ± 31.65	8350.6 ± 52.53	818.05 ± 14.31	27.95 ± 2.03	17977.12 ± 121.51	
4	19.01 ± 0.79	219.76 ± 9.28	0.6 ± 0.21	162.28 ± 5.57	6.96 ± 1.69	2438.34 ± 62.05	129.41 ± 13.49	3450.39 ± 33.75	303.79 ± 8.6	10.63 ± 1.31	6722.18 ± 73.27	
5	18.73 ± 0.79	186.18 ± 10.28	0.49 ± 0.25	105.21 ± 4.47	3.61 ± 1.28	1540.77 ± 52.97	58.11 ± 7.38	1087.02 ± 19.07	108.14 ± 5.04	6.97 ± 1.12	3096.5 ± 58.12	
6	18.68 ± 0.79	144.59 ± 8.75	0.49 ± 0.25	102.47 ± 4.46	1.37 ± 0.61	1298.99 ± 48.31	57.42 ± 8.2	997.91 ± 18.3	97.3 ± 4.81	6.43 ± 1.09	2706.98 ± 53.46	
7	17.27 ± 0.76	39.06 ± 3.89	0.0 ± 0.0	2.06 ± 0.27	0.0 ± 0.0	327.23 ± 19.84	3.26 ± 0.79	177.6 ± 7.66	21.13 ± 2.02	1.24 ± 0.36	571.58 ± 21.73	
8	16.9 ± 0.75	40.03 ± 4.02	0.0 ± 0.0	0.86 ± 0.12	0.0 ± 0.0	326.25 ± 20.0	2.39 ± 0.6	171.05 ± 7.53	20.23 ± 1.97	1.19 ± 0.36	562.0 ± 21.85	
9	13.0 ± 0.65	17.98 ± 2.42	0.0 ± 0.0	-2.15 ± 0.38	0.0 ± 0.0	209.35 ± 15.43	-9.56 ± 3.02	114.75 ± 6.12	14.15 ± 1.66	0.45 ± 0.28	368.4 ± 17.13	
10	13.0 ± 0.65	17.98 ± 2.42	0.0 ± 0.0	-2.15 ± 0.38	0.0 ± 0.0	209.35 ± 15.43	-9.56 ± 3.02	114.75 ± 6.12	14.15 ± 1.66	0.45 ± 0.28	368.4 ± 17.13	
11	12.57 ± 0.63	17.57 ± 2.51	0.0 ± 0.0	-2.7 ± 0.54	0.0 ± 0.0	202.79 ± 16.61	-9.56 ± 3.02	35.89 ± 3.42	6.12 ± 0.99	0.42 ± 0.28	275.05 ± 17.45	
12	11.41 ± 0.6	15.59 ± 2.35	0.0 ± 0.0	-2.7 ± 0.54	0.0 ± 0.0	177.25 ± 14.87	-9.56 ± 3.02	32.91 ± 3.29	6.01 ± 1.02	0.42 ± 0.28	244.45 ± 15.75	
13	10.19 ± 0.56	2.96 ± 0.7	0.0 ± 0.0	-2.94 ± 0.82	0.0 ± 0.0	44.59 ± 5.57	-10.7 ± 4.78	11.02 ± 1.84	2.2 ± 0.64	0.23 ± 0.23	74.64 ± 7.68	
14	7.59 ± 0.49	0.8 ± 0.4	0.0 ± 0.0	-0.18 ± 0.08	0.0 ± 0.0	-2.42 ± 0.67	3.43 ± 1.98	2.08 ± 0.79	0.45 ± 0.32	0.0 ± 0.0	9.36 ± 2.29	
15	7.56 ± 0.48	0.8 ± 0.4	0.0 ± 0.0	-0.38 ± 0.27	0.0 ± 0.0	-2.4 ± 0.8	3.43 ± 1.98	0.63 ± 0.45	0.22 ± 0.22	0.0 ± 0.0	7.85 ± 2.25	
16	7.39 ± 0.48	0.69 ± 0.4	0.0 ± 0.0	-0.38 ± 0.27	0.0 ± 0.0	-2.92 ± 2.06	6.09 ± 4.31	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	10.07 ± 4.8	

Table 23: Impact of final selection on signal, individual major backgrounds and total background for the 10 GeV HNL mass.

Cut	Sample											total background
	signal 20G	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$Z \rightarrow \tau\tau$	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$	$t\bar{t}$	single top	di-bosons		
1	27.27 ± 0.54	366479.47 ± 557.75	871.48 ± 19.84	6328.66 ± 53.49	7768.64 ± 106.1	13682.19 ± 140.96	1834.98 ± 51.55	25847.69 ± 92.71	2639.57 ± 25.69	680.68 ± 5.43	426132.88 ± 597.83	
2	27.21 ± 0.54	614.25 ± 16.45	397.07 ± 18.37	1758.8 ± 26.84	3861.25 ± 98.43	7039.21 ± 104.78	853.28 ± 37.06	9072.78 ± 54.88	927.91 ± 15.15	39.64 ± 2.37	24564.57 ± 163.15	
3	27.04 ± 0.54	353.66 ± 11.68	194.8 ± 12.65	1631.02 ± 25.8	1800.39 ± 62.6	4217.17 ± 76.91	583.54 ± 31.65	8350.6 ± 52.53	818.05 ± 14.31	27.95 ± 2.03	17977.12 ± 121.51	
4	18.32 ± 0.44	219.76 ± 9.28	0.6 ± 0.21	162.28 ± 5.57	6.96 ± 1.69	2438.34 ± 62.05	129.41 ± 13.49	3450.39 ± 33.75	303.79 ± 8.6	10.63 ± 1.31	6722.28 ± 73.27	
5	17.8 ± 0.44	186.18 ± 10.28	0.49 ± 0.25	105.21 ± 4.47	3.61 ± 1.28	1540.77 ± 52.97	58.11 ± 7.38	1087.02 ± 19.07	108.14 ± 5.04	6.97 ± 1.12	3096.5 ± 58.12	
6	17.77 ± 0.44	144.59 ± 8.75	0.49 ± 0.25	102.47 ± 4.46	1.37 ± 0.61	1298.99 ± 48.31	57.42 ± 8.2	997.91 ± 18.3	97.3 ± 4.81	6.43 ± 1.09	2706.98 ± 53.46	
7	16.37 ± 0.42	97.23 ± 7.64	0.22 ± 0.22	22.21 ± 1.63	0.0 ± 0.0	636.02 ± 29.59	9.6 ± 1.92	338.02 ± 10.61	36.74 ± 2.81	3.07 ± 0.69	1143.11 ± 32.58	
8	15.85 ± 0.41	96.35 ± 7.69	0.22 ± 0.22	17.34 ± 1.33	0.0 ± 0.0	634.44 ± 29.81	8.73 ± 1.78	323.61 ± 10.41	34.67 ± 2.72	3.02 ± 0.69	1118.38 ± 32.69	
9	13.37 ± 0.38	47.99 ± 5.47	0.0 ± 0.0	5.11 ± 0.54	0.0 ± 0.0	451.23 ± 26.18	-6.68 ± 1.73	193.78 ± 7.99	21.65 ± 2.14	1.21 ± 0.47	727.65 ± 28.06	
10	13.37 ± 0.38	47.99 ± 5.47	0.0 ± 0.0	5.11 ± 0.54	0.0 ± 0.0	451.23 ± 26.18	-6.68 ± 1.73	193.78 ± 7.99	21.65 ± 2.14	1.21 ± 0.47	727.65 ± 28.06	
11	13.16 ± 0.38	47.44 ± 5.8	0.0 ± 0.0	4.02 ± 0.48	0.0 ± 0.0	437.37 ± 27.61	-7.14 ± 1.98	62.64 ± 4.47	10.37 ± 1.45	1.18 ± 0.47	570.16 ± 28.67	
12	12.2 ± 0.36	45.37 ± 5.81	0.0 ± 0.0	2.66 ± 0.32	0.0 ± 0.0	377.76 ± 24.75	-7.14 ± 1.98	56.08 ± 4.25	9.83 ± 1.45	1.18 ± 0.47	500.02 ± 25.9	
13	11.48 ± 0.35	32.06 ± 5.76	0.0 ± 0.0	0.67 ± 0.11	0.0 ± 0.0	211.61 ± 17.88	-9.05 ± 3.42	28.55 ± 3.01	4.09 ± 0.91	0.99 ± 0.44	287.02 ± 19.36	
14	9.21 ± 0.31	3.43 ± 1.08	0.0 ± 0.0	3.9 ± 0.72	0.0 ± 0.0	72.3 ± 11.43	4.3 ± 2.15	6.74 ± 1.4	1.38 ± 0.36	0.17 ± 0.17	92.21 ± 11.8	
15	9.11 ± 0.31	3.05 ± 1.08	0.0 ± 0.0	2.37 ± 0.53	0.0 ± 0.0	72.39 ± 13.22	4.3 ± 2.15	2.4 ± 0.85	0.44 ± 0.31	0.17 ± 0.17	85.12 ± 13.48	
16	8.81 ± 0.31	2.93 ± 1.2	0.0 ± 0.0	2.26 ± 0.53	0.0 ± 0.0	43.11 ± 10.78	6.96 ± 4.02	0.91 ± 0.53	0.22 ± 0.22	0.0 ± 0.0	56.39 ± 11.59	

Table 24: Impact of final selection on signal, individual major backgrounds and total background for the 20 GeV HNL mass.

Cut	Sample											total background
	signal 30G	$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$Z \rightarrow \tau\tau$	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$	$t\bar{t}$	single top	dibosons		
1	46.82 ± 0.83	366479.47 ± 557.75	871.48 ± 19.84	6328.66 ± 53.49	7768.64 ± 106.1	13682.19 ± 140.96	1834.98 ± 51.55	25847.69 ± 92.71	2639.57 ± 25.69	680.68 ± 5.43	426132.88 ± 597.83	
2	46.75 ± 0.83	614.25 ± 16.45	397.07 ± 18.37	1758.8 ± 26.84	3861.25 ± 98.43	7039.21 ± 104.78	853.28 ± 37.06	9072.78 ± 54.88	927.91 ± 15.15	39.64 ± 2.37	24564.57 ± 163.15	
3	46.34 ± 0.82	353.66 ± 11.68	194.8 ± 12.65	1631.02 ± 25.8	1800.39 ± 62.6	4217.17 ± 76.91	583.54 ± 31.65	8350.6 ± 52.53	818.05 ± 14.31	27.95 ± 2.03	17977.12 ± 121.51	
4	29.57 ± 0.66	219.76 ± 9.28	0.6 ± 0.21	162.28 ± 5.57	6.96 ± 1.69	2438.34 ± 62.05	129.41 ± 13.49	3450.39 ± 33.75	303.79 ± 8.6	10.63 ± 1.31	6722.28 ± 73.27	
5	29.04 ± 0.65	186.18 ± 10.28	0.49 ± 0.25	105.21 ± 4.47	3.61 ± 1.28	1540.77 ± 52.97	58.11 ± 7.38	1087.02 ± 19.07	108.14 ± 5.04	6.97 ± 1.12	3096.5 ± 58.12	
6	28.98 ± 0.65	144.59 ± 8.75	0.49 ± 0.25	102.47 ± 4.46	1.37 ± 0.61	1298.99 ± 48.31	57.42 ± 8.2	997.91 ± 18.3	97.3 ± 4.81	6.43 ± 1.09	2706.98 ± 53.46	
7	26.7 ± 0.62	130.61 ± 8.79	0.28 ± 0.2	48.84 ± 2.65	0.29 ± 0.21	1066.37 ± 43.28	15.69 ± 2.65	575.71 ± 13.9	56.84 ± 3.57	4.76 ± 0.89	1899.4 ± 46.6	
8	25.59 ± 0.61	130.74 ± 8.94	0.28 ± 0.2	44.83 ± 2.58	0.29 ± 0.21	1050.22 ± 43.05	14.82 ± 2.54	545.75 ± 13.51	53.75 ± 3.47	4.59 ± 0.89	1845.26 ± 46.28	
9	20.93 ± 0.55	76.1 ± 7.39	0.0 ± 0.0	19.15 ± 1.56	0.0 ± 0.0	799.67 ± 40.24	-7.72 ± 1.73	317.15 ± 10.3	31.01 ± 2.6	2.59 ± 0.73	1253.4 ± 42.34	
10	20.93 ± 0.55	76.1 ± 7.39	0.0 ± 0.0	19.15 ± 1.56	0.0 ± 0.0	799.67 ± 40.24	-7.72 ± 1.73	317.15 ± 10.3	31.01 ± 2.6	2.59 ± 0.73	1253.4 ± 42.34	
11	20.73 ± 0.55	75.2 ± 7.84	0.0 ± 0.0	17.31 ± 1.61	0.0 ± 0.0	778.57 ± 42.47	-7.44 ± 1.81	108.17 ± 5.97	17.02 ± 1.93	2.49 ± 0.73	1006.21 ± 43.72	
12	19.48 ± 0.54	74.61 ± 8.29	0.0 ± 0.0	6.5 ± 0.63	0.0 ± 0.0	683.07 ± 39.31	-7.63 ± 1.91	94.67 ± 5.56	15.67 ± 1.9	2.41 ± 0.73	884.57 ± 40.65	
13	18.53 ± 0.52	61.78 ± 9.01	0.0 ± 0.0	1.88 ± 0.22	0.0 ± 0.0	495.41 ± 35.03	-9.55 ± 3.02	57.88 ± 4.35	7.54 ± 1.29	1.98 ± 0.69	636.02 ± 36.59	
14	15.34 ± 0.47	21.44 ± 5.05	0.0 ± 0.0	4.09 ± 0.58	0.0 ± 0.0	240.88 ± 28.0	3.34 ± 1.5	25.45 ± 2.86	4.28 ± 1.01	0.21 ± 0.17	299.7 ± 28.66	
15	15.1 ± 0.47	20.95 ± 5.41	0.0 ± 0.0	3.23 ± 0.55	0.0 ± 0.0	233.13 ± 31.15	4.3 ± 2.15	9.73 ± 1.78	2.18 ± 0.73	0.21 ± 0.17	273.72 ± 31.76	
16	14.44 ± 0.46	21.66 ± 6.85	0.0 ± 0.0	2.82 ± 0.52	0.0 ± 0.0	197.72 ± 34.42	6.96 ± 4.02	5.22 ± 1.35	1.74 ± 0.66	0.04 ± 0.03	236.16 ± 35.36	

Table 25: Impact of final selection on signal, individual major backgrounds and total background for the 30 GeV HNL mass.

Cut	signal 50G	Sample										total background
		$Z \rightarrow \mu\mu$	$Z \rightarrow ee$	$Z \rightarrow \tau\tau$	$W \rightarrow e\nu$	$W \rightarrow \mu\nu$	$W \rightarrow \tau\nu$	$t\bar{t}$	single top	di-bosons		
1	47.99 ± 0.69	366479.47 ± 557.75	871.48 ± 19.84	6328.66 ± 53.49	7768.64 ± 106.1	13682.19 ± 140.96	1834.98 ± 51.55	25847.69 ± 92.71	2639.57 ± 25.69	680.68 ± 5.43	426133.36 ± 597.83	
2	47.95 ± 0.69	614.25 ± 16.45	397.07 ± 18.37	1758.8 ± 26.84	3861.25 ± 98.43	7039.21 ± 104.78	853.28 ± 37.06	9072.78 ± 54.88	927.91 ± 15.15	39.64 ± 2.37	24564.2 ± 163.15	
3	47.54 ± 0.69	353.66 ± 11.68	194.8 ± 12.65	1631.02 ± 25.8	1800.39 ± 62.6	4217.17 ± 76.91	583.54 ± 31.65	8350.6 ± 52.53	818.05 ± 14.31	27.95 ± 2.03	17977.12 ± 121.51	
4	24.91 ± 0.5	219.76 ± 9.28	0.6 ± 0.21	162.28 ± 5.57	6.96 ± 1.69	2438.34 ± 62.05	129.41 ± 13.49	3450.39 ± 33.75	303.79 ± 8.6	10.63 ± 1.31	6722.18 ± 73.27	
5	24.38 ± 0.49	186.18 ± 10.28	0.49 ± 0.25	105.21 ± 4.47	3.61 ± 1.28	1540.77 ± 52.97	58.11 ± 7.38	1087.02 ± 19.07	108.14 ± 5.04	6.97 ± 1.12	3096.5 ± 58.12	
6	23.47 ± 0.48	144.59 ± 8.75	0.49 ± 0.25	102.47 ± 4.46	1.37 ± 0.61	1298.99 ± 48.31	57.42 ± 8.2	997.91 ± 18.3	97.3 ± 4.81	6.43 ± 1.09	2706.98 ± 53.46	
7	23.11 ± 0.48	144.48 ± 8.78	0.31 ± 0.18	95.24 ± 4.25	1.28 ± 0.64	1265.92 ± 47.34	57.11 ± 8.42	911.79 ± 17.53	90.77 ± 4.63	6.42 ± 1.09	2573.3 ± 52.32	
8	19.61 ± 0.44	144.43 ± 8.97	0.28 ± 0.2	91.78 ± 4.4	0.49 ± 0.29	1251.49 ± 47.37	44.34 ± 6.76	844.1 ± 16.82	86.03 ± 4.5	6.22 ± 1.09	2469.16 ± 51.91	
9	17.41 ± 0.41	85.42 ± 7.41	0.0 ± 0.0	51.84 ± 3.46	0.2 ± 0.2	893.15 ± 41.46	23.72 ± 4.84	498.05 ± 12.95	47.99 ± 3.27	3.04 ± 0.76	1603.4 ± 44.59	
10	17.41 ± 0.41	85.42 ± 7.41	0.0 ± 0.0	51.84 ± 3.46	0.2 ± 0.2	893.15 ± 41.46	23.72 ± 4.84	498.05 ± 12.95	47.99 ± 3.27	3.04 ± 0.76	1603.4 ± 44.59	
11	17.19 ± 0.41	84.65 ± 7.89	0.0 ± 0.0	47.39 ± 3.62	0.2 ± 0.2	914.38 ± 45.95	24.0 ± 5.24	169.39 ± 7.46	28.82 ± 2.55	2.91 ± 0.76	1271.74 ± 47.72	
12	15.95 ± 0.4	84.74 ± 8.6	0.0 ± 0.0	36.42 ± 2.98	0.0 ± 0.0	846.86 ± 44.88	-5.99 ± 1.38	143.63 ± 6.89	24.35 ± 2.38	2.52 ± 0.73	1144.51 ± 46.4	
13	15.48 ± 0.39	71.88 ± 9.36	0.0 ± 0.0	20.26 ± 2.01	0.0 ± 0.0	642.76 ± 41.06	-7.91 ± 2.19	91.07 ± 5.4	13.86 ± 1.77	2.09 ± 0.69	849.83 ± 42.61	
14	13.42 ± 0.36	27.84 ± 5.57	0.0 ± 0.0	21.38 ± 2.45	0.0 ± 0.0	380.9 ± 37.17	4.6 ± 1.74	51.91 ± 4.04	8.91 ± 1.46	0.24 ± 0.18	495.79 ± 37.95	
15	13.21 ± 0.36	26.91 ± 6.02	0.0 ± 0.0	20.06 ± 2.71	0.0 ± 0.0	338.58 ± 37.62	4.3 ± 2.15	21.66 ± 2.65	4.86 ± 1.09	0.24 ± 0.18	416.61 ± 38.36	
16	12.56 ± 0.35	27.25 ± 7.87	0.0 ± 0.0	19.41 ± 2.86	0.0 ± 0.0	301.93 ± 41.87	6.96 ± 4.02	13.39 ± 2.09	4.42 ± 1.04	0.04 ± 0.03	373.4 ± 42.95	

Table 26: Impact of final selection on signal, individual major backgrounds and total background for the 50 GeV HNL mass.

8 A neural network approach to signal optimization

In this analysis a neural network approach is implemented to distinguish signal from background by exploiting differences in the kinematic topologies. It is expected that the performance of a neural network classifier is better in comparison to the cut-based selection described in the previous section.

Section 8.1 provides an overview of the theory and functionality of a neural network and in Section 8.2 a description of the chosen model is given. The final comparison between the neural network approach and the cut-based selection is performed in Section 8.3.

8.1 Theory and functionality of neural networks

A neural network [32] bears a lot of similarity to a human brain where a complex and interconnected set of neurons and synapses fire electrical signals to form a decision. Thus it becomes a powerful tool for advanced data analysis.

The basic building block of a neural network is called a perceptron. The so-called artificial neuron is shown in figure 34.

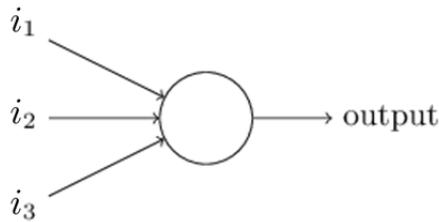


Figure 34: An illustration of a perceptron functioning as an artificial neuron. The perceptron receives a given set of weighted inputs denoted by i_j with $j = \{1, 2, 3\}$ and transforms them to a decision output [32].

In this case the perceptron takes three inputs i_j with $j = \{1, 2, 3\}$ and each input is assigned a corresponding weight w_j according to its importance in determining the output. The weighted sum of i_j is compared to a chosen threshold value and the perceptron output is then set to either 0 or 1 on the basis of this comparison as shown below

$$\text{output} = \begin{cases} 0 & \text{if } \sum_j w_j i_j \leq \text{threshold} \\ 1 & \text{if } \sum_j w_j i_j > \text{threshold} \end{cases}$$

This can also be expressed as

$$\text{output} = \begin{cases} 0 & \text{if } w \cdot i + b \leq 0 \\ 1 & \text{if } w \cdot i + b > 0 \end{cases}$$

where w is a vector of the weights and i is a vector of the inputs. The value b is referred to as the bias. One can employ a handy mathematical trick to get rid of the addition sign in $w \cdot i + b$. Suppose the perceptron takes an extra input $i_0 = 1$ with an assigned weight of $w_0 = b$ then $w \cdot i + b$ becomes $w \cdot i$

where the bias is now accounted for by the inclusion of w_0 in the weight vector w .

To achieve more advanced and nuanced decision-making the perceptrons are combined in a structure that mimics interconnected brain synapses in a network of neurons. This is illustrated by figure 35 where each layer of neurons feeds information to the subsequent layer in an iterative process. This type of structure is a feed-forward/sequential neural network. A perceptron/neuron is referred to as a node, the first layer of nodes is called the input layer, the last layer of nodes makes up the output layer and the layers in between are known as hidden layers [32].

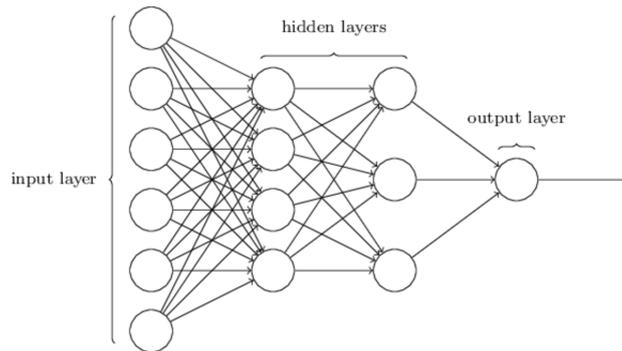


Figure 35: An illustration of the structure of a neural network in which several perceptrons/nodes are combined. An input layer consisting of several nodes are joined to nodes in the output layer through additional nodes in a series of hidden layers. Here the output layer consists of one node but it can be comprised of several depending on the problem [32].

