ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN FISICA

CICLO XXXV

Settore Concorsuale: 02/A1 - FISICA SPERIMENTALE DELLE INTERAZIONI FONDAMENTALI
Settore Scientifico Disciplinare: FIS/01 - FISICA SPERIMENTALE

Search for heavy neutrinos in D_s decays with the CMS detector at LHC

PRESENTATA DA:
Leonardo Lunerti

COORDINATORE DOTTORATO:
Prof. Michele Cicoli

SUPERVISORE:
Prof. Luigi Guiducci

CO-SUPERVISORE:
Dott. Giovanni Abbiendi

Esame Finale - Anno 2023

Contents

Abstract

This thesis presents a search for a sterile right-handed neutrino N produced in D_s meson decays, using proton-proton collisions collected by the CMS experiment at the LHC. The data set used for the analysis, the B-Parking data set, corresponds to an integrated luminosity of $41.7\,\mathrm{fb}^{-1}$ and was collected during the 2018 data-taking period. The analysis is targeting the $D_s^+ \to N(\to \mu^\pm\pi^\mp)\mu^+$ decays, where the final state muons can have the same electric charge allowing for a lepton flavor violating decay. To separate signal from background, a cut-based analysis is optimized using requirements on the sterile neutrino vertex displacement, muon and pion impact parameter, and impact parameter significance. The expected limit on the active-sterile neutrino mixing matrix parameter $|V_\mu|^2$ is extracted by performing a fit of the $\mu\pi$ invariant mass spectrum for two sterile neutrino mass hypotheses, 1.0 and 1.5 GeV. The analysis is currently blinded, following the internal CMS review process. The expected limit ranges between approximately 10^{-4} for a 1.0 GeV neutrino to 7×10^{-5} for a 1.5 GeV neutrino. This is competitive with the best existing results from collider experiments over the same mass range.

Introduction

Since the discovery by ATLAS and CMS of a new boson compatible with the standard model Higgs boson, the evidences supporting the fact that the masses of elementary particle are generated through the electroweak symmetry breaking mechanism piled up [1, 2]. The electroweak symmetry breaking mechanism predicts the existence of a self-interacting $SU(2)_L$ doublet, the Higgs field, which couples with the gauge fields associated with the $SU(2)_L \times U(1)_Y$ standard model symmetry. When the Higgs field acquires a vacuum expectation value three of the generators are broken resulting in mass terms for the $W^{\pm}$ and Z bosons, while no mass term is generated for the photon. The mass of a fermion is generated through the Yukawa interaction in which the Higgs doublet is coupled with both the right- and the left-handed components of a fermion. The right-handed neutrino, which is a singlet under the standard model gauge group, does not appear in the standard model. As a consequence, there is no Yukawa mass term for neutrinos, which are massless in the standard model.

Nonetheless, neutrino flavor oscillations experiments from astrophysical and terrestrial sources provide striking evidences that neutrinos have small but nonzero masses [3]. Nowadays the observation of neutrino flavor oscillations is the only experimental evidence of physics beyond the standard model. Neutrino flavor oscillations are compatible with an active neutrino mixing structure parameterized by the 3×3 Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix [4–6], with at least two massive neutrinos. An extension of the standard model with three right-handed neutrinos can explain simultaneously dark matter and baryon asymmetry of the universe and is consistent with the experiments on neutrino oscillations [7–10]. In this model, the right-handed neutrinos are singlets under the standard model symmetry group but they could mix with the standard model left-handed neutrinos. For this reason these neutrinos are often referred to as "sterile" neutrinos. This model can explain the smallness of the left-handed neutrino masses via the so-called see-saw mechanism, i.e. when the active-sterile neutrino mixing parameter is small and the mass difference is large.

A search for a heavy neutral lepton can be performed with many different experimental setups. Since the theory does not provide any indication on the sterile neutrino mass, various experiments have probed the existence of sterile neutrinos over a wide range of masses, from the eV to the TeV energy scale [11]. Neutrino flavor oscillations measurements and searches for neutrinoless double-β decay are well-suited for probing sterile neutrino masses below $\sim 1\,\text{MeV}$. The sterile neutrino lifetime is predicted to become larger as the active-sterile neutrino mixing parameter gets smaller, and as the sterile neutrino mass becomes smaller. For this reason, the beam dump facilities are particularly well-suited for searches involving the sterile neutrino production in light, strange or charmed mesons decays. On the other hand, collider experiments are more advantageous for sterile neutrino searches above the GeV scale. Many different experiments have contributed so far with results on sterile neutrino searches, but no evidence has been found.

In the context of heavy neutral lepton searches, the LHC experiments have provided many results, but almost all of them covered the mass spectrum above few GeV. Except for a public result from LHCb investigating the B meson decays [12], there is a lack of sterile neutrino searches involving bottom and charmed meson decays at LHC experiments. The CMS experiment has the possibility to give an important contribution to these searches by exploiting the potential of the so-called B-Parking data set, collected during the 2018 data-taking period, using proton-proton collisions at a center of mass energy of 13 TeV. The B-Parking data set has been collected by triggering on low-p_T non-prompt muons most likely produced in semileptonic b hadron decays. The B-Parking data-taking campaign allowed to collect a sample of 10 billion unbiased b hadron decays with a purity of approximately 70%, and it corresponds to an integrated luminosity of $41.7\,\mathrm{fb}^{-1}$. Although being designed especially for lepton flavor universality tests (e.g. the $R(K)$ measurement), this data set is well-suited also for sterile neutrino searches, since the neutrino is expected to decay with a quite large displacement from the interaction point. While a sterile neutrino search exploiting B decays is ongoing in CMS, the aim of this work is to provide a complementary CMS result using charmed meson decays for the first time.

In the work presented in this thesis the CMS B-Parking data set is used to probe the existence of a sterile right-handed neutrino N looking for the $D_s^+ \to N\mu^+$ decay, where the N subsequently decays into a muon and a pion. This particular D_s meson decay appears to be the most promising for a sterile neutrino search from charmed meson decays, and it allows to probe a mass range below $\sim 2\,\mathrm{GeV}$. This thesis is organized as follows: in Chapter 1, starting from the mass generation mechanism in the standard model, a brief summary of the minimal standard model extension allowing to include a sterile right-handed neutrino is presented, as well as all the equations needed to estimate the sterile neutrino yield from the D_s decay; in Chapter 2 there is an overview of the CMS detector, its trigger system and the computing model which allows for the storage and the analysis of the collected data; Chapter 3 is focused on the CMS event reconstruction; Chapter 4 presents the data analysis for the heavy neutral lepton search by means of D_s decays.

Chapter 1

Standard model and heavy neutral leptons

1.1 The standard model

The standard model (SM) of particle physics is the theory describing all the known elementary particles and their electromagnetic, strong and weak interactions. The fundamental constituents of matter are point-like fermions of spin 1/2. Fermionic particles can be arranged in two main families: quarks and leptons. The most important difference between these two families is that quarks, unlike leptons, can experience also the strong force in addition to weak and electromagnetic forces.

Both quarks and leptons are arranged in three generations. Electron (e) and electron neutrino (ν_e) are the first generation leptons, muon (μ) and muon neutrino (ν_μ) belong to the second generation, tau (τ) and tau neutrino (ν_τ) belong to the third one. Electrons, muons, and tau leptons have a unit electric charge, whereas neutrinos are neutral. For what concerns quarks: up (u) and down (d) are first generation quarks, charm (c) and strange (s) are second generation quarks, top (t) and bottom (b) are third generation quarks. The *up-type* quarks (u, c and t) have $+2/3$ the unit electric charge while the *down-type quarks* (d, s and b) have $-1/3$ the unit electric charge. Quarks and leptons have each their own antiparticle partner as well, having opposite charges under the SM forces.

Fundamental interactions are mediated by point-like bosonic particles. Photons (γ) and gluons (g) are massless spin-1 particles that mediate electromagnetic and strong interactions, respectively. The $W^\pm$ and Z^0 are massive spin-1 bosons and they are the weak force carriers. The W boson has a unit electric charge and mass of approximately 80 GeV, while the Z boson has no electric charge and mass of approximately 90 GeV. The Higgs boson (H) is a spin-0 point-like particle with no electric charge and no color charge, it has a mass of about 125 GeV and has direct couplings only with massive particles.

The standard model is a renormalizable, relativistic quantum field theory in a $3+1$ Minkowski space with a local $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry or *gauge* symmetry, where C stands for color, L for left-handedness, and Y for hypercharge.

The quantum chromodynamics (QCD) sector of the SM Lagrangian, which has a local $SU(3)_C$ symmetry, defines the interaction between quarks and gluons. The Lagrangian describing a free massless quark ψ of flavor q and color a is

$$\mathcal{L}_{free} = \bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu)\psi_{q,a}, \tag{1.1}$$

where the color index a runs from 1 to 3. In order to make this Lagrangian invariant under local $SU(3)_C$ transformations eight gauge fields G^C_μ must be introduced, with $C = 1, 2, ..., 8$, which are representing the gluon field mediating the strong interaction. Substituting the partial derivative ∂_μ with the covariant derivative $D_\mu = \partial_\mu + ig_s t^C G^C_\mu$, we obtain the final gauge-invariant QCD Lagrangian:

$$\boxed{\mathcal{L}_{QCD} = \sum_q \bar{\psi}_{q,a}(i\gamma^\mu \partial_\mu \delta_{ab} - g_s \gamma^\mu t^C_{ab} G^C_\mu - m_q \delta_{ab})\psi_{q,b} - \frac{1}{4} G^A_{\mu\nu} G^{A\mu\nu}}. \tag{1.2}$$

In the above equation t^C are 3×3 matrices which are the generators of the $SU(3)_C$ group, the quantity g_s (or $\alpha_s = \frac{g_s^2}{4\pi}$) is the QCD coupling constant and, finally, the field tensor $G^A_{\mu\nu}$ is given by $G^A_{\mu\nu} = \partial_\mu G^A_\nu - \partial_\nu G^A_\mu - g_a f_{ABC} G^B_\mu G^C_\nu$, where $[t^A, t^B] = i f_{ABC} t^C$ is parameterized using f_{ABC}, i.e. the structure constants of the $SU(3)_C$ group. The first term of Eq. 1.2 describes the propagation of free quarks, the second term describes the quark-gluon interaction, the third term is the mass term of the quarks while the last term describes both the free propagation of gluons and the gluon-gluon interactions.

The electroweak sector of the SM Lagrangian describes the electromagnetic and weak interactions. The electroweak Lagrangian can be obtained by requiring the Lagrangian of a free spinor field to be invariant under local $SU(2)_L \times U(1)_Y$ transformations, where the subscript L indicates that the transformation acts only on left-handed fermions. For a massless fermion ψ, the free Lagrangian can be written as

$$\mathcal{L}_{free} = \bar{\psi}(i\gamma^\mu \partial_\mu)\psi. \tag{1.3}$$

When we require the above Lagrangian to be invariant under local $SU(2)_L \times U(1)_Y$ transformations, four gauge fields must be introduced: $\vec{W}_\mu = (W^1_\mu, W^2_\mu, W^3_\mu)$ for $SU(2)_L$ and B_μ for $U(1)_Y$. The gauge invariant Lagrangian can be obtained by substituting the partial derivative ∂_μ with the covariant derivative

$$D_\mu = \partial_\mu + \frac{ig'}{2} Y B_\mu + \frac{ig}{2} \sigma_a W^a_\mu, \tag{1.4}$$

and it can be written as

$$\boxed{\mathcal{L}_{EW} = \bar{\psi} i\gamma^\mu D_\mu \psi - \frac{1}{4} W^{\mu\nu} W_{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu}}, \tag{1.5}$$

where g and g' are the $SU(2)_L$ and $U(1)_Y$ gauge couplings, Y is the hypercharge and σ^a, with $a =$ 1,2,3, are the Pauli matrices. The $B_{\mu\nu}$ and $W_{\mu\nu}$ tensors are defined in the same way as the $G_{\mu\nu}$ tensor using the appropriate $U(1)_Y$ and $SU(2)_L$ generators. The first term in the 1.5 contains the free propagation of the fermions as well as their interaction with the photon and the weak bosons which are expressed as a linear combination of B_μ and $\vec{W}_\mu$ fields:

$$\begin{cases} W^\pm_\mu = \frac{W^1_\mu \pm i W^2_\mu}{\sqrt{2}} & \text{for } W^\pm \text{ bosons} \\ A_\mu = \cos\theta_W B_\mu + \sin\theta_W W^3_\mu & \text{for the photon} \\ Z_\mu = -\sin\theta_W B_\mu + \cos\theta_W W^3_\mu & \text{for the } Z^0 \text{ boson} \end{cases}, \tag{1.6}$$

where θ_W is the Weinberg angle. The last two terms of the electroweak Lagrangian describe the free propagation of W and Z gauge bosons and their self-interactions, and the free propagation of photons. Mass terms for fermions and bosons are forbidden in 1.5 since they are not invariant under local $SU(2)_L$ transformations.

1.2 Spontaneous symmetry breaking

In the SM, the W and Z gauge bosons masses are generated by the so-called electroweak symmetry breaking (EWSB) mechanism [13–16]. As a consequence the weak interaction also becomes short-ranged. The idea of spontaneous symmetry breaking is graphically represented in Fig. 1.1 which shows the Goldstone model adopted by the SM. The spontaneous symmetry breaking mechanism can be introduced by including the following renormalizable Higgs potential in the SM

$$V(\Phi) = m^2\Phi^\dagger\Phi + \lambda(\Phi^\dagger\Phi)^2\,, \tag{1.7}$$

with the Higgs field Φ being a $SU(2)_L$ doublet of complex fields with weak hypercharge $Y = 1$. For $\lambda > 0$ and $m^2 < 0$ the potential takes the "Mexican hat" shape, see Fig. 1.1. This $SU(2)_L$ doublet can be written as a linear combination of four real fields [17]:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \tag{1.8}$$

where $\phi^+ \equiv (\varphi_1 + i\varphi_2)/\sqrt{2}$ and $\phi^0 \equiv (\varphi_3 + i\varphi_4)/\sqrt{2}$. Assuming a non-zero vacuum expectation value (VEV) of the Higgs field:

$$\langle\Phi\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{1.9}$$

where $v^2 \equiv -m^2/\lambda$, we induce the spontaneous symmetry breaking of the SM gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ into $SU(3)_C \times U(1)_{EM}$.

Three of the four generators of the $SU(2)_L \times U(1)_Y$ group are broken, resulting in three massless Goldstone bosons and a massive Higgs field. The three Goldstone bosons disappear when choosing the right gauge transformation. The Higgs field couples to the $\vec{W}_\mu$ and B_μ gauge fields associated with the $SU(2)_L \times U(1)_Y$ local symmetry through the covariant derivative (Eq. 1.4) appearing in the kinetic term of the Higgs Lagrangian

$$\boxed{\mathcal{L}_{Higgs} = (D_\mu\Phi)^\dagger(D^\mu\Phi) - V(\Phi)}\,. \tag{1.10}$$

The gauge boson mass terms come from the first term in Eq. 1.10, evaluated at the VEV of the scalar field 1.9. The relevant terms are [18]:

$$\Delta\mathcal{L} = \frac{1}{2}\frac{v^2}{4}\left[g^2W^{1\mu}W^1_\mu + g^2W^{2\mu}W^2_\mu + (-gW^{3\mu} + g'B^\mu)(-gW^3_\mu + g'B_\mu)\right]\,. \tag{1.11}$$

The $W^\pm_\mu$ and Z_μ bosons acquire masses $m_W = gv/2$ and $m_Z = v\sqrt{g^2 + g'^2}/2$, while the photon field A_μ remains massless.

Fermions mass arises from the Yukawa interactions which couple a right-handed fermion (here denoted as u_R, d_R and e_R respectively for up-type quarks, down-type quarks, and leptons) with its left-handed doublet (here denoted as q_L and l_L for quarks and leptons) and the Higgs field:

$$\mathcal{L}_{\text{Yukawa}} = -\hat{h}^d_{ij}\bar{q}_{L_i}\Phi d_{R_j} - \hat{h}^u_{ij}\bar{q}_{L_i}\tilde{\Phi}u_{R_j} - \hat{h}^l_{ij}\bar{l}_{L_i}\Phi e_{R_j} + \text{h.c.}, \tag{1.12}$$

where $\tilde{\Phi} = i\sigma_2\Phi^*$. In the above Lagrangian, each term is parameterized by a 3×3 matrix, $\hat{h}^f$, in the family space, where the superscript $f = u, d, l$ refers to up-type quarks, down-type quarks, and leptons. After the Higgs field acquires a VEV, and

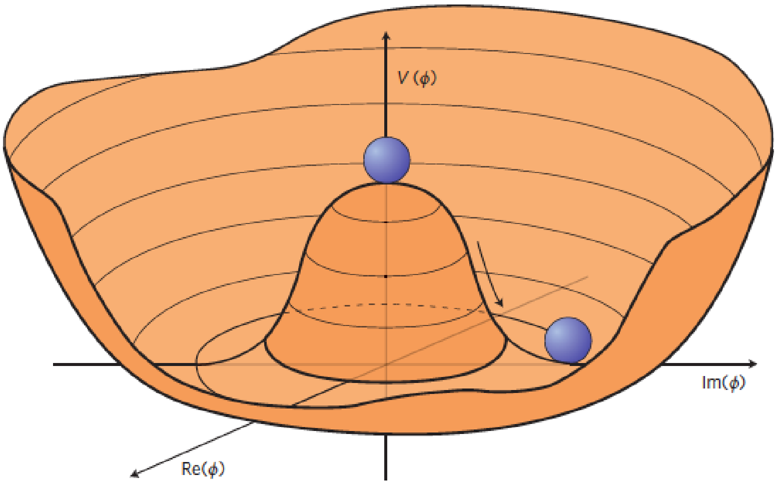

Figure 1.1: Graphical representation of the Higgs potential according to the Goldstone model.

after the rotation to the fermion mass eigenstate basis that also diagonalizes the Higgs-fermion interactions, $\hat{h}^f_{ij} \to y^f_i \delta_{ij}$, all fermions f, except the neutrinos, acquire a mass that is proportional to the Yukawa coupling y^f_i,

$$m^f_i = y^f_i \frac{v}{\sqrt{2}}. \tag{1.13}$$

In the SM, neutrinos have only the left-handed helicity state, thus there cannot be any Yukawa term yielding a Dirac mass term for the neutrino, unlike the charged fermions. Therefore, neutrinos are massless in the SM. However, the observation of neutrino oscillations in solar, atmospheric, reactor and accelerator neutrino data demonstrates that at least two of the three active neutrinos have a non-zero mass and that individual lepton flavor is violated [3]. In principle, the mass term for neutrinos could be added in an analogous manner to the up-type quarks if right-handed neutrinos (ν_R) are added to the SM particles by a term

$$\Delta\mathcal{L} = -\hat{h}^l_{ij} \bar{l}_{L_i} \tilde{\Phi} \nu_{R_j} \tag{1.14}$$

to Eq. 1.12. Nevertheless, to get sub-eV left-handed neutrino masses as required by the neutrino oscillations data, the Yukawa couplings must be extremely small, i.e. $\lesssim 10^{-12}$. There is no theoretical justification for the large disparity between such small neutrino Yukawa couplings and other SM Yukawa coupling [11]. As a consequence, no terms like 1.14 exist in the Yukawa interaction term of the SM Lagrangian, leaving the neutrino massless: beyond standard model physics might be responsible for the observed smallness of neutrino masses.

1.3 Neutrinos in the standard model: massless neutrinos

In the SM, neutrinos are fermions that do not have strong nor electromagnetic interactions. Consequently, they are singlets of the subgroup $SU(3)_C \times U(1)_{EM}$. Only left-handed neutrinos exist in the SM: the interaction with their corresponding charged lepton has the form of a weak charged current (CC), being mediated by the

$W^{\pm}$ boson, while the interaction among themselves has the form of a neutral current (NC), being mediated by the Z boson. These interactions are described respectively by

$$-\mathcal{L}_{CC} = \frac{g}{\sqrt{2}} \sum_l \bar{\nu}_{lL} \gamma^\mu l_L^- W_\mu^\pm + \text{h.c.}\,, \text{ and} \tag{1.15}$$

$$-\mathcal{L}_{NC} = \frac{g}{\sqrt{2}\cos\theta_W} \sum_l \bar{\nu}_{lL} \gamma^\mu \nu_{lL} Z_\mu\,. \tag{1.16}$$

The left-handed neutrinos participating to the above interactions are said to be *active* neutrinos. Equation 1.16 determines the width of the decay of the Z boson into left-handed neutrinos with a mass smaller than half the mass of the Z. From the measurement of the total decay width of the Z boson, the number of light neutrinos has been found to be compatible with 3. As a consequence, any extension of the SM should contain only three light active neutrinos. On the other hand, *sterile* neutrinos are defined as having no SM gauge interactions, that is, they are singlets of the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group.

An important feature of the SM, which is relevant to the question of the neutrino mass, is the fact that the SM presents an accidental global symmetry, which is not imposed but appears as a consequence of the gauge symmetry and the representation of the matter fields:

$$G_{\text{SM}}^{\text{global}} = U(1)_B \times U(1)_{L_e} \times U(1)_{L_\mu} \times U(1)_{L_\tau}\,, \tag{1.17}$$

where $U(1)_B$ and $U(1)_{L_e, L_\mu, L_\tau}$ are the baryon number symmetry and the three lepton flavor symmetries. The lepton number L_i is defined as the quantum number carried by leptons of the i-th generation and is +1 for particles and -1 for antiparticles. The baryon number is defined by means of the difference of the number of quarks and antiquarks within a baryon as $(n_q - n_{\bar{q}})/3$. The total lepton number $L = L_e + L_\mu + L_\tau$ is an accidental symmetry as well, since it is a subgroup of $G_{\text{SM}}^{\text{global}}$.

As described in Sec. 1.2, the term corresponding to the neutrino mass could be included in the SM Lagrangian by adding a Yukawa interaction term that involves a right-handed neutrino. Since the SM does not contain right-handed neutrinos, no such Yukawa interaction can be built for the neutrinos, which are consequently massless.

In principle a term for the neutrino mass could arise from loop corrections. The only neutrino mass term that can be constructed with the SM field is the bilinear $\bar{l}_L l_L^C$, where the superscript C indicates charge conjugation. Such a mass term would not be allowed in the SM since it would violate the total lepton symmetry by two units. Furthermore, the $U(1)_{B-L}$ subgroup of $G_{\text{SM}}^{\text{global}}$ is non-anomalous, meaning that the bilinear $\bar{l}_L l_L^C$ cannot be induced by nonperturbative corrections either since it breaks $B - L$ [3].

Therefore, neutrinos are exactly massless in the SM. This means that to add a mass to the neutrino, one must go beyond the SM.