The goal is to train the neural network by adjusting the weights to correctly classify a known target. This is difficult when using a discrete threshold function since it is not differentiable. Hence, another artificial neuron is introduced that uses a continuous threshold function instead. This modified perceptron is called a sigmoid neuron. It applies the sigmoid function on the product $p = w \cdot i$ and returns the observed output given by o as shown below

$$o = \frac{1}{1 + e^{-p}}.$$

Other alternatives to the sigmoid also exist and these types of output transformers are all called activation functions. The sigmoid function is often chosen because it's mathematically convenient e.g. its derivative can be expressed in terms of itself as follows

$$\frac{do}{dp} = o(1 - o).$$

This comes in handy when tuning the weights as will be explained shortly. Another attractive property is that the sigmoid function returns 0 when p is very negative and 1 when p is very positive as shown in figure 36. Thus the sigmoid neuron bears the desired similarity to the original perceptron while maintaining continuity.

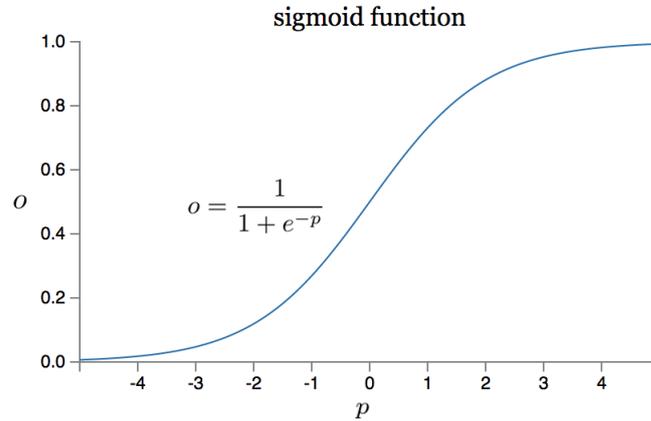


Figure 36: A distribution of the sigmoid activation function which exhibits a smooth curve and is 0 when p is very negative and 1 when p is very positive. These features make the sigmoid neuron an attractive alternative to the perceptron.

The performance of a neural network is often quantified by [33]

$$P = \frac{1}{2}(d_{\text{sample}} - o_{\text{sample}})^2, \quad (18)$$

where d_{sample} denotes the desired output for a given sample and o_{sample} denotes the observed output of the neural network for the same sample. Using a desired output to gauge performance is known as supervised learning. This type of performance function is also sometimes referred to as the cost or loss and its specific shape can vary but Eq. 18 shows a simple example, namely a quadratic function of d_{sample} and o_{sample} with a factor of 1/2 introduced to obtain a nice derivative. The quadratic function is a good choice since the loss is smallest when $d_{\text{sample}} = o_{\text{sample}}$ and monotonically increases as d_{sample} and o_{sample} deviate. The phrase "training a neural network" actually refers to the process of minimizing the performance function which is achieved through gradient descent. The observed output o_{sample} depends on the weights as previously described and therefore the performance function does too. For a set of weights given by $\{w_1, \dots, w_m\}$ for m nodes one can write [32]

$$\Delta P \approx \nabla P \cdot \Delta w,$$

where the gradient ∇P is defined by

$$\nabla P \equiv \left(\frac{\partial P}{\partial w_1}, \dots, \frac{\partial P}{\partial w_m} \right)^T.$$

The term ΔP can be made negative by introducing a small and positive learning rate, denoted by α , such that

$$\Delta w = -\alpha \nabla P.$$

The necessary weight adjustment to change P by a small amount is then

$$w \rightarrow w' = w - \alpha \nabla P.$$

Finding $\{\frac{\partial P}{\partial w_1}, \dots, \frac{\partial P}{\partial w_m}\}$ obviously isn't trivial and relies on two pillars of machine learning, namely forward and backpropagation [33]. The weights are assigned during forward propagation and in the first iteration of this process the assignment is performed at random. The weights are only updated after the neural network finds the necessary changes during backpropagation. The easiest way to explain how backpropagation works is to consider the simplest neural network possible in which just two neurons are joined together. This structure is shown in figure 37 and the two neurons are referred to as left and right. The left neuron takes an input i_l to which it assigns a weight w_l . The product $p_l = i_l \cdot w_l$ is then passed through the sigmoid activation function and the final output of the left neuron is denoted by o_l . The left neuron output then becomes the right neuron input i.e $o_l = i_r$. The process is repeated and w_r , p_r and o_r denote the weight, product and final output of the right neuron respectively.

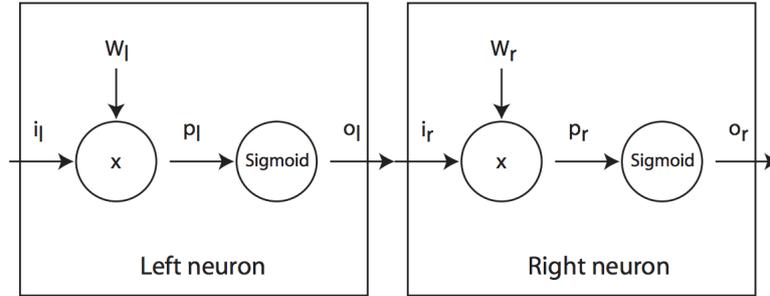


Figure 37: The simplest neural network possible in which just two neurons (denoted by left and right) are joined together. The left o_l becomes the right input i_r [33].

One can express the partial derivatives necessary to find the weight updates in terms of known derivatives. Starting with the right neuron: the partial derivative $\frac{\partial P}{\partial w_r}$ can be computed by applying the chain rule two times in a row as shown below

$$\begin{aligned} \frac{\partial P}{\partial w_r} &= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial w_r} \\ &= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial p_r} \times \frac{\partial p_r}{\partial w_r} \\ &= [(d - o_r)] \times [o_r(1 - o_r)] \times [i_r] \end{aligned}$$

Here $\frac{\partial P}{\partial o_r} = d - o_r$ is found by differentiating the performance function in Eq. 18 with respect to o_r , $\frac{\partial o_r}{\partial p_r} = o_r(1 - o_r)$ is obtained by differentiating the sigmoid function with respect to p_r and $\frac{\partial p_r}{\partial w_r} = i_r$ is simply obtained from $p_r = i_r \cdot w_r$. A similar expression can be obtained for the left neuron: now the partial derivative $\frac{\partial P}{\partial w_l}$ is computed by applying the chain rule four times as shown below

$$\begin{aligned}
\frac{\partial P}{\partial w_l} &= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial w_l} \\
&= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial p_r} \times \frac{\partial p_r}{\partial w_l} \\
&= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial p_r} \times \frac{\partial p_r}{\partial o_l} \times \frac{\partial o_l}{\partial w_l} \\
&= \frac{\partial P}{\partial o_r} \times \frac{\partial o_r}{\partial p_r} \times \frac{\partial p_r}{\partial o_l} \times \frac{\partial o_l}{\partial p_l} \times \frac{\partial p_l}{\partial w_l} \\
&= [(d - o_r)] \times [o_r(1 - o_r)] \times [w_r] \times [o_l(1 - o_l)] \times [i_l].
\end{aligned}$$

The partial derivatives $\frac{\partial P}{\partial o_r} = d - o_r$ and $\frac{\partial o_r}{\partial p_r} = o_r(1 - o_r)$ are found as before. The term $\frac{\partial p_r}{\partial o_l} = w_r$ comes from $p_r = i_r \cdot w_r$ since $o_l = i_r$, and $\frac{\partial o_l}{\partial p_l} = o_l(1 - o_l)$ is found by differentiating the sigmoid function with respect to p_l . The last factor $\frac{\partial p_l}{\partial w_l} = i_l$ is found from $p_l = i_l \cdot w_l$. It's important to note that the two partial derivatives are now completely defined in terms of known quantities. One can shorten the notations by defining two quantities δ_r and δ_l such that

$$\delta_r \equiv o_r(1 - o_r) \times (d - o_r)$$

and

$$\delta_l \equiv o_l(1 - o_l) \times w_r \times \delta_r.$$

The partial derivatives can then be expressed as

$$\frac{\partial P}{\partial w_r} = i_r \times \delta_r$$

and

$$\frac{\partial P}{\partial w_l} = i_l \times \delta_l.$$

Now suppose the simple model is extended to include several hidden layers consisting of one neuron each then the weight update of the final layer f will be given by

$$\Delta w_f = \alpha \times i_f \times \delta_f$$

where

$$\delta_f = o_f(1 - o_f) \times (d - o_f).$$

Knowing δ_f and the old weight w_f , it is a simple task of propagating backwards through the neural network to obtain the weight updates associated with the left layers, namely $\{\Delta w_{(f-1)}, \Delta w_{(f-2)}, \Delta w_{(f-3)}, \dots, \Delta w_0\}$. Therefore, these weight changes are simply given by

$$\Delta w_l = \alpha \times i_l \times \delta_l,$$

where

$$\delta_l = o_l(1 - o_l) \times w_r \times \delta_r.$$

Here $\Delta w_{(f-1)}$ is calculated from δ_f and the old w_f , whereas $\Delta w_{(f-2)}$ is calculated from $\delta_{(f-1)}$ and the old $w_{(f-1)}$. To extend the model even further by including multiple neurons in each layer one has to modify the equations slightly since now the neurons have several inputs and weights.

It should be noted that the frequency of weight updates depends on the variant of gradient descent used. A common model is stochastic gradient descent [32]. In this model an average gradient is computed for a set of training examples known as a batch (the entire training sample is split into batches). The neural network will go through all training examples in a batch, compute the average gradient, perform weight updates and start the next batch using the updated weights.

When the neural network has gone through all batches and seen every example in the entire training sample a complete cycle is reached. This is also known as an epoch. Upon reaching the end of an epoch the neural network starts over and begins the next epoch and is only fully trained when the specified number of epochs is completed.

Both a neural network approach and a cut-based selection have their pros and cons. A neural network can exploit correlations between variables in a more intelligent way than a simple cut-based procedure. If there were only a few variables one could simply use a hand-full of two-dimensional plots to intelligently guide the cut-based selection but in this analysis there are many variables to consider and a neural network will therefore be able to learn things that one cannot gauge from a simple cut-based selection. Both methods are heavily limited by the statistical size of the simulation samples currently available but a neural network approach in particular relies on decent samples sizes to successfully learn about different topologies within signal and background. The neural network performance presented in the subsequent sections is therefore not optimal but has been attempted anyway to get an idea of how one might improve with the current statistics.

8.2 The neural network model

A neural network method will now be applied to the HNL analysis. Each HNL mass signal is treated individually which means that different neural networks are trained for different masses. Statistics has been a recurrent limiting factor throughout the analysis and it is crucial to have large enough samples to train on in order to obtain an optimal neural network. Otherwise, the neural network will not have enough examples of different signal and background topologies to learn about all the possible characteristics. Thus, in this attempt the focus will be on constructing a neural network method centered on the 50 GeV mass as it has the largest number of unweighted events after basic pre-selection cuts. It is also the mass value for which the significance is currently lowest in the cut-based approach.

The most troublesome backgrounds at the end of the signal optimization procedure were $t\bar{t}$, W + jets and Z + jets with the most predominant background

being $W + \text{jets}$ (see Section 7.5). Three classifiers are constructed to handle each of the backgrounds separately. Thus, one ends up with a neural network to classify the 50 GeV signal from $t\bar{t}$, another to classify the 50 GeV signal from $W + \text{jets}$ and a third to classify the 50 GeV signal from $Z + \text{jets}$. A 2-fold cross validation procedure is used which means that the sample statistics is split up into two and used to train two separate neural networks for each classifier. In this analysis the training data is split up into even and odd events based on the parity of the event number in the ntuple. The odd events are used to test and validate a classifier trained on even events and vice versa. Detailed information on the sample splitting can be found in Section 8.2.2.

All classifiers are based on the same model and are subject to the same procedure. A sequential model with dense layers is constructed by using the python deep learning and open source library Keras 1.1.1 with the open source library Tensorflow 0.11 as a backend [34]. A dense layer is a hidden layer where each node is connected to all nodes in the input and output layers, and sequential simply refers to the feed-forward structure described in the previous section. Furthermore, the neural network approach described here uses supervised learning and the Adam optimizer [35] which uses a variation of stochastic gradient descent is employed. The neural network's number of layers in the model, the nodes in each layer, and the number of epochs (iterations over the entire training sample), are set to

$$\begin{aligned} \text{layers}_{t\bar{t}} &= 1, \text{ nodes}_{t\bar{t}} = 12, \\ \text{layers}_{Z+\text{jets}} &= 1, \text{ nodes}_{Z+\text{jets}} = 12, \\ \text{layers}_{W+\text{jets}} &= 1, \text{ nodes}_{W+\text{jets}} = 12, \\ \text{epochs} &= 200. \end{aligned}$$

This yields a total of 181 training parameters for each classifier. This number comes from the fact that here each node in a dense layer takes 14 inputs (including the bias) and when multiplying this with the total number of nodes in this layer one gets 168 trainable parameters. To get the total number of these parameters one has to add the number of inputs of the final node in the output layer (i.e 12 inputs from the dense layer and 1 bias value) which gives 181 trainable parameters. There is no formula for choosing the number of layers, nodes and epochs, so one simply decides upon reasonable values through trial and error when looking at the resulting consistency, loss and accuracy plots as shown in Section 8.2.3. It should be noted that one could have chosen to combine all three backgrounds. However, the separate binary classification methods are chosen to avoid mixing the background topologies in the hope that it leads to a better performance overall.

8.2.1 The input layer

The kinematic topologies used to distinguish signal and background are based on the four-vectors of the leading muon, the sub-leading muon, the electron and the four-vector sum of neutrinos. However, these 16 inputs can be reduced by considering the kinematic constraints of the system. Suppose one boosts the system from the laboratory frame of reference to the frame where the sum of the four-vectors mentioned is zero. One can rotate the leading muon such

that its momentum is aligned along the z-axis. After this transformation the four-vectors are given by

$$\text{Leading muon : } 0, 0, p_z(\mu_{\text{lead}}), E(\mu_{\text{lead}}),$$

$$\text{Sub-leading muon : } p_x(\mu_{\text{sub-lead}}), p_y(\mu_{\text{sub-lead}}), p_z(\mu_{\text{sub-lead}}), E(\mu_{\text{sub-lead}}),$$

$$\text{Electron : } p_x(e), p_y(e), p_z(e), E(e),$$

and

$$\text{Neutrino : } p_x(\nu), p_y(\nu), p_z(\nu), E(\nu).$$

The system can then be rotated around the z-axis to align the momentum of the sub-leading muon to the xz-plane after which the four-vectors become

$$\text{Leading muon : } 0, 0, p_z(\mu_{\text{lead}}), E(\mu_{\text{lead}}),$$

$$\text{Sub-leading muon : } p_x(\mu_{\text{sub-lead}}), 0, p_z(\mu_{\text{sub-lead}}), E(\mu_{\text{sub-lead}}),$$

$$\text{Electron : } p_x(e), p_y(e), p_z(e), E(e),$$

and

$$\text{Neutrino : } p_x(\nu), p_y(\nu), p_z(\nu), E(\nu).$$

Hence, by using this approach three components can be made redundant. Moreover, the soft term of the E_T^{miss} (track-based) can be considered negligible which means that the x-y E_T^{miss} vector is very close to the x-y component of the neutrino four-vector. Three additional constraints then apply to the electron and neutrino four-vectors:

$$p_y(e) = -p_y(\nu), \tag{19}$$

$$p_x(\mu_{\text{sub-lead}}) = -p_x(e) - p_x(\nu), \tag{20}$$

and

$$p_z(\mu_{\text{lead}}) = -p_z(\mu_{\text{sub-lead}}) - p_z(e) - p_z(\nu). \tag{21}$$

The terms $p_y(e)$, $p_x(\mu_{\text{sub-lead}})$ and $p_z(\mu_{\text{lead}})$ in Eq. 19, 20 and 21 are completely determined provided that one knows the terms on the right-hand side. The final transformed inputs to the neural network are therefore given by

1. $E(\mu_{\text{lead}})$
2. $p_z(\mu_{\text{sub-lead}}), E(\mu_{\text{sub-lead}})$

3. $p_x(e), p_z(e), E(e)$
4. $p_x(\nu), p_y(\nu), p_z(\nu), E(\nu)$
5. $p_T(\text{boost}), \text{number of b jets, charge}$

Thus, the number of four-vector inputs have been reduced from 16 to 10 in going from the laboratory frame to a more optimal choice. Note that three additional inputs are listed in line 5 above. When switching between frames of reference there is a loss of information that could potentially help the neural network, namely the boost between frames. Therefore the transverse momentum of this boost is given as an additional input denoted by $p_T(\text{boost})$. The charge and the number of b jets are also provided as inputs but it should be noted that only one charge input is given such that

$$\text{charge} = \begin{cases} 1 & \text{if no OSSF pairs} \\ -1 & \text{otherwise} \end{cases}$$

There may also be an advantage to giving charge as an input compared to implementing the no OSSF cuts at pre-selection. The correlations of the input variables (only momenta and energies) are shown in figure 38 to 40. Here a scatter plot of the signal (green) is overlaid on a scatter plot of each individual background (blue) for different variable combinations. One-dimensional histograms are also shown for each variable. Linear regression lines are shown in the scatter plots for both signal and background to gauge correlation discrepancies and since linear regression depends on interchanging axes the plots show e.g. $p_z(\mu_2)$ vs. E_{μ_1} in addition to E_{μ_1} vs. $p_z(\mu_2)$ (it should be noted that μ_1 and μ_2 represent the leading and sub-leading muon respectively). In comparing the different correlations of the signal inputs to those of the three backgrounds one can see different shapes and patterns emerging, and there are certainly discrepancies in the linear regression lines for signal and background which suggests that there is potential for a neural network to learn about these topologies.

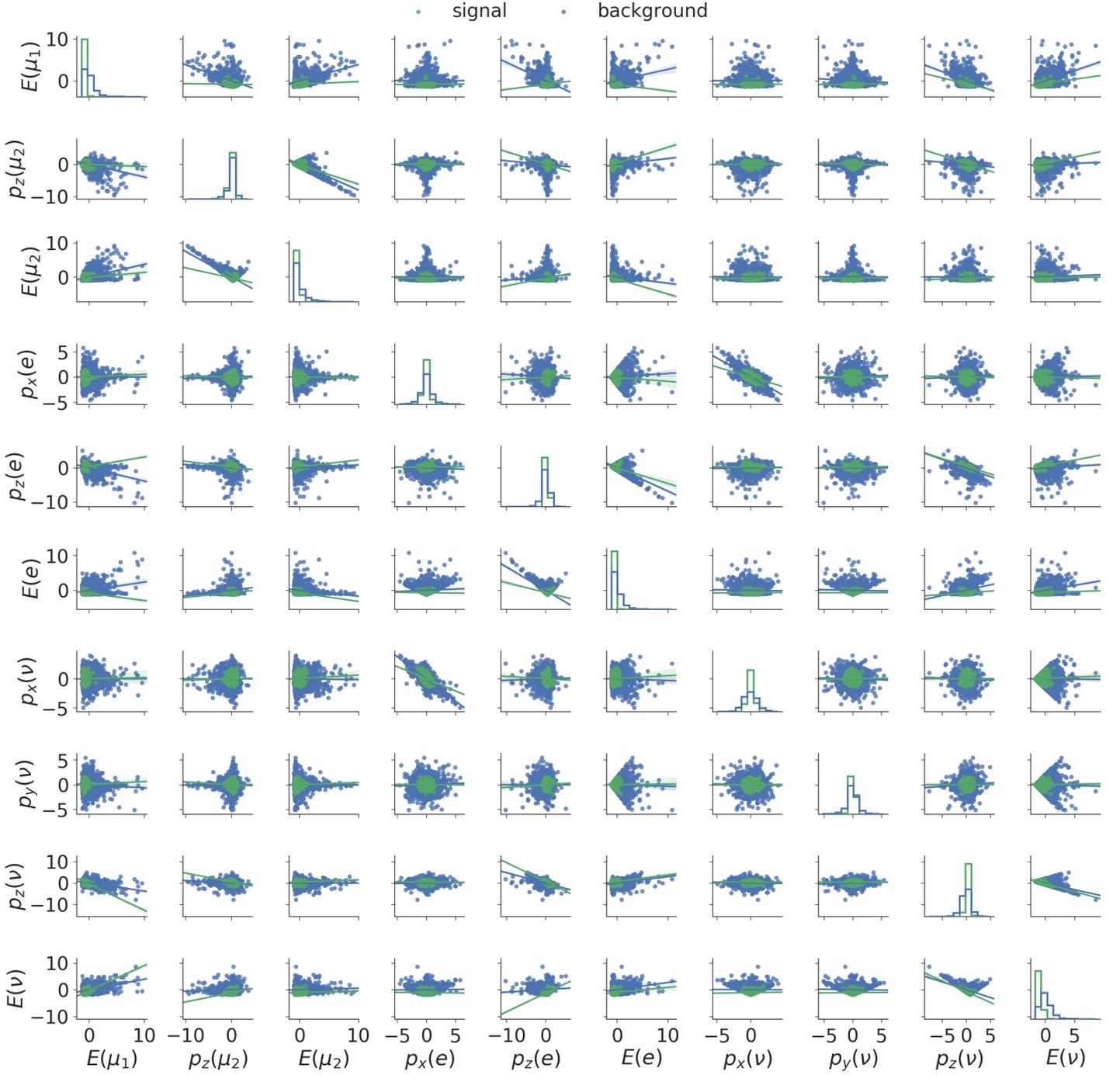


Figure 38: The figure shows correlations between inputs for $t\bar{t}$ (represented by blue) and the 50 GeV signal (represented by green). All inputs are given in GeV.