1.4 Introducing massive neutrinos in the standard model

From the previous section, it appears that it is not possible to construct a renormalizable mass term for neutrinos having the fermionic content and the gauge symmetry of the SM. In order to introduce a neutrino mass in the theory, one must either extend the particle contents of the model or depart from gauge invariance and/or renormalizability, or do both.

If an arbitrary number n of sterile neutrinos ν_{si} $(i = 1, \dots n)$ is introduced in the model, one can construct two gauge invariant renormalizable operators leading to two mass terms. This would mean adding the following term to the Lagrangian:

$$-\mathcal{L}_{M_\nu} = M^D_{ij}\bar{\nu}_{si}\nu_{Lj} + \frac{1}{2}M^N_{ij}\bar{\nu}_{si}\nu^c_{sj} + \text{h.c.}\,. \tag{1.18}$$

In the above equation, M^D is a complex $n \times 3$ matrix while M^N is a symmetric $n \times n$ matrix. The first term in Eq. 1.18 arise from the following Yukawa interaction term

$$y_{ij}\bar{\nu}_{si}\left(l_{Li} \cdot \tilde{\Phi}\right)\,. \tag{1.19}$$

After the spontaneous electroweak symmetry breaking, we can define the M^D matrix as

$$M^D_{ij} = y_{ij}\frac{v}{\sqrt{2}}, \tag{1.20}$$

similarly to Eq. 1.12 for charged fermion masses. The first term of 1.18 will be referred to as Dirac mass term. This term allows for the conservation of the total lepton number L but it can break either L_e, L_μ or L_τ individually.

The second term in Eq. 1.18 will be referred to as Majorana mass term. This term is a singlet of the SM gauge group, therefore, it can appear as a bare mass term in the Lagrangian. Moreover, since it involves two neutrino fields it breaks lepton number by two units. Moreover, this term is allowed in the Lagrangian only if the neutrinos do not carry any additive conserved charge.

Equation 1.18 can be written in a compact form as:

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\left(\bar{\nu}^c_L, \bar{\nu}_s\right)\begin{pmatrix} 0 & (M^D)^T \\ M^D & M^N \end{pmatrix}\begin{pmatrix} \nu_L \\ \nu^c_s \end{pmatrix} + \text{h.c.} \equiv \bar{\nu}^c M_\nu \nu + \text{h.c.}\,, \tag{1.21}$$

where $\nu = (\bar{\nu}^c_L, \bar{\nu}_s)$ is a vector of dimension $3+n$, and M_ν is a $(3+n)\times(3+n)$ complex symmetric matrix. Therefore, M_ν can be diagonalized by an unitary $(3+n)\times(3+n)$ matrix V^ν:

$$\left(V^\nu\right)^T M_\nu V^\nu = \text{diag}(m_1, m_2, \dots, m_{3+n})\,, \tag{1.22}$$

where $m_1, m_2, \dots, m_{3+n}$ represent the masses of the $3+n$ mass eigenstates, which can be expressed as:

$$\nu_{\text{mass}} = \left(V^\nu\right)^\dagger \nu\,. \tag{1.23}$$

Equation 1.18 can be rewritten in the mass eigenstate basis as

$$-\mathcal{L}_{M_\nu} = \frac{1}{2}\sum_{k=1}^{3+n} m_k\left(\bar{\nu}^c_{\text{mass},k}\nu_{\text{mass},k} + \bar{\nu}_{\text{mass},k}\nu^c_{\text{mass},k}\right) = \frac{1}{2}\sum_{k=1}^{3+n} m_k\bar{\nu}_{Mk}\nu_{Mk}\,, \tag{1.24}$$

where the state $\nu_{Mk} = \nu^c_{\text{mass},k}\nu_{\text{mass},k}$. Being its own charge-conjugate state, ν_{Mk} is a Majorana particle.

In the so-called "see-saw" limit, i.e. $M^D \ll M^N$ at the level of eigenvalues, the full a $(3+n)\times(3+n)$ M_ν matrix for ν_L and ν_s has two distinct sets of eigenvalues: n eigenvalues M_I are of the order of the eigenvalues of M^N, while the remaining three eigenvalues m_i are suppressed by a power of two of the *active-sterile mixing matrix*

$$U = M^D(M^N)^{-1}\,. \tag{1.25}$$

The see-saw hierarchy $U \ll 1$ separates two distinct sets of mass eigenstates. Three mass states have light masses $m_i \sim \mathcal{O}\left[U^2 M^N\right]$ and are mainly mixings of the $SU(2)_L$ charged fields ν_{Li}. The remaining n mass states have a mass $M_I \sim \mathcal{O}\left[M^N\right]$ and are mainly mixing of the singlet fields ν_s.

Among the physics phenomena that are currently not described by the standard model there are the nature of the dark matter (DM), the baryon asymmetry of the Universe, and the smallness of the neutrino masses. The so-called neutrino minimal standard model (νMSM) could explain the smallness of the neutrino masses through the see-saw mechanism while also providing a DM candidate and giving a possible explanation for the baryon asymmetry.

The νMSM involve three right-handed heavy sterile neutrinos [7–10]: the lightest neutrino, N_1, can explain the DM in the Universe, while the heavier neutrinos, N_2 and N_3, can be responsible for the baryon asymmetry through leptogenesis [10, 19–23] or neutrino oscillation [9, 10]. The production cross section, decay width and lifetime of the heavy neutrinos depend on $|U|^2$ and on the mass of the neutrino. In the νMSM, N_1 is expected to be too light and long-lived to produce an unambiguous signal at collider experiments, but N_2 and N_3 could decay to Wl, $Z\nu$, or $H\nu$ and are therefore potentially detectable.

When sterile neutrinos are not degenerate in mass, in all processes they are produced and decay independently, without oscillations between themselves, unlike active neutrinos [24–26]. Therefore, from the phenomenological point of view, it is enough to describe only one sterile neutrino. In the scenario of a single heavy neutral lepton (HNL) or sterile neutrino N [27], the mass eigenstates of active-plus-sterile sector are the mixture of ν_L and N, having small mixing angles and large splitting between mass scale of sterile and active neutrinos. The active-sterile mixing matrix reduces to $U_\alpha = M^D_\alpha \left(M^N\right)^{-1}$, where $\alpha = e, \mu, \tau$. As a result of mixing, HNL couples to the SM fields in the same way as active neutrinos do

$$\mathcal{L}_{\text{int}} = \frac{g}{2\sqrt{2}} W^+_\mu \bar{N}^c \sum_\alpha U^*_\alpha \gamma^\mu (1-\gamma_5) l^-_{L\alpha} + \frac{g}{4\cos\theta_W} Z_\mu \bar{N}^c \sum_\alpha U^*_\alpha \gamma^\mu (1-\gamma_5)\nu_{L\alpha} + \text{h.c.} \tag{1.26}$$

Therefore, to describe a single sterile neutrino only 4 parameters are needed: the mass m_{HNL} and three mixing angles U_α with active neutrinos.

1.5 Status of the heavy neutral lepton searches

Many experimental searches have put stringent constraints on sterile neutrino mixing with active ones in a range of HNL mass from few eV to TeV scale. A complete review of such limits can be found in Ref. [11].

Heavy neutrinos in the sub-eV mass range have been excluded by neutrino-oscillations experiments [28–31]. In the mass range between $\sim 10\,\text{eV}$ and $1\,\text{MeV}$ the mixing of the sterile neutrino with electron neutrino has been constrained in neutrino-less double-β decays and precision measurement of the β-decay energy spectra. For the $\sim$ 1 MeV to $\sim$ 1 GeV mass range, both peak searches at collider experiments and beam dump facilities are able to constraint the sterile neutrino mixing with both electrons and muons targeting the decays of π and K. For masses well above 1 $\sim$ GeV up to 1 $\sim$ TeV only collider experiments are able to probe the active-sterile neutrino mixing matrix. Upper limits on the active-sterile neutrino mixing elements have also been derived from cosmological bounds on sterile neutrino lifetimes as required for the success of the big bang nucleosynthesis [32–34]. A summary on present and future searches for sterile neutrinos is summarized in Fig. 1.2, which shows as a shaded area the present constraints on $|U_e|^2$ (top left), $|U_\mu|^2$ (top right) and $|U_\tau|^2$ (bottom) in the 0.1-1000 GeV sterile neutrino mass, as well as the potential constraints from future planned experiments as solid lines.

In the $\sim$ 0.1-1 GeV mass range, the most stringent limits come from the measurement of the energy spectrum of electrons or muons coming from the leptonic decays of a light meson, typically a pion or a kaon. The mass spectrum can be extended to larger masses by probing the decay of heavier mesons, e.g. the Belle experiment used the $B \to XlN$ decay to put limits on $|U_e|^2$ and $|U_\mu|^2$ [35]. Searches for $\tau \to XN$ decays are promising for constraining $|U_\tau|^2$ in the mass region below the tau lepton mass: the solid (dashed) red line in Fig. 1.2 show the potential constraint on $|U_\tau|^2$ using the $\tau^- \to N\pi^-\pi^+\pi^-$ in a conservative (optimistic) scenario at B-Factories [36].

Limits on sterile neutrinos can be put also searching for their decay products. A sterile neutrino could be produced in a semileptonic meson decay and then subsequently decay into visible final states such as charged leptons, pions and kaons. These visible products can be searched for in beam dump experiments by placing the detector some distance away from the production site.

The existence of heavy Majorana neutrinos can be probed also by looking at lepton number violation decays where a hadron $X_1^\pm$ decays to two same sign leptons, e.g. $X_1^\pm \to l^\pm N$ and $N \to l^\pm X_2^\mp$. For small neutrino masses the $K^+ \to l^+l^+\pi^-$ decay is the most promising channel [37], while at heavier masses the LHCb collaboration put a limit on $|U_\mu|^2$ using the $B^- \to \pi^+\mu^-\mu^-$ decay [12].

Heavy neutrinos with masses of the order of electroweak scale can be directly produced on-shell at colliders. Heavy neutrino can be produced e.g. in $e^+e^- \to N\nu_l$ at e^+e^- colliders or in $q\bar{q}' \to Nl^+$ at hadron colliders. The search can be performed by targeting $N \to lW$, $N \to \nu_l Z$ or $N \to \nu_l H$ decays. Recent searches by CMS [38] and ATLAS [39] allowed to set the most stringent limits for sterile neutrinos masses up to $\sim$ 100 GeV.

For Majorana neutrinos having a mass lighter than the Z boson mass, the $Z \to \nu_l N$ decay has been used to put a limit on the active-sterile neutrino mixing parameters by L3 [40] and DELPHI [41] collaborations from a reanalysis of LEP data. Such searches are expected to probe $|U_\alpha|^2$ value down to $\sim 10^{-12}$ at future high luminosity Z-factory and covering a large phase space for heavy neutrino masses in the 10-80 GeV range [42, 43], labeled in Fig. 1.2 as FCC-ee (electron-positron future circular collider [44]).

At LHC experiments, compared to searches for HNL in the $W^\pm/Z$ channels, results from hadron decays are only few ones. Except of the $B^- \to \pi^+\mu^-\mu^-$ search by LHCb [12], no public results are available on on-shell production of heavy neutral leptons at LHC from hadron decays. While HNL searches targeting B meson decays are ongoing at CMS, the present work aims at investigating the existence of sterile neutrinos from the D_s decay.

1.6 Heavy neutral leptons from meson decays

The interactions between HNLs and SM particles are defined in Eq. 1.26. For the purpose of this work, this section is focusing on the production of heavy neutral leptons from charm mesons decays.

Heavy neutral leptons can be produced in meson decays via 2-body purely leptonic decays (left diagram of Fig. 1.3) or semileptonic decays (right diagram of Fig.

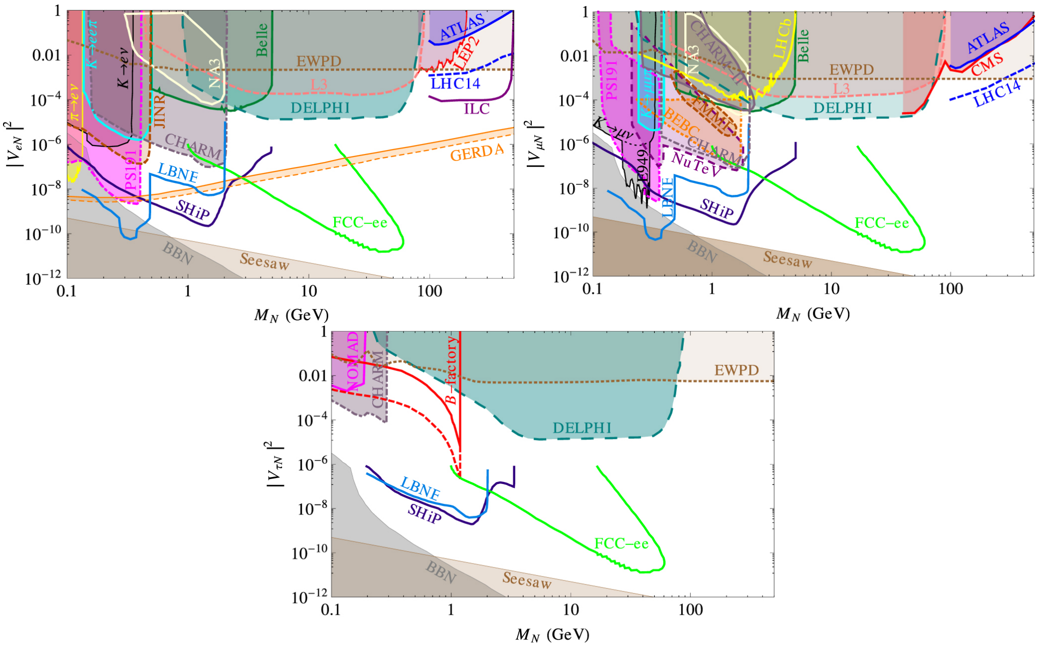

Figure 1.2: Existing limits on HNL masses and their couplings to electrons (top left), muons (top right), tau leptons (bottom), and projected sensitivity achievable in planned future facilities [11].

1.3). The most relevant charmed mesons for the HNL productions are: D^0, D^+ and D_s. The branching ratios for these charmed mesons decaying into a heavy neutral lepton N are shown in Fig. 1.4, where the sterile neutrino is considered to mix only with the electron neutrino. However, one can expect the same relative branching fraction proportions between different decay channels when the final state lepton is a muon instead of an electron.

The D^0 meson is neutral, thus, at tree level, it can only decay through charged current interaction and will involve a meson in the final state. Since the largest branching ratio of D^0 decays is to K meson a search for HNLs targeting D^0 decays can explore a mass range $M_N < M_D - M_K \simeq 1.4\,\text{GeV}$. The same argument can be applied to charmed baryons: they should decay into baryons, and the most probable is a decay into strange baryons, limiting the mass range to $M_N < M_{\Lambda_c} - M_{\lambda_0} \simeq 1.4\,\text{GeV}$. Therefore, these channels can explore HNL masses below $\sim 1.4\,\text{GeV}$.

The charged charmed mesons $D^{\pm}$ and $D_s^{\pm}$ could decay into a HNL and a charged lepton. These decays could produce HNLs with a mass relatively close to the meson itself. The branching ratio of the decay of a D_s meson to a HNL and a lepton is Cabibbo-favored, as a result is a factor 10 larger than any similar decay of other $D^{\pm}$ mesons. The number of D_s mesons is expected to be suppressed with respect to $D^{\pm}$ and D^0 mesons, but only by a small factor. For example, the production fractions of the charmed mesons are given by $f(D^+) = 0.204$, $f(D^0) = 0.622$ and $f(D_s) = 0.104$ at $\sqrt{s} = 400\,\text{GeV}$ [45], while at LHC energy ($\sqrt{s} = 5.02\,\text{TeV}$) we have $f(D^+) = 0.173$, $f(D^0) = 0.622$ and and $f(D_s) = 0.073$ [46]. As a result, the two-body decays of D_s mesons are expected to dominate the HNL production from charmed meson decays, and, in addition, they allow to explore a wider HNL mass range with respect to semileptonic decays.

The D_s two-body decay to a HNL represents the focus of this work, in particular the $D_s^+ \to N\mu^+$ decay followed by the $N \to \mu^+\pi^-$ decay, which has a clear experimental signature and allow to explore a wide HNL mass range. The goal of the present section and the next one is to give a theoretical description of the D_s

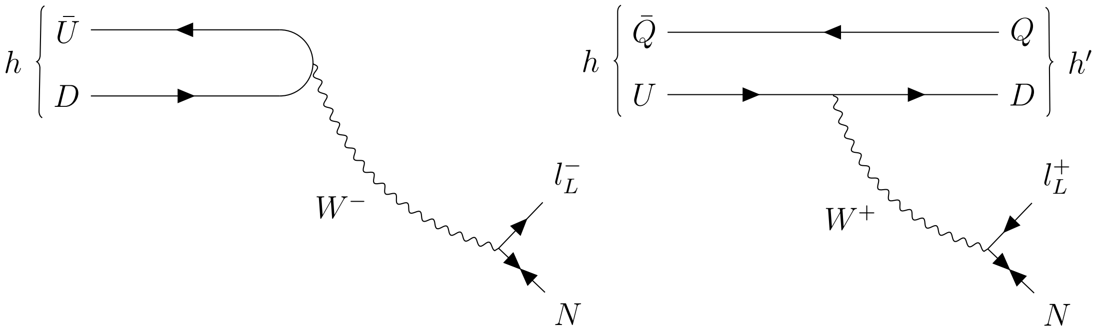

Figure 1.3: Left: the diagram of the leptonic decay of a meson h to a lepton l_L^- and a heavy neutral lepton N. Right: the diagram of the semileptonic decay of a meson h to a meson h', a lepton l_L^+ and a heavy neutral lepton N. Up- and down-type quarks are denoted with U and D, while Q indicates a generic quark.

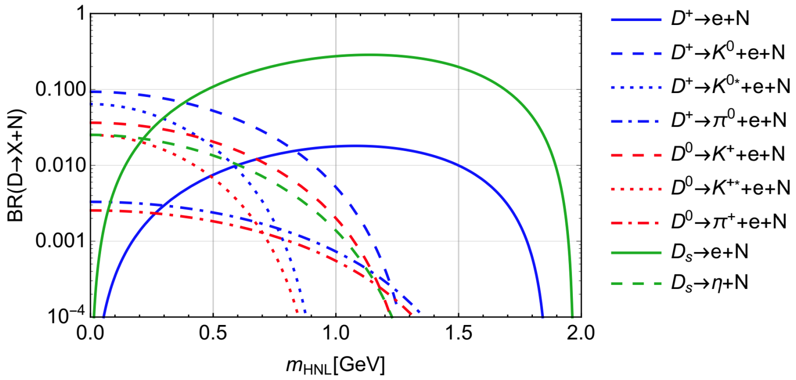

Figure 1.4: Dominant branching ratios of HNL production from different charmed mesons in the scenario where $U_e = 1$, $U_\mu = U_\tau = 0$ [27]. For charged mesons two-body leptonic decays are shown, while for the neutral mesons decays are necessarily semi-leptonic.

decay rate to a heavy neutral lepton, that is crucial for the HNL yield estimate

Taking into account the leptonic decay of a pseudoscalar hadron h of mass m_h into a lepton l_α, where $\alpha = e, \mu, \tau$, of mass m_l and a heavy neutrino N, the total decay width can be calculated as:

$$\Gamma\left(h \to l_\alpha N\right) = \frac{G_F^2 f_h^2 m_h^3}{8\pi} |V_{UD}|^2 |U_\alpha|^2 \left[y_N^2 + y_l^2 - (y_N^2 - y_l^2)^2\right] \sqrt{\lambda(1, y_N^2, y_l^2)}\,, \tag{1.27}$$

where G_F^2 is the Fermi constant, f_h is the meson form factor, V is the Cabibbo-Kobayashi-Maskawa (CKM) matrix, $y_l = m_l/m_h$ and $y_N = m_N/m_h$. The function λ is the Källén function [47]:

$$\lambda(a, b, c) = a^2 + b^2 + c^2 - 2ab - 2ac - 2bc\,. \tag{1.28}$$

The process $h \to h' l N$, where h' can be either a pseudoscalar or a vector meson, has not been taken into account in this work, thus the decay width for such processes is not being reported here. The interested reader can found a complete review in Ref. [27].

1.7 Heavy neutral lepton decay modes

This section is meant to give an overview of the heavy neutrino decay rate calculation with a focus on the case of interest of this study: the $D_s^+ \to N\mu^+$ decay followed by the $N \to \mu^+\pi^-$ decay. The heavy neutral lepton is expected to undergo a weak decay. The two basic processes contributing to all decays are presented in Fig. 1.5. For the charged current-mediated decay, the final state particles U and D are part of an $SU(2)_L$ doublet and can be either a lepton pair (ν_l, l) or a pair of up-type and down-type quarks (u, d). For the neutral current-mediated decay f can be any fermion.

The total decay width for the heavy neutral lepton can be computed as the sum of all possible diagrams. The general formula for the charged current-mediated processes $N \to l_i^- \nu_j l_j^+$, $i \neq j$ (the $i = j$ case is treated separately since it interfere with the neutral current diagram), and $N \to l_i u_j \bar{d}_j$ is [27]:

$$\Gamma(N \to l_i^- U\bar{D}) = N_W \frac{G_F^2 m_N^5}{192\pi^3} |U_i|^2 I(x_u, x_d, x_l)\,, \tag{1.29}$$

where (U, D) can be either an up-type-down-type quark doublet or a (ν_l, l) doublet of $SU(2)_L$, N_W is 1 in case of leptons and $3|V_{UD}|^2\,(1 + \Delta_{\mathrm{QCD}})$ in case of quarks, and $|V_{UD}|$ is the CKM matrix element. The $I(x_u, x_d, x_l)$ function is parameterized using $x_l = \frac{m_{l_i}}{M_N}$, $x_u = \frac{m_U}{M_N}$ and $x_d = \frac{m_D}{M_N}$, where M_N is the heavy neutrino mass. The contribution from bottom and top quarks is not taken into account since it is not kinematically accessible. The $(1 + \Delta_{\mathrm{QCD}})$ term corresponds to the QCD loop corrections to the tree-level decay into quarks and it has been estimated in case of τ lepton hadronic decays up to three loops [48]:

$$\Delta_{\mathrm{QCD}} = \frac{\alpha_s}{\pi} + 5.2\frac{\alpha_s^2}{\pi^2} + 26.4\frac{\alpha_s^3}{\pi^3}\,, \tag{1.30}$$

where the strong coupling constant α_s has been evaluated at the mass of the tau lepton. The $I(x_u, x_d, x_l)$ function describes the corrections due to the finite masses of final-state fermions and it is given by

$$I(x_u, x_d, x_l) = 12\int_{(x_d+x_l)^2}^{(1-x_u)^2} \frac{dx}{x}(x - x_l^2 - x_d^2)(1 + x_u^2 - x\sqrt{\lambda(x, x_l^2, x_d^2)\lambda(1, x, x_u^2)})\,, \tag{1.31}$$

where $\lambda(a, b, c)$ is the Källén function (Eq. 1.28). The $I(a, b, c)$ function has the following properties:

1. $I(0, 0, 0) = 1$;
2. The function $I(a, b, c)$ is symmetric under any permutation of its arguments;
3. In the case of mass hierarchy $m_a, m_b \ll m_c$ (where a, b, c are leptons and/or quarks in some order) one can use the following approximation

   $$I(x, 0, 0) = (1 - 8x^2 + 8x^6 - x^8 - 12x^4 \log x^2)\,, \tag{1.32}$$

   where $x = m_c/M_N$.