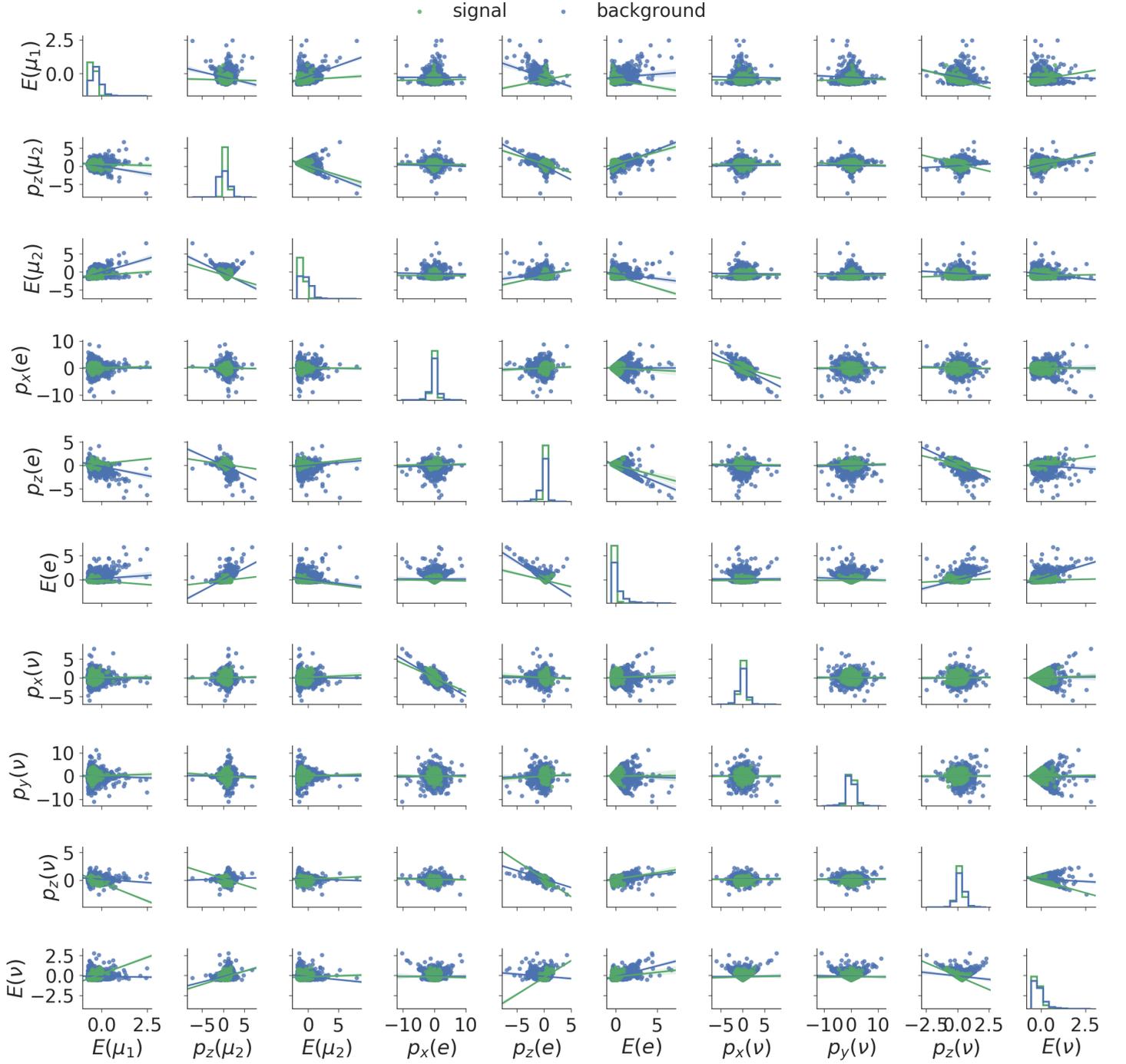


Figure 39: The figure shows correlations between inputs for $Z + \text{jets}$ (represented by blue) and the 50 GeV signal (represented by green). All inputs are given in GeV.

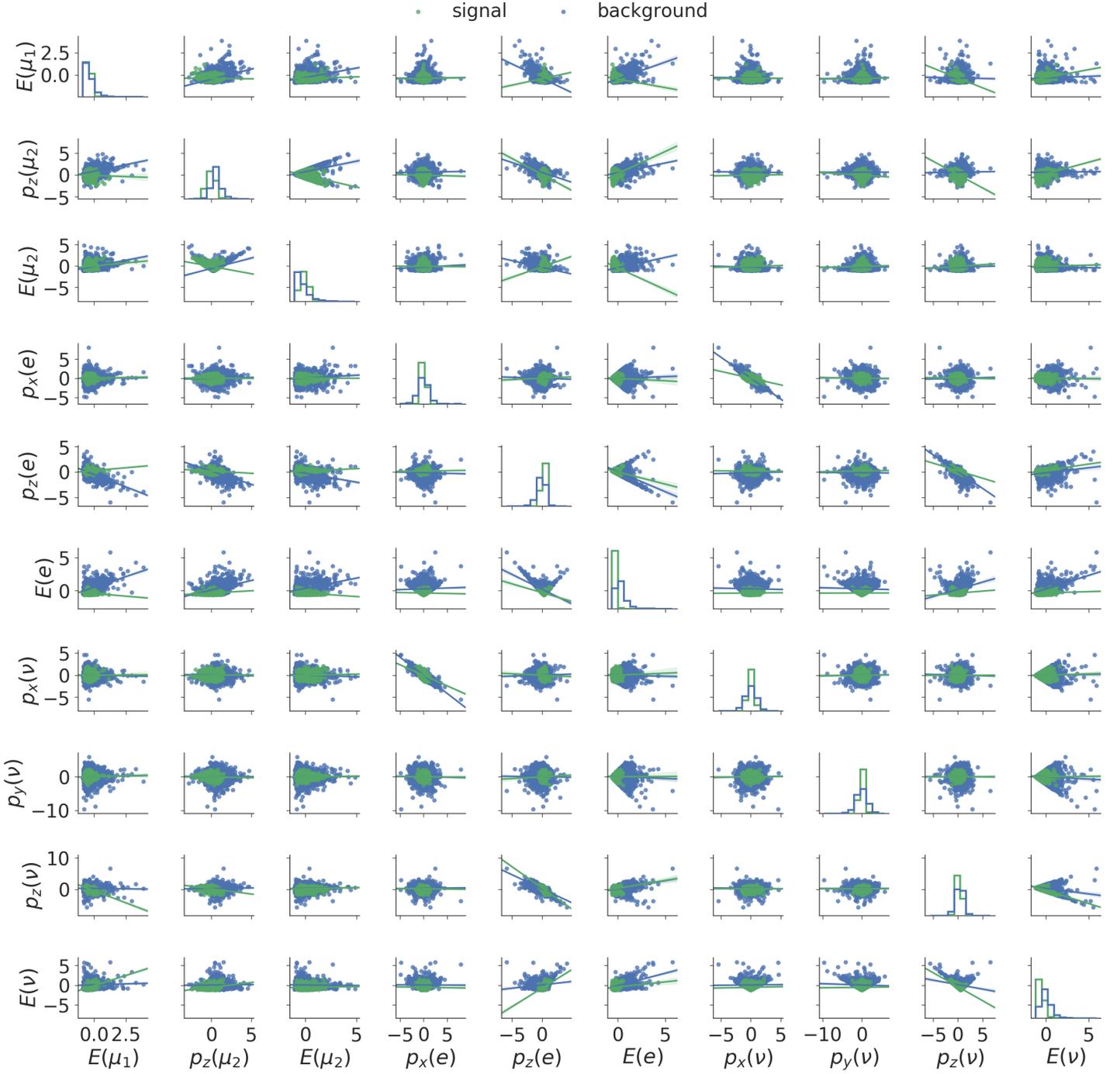


Figure 40: The figure shows correlations between inputs for $W + \text{jets}$ (represented by blue) and the 50 GeV signal (represented by green). All inputs are given in GeV.

8.2.2 Sample preparation

The inputs given to the neural networks are arranged in a matrix where each column represents a different input and each row corresponds to the event number. To perform supervised learning the classifiers are also given a true label (desired output) for each event: 0 is given for background and 1 is given for signal. Furthermore, a training sample consisting of signal and one of the backgrounds is shuffled since a classifier would develop a bias otherwise. Furthermore, not all available statistics is given to the neural networks. Basic pre-selection cuts are implemented first in the sample preparation stage so the classifiers will focus on the most signal-like events during training.

Cut 1 to 4 in table 27 were initially implemented as pre-selection cuts but the remaining statistics (column two in table 28) was too low to complete proper training of the classifiers. Information about the number of b jets and the lepton charges was given as inputs instead, considering also that when it comes to the $W + \text{jets}$ background, what there is to learn about kinematics should be equally learnable from all charge combinations for example. The resulting numbers are shown in column three in table 28.

The samples are split into two training sets, two validation sets and two test sets according to the parity of the event number. Two large training samples were made to contain either even or odd numbered events. Each training sample was then further sub-divided into a validation and test sample. Here 0 will be used to denote even events and 1 will be used to denote odd events when referring to the training sets. For the validation and test sets the reverse is true. The purpose of this is to train a classifier on even events but validate and test on odd events and vice versa. This ensures that the validation and test sets are unknown to the neural network and training a classifier on two different data sets in this way is called 2-fold cross validation. The statistics of each sample is shown in table 29.

Each sample type serves a different purpose. A training set is used to calibrate the weights as described in Section 8.1. A validation set isn't used for weight calibration but is instead used to keep track of the accuracy and loss. Accuracy refers to the percentage of correctly classified events and loss is another term for the performance function described in Section 8.1. If a weight adjustment results in an improved accuracy or loss value for training but not for validation, then one is essentially over-training the neural network. This will manifest itself as diverging graphs when plotting the accuracy and loss vs. the number of epochs for the training and validation sets. Lastly, a test set is used to probe the final predictive power of the neural network when it is given something unknown.

Finally, before a sample is fed to a classifier each input, x , is scaled by $(x - \bar{x})/\sigma$, where the input mean, represented by \bar{x} , and standard deviation, represented by σ , are calculated from all events used to train that classifier. This is done to standardize the input range.

number	Cut
1	2 muons (loose and combined) and 1 electron (loose)
2	$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
3	$q_{\mu,\text{lead}} = -q_e$
4	0 bjets

Table 27: Cuts used for pre-selection. Cuts 1 to 4 were initially used for pre-selection but reduced the statistics too much. Therefore, only cut 1 is used now.

sample	total	cut 1 to 4	cut 1
signal50G	5326	4729	4840
$t\bar{t}$	200699	7727	77786
W + jets	51603	3540	16053
Z + jets	1201026	4013	447943

Table 28: Statistics for the two different sets of cuts tried at the pre-selection stage.

sample	train 0	train 1	valid 0	valid 1	test 0	test 1
signal50G	2481	2359	1190	1248	1169	1233
$t\bar{t}$ bar	38647	39139	19701	19242	19438	19405
W + jets	8070	7983	3993	4053	3990	4017
Z + jets	223917	224026	112182	111635	111844	112282

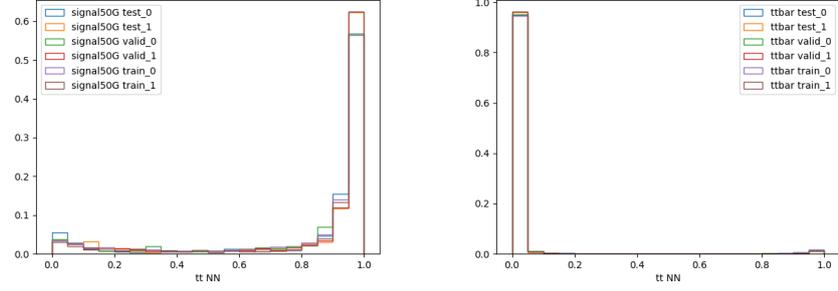
Table 29: Statistics for even and odd training, validation and test samples.

8.2.3 Sanity checks

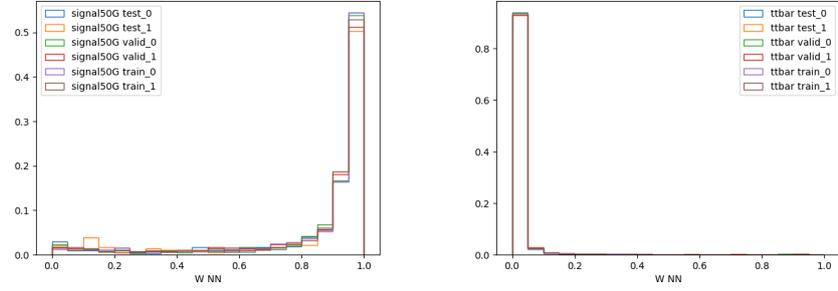
The output of the even and odd training, validation and test samples for signal and background can be compared in so-called consistency plots. Since the samples are divided at random as described in Section 8.2.2 one would expect the individual subsets to perform similarly in terms of output. Thus, consistency plots function as tests of anything symptomatic in the classifier outputs. If significant deviations are observed it could indicate over-training.

Figure 41 and 42 show consistency plots of the signal and background performance in the $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ neural networks for weighted events. All testing, training and validation subsets for each sample are consistent within each classifier and hence there are no signs of over-training.

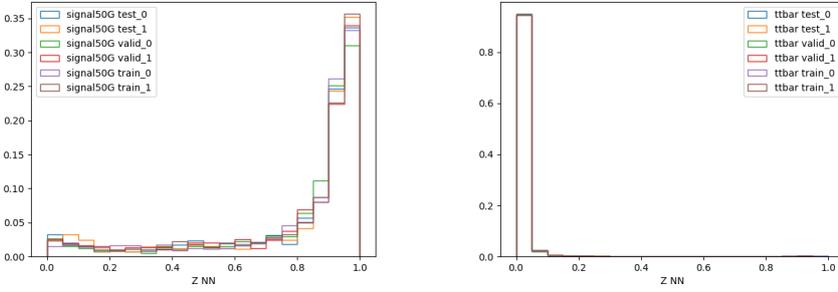
Figure 43 show plots of the accuracy and loss values as they vary with the number of epochs. This is shown for unweighted events. Here one expects the training and validation samples to converge as described in Section 8.2.2. Good agreement is shown in terms of both accuracy and loss for all three classifiers.



(a) Consistency of 50 GeV signal in $t\bar{t}$ classifier. (b) Consistency of $t\bar{t}$ background in $t\bar{t}$ classifier.

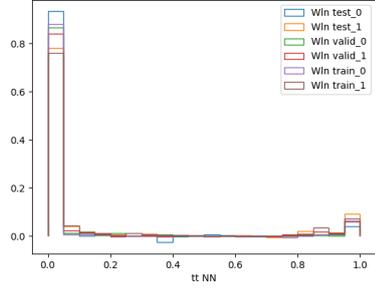


(c) Consistency of 50 GeV signal in $W + \text{jets}$ classifier. (d) Consistency of $t\bar{t}$ background in $W + \text{jets}$ classifier.

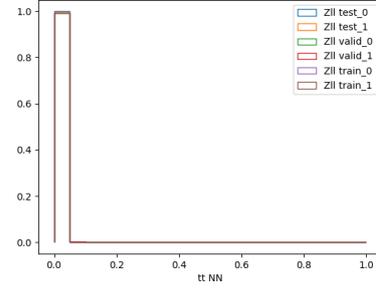


(e) Consistency of 50 GeV signal in $Z + \text{jets}$ classifier. (f) Consistency of $t\bar{t}$ background in $Z + \text{jets}$ classifier.

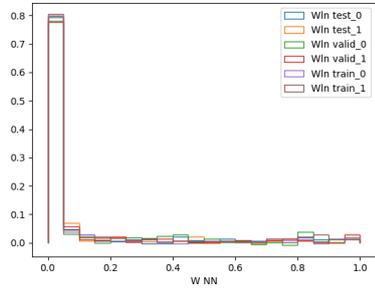
Figure 41: The figure shows consistency plots for the 50 GeV signal and the $t\bar{t}$ background in the $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ classifiers.



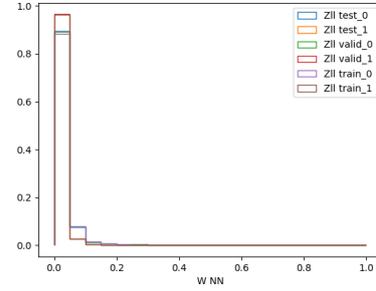
(a) Consistency of $W + \text{jets}$ background in $t\bar{t}$ classifier.



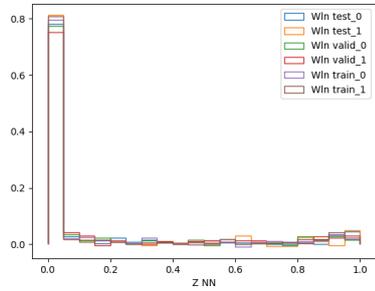
(b) Consistency of $Z + \text{jets}$ background in $t\bar{t}$ classifier.



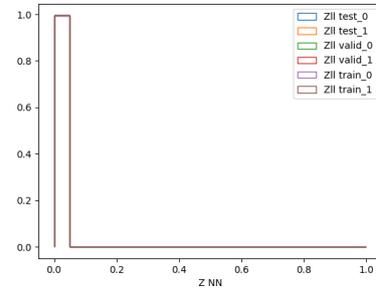
(c) Consistency of $W + \text{jets}$ background in $W + \text{jets}$ classifier.



(d) Consistency of $Z + \text{jets}$ background in $W + \text{jets}$ classifier.



(e) Consistency of $W + \text{jets}$ background in $Z + \text{jets}$ classifier.



(f) Consistency of $Z + \text{jets}$ background in $Z + \text{jets}$ classifier.

Figure 42: The figure shows consistency plots for the $W + \text{jets}$ and $Z + \text{jets}$ backgrounds in the $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ classifiers.

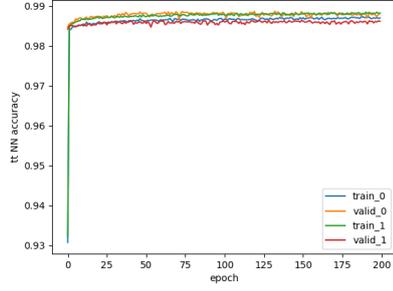
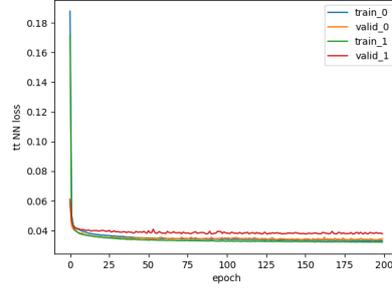
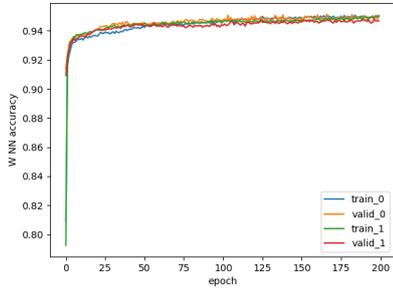
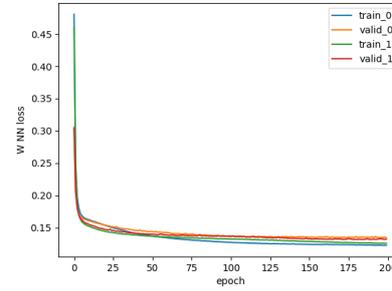
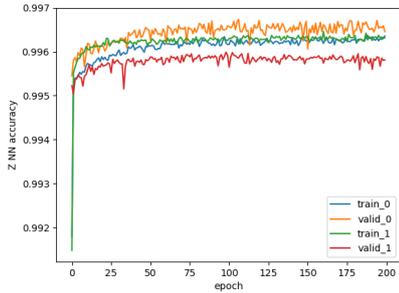
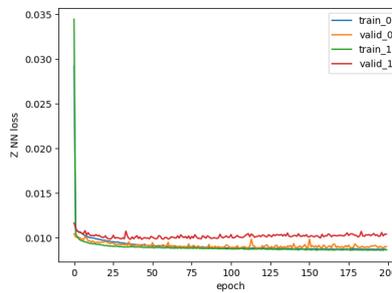
(a) Accuracy of $t\bar{t}$ classifier.(b) Loss of $t\bar{t}$ classifier.(c) Accuracy of $W + \text{jets}$ classifier.(d) Loss of $W + \text{jets}$ classifier.(e) Accuracy of $Z + \text{jets}$ classifier.(f) Loss of $Z + \text{jets}$ classifier.

Figure 43: The figure shows loss and accuracy plots for the $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ classifiers.

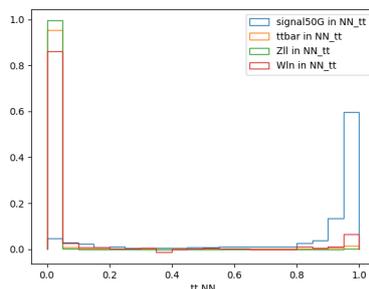
8.2.4 Predictions

The predictions of the $t\bar{t}$, $Z + \text{jets}$ and $W + \text{jets}$ classifiers are shown in figure 44 for weighted events. Here the results for test 0 and test 1 are concatenated for each sample. The three classifiers provide good discrimination power between signal and all backgrounds. In all three cases a slight contamination of $W + \text{jets}$ is seen within the signal peak. Moreover, a small amount of signal is misclassified as background in each case. This is very likely due to the fact that these samples are lowest in statistics.

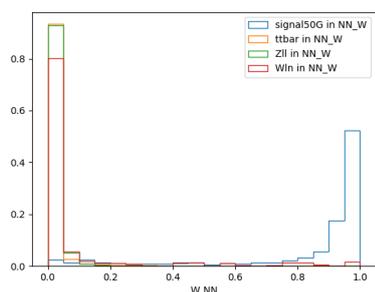
In the subsequent analysis of the neural network results a combination of the three outputs will be combined into a single value such that

$$O_{\text{combined}} = 1 - \frac{\sqrt{(1 - O_{t\bar{t}})^2 + (1 - O_W)^2 + (1 - O_Z)^2}}{\sqrt{3}} \quad (22)$$

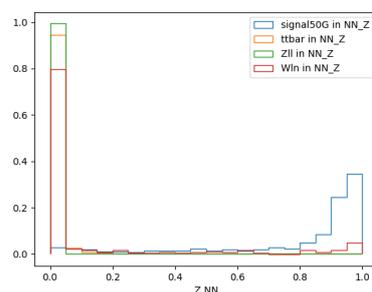
Consider the three-dimensional space of all three outputs then Eq. 22 can be thought of as finding the distance to the point (1,1,1) where the signal peaks in all three classifiers.



(a) Predictions of $t\bar{t}$ classifier.



(b) Predictions of $W + \text{jets}$ classifier.



(c) Predictions of $Z + \text{jets}$ classifier.

Figure 44: Predicted distributions of the $t\bar{t}$, $W + \text{jets}$ and $Z + \text{jets}$ classifiers for signal and background.

8.2.5 The receiver operating characteristic curve

An important feature in determining the success of a neural network is the receiver operating characteristic (ROC) curve [36]. In a binary classification problem a neural network will make predictions about instances of two classes (e.g signal and background) and label them as either positive or negative. In this analysis the outcome is of a probabilistic nature and thus the positive and negative classification is defined with respect to a given threshold. In other words a cut on the neural network is chosen by looking at the outcome distributions of signal and background. There are four possible scenarios for the outcome. If the neural network predicts the instance to be positive with respect to the threshold but the true label of the instance is actually negative it is called a false positive. If on the other hand the prediction is negative but the true label is

positive the instance is called a false negative. True positives and true negatives occur when the the neural network prediction matches the true label. This can be summarized in a confusion matrix as shown in figure 45.

		True class	
		p	n
Hypothesized class	Y	True Positives	False Positives
	N	False Negatives	True Negatives

Figure 45: The confusion matrix for a binary classification model [36].

Here the elements on the diagonal represent correct classification whereas the off diagonal elements represent the confusion in the classification model. The definition of these labels in relation to the chosen threshold is illustrated in figure 46.

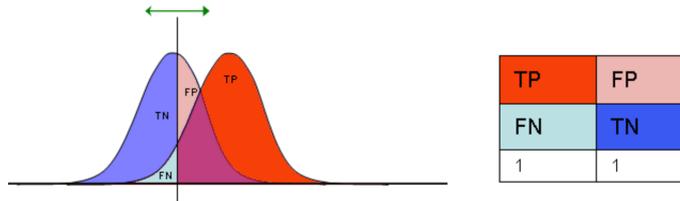


Figure 46: The classification definitions of the true positive, the false positive, the true negative and the false negative outcomes with respect to a given threshold chosen to distinguish background (blue) from signal (red) [37].