The neutral current decay rate to a fermion f is different depending on its electric charge Q_f and on whether it is left- or right-handed, and it is expressed by the following factors

$$C_{fL} = \pm\frac{1}{2} - Q_f \sin^2\theta_W \quad \text{and} \quad C_{fR} = -Q_f \sin^2\theta_W\,, \tag{1.33}$$

where the + sign applies to up-type quarks and neutrinos, and the − sign applies to down-type quarks and charged leptons. For more compact expressions, let

$$S_f^2 = C_{fL}^2 + C_{fR}^2 = \frac{1}{4} - |Q_f| \sin^2\theta_W + 2Q_f^2 \sin^4\theta_W \,. \tag{1.34}$$

In the case where $f \neq \nu_i$, the neutral-current-mediated width is

$$\Gamma(N \to \bar{\nu}_i f \bar{f}) = \frac{G_F^2 m_N^5}{192\pi^3} |U_i|^2 S_f^2 \,, \tag{1.35}$$

and the decay rate for $N \to \nu_i \nu_j \bar{\nu}_j$ is

$$\Gamma(N \to \nu_i \nu_j \bar{\nu}_j) = (1 + \delta_{ij}) \frac{G_F^2 m_N^5}{768\pi^3} |U_i|^2 \,, \tag{1.36}$$

where δ_{ij} is the Kroneker delta.

Finally, for the charged lepton decay with $i = j$:

$$\Gamma(N \to l_i^- l_i^+ \nu_i) = \frac{G_F^2 m_N^5}{192\pi^3} |U_i|^2 \left(\frac{1}{4} + \sin^2\theta_W + 2\sin^4\theta_W \right) . \tag{1.37}$$

Following the approach in Ref. [49], the total decay width will be considered in the approximation of massless first and second generation quarks, and ignoring the CKM mixing. Moreover, the case of interest where $U_\mu = 1$ and $U_e = U_\tau = 0$ is being considered.

The contribution of $\nu_e e\mu$, $\nu_\mu ee$ and $\nu_\mu e\mu$ decays to the total decay width are summed separately in the following expression:

$$\Gamma(N \to ll\nu) = \frac{G_F^2 m_N^5}{96\pi^3} |U_\mu|^2 \times (1 + 0.13 + 0.59) \,. \tag{1.38}$$

The sum of the decay widths of heavy neutral lepton decay to quarks gives

$$\Gamma(N \to \text{quarks}) = \frac{G_F^2 m_N^5}{96\pi^3} |U_\mu|^2 \times (8.24) \,, \tag{1.39}$$

and the total decay width to three neutrinos is

$$\Gamma(N \to \nu\nu\nu) = \frac{G_F^2 m_N^5}{96\pi^3} |U_\mu|^2 \,. \tag{1.40}$$

The sum of all the aforementioned contributions gives the heavy neutral lepton total width:

$$\Gamma_N = \frac{G_F^2 m_N^5}{96\pi^3} |U_\mu|^2 \times (10.95) \,. \tag{1.41}$$

In addition to Eq. 1.41, to calculate the branching ratio of the HNL decay to $\mu\pi$, the decay width the $N \to \mu^+\pi^-$ process must be computed. This can be done using the sterile neutrino decay width to a charged pseudoscalar meson h, given by [27]:

$$\Gamma(N \to l_i^- h^+) = \frac{G_F^2 f_h^2 |V_{UD}|^2 |U_i|^2 m_N^3}{16\pi} \left[(1 - x_l^2 - x_h^2(1 + x_l^2))\right] \sqrt{\lambda(1, x_h^2, x_l^2)} \,, \tag{1.42}$$

where f_h is the final state meson form factor, x_l and x_h are defined as in 1.29, and $\lambda(a, b, c)$ is the Källén function (Eq. 1.28).

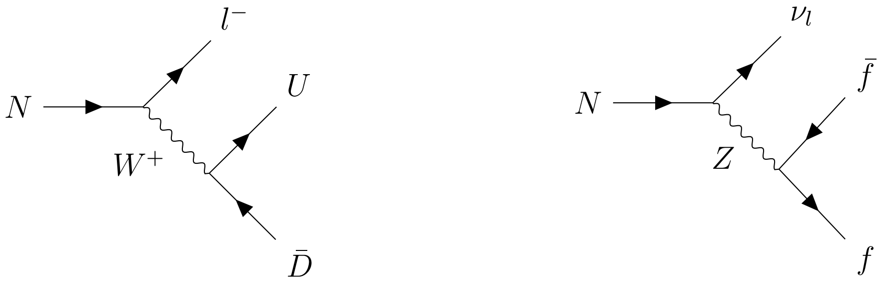

Figure 1.5: Diagram for the HNL decays mediated by charged (left) and neutral (right) currents.

1.8 Theoretical inputs to the HNL search from D_s decays

In the previous sections an overview of the theory concerning heavy neutral leptons has been presented, as well as the equations needed to predict the D_s decay rate in the channel of interest for this analysis. The focus of this work is the search for heavy neutral leptons from D_s decays. Namely, the analysis is targeting the $D_s^+ \to N\mu^+$ decay followed by the $N \to \mu^+\pi^-$ decay. As discussed in Sec. 1.6, this particular decay mode is the preferred one among all decays of the charmed mesons. The theoretical predictions on the branching ratios of the various decay channels are crucial for this analysis when it comes to the signal yield estimation, which will be discussed in Sec. 4.1.

The branching fraction of the $D_s^+ \to N\mu^+$ decay can be obtained by dividing the width of the decay, derived from Eq. 1.27 where $h = D_s$ and $\alpha = \mu$, and the total D_s width $\Gamma(D_s) = \hbar/\tau(D_s) \simeq 1.3 \times 10^{-22}\,\mathrm{MeV\ s}$, which is derived from the average of the D_s lifetime measurements currently available, namely $\tau(D_s) = 504 \pm 4 \times 10^{-15}\,\mathrm{s}$ [3]. The branching fraction of the $N \to \mu^+\pi^-$ decay is fully based on the theoretical predictions presented in Sec. 1.7: it can be calculated as the ratio of Eq. 1.42, where $i = \mu$ and $h = \pi$, and Eq. 1.41.

All the physical constants needed to compute the branching fractions involving the heavy neutral lepton have been taken from *The Review of Particle Physics* [3], while the meson form factors have been taken from the review of *Leptonic Decays of Charged Pseudoscalar Mesons* [50].

Chapter 2

The CMS experiment at the CERN LHC

2.1 The Large Hadron Collider at CERN

2.1.1 Particle collider fundamentals

The most important figures of merit for a particle collider are the collision energy and the luminosity. For two beams of energy E_1 and E_2 colliding head-on, and assuming a small or zero crossing angle, the energy available for the interaction at the center-of-mass is

$$\sqrt{s} \approx 2\sqrt{E_1 E_2}\,. \tag{2.1}$$

The instantaneous luminosity $\mathcal{L}_{\text{inst}}$ is related to the rate of events at the collision point as

$$\frac{dN}{dt} = \sigma \mathcal{L}_{\text{inst}}(t)\,, \tag{2.2}$$

where N is the number of events and σ is the process cross section. The number of expected events in a period of time T can be obtained by integrating Eq. 2.2:

$$N_{\text{tot}} = \sigma \int_{t=0}^{T} \mathcal{L}_{\text{inst}}(t) dt = \sigma \mathcal{L}_{\text{int}}\,, \tag{2.3}$$

where $\mathcal{L}_{\text{int}}$ is referred to as integrated luminosity. Since low cross section processes are usually of higher physics interest, being able to reach a high luminosity is fundamental. The luminosity for colliding bunched beams in a circular accelerator is given by

$$\mathcal{L} = \frac{n_b N_1 N_2 f_{\text{rev}}}{A}\,, \tag{2.4}$$

where N_1 and N_2 are the number of particles in the colliding bunches, n_b is the number of circulating bunches, f_{rev} is the revolution frequency and A is the overlap area of the beams. Assuming a Gaussian profile for the beams colliding head-on, $A = 4\pi\sigma_x\sigma_y$, where σ_x and σ_y are the r.m.s. transverse beam sizes at the interaction point. For real detectors, bunches collide at a crossing angle θ_c, thus the luminosity is reduced by a factor $\mathcal{F} \simeq 1/\sqrt{1+\phi^2}$, where the parameter $\phi = \theta_c\sigma_z/2\sigma_x$ is called the Piwinski parameter and σ_z is the longitudinal r.m.s. beam size.

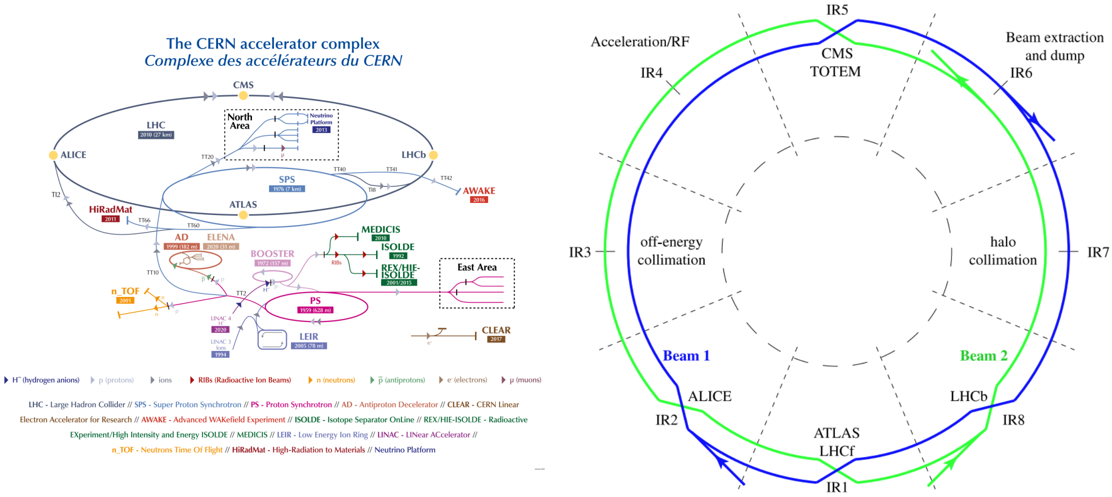

Figure 2.1: Left: The CERN acceletator complex layout in 2022. Right: Schematic layout of the LHC with its eight arc sections, two-beam design and its four experimental sites.

2.1.2 The Large Hadron Collider

The Large Hadron Collider (LHC) is currently the world's largest and highest-energy particle accelerator and collider [51, 52]. It is installed in the tunnel previously instrumented with the Large Electron Positron collider (LEP) at CERN, near Geneva. The tunnel has a length of 27 km and lies between 45 and 170 meters below the surface on a plane inclined at 1.4% sloping toward the *Léman* lake.