The true positive rate is given by

$$\text{TP rate} = \frac{\text{positives correctly classified}}{\text{total positives}}$$

and the false positive rate is given by

$$\text{FP rate} = \frac{\text{negatives incorrectly classified}}{\text{total negatives}}.$$

In the ROC curve the TP rate is plotted on the y-axis and the FP rate is plotted on the x-axis. The function $y = x$ is always used to represent the random guess, where the classification model in question predicts an instance to be positive half the time. One can then gauge the trade-off between these numbers. Clearly the most attractive case is the top left-hand corner seeing as there are significantly more true positives than false positives.

Figure 47 shows the ROC curves for the $t\bar{t}$, W +jets and Z +jets neural networks using weighted events. The results for test 0 and test 1 are again concatenated for each sample. Each ROC curve is done using only signal and the background that the classifier was originally trained on, e.g the black ROC curve shows the classification power of the $t\bar{t}$ neural network with respect to signal and $t\bar{t}$ only. The dotted red line represents the random guess. The curves migrate towards the top left-hand corner which indicates a strong classification power of all three classifiers. It should be noted that the slight fluctuations in the ROC curve of the W +jets classifier is a feature of the limited statistics. The classifier accuracies given by area under the curve (AUC) values are also shown. Here a score of 1 represents perfect discrimination and is obtained for Z +jets. Near perfect values are also obtained for W +jets and $t\bar{t}$.

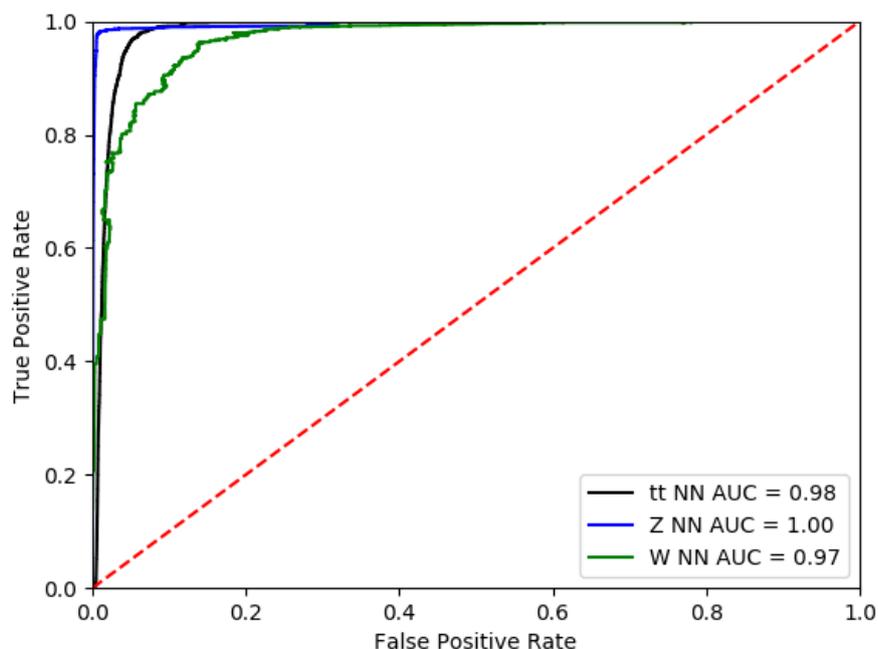


Figure 47: ROC curves for the $t\bar{t}$, W +jets and Z +jets neural networks. Each ROC curve is done using only signal and the background that the classifier was originally trained on, e.g the black ROC curve shows the classification power of the $t\bar{t}$ neural network with respect to signal and $t\bar{t}$ only. The discrimination power of each classifier is represented by an area under the curve (AUC) value.

8.3 Comparison of neural network approach with standard cut-based procedure

To determine whether a neural network approach is indeed better than a standard cut-based strategy for selecting events the lepton and trigger cuts (number 1 and 4 respectively) in table 20 are used as common cuts for both procedures.

Figure 48 shows the optimization procedure for finding the optimal cut on the combined output of the $t\bar{t}$, $Z + \text{jets}$ and $W + \text{jets}$ neural networks.

The fully trained classifier models are used to make predictions on all signal and background events (including single-top and diboson), and a combined output is then assigned to each event. The combined output distributions for signal and background after applying the lepton and trigger cuts are shown in figure (a), (b) and (c). The signal and background efficiencies for cut steps of 0.1 are shown in figure (d) and (e). The corresponding plot of the significance is shown in figure (f).

Table 30 shows the signal and background efficiency in addition to the significance for the two procedures. For roughly the same signal efficiency the neural network performs significantly better in terms of background efficiency and one can improve quite a bit in terms of significance.

quantities	optimized cut-based	optimized combined NN
signal efficiency	0.50	0.45
background efficiency	$3\text{-}4 \cdot 10^{-3}$	10^{-3}
significance	0.65	1.44

Table 30: Comparison of performance for the cut-based and NN procedures, where the lepton and trigger cuts in table 20 are common cuts.

Figure 49 shows plots of the trilepton and dilepton invariant mass upon implementing the optimal cut on the neural network. It can be seen that both masses fall within the desired range as specified in 20. This is a very important testament to the classification power of the neural network approach and additionally demonstrates that no further gain in significance can be achieved by cutting on either mass.

Due to the large increase in significance the neural network approach is more effective than the cut-based selection for the 50 GeV HNL mass. The two methods are limited by statistics and an increase would benefit both. However, the neural network approach, despite not being optimal, proves to be very powerful and is therefore recommended as the general method for signal selection in this analysis assuming that a corresponding improvement would be seen for the 5, 20, 10 and 30 GeV HNL masses given the right statistics.

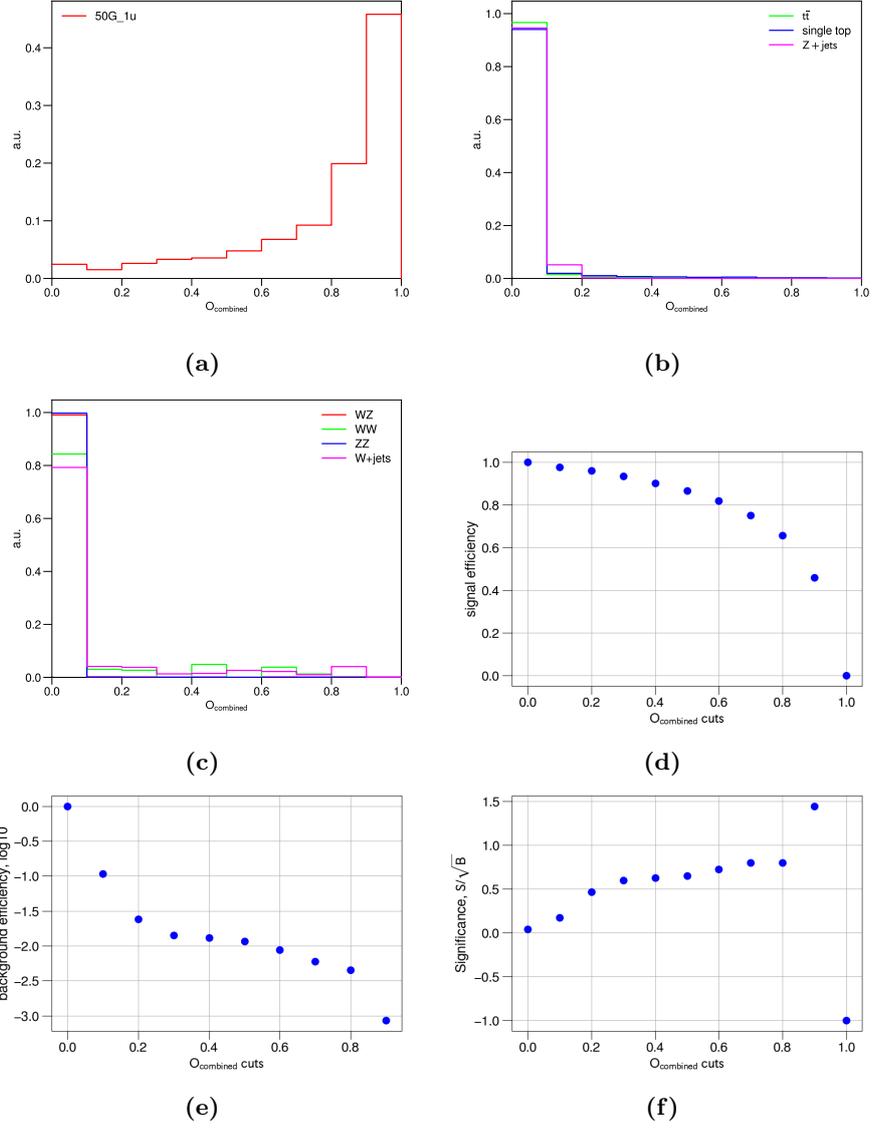


Figure 48: Figure (a), (b) and (c) show the signal and background distributions of the combined output after applying lepton and trigger cuts. Figure (d) and (e) show the signal and background efficiency versus cuts on the combined output in steps of 0.1. Figure (f) shows the corresponding plot of the significance.

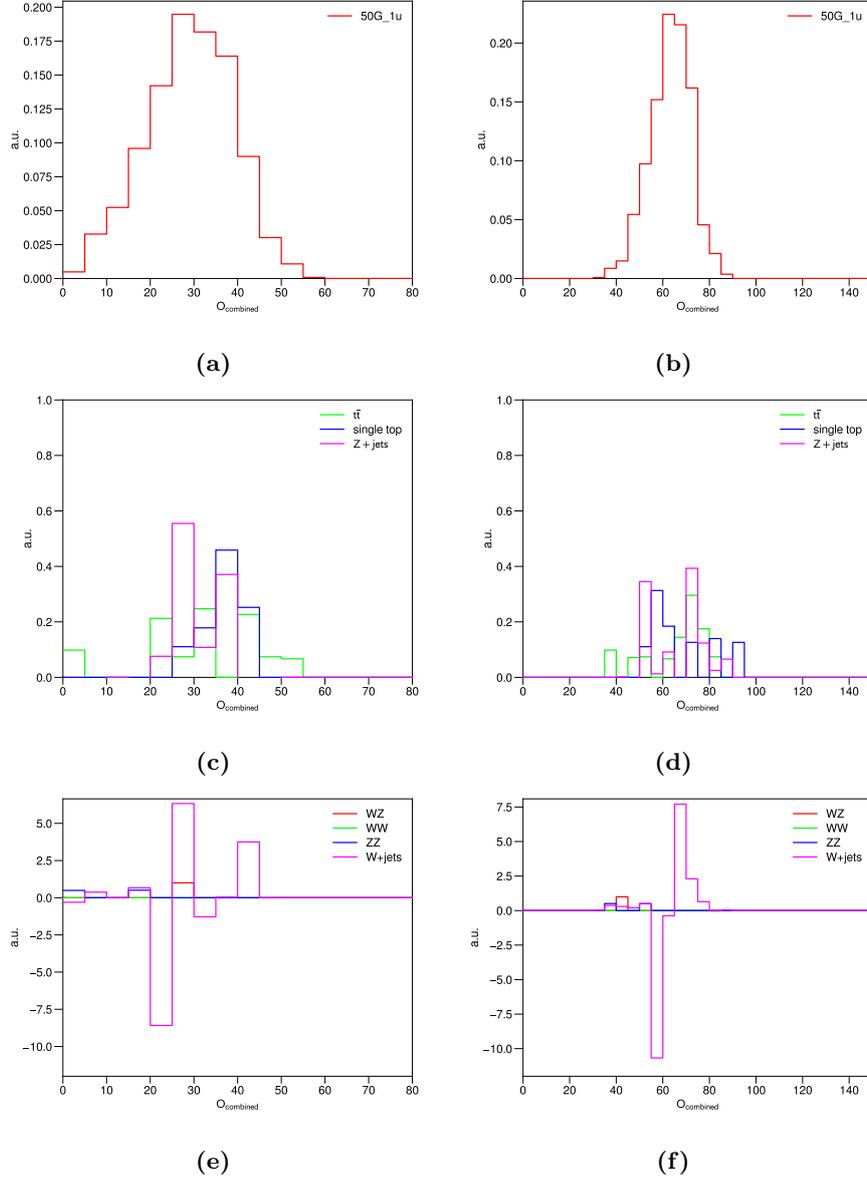


Figure 49: Dilepton (first column) and trilepton (second column) invariant mass distributions for the 50 GeV HNL mass after implementing the optimal neural network cut. Figure (a) and (b) show signal distributions, figure (c) and (d) show the first set of backgrounds and figure (e) and (f) show the second set of backgrounds.

9 Conclusion

In this analysis I have explored what the ATLAS experiment can do to investigate the ν MSM parameters by looking at the prompt trilepton signature of right-handed neutrinos. Based on restrictions imposed on the active-sterile Yukawa couplings in the ν MSM it was decided to focus on the regime in which a single sterile neutrino mixes with a muon neutrino. Furthermore, no OSSF cuts were required in order to focus on Majorana neutrinos. Five mass points for 5, 10, 20, 30 and 50 GeV were considered as listed in table 2.

This is a very complex and relatively unexplored analysis. In fact the background estimation is still ongoing within the analysis subgroup. Therefore I decided to focus on optimization aspects alone.

The analysis uses a contact interaction vertex for the HNL and its decay products in the Monte Carlo simulations of signal. However, the correct physical model in which the HNL decays via an intermediate W boson was tested and compared to the contact interaction model. The two models were found to be similar in terms of efficiencies and kinematic spectra, and thus there is no concern for the validity of the results presented in this thesis. However, based on the tests conducted in this thesis and additional suggestions to implement a filter to solve issues with low statistics, the intermediate W decay mode is now the default used in ATLAS.

Two different approaches were explored with respect to signal selection, namely a cut-based method and a neural network method. The cut-based procedure was challenged by the selection strategy for the prompt and displaced muons in which the former is required to have the highest muon- p_T and the latter is required to have the second highest muon- p_T . The mislabelling was found to be almost insignificant at low mass but quite dominant at high mass. E.g 30% of displaced muons were mislabelled in the case of a 50 GeV mass. Different selection strategies were considered based on kinematic distributions but the issue could not be resolved for the 50 GeV HNL mass.

To address the challenges of the cut-based selection I tried to investigate a neural network approach to signal selection. I expected that one could exploit the different kinematic topologies within signal and background and therefore improve the final significance while avoiding some of the pitfalls of the cut-based method. Due to statistics only the 50 GeV HNL mass was considered which was also the mass with the lowest significance in the cut-based method. The neural network resulted in an improvement from 0.65 to 1.44 in terms of significance and similar results are expected for the other masses provided enough statistics is available. The neural network study is therefore very important in preparation for the best strategy and the best result for a public statement on this search at the end of Run 2.

Both the cut-based and neural network methods are not optimally exploited due to low statistics of background Monte Carlo samples and this was found to be a challenge throughout the analysis. Nevertheless, given the current significances obtained with the cut-based and NN methods it is possible to perform a rough calculation of which values of $|V_{\mu N}|^2$ can be excluded within 36 fb^{-1} , namely those of table 2 or larger/smaller ones. The results for the HNL mass

points are shown in table 31.

sample	significance (36 fb^{-1}) from cut-based (NN-based)	signal exclusion, μ (36 fb^{-1})	luminosity for 5σ discovery, fb^{-1}
5 GeV	79.94	-	0.14
10 GeV	2.33	0.86	165.78
20 GeV	1.17	1.71	657.46
30 GeV	0.94	2.13	1018.56
50 GeV	0.65 (1.44)	3.08 (1.39)	2130.18 (434.03)

Table 31: Estimates for the signal strengths, represented by μ , needed for signal exclusion (2σ) and the luminosity needed for discovery (5σ). The results refer to the significances obtained for the cut-based method except for the numbers in parentheses which refers to the NN approach.

Here the parameter μ is defined as the signal strength in $\mu \cdot S + B$, where S and B denote the signal and background respectively, and $\mu = 1$ is the value shown in table 2. Such a μ value is found by requiring $\mu \cdot S + B$ to be at least 2σ above the background only expectation. Table 31 also shows the luminosity required to achieve 5σ which is needed for discovery. All these values are calculated assuming no systematic errors and that sufficiently large MC samples are available so statistical errors do not influence the results.

References

- [1] M. Thomson, *Modern Particle Physics*, 3rd ed., 2013, United Kingdom: Cambridge University Press.
- [2] Cern website, the Standard Model, <https://home.cern/about/physics/standard-model>. [30 July 2017].
- [3] A. Boyarsky et al., *The role of sterile neutrinos in cosmology and astrophysics*, Ann.Rev.Nucl.Part.Sci.59:191-214 (2009), arXiv: 0901.0011 [**hep-ph**].
- [4] Wikipedia, Standard Model, https://en.wikipedia.org/wiki/Standard_Model. [30 July 2017].
- [5] C. Patrignani et al. (Particle Data Group), *Chapter 14: Neutrino Mass, Mixing and Oscillations*, Chin. Phys. C, 40, 100001 (2016) and 2017 update.
- [6] JUNO Collaboration, F. An et al., *Neutrino Physics with JUNO*, Journal of Physics G, arXiv: 1507.05613 [**physics.ins-det**].
- [7] C. Giunti and M. Laveder, *Neutrino Mixing*, Developments in Quantum Physics - 2004, arXiv: 0310238v2 [**hep-ph**].
- [8] S. N. Gninenko et al., *Search for GeV-scale sterile neutrinos responsible for active neutrino oscillations and baryon asymmetry of the Universe*, Invited Review article for the Special Issue on Neutrino Physics, arXiv:1301.5516 [**hep-ph**].
- [9] O. Ruchayskiy and A. Ivashko, *Experimental bounds on sterile neutrino mixing angles*, JHEP 1206 (2012) 100, arXiv: 1112.3319v2 [**hep-ph**].
- [10] T. Asaka and M. Shaposhnikov., *The ν MSM, Dark Matter and Baryon Asymmetry of the Universe*, Phys.Lett.B620:17-26 (2005), arXiv:0505013 [**hep-ph**].
- [11] D. Gorbunov and M. Shaposhnikov, *How to find neutral leptons of the ν MSM?*, JHEP0710:015 (2007), arXiv: 0705.1729v2 [**hep-ph**].
- [12] E. Izaguirre and B. Shuve, *Multilepton and Lepton Jet Probes of Sub-Weak-Scale Right-Handed Neutrinos*, Phys. Rev. D 91, 093010 (2015), arXiv: 1504.02470 [**hep-ph**].
- [13] ATLAS Collaboration, G. Aad et al., *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST 3 (2008) S08003, inSPIRE: 796888.
- [14] CERN, The accelerator complex, <http://home.web.cern.ch/about/accelerators>. [3 August 2017].
- [15] CERN images, the LHC lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/. [3 August 2017].

- [16] ATLAS Collaboration, G. Aad et al., *Pile-up Suppression in Missing Transverse Momentum Reconstruction in the ATLAS Experiment in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV*, ATLAS-CONF-2014-019, inSPIRE: 1296685.
- [17] ATLAS Experiment-Public Results; Luminosity Public Results Run 2, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>. [28 August 2017].
- [18] The ATLAS detector <http://inspirehep.net/record/834202/plots>. [3 August 2017].
- [19] P. Hansen, *Particle Detectors and Accelerators Lecture Notes*, 2nd ed., 2015, University of Copenhagen.
- [20] ATLAS Collaboration, M. Capeans et al., *ATLAS Insertable B-Layer Technical Design Report*, CERN-LHCC-2010-013; ATLAS-TDR-19, CDS: 1291633.
- [21] CDF Collaboration, T. Aaltonen et al., *Search for Standard Model Higgs Boson Production in Association with a W Boson Using a Matrix Element Technique at CDF in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV*, Phys. Rev. D. 85, 072001 (2012), arXiv: 1112.4358 [**hep-ex**].
- [22] F. Thiele, *An ATLAS search for sterile neutrinos*, slides available at, <http://indico.cp3-origins.dk/event/15/session/4/contribution/47>. [4 September 2017].
- [23] S. Frixione and B.R. Webber, *Matching NLO QCD computations and parton shower simulations*, JHEP 0206:029,2002, arXiv: 0204244 [**hep-ph**].
- [24] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, CERN-PH-TH-2014-190, arXiv: 1410.3012 [**hep-ph**].
- [25] T. Gleisberg et al., *Event generation with SHERPA 1.1*, JHEP 0902:007,2009, arXiv: 0811.4622 [**hep-ph**].
- [26] S. Xella, *Physics object reconstruction in the ATLAS experiment*, ATLAS-GEN-PROC-2013-001, CDS:1519113.
- [27] ATLAS Collaboration, G. Aad et al., *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s}=13$ TeV*, ATLAS-PHYS-PUB-2015-015, CDS: 2037613.
- [28] ATLAS Collaboration, G. Aad et al., *Pile-up subtraction and suppression for jets in ATLAS*, ATLAS-CONF-2013-083, inSPIRE: 1260963.
- [29] ATLAS Collaboration, G. Aad et al., *Machine Learning Algorithms for $b\bar{b}$ -Jet Tagging at the ATLAS Experiment*, ATLAS-PHYS-SLIDE-2017-677, CDS: 2280861.
- [30] ATLAS Collaboration, G. Aad et al., *Optimisation of the ATLAS $b\bar{b}$ -tagging performance for the 2016 LHC Run*, ATLAS-PHYS-PUB-2016-012, CDS: 2160731.

- [31] ATLAS Collaboration, G. Aad et al., *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C (2016) 76:292, arXiv: 1603.05598 [**hep-ex**].
- [32] M. Nielsen, *Neural Networks and Deep Learning*, online book on neural networks, <http://neuralnetworksanddeeplearning.com/index.html>. [24 July 2017].
- [33] P. Winston, *Artificial Intelligence*, 3rd ed., 1992, USA: Addison-Wesley, relevant excerpt: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/readings/MIT6_034F10_netmath.pdf. [24 August 2017].
- [34] Keras Documentation, <https://keras.io/>. [12 August 2017].
- [35] D. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, Learning (cs.LG), Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015, arXiv: 1412.6980v8 [**cs.LG**].
- [36] T. Fawcett, *ROC Graphs: Notes and Practical Considerations for Data Mining Researchers*, <http://www.hpl.hp.com/techreports/2003/HPL-2003-4.pdf>. [24 July 2017].
- [37] Wikipedia, Receiver operating characteristic, https://en.wikipedia.org/wiki/Receiver_operating_characteristic. [24 July 2017].

Appendix A Kinematic distributions used in truth analysis

A.1 Truth-level analysis distributions for a 5 GeV HNL mass

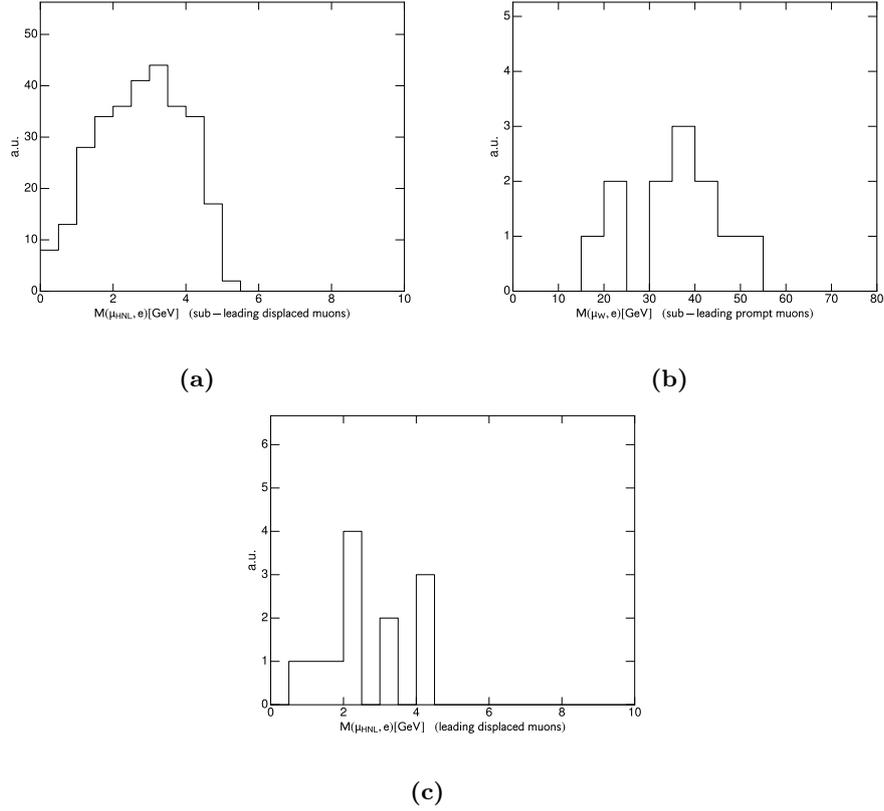


Figure A.1: HNL invariant mass distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 5 GeV mass sample.

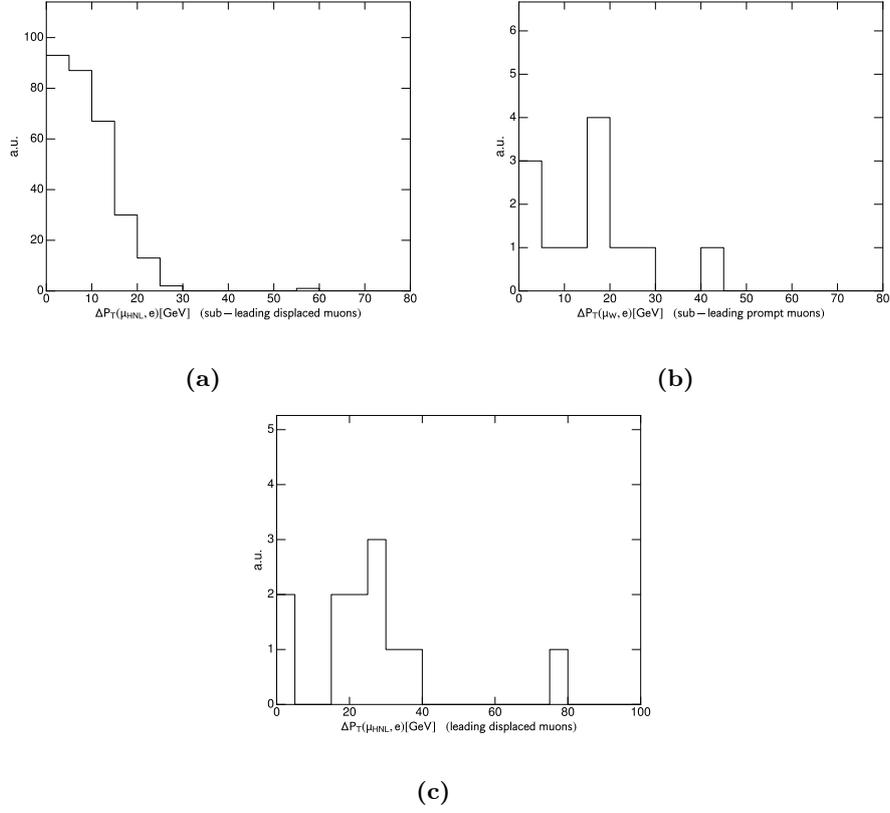


Figure A.2: Δp_T distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 5 GeV mass sample.

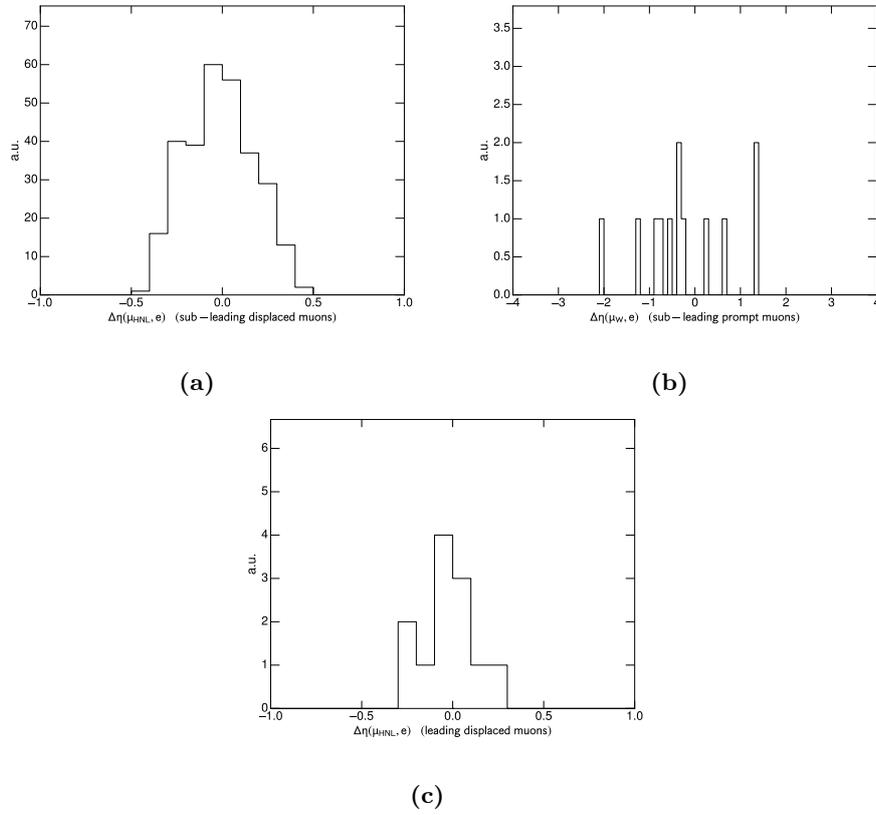


Figure A.3: $\Delta\eta$ distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 5 GeV mass sample.

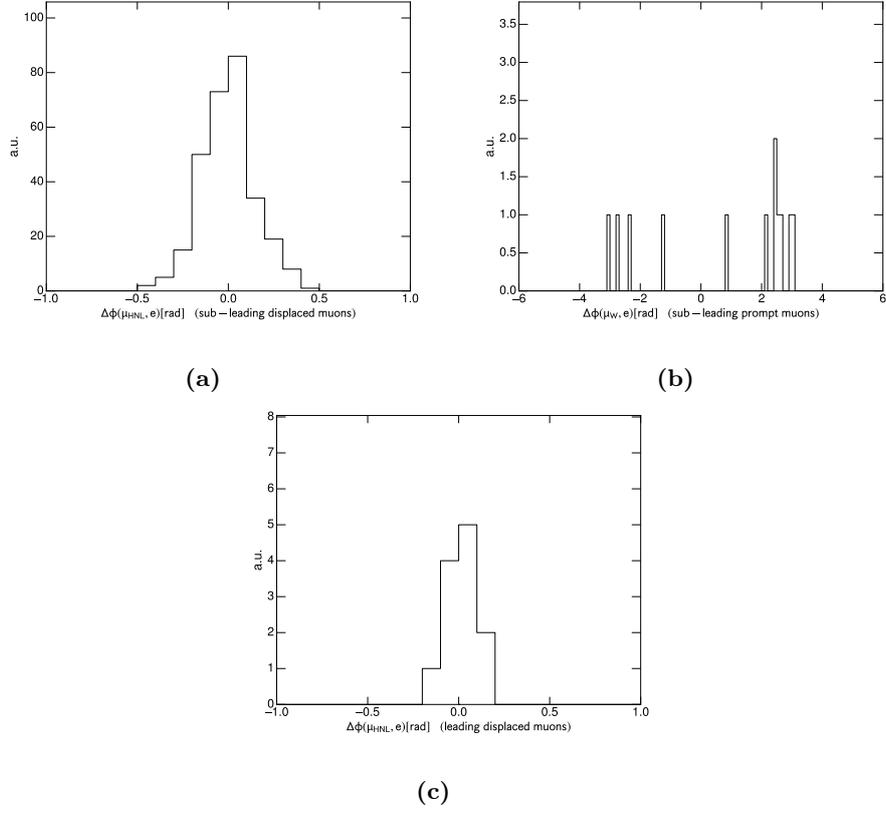


Figure A.4: $\Delta\phi$ distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 5 GeV mass sample.

A.2 Truth analysis distributions for a 50 GeV HNL mass

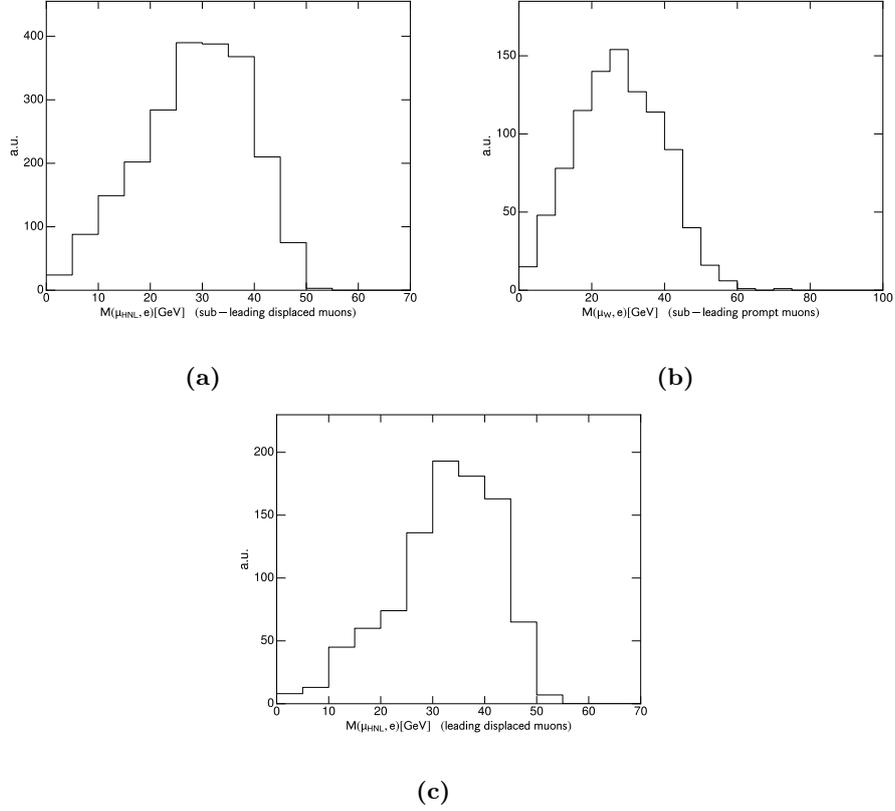


Figure A.5: HNL invariant mass distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 50 GeV mass sample.

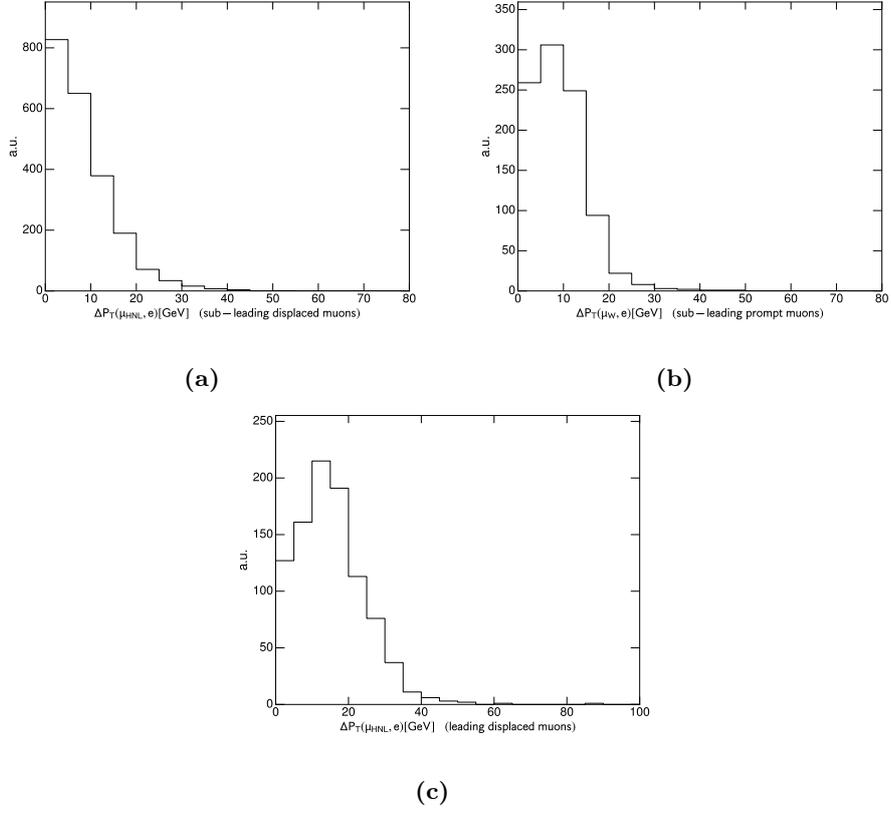


Figure A.6: Δp_T distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 50 GeV mass sample.

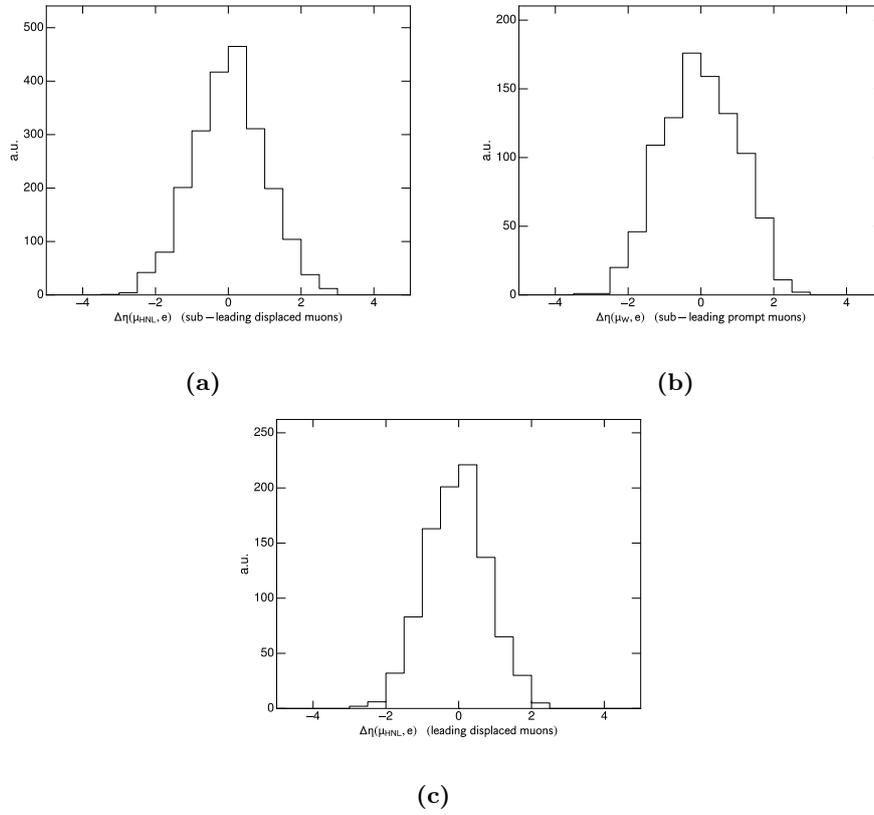


Figure A.7: $\Delta\eta$ distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 50 GeV mass sample.

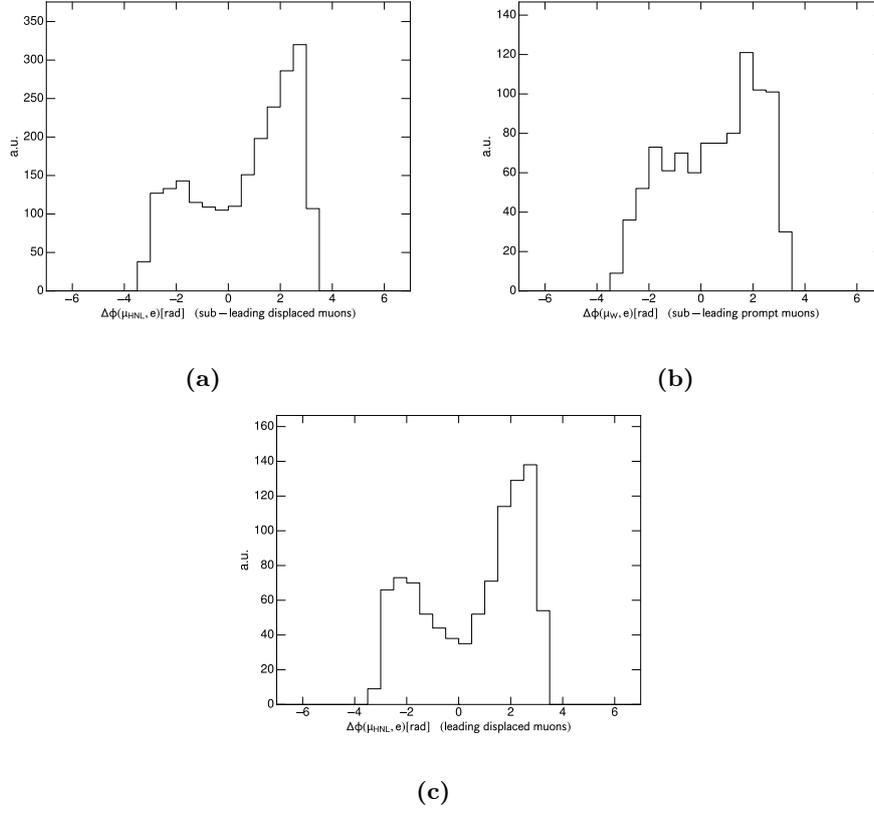


Figure A.8: $\Delta\phi$ distributions of the sub-leading displaced muon and the electron (a), the sub-leading prompt muon and the electron (b) and the leading displaced muon and the electron (c) for the 50 GeV mass sample.

Appendix B Invariant mass plots

B.1 Dilepton invariant mass plots

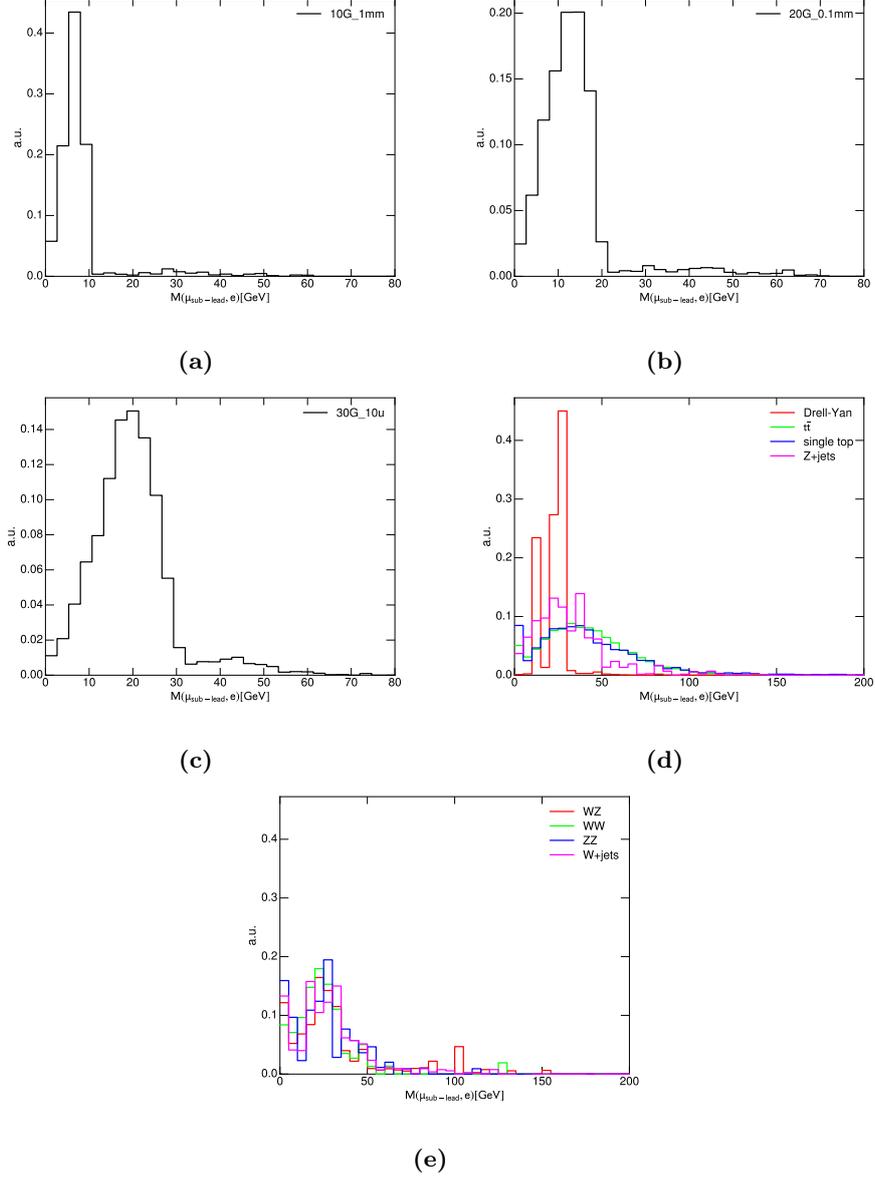


Figure B.1: Dilepton invariant mass distributions for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses. The backgrounds are shown in (d) and (e).