The LHC is built with two beam-pipes which cross at four interaction regions (IR), as displayed in Fig. 2.1 (right). This allows protons or heavy ions to be accelerated and collided. According to its design, LHC was expected to be able to collide proton beams with a center-of-mass energy of 14 TeV and reach instantaneous luminosity values of $10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The LHC was also designed to collide heavy ions with an energy of 2.8 TeV per nucleon and a peak luminosity of $10^{27}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. The LHC is composed of eight arcs connected by eight straight sections, which are labeled as IR 1 to IR 8. The beams cross in 4 of these, IR 1, 2, 5 and 8, where the experiments ATLAS (*A Toroidal LHC Apparatus*), ALICE (*A Large Ion Collider Experiment*), CMS (*Compact Muon Solenoid*) and LHCb (*LHC-beauty*) are located, respectively.

Two transfer tunnels link the LHC to the CERN accelerator complex that acts as injector. The CERN accelerator complex, shown in Fig. 2.1 (left), gradually accelerates protons before the LHC acceleration stage. The protons are provided to the accelerator chain from hydrogen atoms after the orbiting electron is removed, by means of the so-called stripping process. Protons are then injected into the first accelerator, the Linear Accelerator (LINAC), followed by the Proton Synchrotron Booster (PSB), the Proton Synchtrotron (PS), and the Super Proton Synchtrotron before being injected in the LHC. The size of these accelerators as well as their beam momentum at injection are listed in Tab. 2.1, where all the reported quantities are the design values.

After a major technical problem occurred in 2008, LHC started operating regularly since the spring of 2010. The LHC operations are organized so that long data-taking periods are alternated with long shutdown (LS) periods where no collisions are delivered by the machine. The first data-taking period (Run 1) spanned over the 2010-2012 years. During Run 1 LHC provided proton-proton collisions at

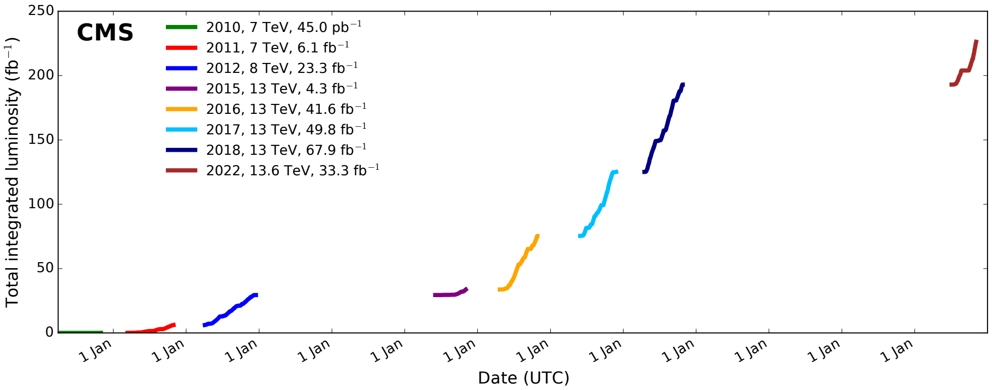

Figure 2.2: Cumulative luminosity versus day delivered to CMS during stable beams for proton-proton collisions at nominal center-of-mass energy [53].

Machine	L (m)	ρ (m)	Beam momentum (GeV)	Bunches
LINAC	30	-	10^{-4}	4×2
PSB	157	8.3	0.05	4×2
PS	628.318	70.676	1.4	72
SPS	6911.56	741.257	26	4×72
LHC	26658.883	2803.98	450	2×2808

Table 2.1: Length (circumference), bending radius ρ, beam momentum at injection and the number of proton bunches by design of the main accelerators in the LHC injection chain.

a center-of-mass energy of 7 TeV until 2011 and 8 TeV in 2012. The Run 1 data-taking period delivered a total integrated luminosity of about $30\,\mathrm{fb}^{-1}$. After the first LS, LHC resumed its operations for the Run 2 data-taking period from 2015 to 2018. During Run 2 LHC collided protons at a center-of-mass energy of 13 TeV and delivered a total integrated luminosity of about $160\,\mathrm{fb}^{-1}$. After the second LS period, LHC resumed delivering proton-proton collisions in the summer of 2022 for the Run 3 data-taking period. Although data-taking operations were ongoing at the time this thesis was written, it should be noted that LHC was able to provide proton-proton collisions at the unprecedented center-of-mass energy of 13.6 TeV, reaching a maximum instantaneous luminosity of approximately $2 \times 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, corresponding to twice the design value, in 2022. A summary of the total integrated luminosity delivered by LHC and collected by the CMS experiment is shown in Fig. 2.2.

The LHC performance in terms of delivered luminosity opens to LHC experiments the possibility to explore rare physics processes. On the other hand, larger instantaneous luminosity means having higher number of collisions per bunch crossing, i.e. a higher *pileup*, which represents a crucial challenge for data-taking operations and reconstruction, especially for high luminosity experiments as ATLAS and CMS.

2.1.3 The coordinate system

The coordinate system adopted by CMS has its origin centered at the nominal collision point inside the experiment. The y-axis is oriented vertically upward with respect to the ideal plane where the LHC lies, while the x-axis points radially inward toward the center of LHC. Therefore, the z-axis points along the beam direction toward the Jura mountains from the LHC IR 5. The azimuthal angle ϕ is measured from the x-axis in the xy plane and the radial coordinate in this plane is denoted by r, while the polar angle θ is measured from the z-axis.

The so-called *pseudorapidity* is defined as $\eta = -\ln(\theta/2)$. The pseudorapidity ranges from $-\infty$ to ∞ when θ ranges from π to 0, and it is 0 when $\theta = \pi/2$. The pseudorapidity which defines the direction of a particle is often combined to its azimuthal angle to separate it from another particle using the quantity $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where $\Delta\eta$ and $\Delta\phi$ represent the differences in η and ϕ between the particles' direction. The momentum and energy transverse to the beam direction are denoted by p_T and E_T, while the so-called missing transverse momentum is defined as $\vec{E}_T^{\mathrm{miss}} = -\sum \vec{p}_T$, where the sum is extended to all reconstructed particles.

2.2 The CMS detector

The Compact Muon Solenoid (CMS) detector is a multipurpose apparatus operating at the CERN LHC [54]. The overall layout of the CMS detector is shown in Fig. 2.3. The CMS design aims to provide a full 4π coverage of the final state particles generated in the hard interaction, and it is ideally separated in two η regions: the central $|\eta| < 1.4$ region is called *barrel*, while, on both sides, the forward ($|\eta| > 1.4$) regions are referred to as *endcaps*. The basic feature which drove the detector design is the 3.8 T superconducting solenoid.

Inside the magnetic coil the inner tracker and the calorimeter detectors are installed. The tracking volume is a cylindrical region with 5.8 m length and 2.5 m diameter, with its axis aligned to the LHC beam line. In order to deal with high track multiplicities, the inner tracker is a combination of silicon microstrip detectors and silicon pixel detectors, the latter being placed close to the interaction point to improve the measurement of the impact parameter of charged-particle tracks, as well as the position of secondary vertices.

The tracking volume is surrounded by a lead tungstate electromagnetic calorimeter (ECAL), followed by a brass and scintillator hadron calorimeter (HCAL). Both ECAL and HCAL are composed of a barrel and two endcap sections. Forward calorimeters (HF) extend the η coverage provided by the barrel and endcap detectors.

Outside the superconducting solenoid, the muon system is installed. Muons are detected in gas-ionization chambers that are embedded in the steel flux-return yoke, and are organized in 4 muon *stations*. The CMS muon system is composed of different subsystems: drift tubes (DT) detectors instrument the barrel region; cathode strip chambers (CSC) are installed in the endcaps; resistive plate chambers (RPC) are present in both the barrel and the endcaps.

New gas electron multiplier (GEM) detectors will be fully installed and operating starting from the High-Luminosity phase of LHC (HL-LHC), foreseen after the LS3 [55]. The GEM subsystem will instrument the $1.6 < |\eta| < 2.8$ region, allowing to increase the redundancy for muon track reconstruction in the $|\eta| < 2.4$ region and to extend the muon system coverage up to $|\eta| < 2.8$. A first batch of 144 GEM chambers was installed during LS2 on the first disk of the two endcaps. These chambers contributed to data-taking operations at the start of Run 3 [56]. Moreover, two more disks of GEM chambers are going to be installed in each endcap before the HL-LHC phase.

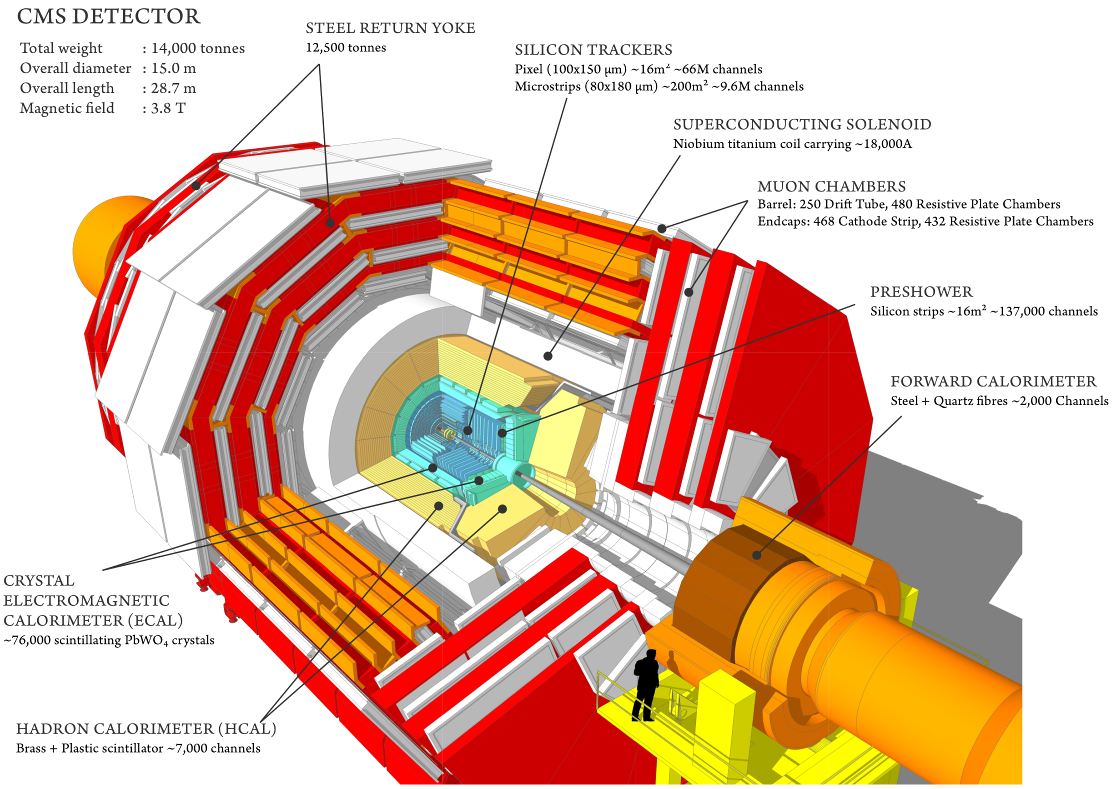

Figure 2.3: A perspective view of the CMS detector.