B.2 Trilepton invariant mass plots

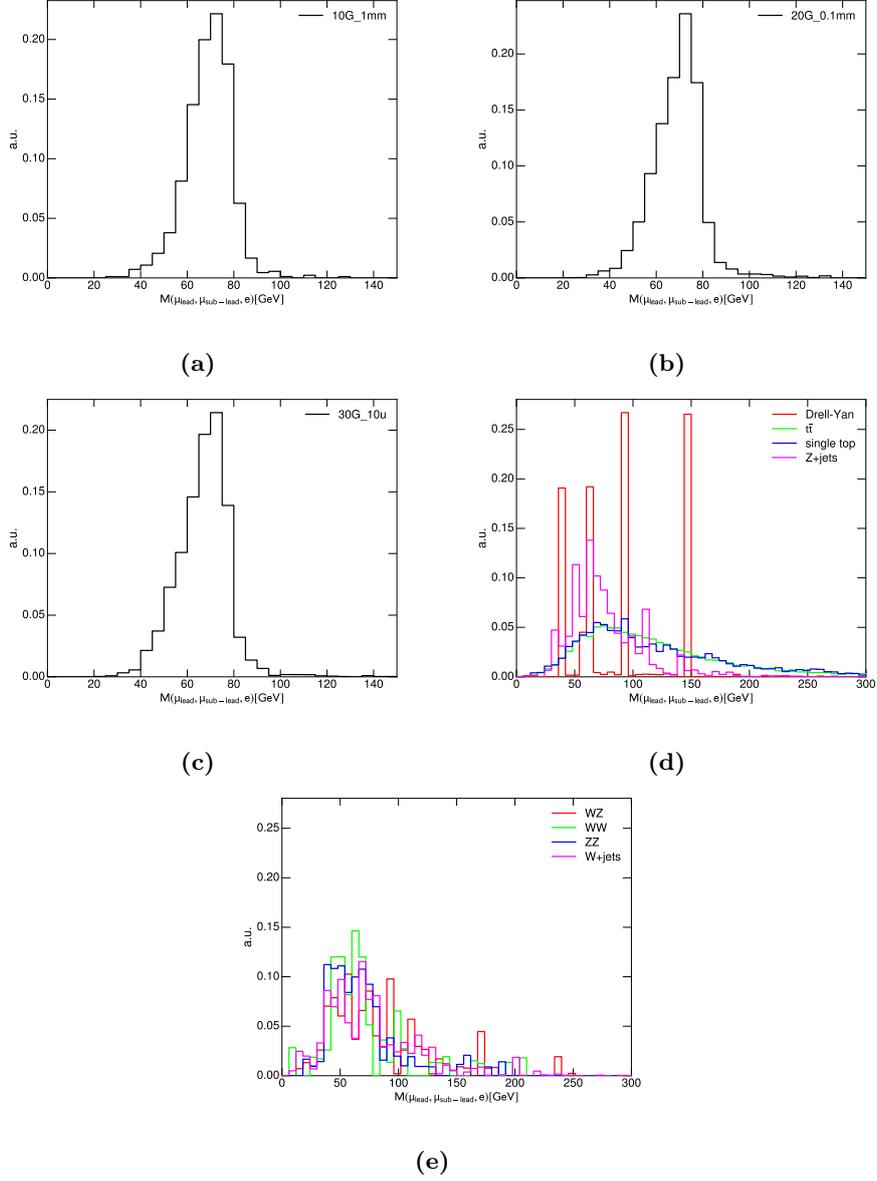


Figure B.2: Trilepton invariant mass distributions for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses. The backgrounds are shown in (d) and (e).

Appendix C Cut optimization of p_T for all three leptons

C.1 Leading muon p_T

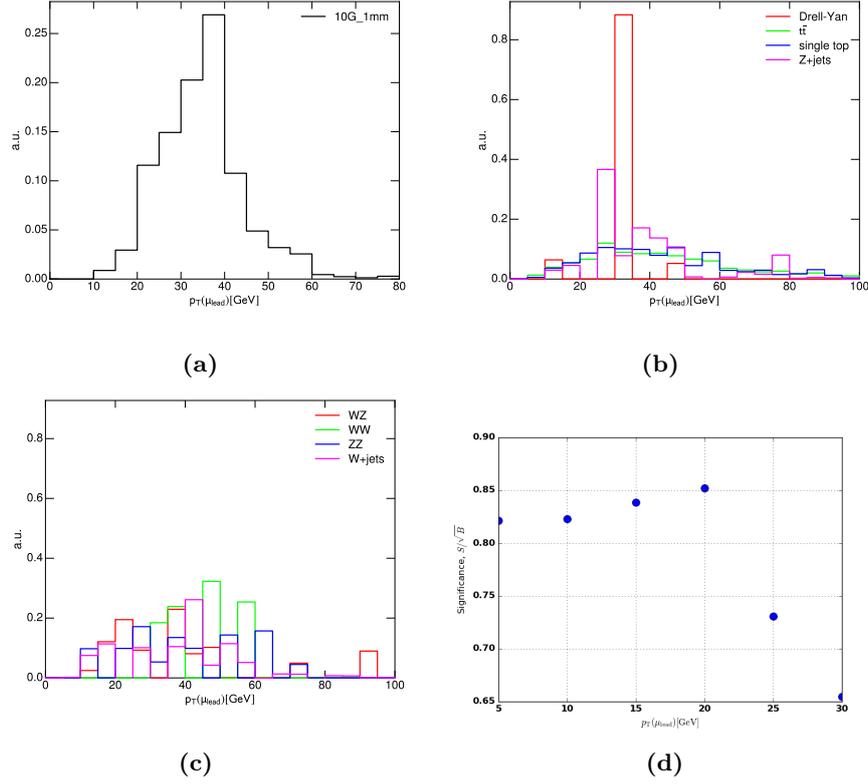


Figure C.1: Choice of the leading muon p_T cut for the 10 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

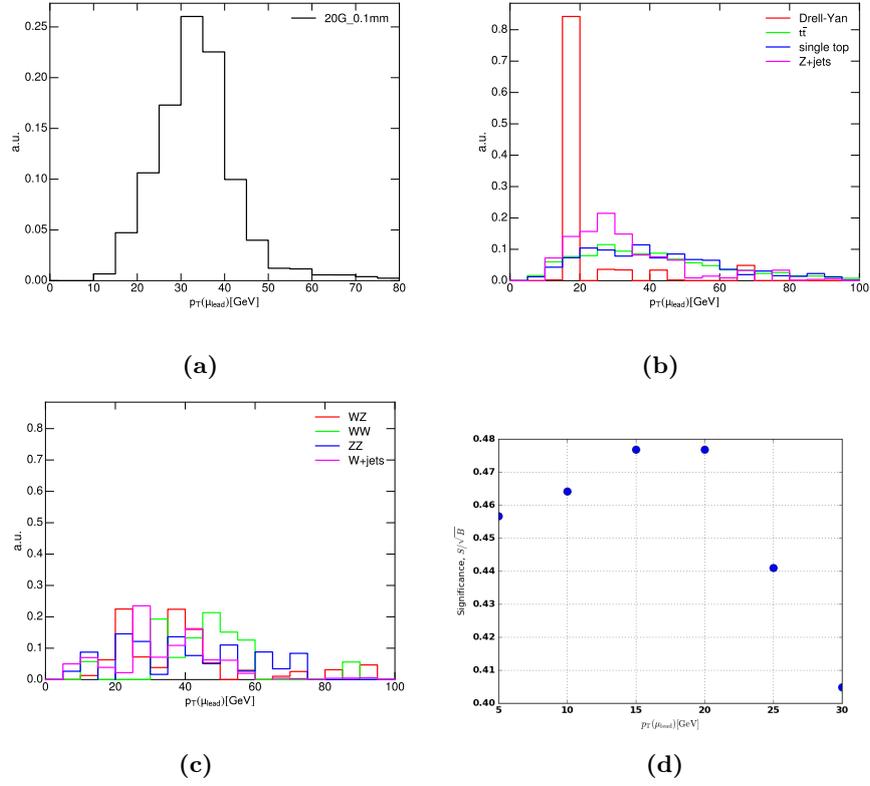


Figure C.2: Choice of the leading muon p_T cut for the 20 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

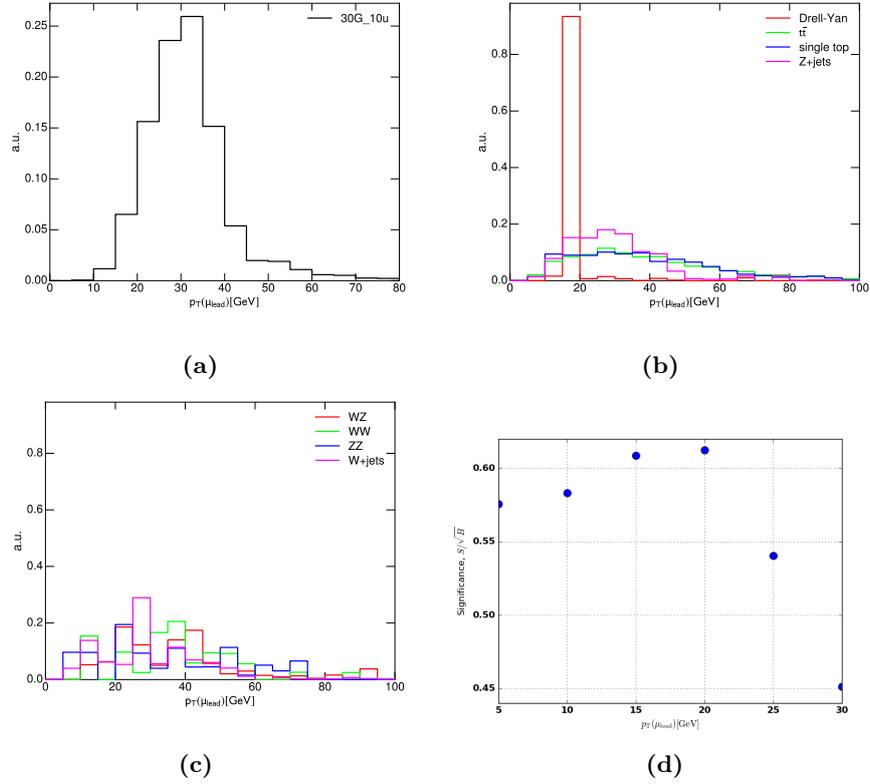


Figure C.3: Choice of the leading muon p_T cut for the 30 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

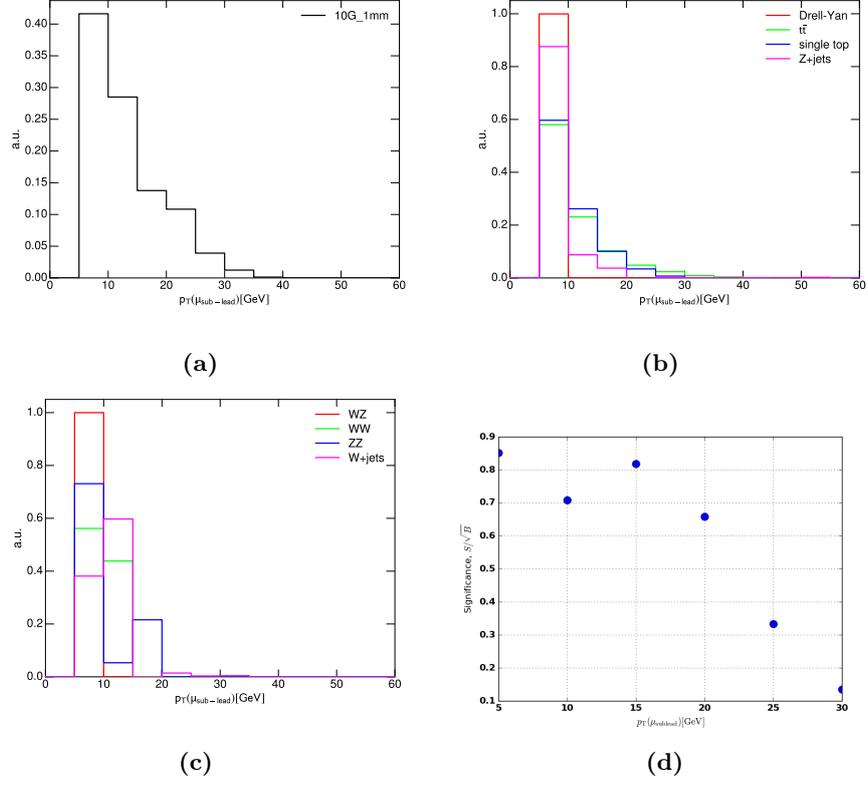
C.2 Sub-leading muon p_T 

Figure C.4: Choice of the sub-leading muon p_T cut for the 10 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

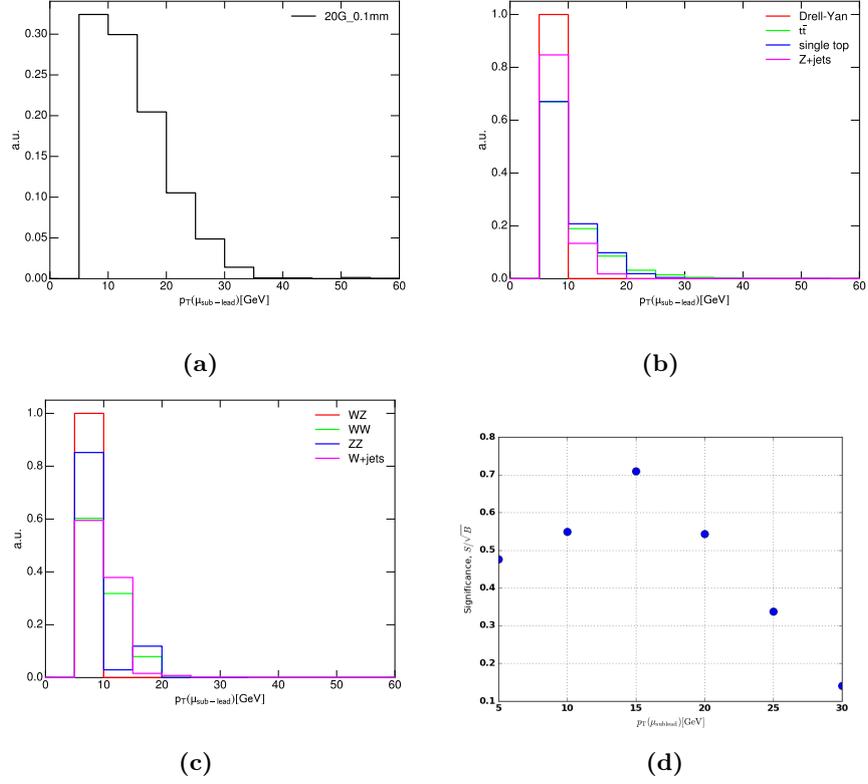


Figure C.5: Choice of the sub-leading muon p_T cut for the 20 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

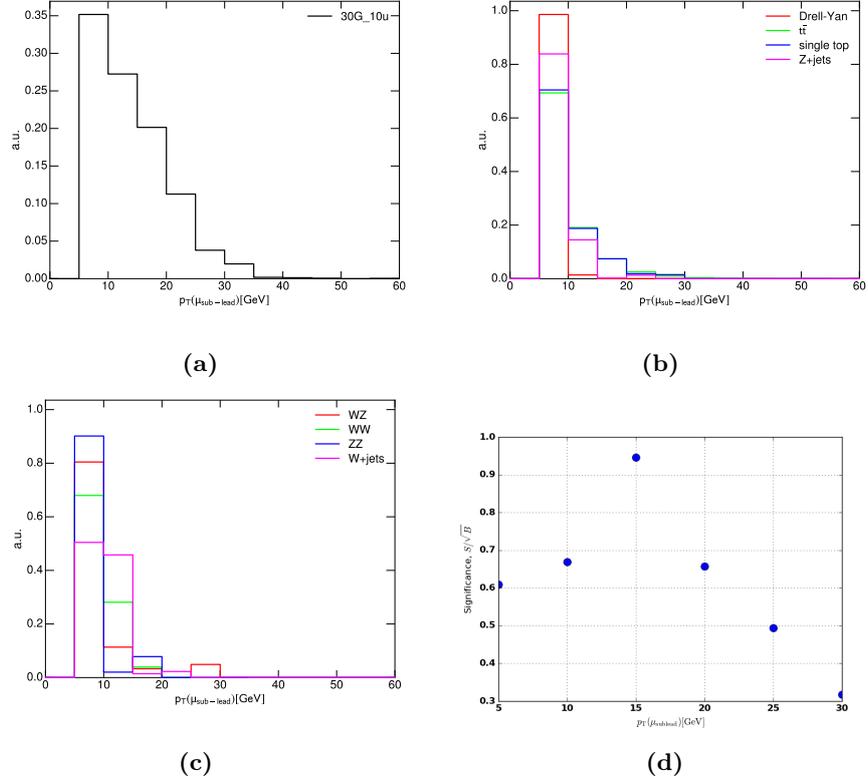


Figure C.6: Choice of the sub-leading muon p_T cut for the 30 GeV mass. Figure(a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

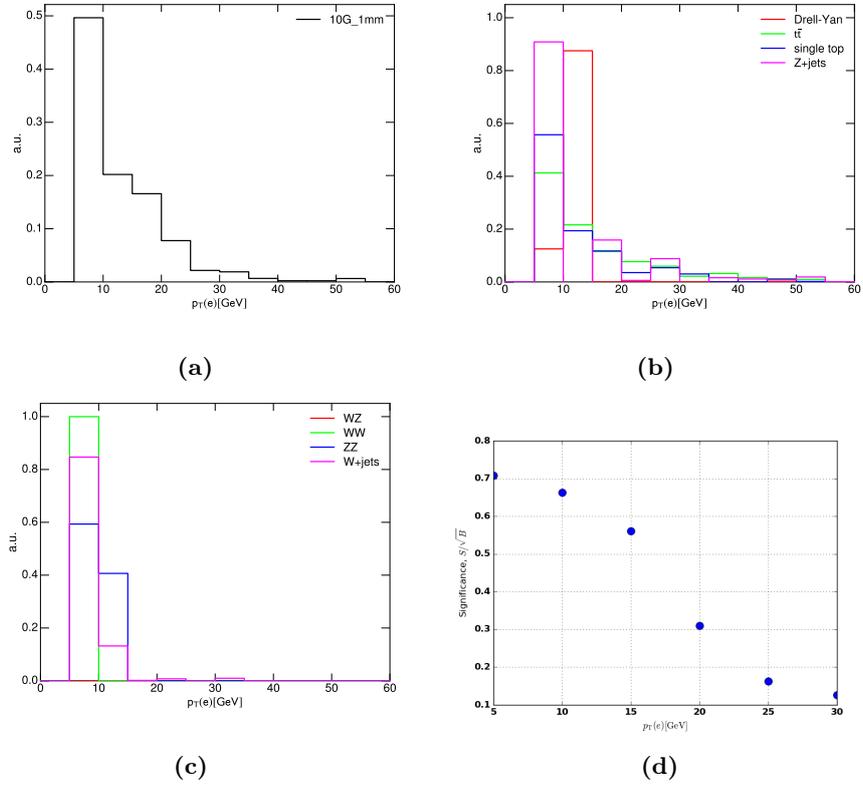
C.3 Electron p_T 

Figure C.7: Choice of the electron p_T cut for the 10 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

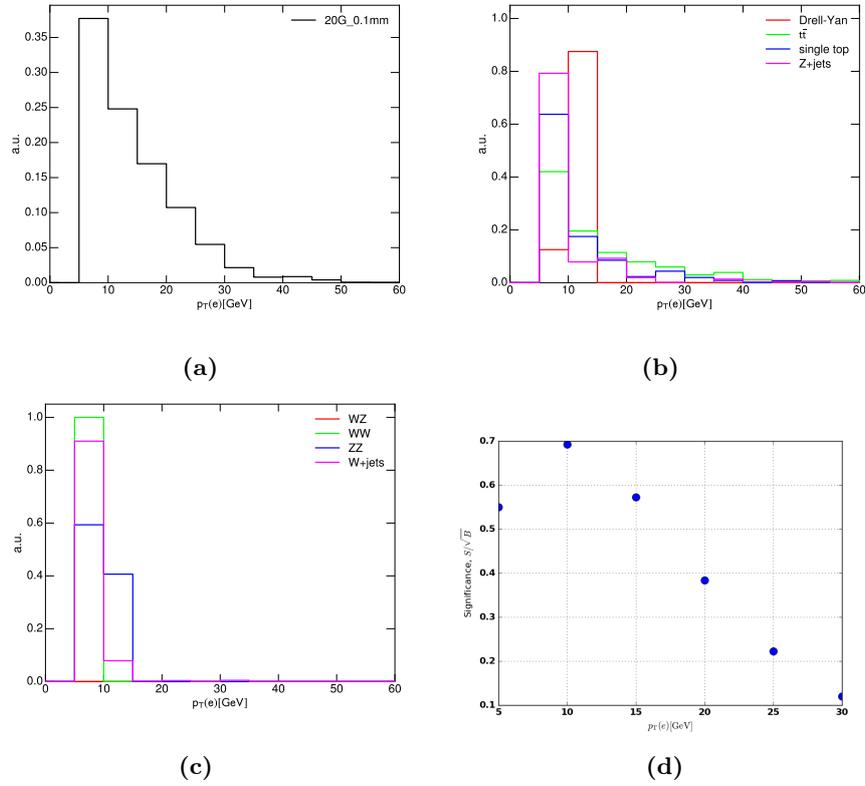


Figure C.8: Choice of the electron p_T cut for the 20 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

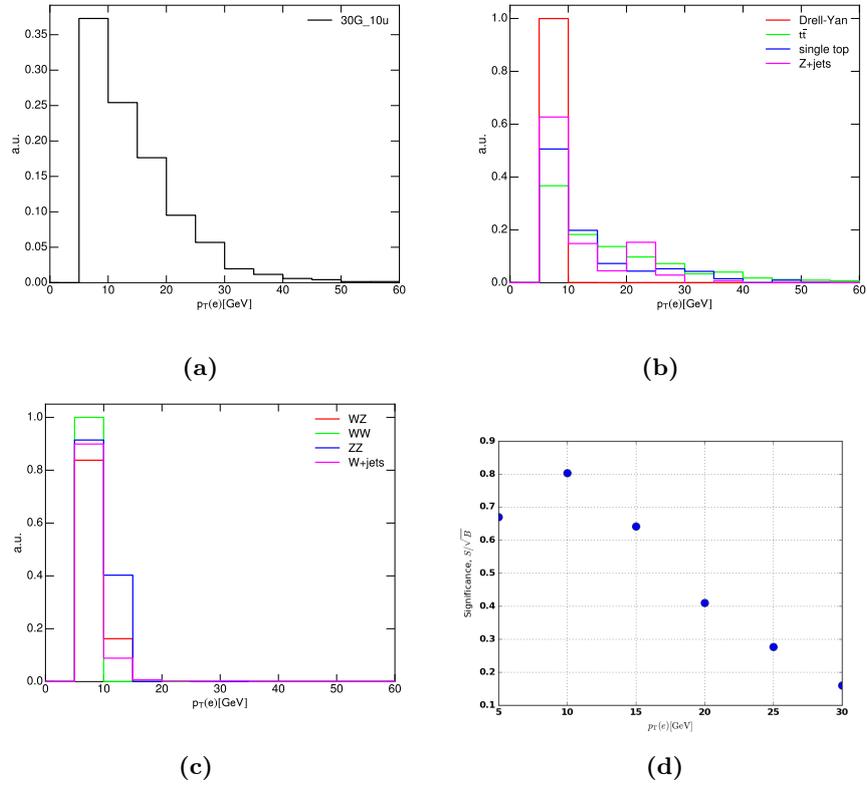


Figure C.9: Choice of the electron p_T cut for the 30 GeV mass. Figure (a) shows the signal p_T distribution, figure (b) and (c) show the background p_T distributions and figure (d) shows the p_T cut dependence of the significance.