2.2.1 Inner tracking system

The CMS tracker is composed of two main subsystems: a large silicon strip tracker with a small silicon pixel tracker located inside it [57]. In the central pseudorapidity region, the pixel tracker consists of four co-axial barrel layers at a radial distance between 2.9 cm and 16.0 cm from the beam axis, while the strip tracker consists of ten co-axial barrel layers extending outwards to a radius of 110 cm. The pixel subdetector is completed in each endcap by three disks. Similarly, three small plus nine large disks complete the strip tracker in each endcap.

The endcaps extend the tracker acceptance up to $|\eta| < 3.0$. The current pixel tracker setup is the result of a major upgrade concluded in 2017 (also referred to as Phase-1 pixel upgrade), in which a new pixel subdetector was installed. It added a fourth pixel layer in the barrel and extended the pseudorapidity range from $|\eta| < 2.5$ to $|\eta| < 3.0$ [58]. A schematic view of the CMS tracker system before the Phase-1 pixel upgrade is displayed in Fig. 2.4, where only the top half of the tracker is shown since it is symmetric about the horizontal line $r = 0$. Figure 2.5 shows the layout of the pixel detector before and after the Phase-1 upgrade.

The upgraded pixel detector consists of cylindrical barrel layers positioned at radii of 2.9, 6.8, 10.9 and 16.0 cm and three pairs of endcap disks at distances of 29.1, 39.6 and 51.6 cm from the center of the detector. The pixel cell has a rectangular shape with size of $100 \times 150\,\mu\text{m}^2$. The hit position resolution is of approximately $10\,\mu$m on the transverse coordinate (short side) and 20-40 μm on the longitudinal coordinate (long side). The pixel modules provide three-dimensional position measurement of the charged particles interaction points, where the third coordinate is given by the sensor plane position. The resolution on the track impact parameter in the xy plane is of about 20-75 μm for particles of p_T between 1 and 10 GeV, and it improves as the particle transverse momentum gets larger [59]. In

Subsystem	Layers	Pixel/strip size	Location (cm)
Pixel tracker barrel	4 cylindrical	$100 \times 150\,\mu\text{m}^2$	$2.9 < r < 16.0$
Strip tracker inner barrel	4 cylindrical	80-120 μm	$20 < r < 55$
Strip tracker outer barrel	6 cylindrical	122-183 μm	$55 < r < 116$
Pixel tracker endcaps	3 disks	$100 \times 150\,\mu\text{m}^2$	$29.1 < \lvert z\rvert < 51.6$
Strip tracker inner disks	3 disks	100-141 μm	$58 < \lvert z\rvert < 124$
Strip tracker endcap	9 disks	97-184 μm	$124 < \lvert z\rvert < 282$

Table 2.2: A summary of the principal characteristics of the various tracker subsystems. The number of disks corresponds to that in a single endcap. The location specifies the region in r (z) occupied by each barrel (endcap) subsystem.

total, the pixel detector is made of 1856 segmented silicon sensor modules, where 1184 modules are used in the barrel pixel detector and 672 modules are used for the forward disks.

The strip tracker is composed of four subsystems: the tracker inner barrel (TIB) and the tracker inner disks (TID) cover the $r < 55$ cm and $|z| < 118$ cm. The TIB is composed of four concentric barrel layers, while the TID is composed of three disks oriented on the transverse plane. These provide position measurements in $r\phi$ with a resolution of approximately 13-38 μm. The tracker outer barrel (TOB), that covers the $r > 55$ cm and $|z| < 118$ cm region, consists of six barrel layers providing position measurements in $r\phi$ with a 18-47 μm resolution. The tracker endcaps (TEC) cover the region $124 < |z| < 282$ cm. Each TEC is composed of nine disks, each containing up to seven concentric rings of silicon strip modules with a resolution similar to the TOB.

The modules in the innermost two layers of both the TIB and the TOB, the modules in rings 1 and 2 of the TID and 1,2, and 5 of the TEC carry a second strip detector module which is mounted back-to-back to the first and rotated by a “stereo” angle of 100 mrad in the plane of the module. These modules provide an additional “stereo hit” which can be combined into “matched hits” that provide a measurement of the second coordinate: z in the barrel and r on the disk. However, the achieved single-point resolution of this measurement is an order of magnitude worse than in $r\phi$.

A summary of the characteristics of the various tracker subsystems is displayed in Tab. 2.2.

2.2.2 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) consists of 75,848 $PbWO_4$ crystals, which cover the range $|\eta| < 1.48$ in the barrel (EB) and $1.48 < |\eta| < 3.0$ in the two endcap regions (EE) [60]. In terms of radiation lengths X_0, the depth of the crystals is $25.8X_0$ in the barrel and $24.7X_0$ in the endcaps. The crystals have a tapered shape, slightly varying with position in η. They are mounted in a quasi-projective geometry to avoid cracks aligned with particle trajectories, so that their axes make a small angle with respect to the vector from the nominal interaction vertex, in both the ϕ and η projections. A preshower detector, installed in front of the endcap crystals, consists of two planes of silicon sensors interleaved with a total of $3X_0$ of lead and it is fundamental for π^0 rejection.

Crystals are contained in a thin-walled alveolar structure called *submodule*. The submodules are assembled into *modules* of different types, each containing 400 or

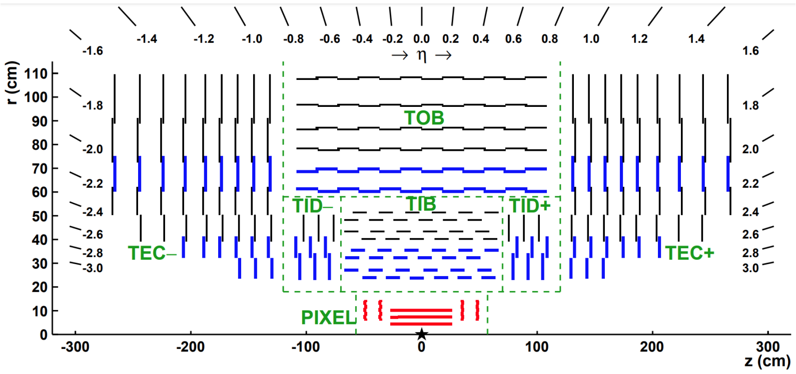

Figure 2.4: Schematic cross section through the CMS tracker in the rz plane (top half), before the Phase-1 pixel upgrade. The position corresponding to the collision point is indicated by a star. Green dashed lines separate each tracker subsystem. Strip tracker modules that provide 2-D hits are shown by thin, black lines, while those permitting the reconstruction of hit positions in 3-D are shown by thick, blue lines. The pixel modules are shown by the red lines.

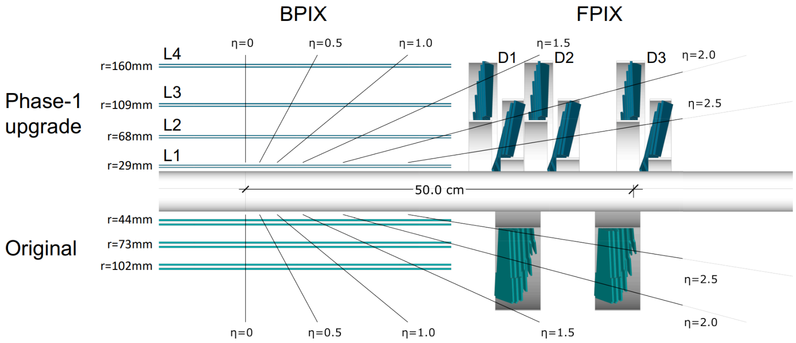

Figure 2.5: Layout of the CMS pixel detector before and after the Phase-1 upgrade, in longitudinal view.

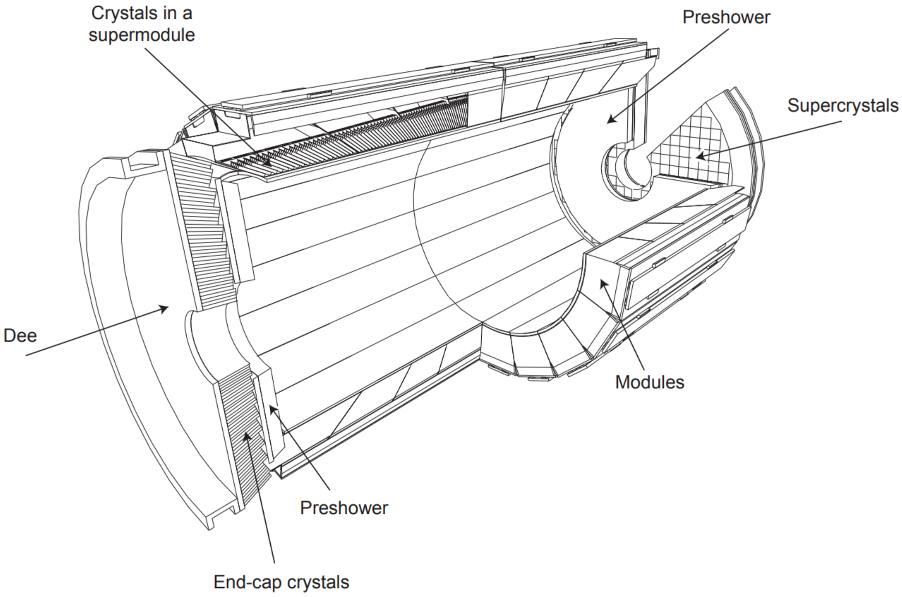

Figure 2.6: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front.

500 crystals depending on their position in η. Four modules are assembled in a *supermodule*, which contains 1700 crystals. In the endcaps, crystals are grouped in supercrystals (SCs), i.e. mechanical units of 5×5 crystals. Each endcap is divided into 2 halves, or *Dees*. Each Dee holds 3 662 crystals. These are contained in 138 standard SCs and 18 special partial SCs on the inner and outer circumference. The overall layout of ECAL is shown in Fig. 2.6.

The energy deposited in the ECAL crystals is detected in the form of scintillation light by avalanche photodiodes in the EB and by vacuum phototriodes in the EE. The electrical signal from the photodetectors is amplified and shaped using a multigain preamplifier, which provides three simultaneous analogue outputs that are shaped to have a rise time of approximately 50 ns and fall to 10% of the peak value in 400 ns. The shaped signals are sampled at the LHC bunch crossing frequency of 40 MHz and digitized by a system of three channels of floating-point analog-to-digital-converters (ADCs). To maximize the dynamic range (40 MeV to ~1.5-3 TeV), three different preamplifiers with different gain settings are used for each of the ECAL crystals, each with its own ADC. The largest unsaturated digitization from the 3 ADCs is used to reconstruct electromagnetic objects.

2.2.3 Hadron calorimeter

The hadron calorimeter (HCAL) is a sampling calorimeter which allows CMS to measure energy and direction of neutral and charged hadrons. The HCAL is composed of four major subdetectors: the hadron barrel (HB), the hadron endcap (HE), the hadron forward (HF) and the hadron outer (HO) calorimeters [61, 62]. A schematic view of the HCAL is shown in Fig 2.7.