C.4 Sub-leading muon significance plots of individual backgrounds for 10 GeV

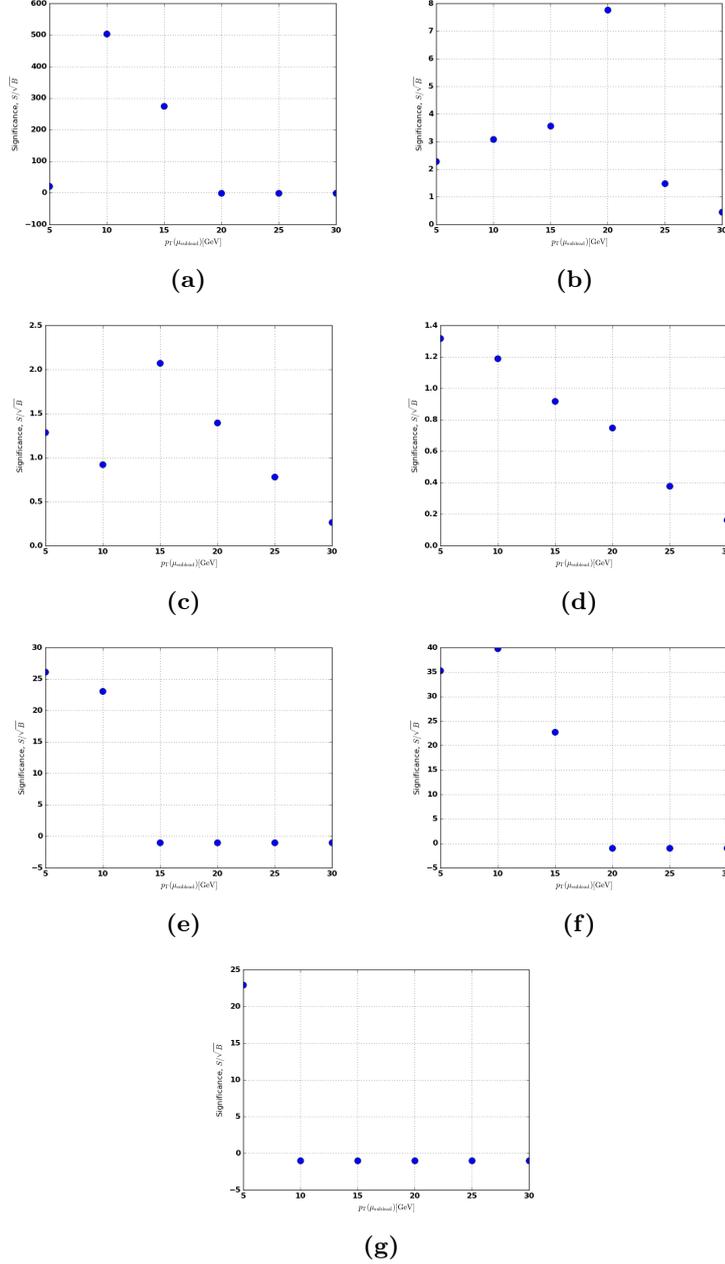


Figure C.10: The figure shows significance plots for testing the optimal cut on the sub-leading muon p_T . The test is performed for the 10 GeV signal sample. The total background in S/\sqrt{B} is replaced by individual background counts for Drell-Yan (a), Z + jets (b), W + jets (c), $t\bar{t}$ and single-top (d), WW (e), ZZ (f) and finally WZ (g).

Appendix D Distributions of the impact parameter

D.1 d_0/σ_{d_0} distributions for the sub-leading muon

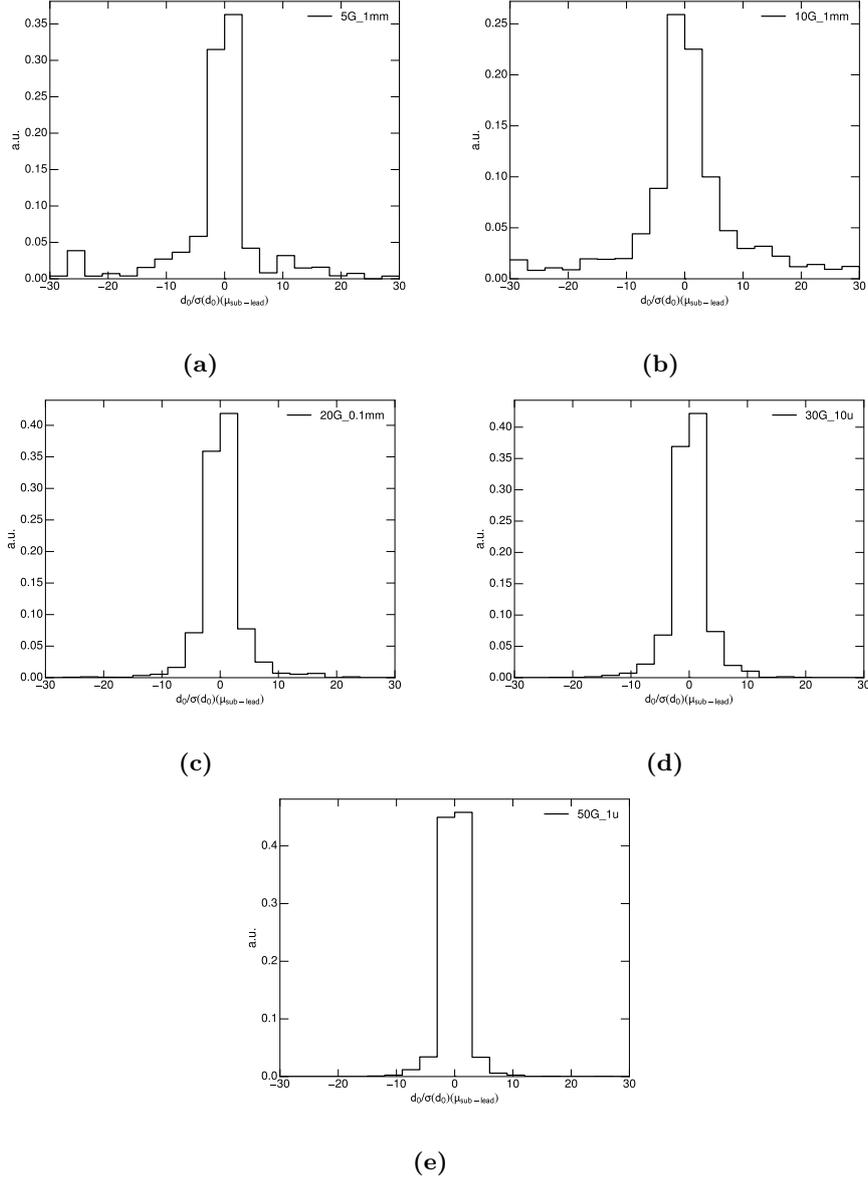
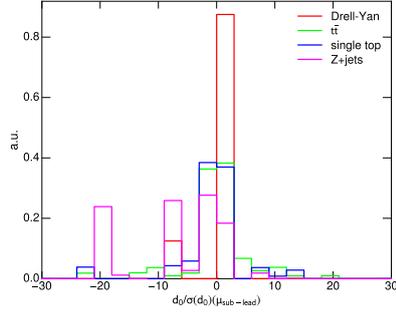
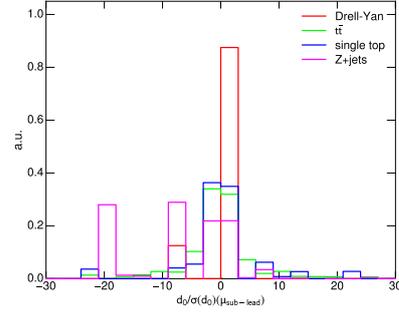


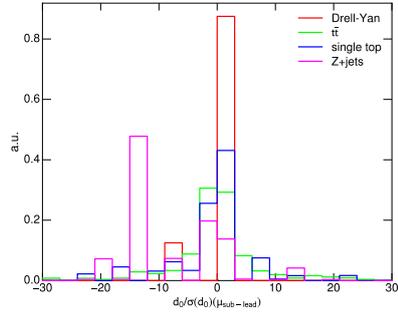
Figure D.1: The figure shows signal distributions of d_0/σ_{d_0} for the sub-leading muon. The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.



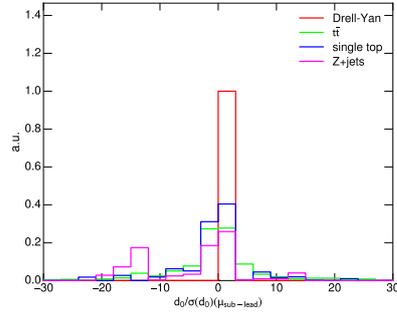
(a)



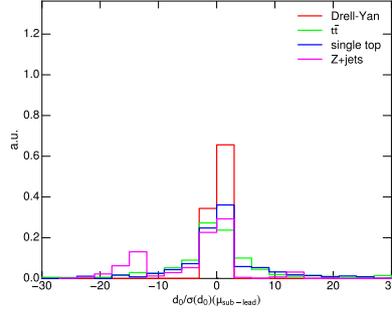
(b)



(c)



(d)



(e)

Figure D.2: The figure shows background distributions of d_0/σ_{d_0} for the sub-leading muon. Here the background consists of Drell-Yan, $t\bar{t}$, single-top and $Z + \text{jets}$. The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.

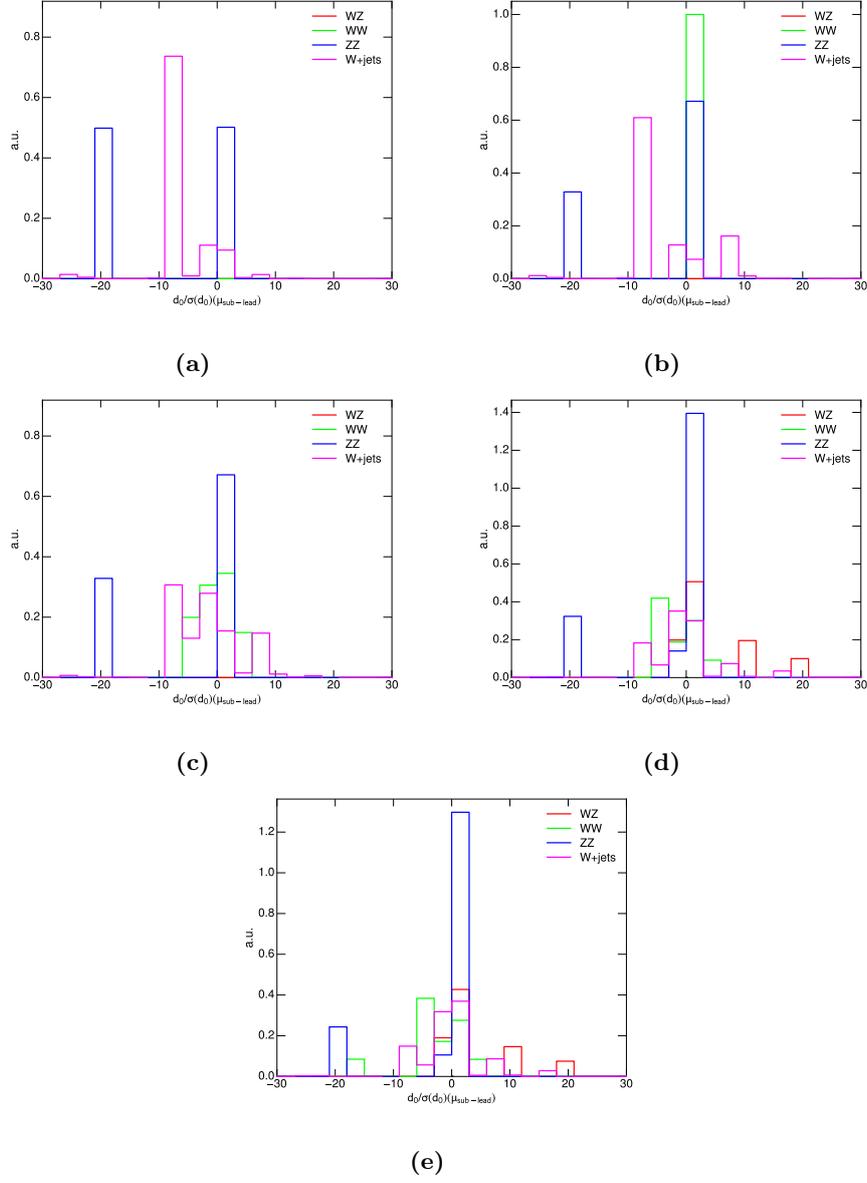


Figure D.3: The figure shows background distributions of d_0/σ_{d_0} for the sub-leading muon. Here the background consists of $W + \text{jets}$, WW , ZZ and WZ . The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.

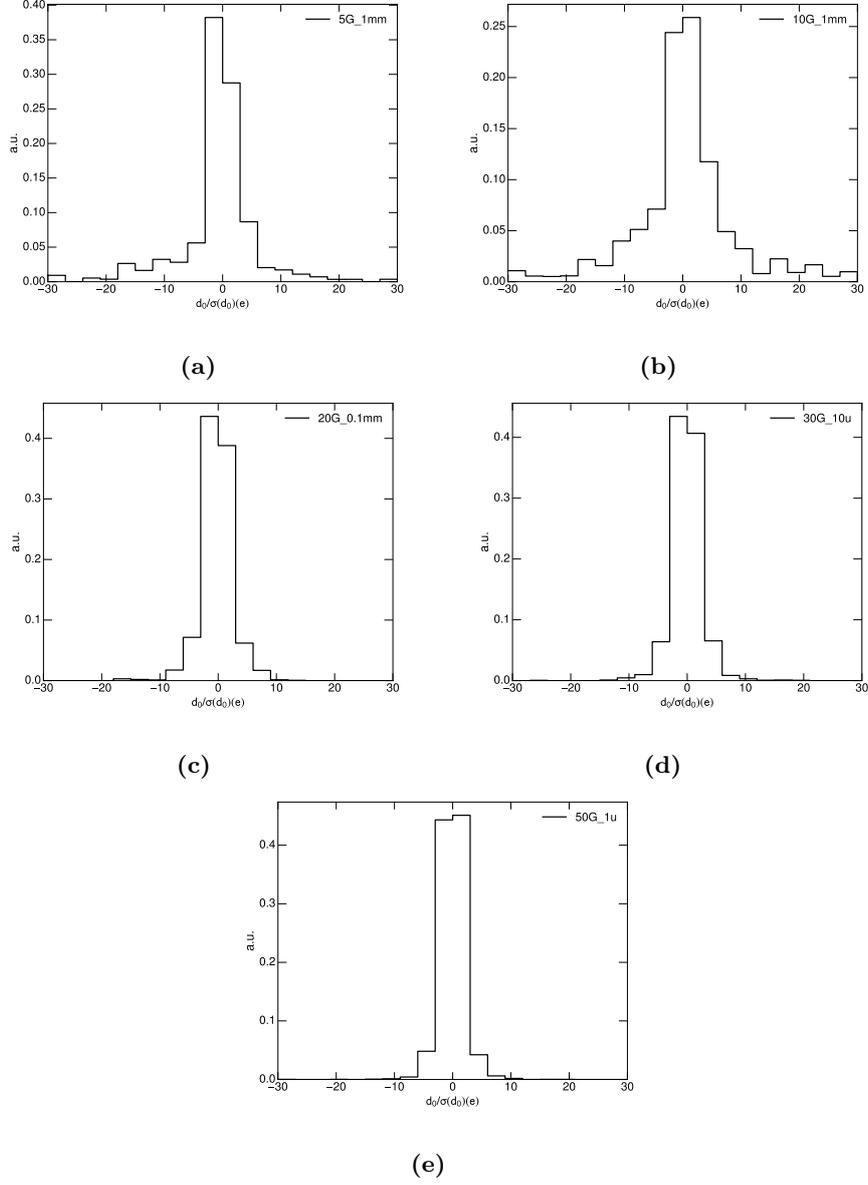
D.2 d_0/σ_{d_0} distributions for the electron

Figure D.4: The figure shows signal distributions of d_0/σ_{d_0} for the electron. The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.

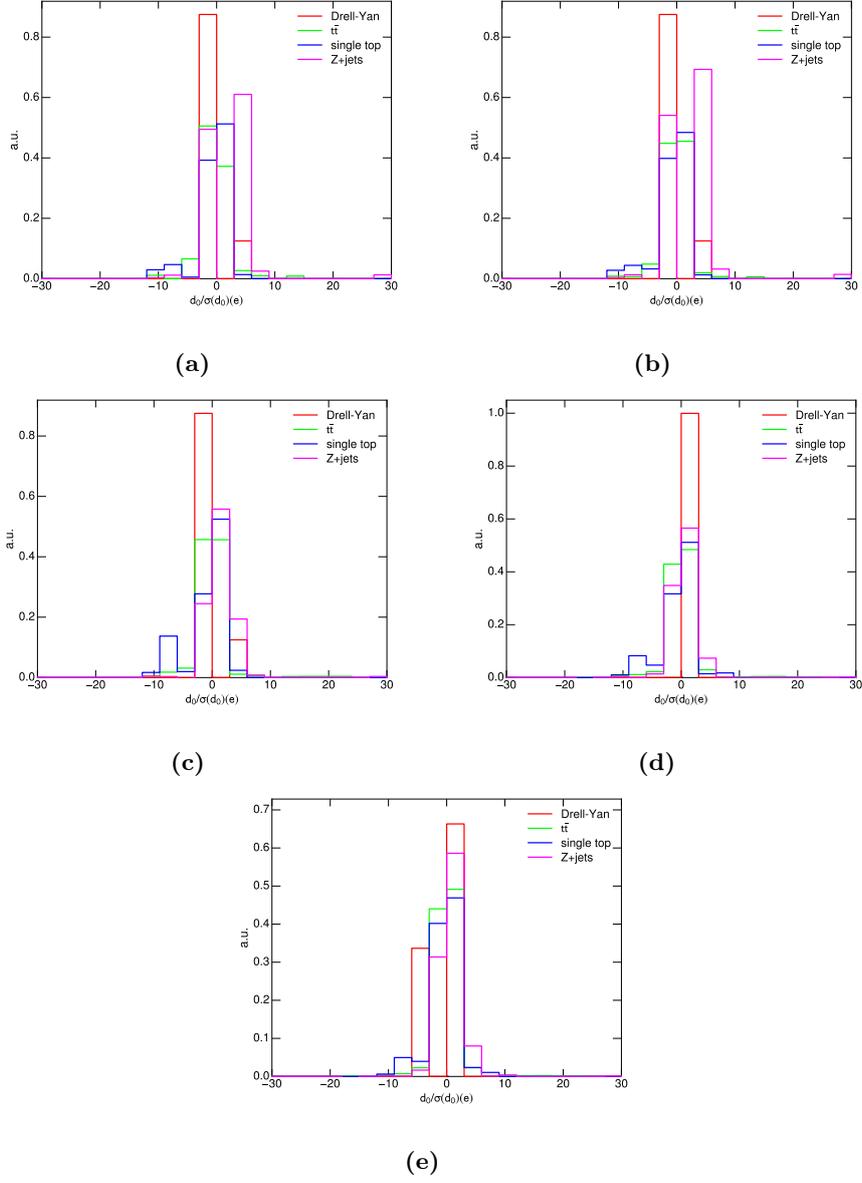


Figure D.5: The figure shows background distributions of d_0/σ_{d_0} for the electron. Here the background consists of Drell-Yan, $t\bar{t}$, single-top and $Z + \text{jets}$. The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.

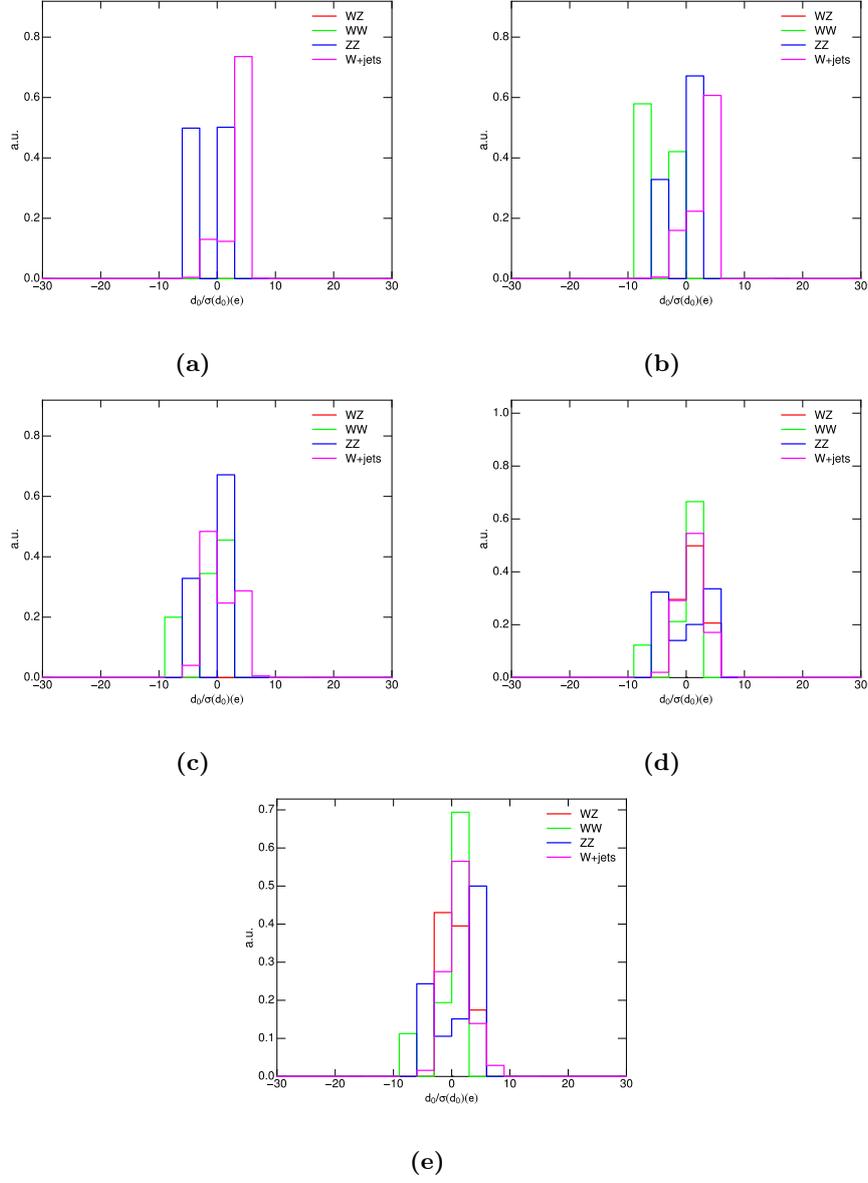


Figure D.6: The figure shows background distributions of d_0/σ_{d_0} for the electron. Here the background consists of $W + \text{jets}$, WW , ZZ and WZ . The plots are shown for the 5 GeV (a), 10 GeV (b), 20 GeV (c), 30 GeV (d) and 50 GeV (e) HNL masses.

Appendix E Optimization of WP combinations

E.1 Summary of cuts used for isolation

Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
$\text{Invmass}(\mu, \mu, e) < 90 \text{ GeV}$
$\text{Invmass}(\mu, \mu, e) > 40 \text{ GeV}$
$\text{Invmass}(\mu, e) < 10 \text{ GeV}$
Leading $p_{T,\mu} \geq 20 \text{ GeV}$
Suleading $p_{T,\mu} \geq 10 \text{ GeV}$
Number of b jets is 0
Gradient isolation for leading muon

Table E.1: A summary of cuts used for testing isolation WPs on the sub-leading muon and the electron. The cuts refer to a study on an HNL signal sample with a mass of 10 GeV.

Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
$\text{Invmass}(\mu, \mu, e) < 90 \text{ GeV}$
$\text{Invmass}(\mu, \mu, e) > 40 \text{ GeV}$
$\text{Invmass}(\mu, e) < 20 \text{ GeV}$
Leading $p_{T,\mu} \geq 20 \text{ GeV}$
Suleading $p_{T,\mu} \geq 10 \text{ GeV}$
Number of b jets is 0
Gradient isolation for leading muon

Table E.2: A summary of cuts used for testing isolation WPs on the sub-leading muon and the electron. The cuts refer to a study on an HNL signal sample with a mass of 20 GeV.

Cut
2 muons (loose and combined) and 1 electron (loose)
$q_{\mu,\text{lead}} = q_{\mu,\text{sub}}$
$q_{\mu,\text{lead}} = -q_e$
Invmass $(\mu, \mu, e) < 90$ GeV
Invmass $(\mu, \mu, e) > 40$ GeV
Invmass $(\mu, e) < 30$ GeV
Leading $p_{T,\mu} \geq 15$ GeV
Suleading $p_{T,\mu} \geq 10$ GeV
Number of b jets is 0
Gradient isolation for leading muon

Table E.3: A summary of cuts used for testing isolation WPs on the sub-leading muon and the electron. The cuts refer to a study on an HNL signal sample with a mass of 30 GeV.

E.2 WP combinations for 10 GeV

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	0.89 ± 0.06	1.46 ± 0.14	1.52 ± 0.15	1.54 ± 0.16	1.69 ± 0.16
Loose	0.9 ± 0.06	1.49 ± 0.15	1.48 ± 0.15	1.49 ± 0.15	1.63 ± 0.16
Tight	1.76 ± 0.16	1.95 ± 0.26	1.95 ± 0.26	2.01 ± 0.29	2.45 ± 0.29
Gradient Loose	1.91 ± 0.21	2.22 ± 0.42	2.18 ± 0.41	2.16 ± 0.4	2.41 ± 0.48
Gradient	2.35 ± 0.38	2.19 ± 0.47	2.15 ± 0.46	2.12 ± 0.45	2.4 ± 0.55

Table E.4: The significance given by S/\sqrt{B} after testing different working point combinations of the sub-leading muon and the electron for the 10 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	11.62 ± 0.61	9.55 ± 0.55	9.36 ± 0.55	9.27 ± 0.54	9.03 ± 0.54
Loose	11.58 ± 0.61	9.51 ± 0.55	9.33 ± 0.55	9.23 ± 0.54	8.99 ± 0.54
Tight	10.97 ± 0.59	8.93 ± 0.53	8.74 ± 0.52	8.65 ± 0.52	8.41 ± 0.51
Gradient Loose	10.41 ± 0.57	8.4 ± 0.51	8.24 ± 0.51	8.15 ± 0.51	7.91 ± 0.5
Gradient	9.68 ± 0.55	7.74 ± 0.49	7.59 ± 0.49	7.49 ± 0.48	7.29 ± 0.48

Table E.5: The number of signal events after testing different working point combinations of the sub-leading muon and the electron for the 10 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	171.16 ± 15.47	42.6 ± 6.35	37.94 ± 5.99	36.16 ± 5.93	28.46 ± 4.22
Loose	165.51 ± 15.55	40.76 ± 6.38	39.94 ± 6.37	38.16 ± 6.32	30.46 ± 4.78
Tight	38.65 ± 5.48	20.87 ± 4.88	20.05 ± 4.86	18.56 ± 4.8	11.77 ± 2.34
Gradient Loose	29.64 ± 5.74	14.28 ± 5.05	14.28 ± 5.05	14.28 ± 5.05	10.81 ± 4.07
Gradient	17.01 ± 5.14	12.45 ± 5.04	12.45 ± 5.04	12.45 ± 5.04	9.26 ± 4.05

Table E.6: The number of background events after testing different working point combinations of the sub-leading muon and the electron for the 10 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

E.3 WP combinations for 20 GeV

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	0.69 ± 0.03	0.97 ± 0.05	0.98 ± 0.06	0.98 ± 0.06	0.98 ± 0.06
Loose	0.73 ± 0.03	0.99 ± 0.06	0.99 ± 0.06	0.99 ± 0.06	0.99 ± 0.06
Tight	0.9 ± 0.04	1.08 ± 0.07	1.08 ± 0.07	1.08 ± 0.07	1.09 ± 0.07
Gradient Loose	1.11 ± 0.06	1.68 ± 0.11	1.68 ± 0.11	1.67 ± 0.11	1.69 ± 0.12
Gradient	1.2 ± 0.09	1.94 ± 0.16	1.95 ± 0.16	1.93 ± 0.17	1.96 ± 0.18

Table E.7: The significance given by S/\sqrt{B} after testing different working point combinations of the sub-leading muon and the electron for the 20 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	13.55 ± 0.38	11.75 ± 0.36	11.6 ± 0.35	11.39 ± 0.35	11.13 ± 0.35
Loose	13.49 ± 0.38	11.69 ± 0.35	11.54 ± 0.35	11.33 ± 0.35	11.07 ± 0.35
Tight	13.15 ± 0.38	11.37 ± 0.35	11.22 ± 0.35	11.04 ± 0.34	10.78 ± 0.34
Gradient Loose	12.82 ± 0.37	11.09 ± 0.34	10.95 ± 0.34	10.78 ± 0.34	10.53 ± 0.34
Gradient	12.41 ± 0.37	10.74 ± 0.34	10.62 ± 0.34	10.45 ± 0.34	10.21 ± 0.33

Table E.8: The number of signal events after testing different working point combinations of the sub-leading muon and the electron for the 20 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	384.74 ± 25.97	147.45 ± 13.6	138.64 ± 13.24	135.98 ± 13.21	129.91 ± 13.33
Loose	342.95 ± 23.69	139.27 ± 13.49	134.69 ± 13.41	132.03 ± 13.37	126.25 ± 13.54
Tight	214.51 ± 17.23	111.7 ± 11.8	107.12 ± 11.69	104.75 ± 11.65	97.57 ± 11.53
Gradient Loose	134.16 ± 13.38	43.74 ± 5.13	42.75 ± 5.11	41.84 ± 5.08	38.89 ± 5.04
Gradient	106.59 ± 14.05	30.75 ± 4.69	29.76 ± 4.67	29.15 ± 4.66	27.25 ± 4.62

Table E.9: The number of background events after testing different working point combinations of the sub-leading muon and the electron for the 20 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

E.4 WP combinations for 30 GeV

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	0.78 ± 0.03	0.91 ± 0.04	0.91 ± 0.04	0.91 ± 0.04	0.89 ± 0.04
Loose	0.81 ± 0.03	0.94 ± 0.04	0.94 ± 0.04	0.94 ± 0.04	0.92 ± 0.04
Tight	0.88 ± 0.03	0.94 ± 0.04	0.94 ± 0.04	0.93 ± 0.04	0.92 ± 0.05
Gradient Loose	0.95 ± 0.04	1.02 ± 0.05	1.01 ± 0.05	1.01 ± 0.05	0.99 ± 0.05
Gradient	0.98 ± 0.05	1.05 ± 0.06	1.04 ± 0.06	1.04 ± 0.06	1.02 ± 0.07

Table E.10: The significance given by S/\sqrt{B} after testing different working point combinations of the sub-leading muon and the electron for the 30 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	23.96 ± 0.59	21.26 ± 0.56	20.95 ± 0.55	20.79 ± 0.55	20.17 ± 0.54
Loose	23.79 ± 0.59	21.09 ± 0.55	20.8 ± 0.55	20.65 ± 0.55	20.08 ± 0.54
Tight	23.16 ± 0.58	20.49 ± 0.55	20.23 ± 0.54	20.08 ± 0.54	19.51 ± 0.53
Gradient Loose	22.81 ± 0.58	20.19 ± 0.54	19.93 ± 0.54	19.77 ± 0.54	19.22 ± 0.53
Gradient	22.05 ± 0.57	19.51 ± 0.53	19.29 ± 0.53	19.14 ± 0.53	18.62 ± 0.52

Table E.11: The number of signal events after testing different working point combinations of the sub-leading muon and the electron for the 30 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Subleading muon WP	Electron WP				
	Loose Track Only	Loose	Tight	Gradient Loose	Gradient
Loose Track Only	949.75 ± 48.26	544.29 ± 36.14	529.59 ± 36.29	524.83 ± 36.7	517.06 ± 38.23
Loose	870.04 ± 44.87	502.39 ± 33.91	491.92 ± 34.08	487.16 ± 34.48	479.67 ± 35.97
Tight	685.22 ± 40.49	476.35 ± 35.34	466.53 ± 35.63	462.05 ± 36.11	450.61 ± 37.37
Gradient Loose	575.98 ± 40.66	392.54 ± 33.54	387.08 ± 33.84	383.86 ± 34.37	377.1 ± 35.59
Gradient	506.65 ± 44.0	343.4 ± 35.76	341.04 ± 36.25	338.43 ± 36.96	331.78 ± 38.01

Table E.12: The number of background events after testing different working point combinations of the sub-leading muon and the electron for the 30 GeV HNL mass sample. The luminosity corresponds to 36 fb^{-1} .

Appendix F Distributions of E_T^{miss} and H_T

F.1 Distributions of E_T^{miss}

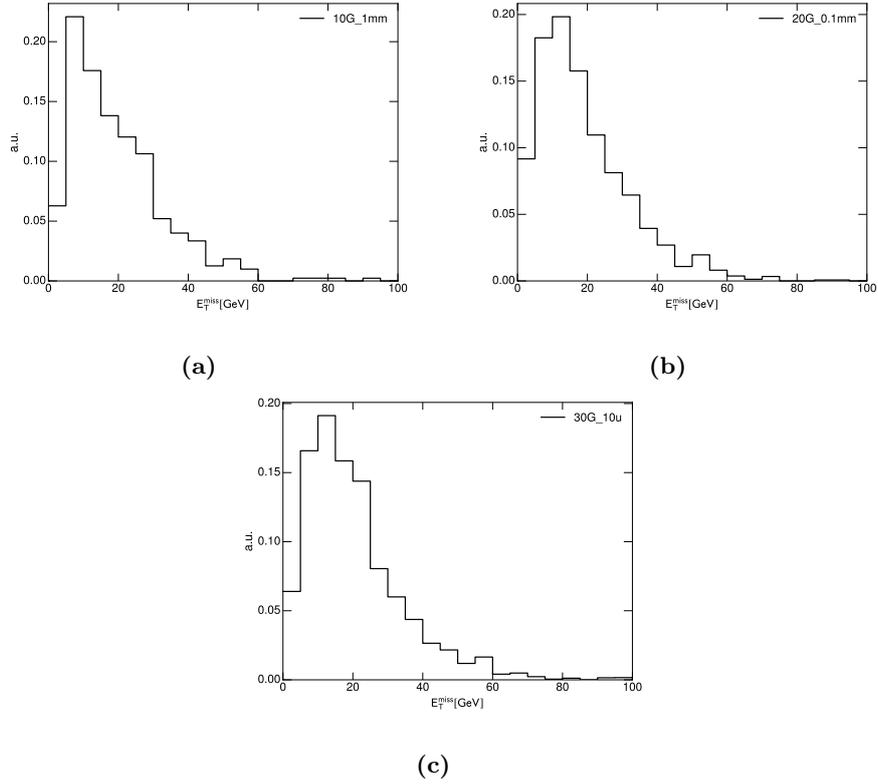


Figure F.1: Signal distributions of E_T^{miss} for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.

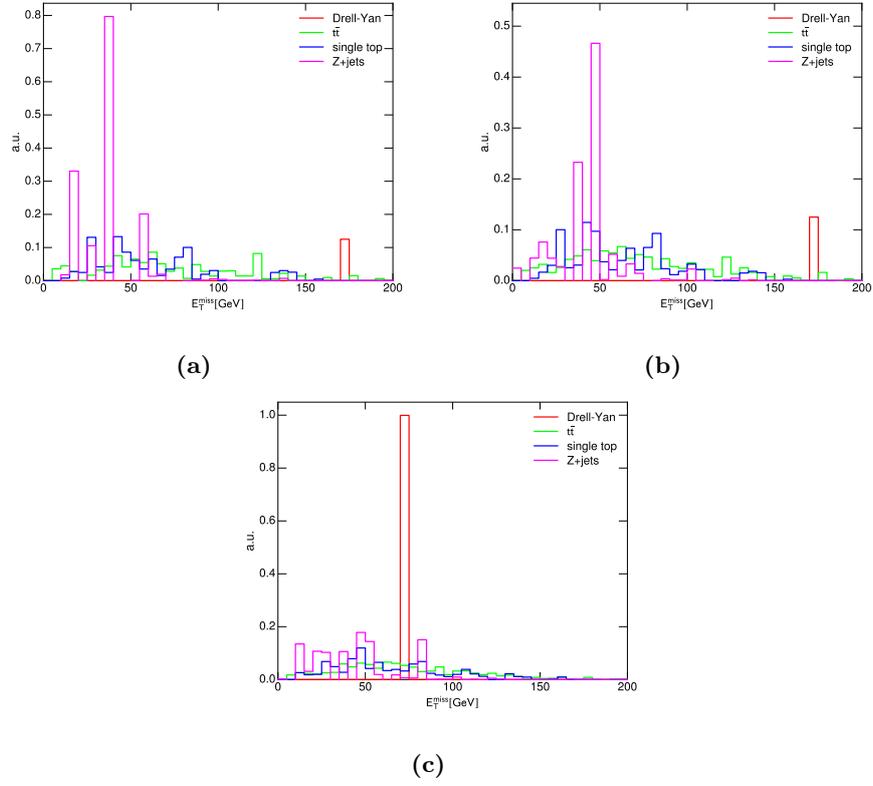


Figure F.2: Background distributions of E_T^{miss} for Drell-Yan, $t\bar{t}$, single top and $Z + \text{jets}$. The plots are shown for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.

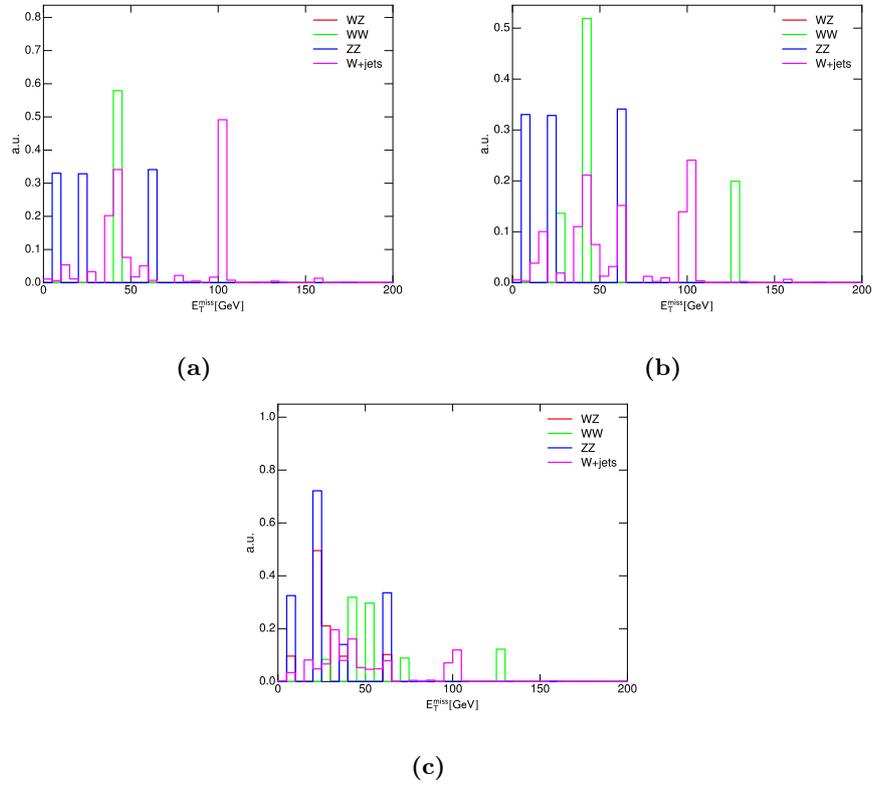


Figure F.3: Background distributions of E_T^{miss} for $W + \text{jets}$, WW , ZZ and WZ . The plots are shown for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.

F.2 Distributions of H_T

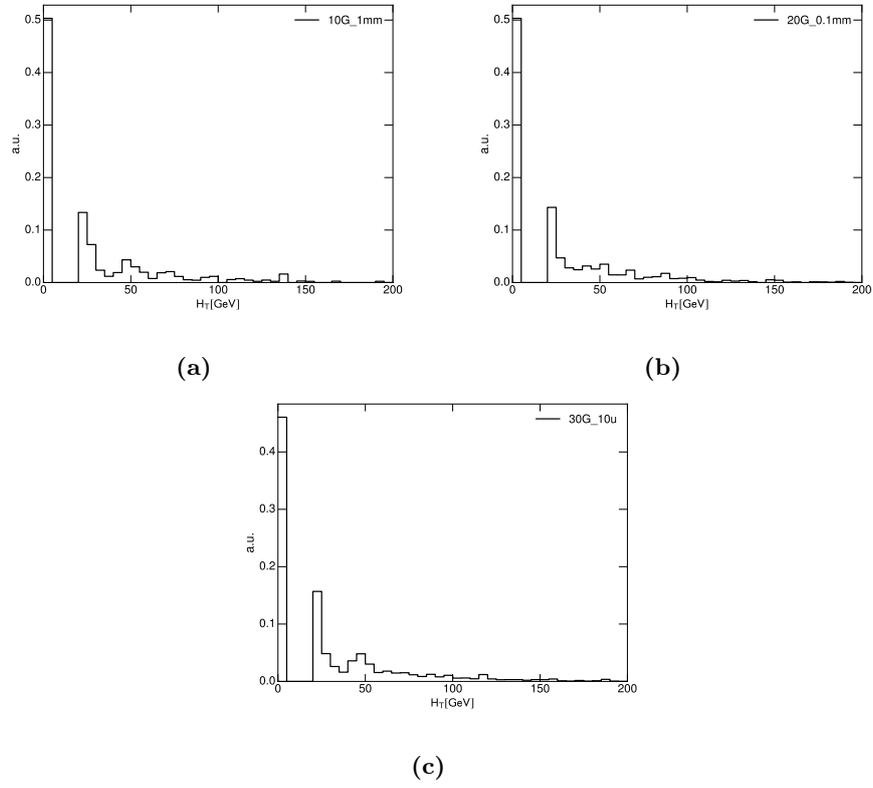


Figure F.4: Signal distributions of H_T for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.

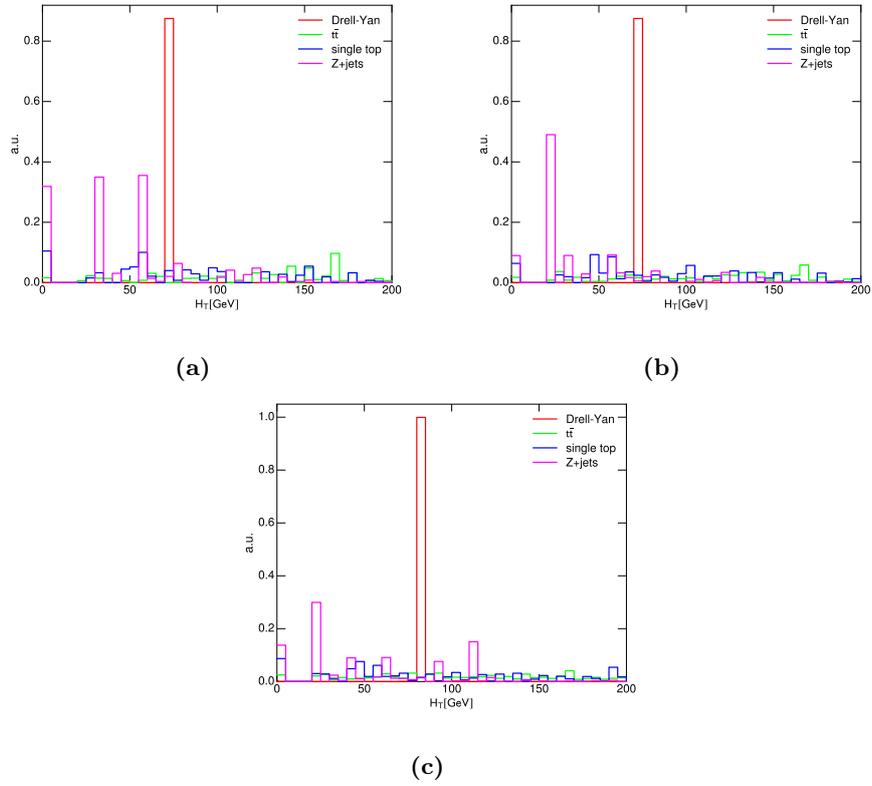


Figure F.5: Background distributions of H_T for Drell-Yan, $t\bar{t}$, single top and $Z + \text{jets}$. The plots are shown for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.

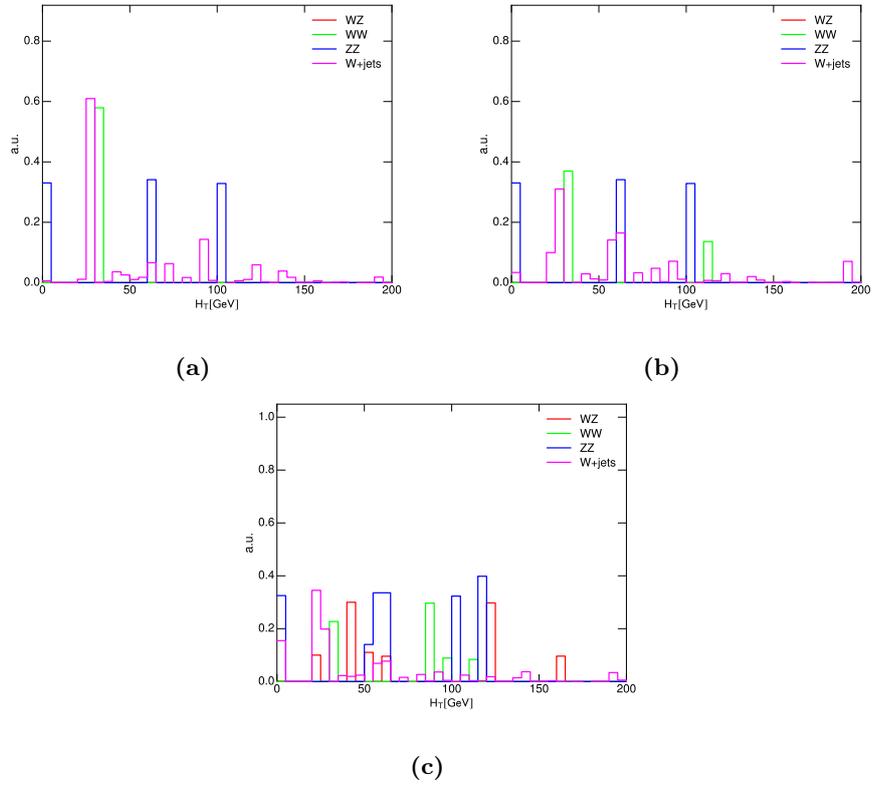


Figure F.6: Background distributions of H_T for $W + \text{jets}$, WW , ZZ and WZ . The plots are shown for the 10 GeV (a), 20 GeV (b) and 30 GeV (c) HNL masses.