The HB covers the $|\eta| < 1.39$ region and it is divided in two half-barrels in the direction along the beam, each assembled from 18 wedges. Each wedge subtends 20° in ϕ. A wedge contains absorber plates made of brass (an alloy with 70% copper and 30% zinc) that are bolted together. The inner and outer plates are made out of stainless steel. Each wedge is divided radially in 17 slots that house the plastic

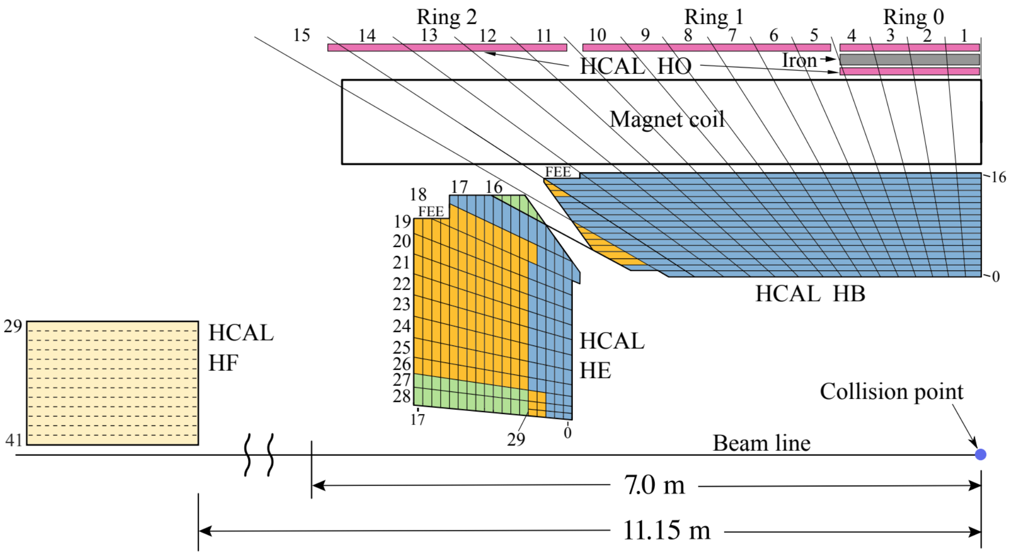

Figure 2.7: Cross section on the longitudinal plane of one quarter of the CMS HCAL during 2016 LHC operation (from Ref. [62]). The positions of the hadron barrel (HB), the hadron endcap (HE), the hadron forward (HF) and the hadron outer (HO) calorimeters are shown. The layers marked in blue are grouped together as depth 1, while the ones in yellow, green and magenta are combined as depths 2, 3 and 4.

scintillator tiles. The HB has about 40 000 scintillator tiles: to limit the number of individual physical elements, the tiles at the same ϕ and depth are grouped into a single scintillator unit, referred to as megatile, covering roughly 5° in ϕ. The megatiles are divided into 16 sections along z, denoted by $|i\eta| = 1, \ldots, 16$, so that each tile covers a η range of 0.087.

The HE calorimeter is also made of brass absorber plates with sampling layers of plastic scintillators. The HE is located at about 4 m from the interaction point and it covers the region $1.3 < |\eta| < 3.0$. The top edge of the front part of the endcap module has a slope of 53° corresponding to the gap angle between the HB and HE calorimeters. It also has a nose-like structure, with an additional layer of absorber for $|i\eta| = 18$, to increase the total interaction length for that tower. The endcap on either side is divided into 14 parts along $|\eta|$, and the 18 layers are combined respectively as depths 2, 3 and 4, as shown in Fig 2.7.

The HO calorimeter consists of one or two layers of scintillators outside the magnet coil and it covers the region $|\eta| < 1.26$. The HO was designed to ensure that hadronic showers are sampled with nearly 11 hadronic interaction lengths. The entire assembly is divided into five rings, each having 12 sectors in ϕ. The central ring has two layers of 10 mm thick scintillator on either side of a stainless steel block. All the other rings are made of a single layer of scintillator.

The front faces of the HF calorimeters are located at about 11 m distance from the interaction point and cover the region $2.85 < |\eta| < 5.19$. Each HF module is composed of 18 wedges made of steel with quartz fibers embedded along its length. The detection technique utilizes emission of Cherenkov light by secondary charged particles going through the quartz fibers.

2.2.4 The muon system

For muon detection CMS is instrumented with three different types of gaseous detectors, which are embedded in the layers of the steel flux-return yoke. The CMS muon system covers the $|\eta| < 2.4$ region. A schematic view of the CMS detector highlighting the muon system structure is shown in Fig. 2.8. The DT chambers are installed the region $|\eta| < 1.2$, the CSC chambers in the $0.9 < |\eta| < 2.4$ and the RPC in the $|\eta| < 1.8$ region.

The basic unit of a DT chamber is the drift cell. The anode wire of a drift cell has a length which may vary between approximately 2 and 3 m. The drift cell cross section is of $13 \times 42\,\mathrm{mm}^2$, thus, the maximum drift path is of 21 mm, which corresponds to about 380 ns of drift time with electric field shape and gas mixture used in the DT, i.e. 85% Ar and 15% CO_2. A DT chamber is made of three or two superlayers (SL), each made of four layers of rectangular drift cells staggered by half cell. In a DT chamber two of the superlayers are oriented so that the wires are parallel to the beam direction and provide a measurement in the magnetic bending plane ($r\phi$), while the other SL is oriented perpendicular to the other two and measures the position along z. The DT chambers are arranged in four stations having about the same radial distance from the beam pipe and installed in five “wheels” along the z direction. For each station the drift tubes are divided in 12 sectors around the ϕ direction, with one chamber in each sector with the exception of two sectors of the station 4. The DT chambers on the outermost station, i.e. the fourth station, are made of 2 SL, therefore they cannot measure the θ coordinate. The DT chambers are labeled as MB/Z/X/Y, where $\mathrm{Z} = 0, \pm1, \pm2$ refers to the wheel, $\mathrm{X} = 1, \ldots, 4$ to the station and $\mathrm{Y} = 1, \ldots, 12$ to the sector, or more simply MB1, 2, 3 and 4 depending on the station where the DT chamber is installed.

The CSC are multiwire proportional chambers made of six anode wire planes interleaved among seven cathode panels. The cathode panels are segmented in strips that run lengthwise at constant $\Delta\phi$ width, while the wires run azimuthally and allow to measure the radial coordinate. The ϕ coordinate is measured by interpolating the position of the wire with the center of the charge distribution induced on the strips, exploiting the fact that the induced signal is expected to be larger in a strip closer to the muon trajectory. The chambers are trapezoidal and cover either 10° or 20° in ϕ. The CSC chambers are arranged in four stations: a station consists of CSC chambers that shares roughly the same position along z. In each station, the CSC chambers are arranged in rings with different radial distance: three rings of CSC in the first stations and two in the remaining ones. The CSC chambers are labeled as ME±X/Y, where $\mathrm{X} = 1, \ldots, 4$ refers to the station, $\mathrm{Y} = 1, \ldots, 3$ to the ring and the ± sign specify whether the chamber is positioned on the positive or negative z side.

The RPC are double gap chambers operated in avalanche mode consisting of two gaps with common pick-up read-out strips in between. The total induced signal is the sum of the two single-gap signals. In the barrel the RPC have the strips oriented along the beam direction, and are arranged in six coaxial cylinders around the beam axis forming four stations. In the fist two station, two layers of RPC are located internally (RB1in and RB2in) and externally (RB1out and RB2out) with respect to the DT chambers. In the third and fourth station there are again two RPC chambers in a single sector but arranged side-by-side on the same DT face, and they are named RB3± and RB4±. Special cases are RB4 in sector 4, consisting of four chamber in the same side, and in sectors 9 and 11, where there is only a single RPC chamber. In the endcaps RPC have a trapezoidal shape, with strips oriented parallel to the CSC strips, and are arranged in four stations (RE1-4). Similarly to

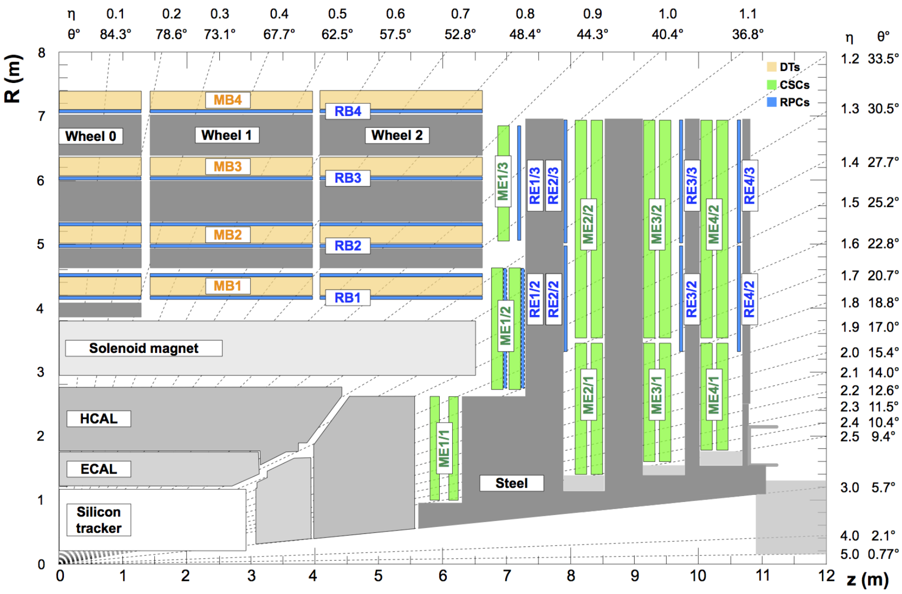

Figure 2.8: Schematic cross section of the CMS detector during Run 2 in the rz plane highlighting the position of the muon system subdetectors: DT (orange), CSC (green) and RPC (blue) embedded in the steel flux-return yoke (dark grey areas).

CSC, an RPC chamber in the endcaps covers an angle of 10°-20° in ϕ.

2.3 The CMS trigger system

The CMS physics program targets many areas of interest to the high-energy physics community. Soon after its discovery [63, 64], the measurement of the Higgs boson properties has become of central importance for the CMS physics program, together with precision measurements of SM properties in the electroweak, top quark and QCD sectors. In addition, searches for beyond standard model particles saw an increasing interest over the years. Heavy ion collisions are included in the CMS physics program as well, allowing to explore the quark-gluon plasma dynamics.

The LHC collides bunches of particles in CMS at a maximum rate of about 40 MHz, where the bunches are spaced 25 ns apart. At the design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$ the proton-proton interaction frequency exceeds 1 GHz. Only a small fraction of these collisions contains events of interest to the CMS physics program, and moreover only a small fraction of those can be stored for later offline analysis. This selection task is performed by the trigger system. The trigger system is expected to be able to select only the interesting events for offline storage from the bulk of the inelastic collision events.

The CMS trigger system is organized in two levels [65]: the Level-1 (L1) trigger system uses custom hardware processors to select up to 100kHz of the most interesting events with a latency of 4 μs, the high level trigger (HLT) performs a more detailed reconstruction of events selected at L1 on a computing processor farm, reducing the rate by another factor of 100 in few hundred milliseconds.

The L1 trigger gathers information from the muon and calorimeter detectors with coarse granularity and precision to select collision events for readout to HLT. The selection is performed using a list of algorithms, known as “seeds”, which check events against predetermined criteria, that are collectively called a “menu”. Any event that satisfies the conditions of at least one seed in a given menu is accepted for further processing in the trigger chain. This initiates a readout of the complete detector information from the data acquisition system, and the data are sent to the HLT. Level-1 trigger menus reflect the variety of research interests of CMS and evolve depending on the CMS physics priorities, beam conditions and detector performance.

Trigger algorithms usually consist of criteria applied to one or more objects of a single type, e.g. muons, hadronic jets, tau leptons, photons or electrons, scalar sum of transverse energy or missing transverse momentum, or a combination of different objects ("cross" seed). Finally, multi-object trigger algorithms may also apply selection criteria based on correlations between the kinematics the different objects they involve. These criteria typically include thresholds on the objects p_T, E_T or η, or a combination of them. This allows to roughly separate candidate events from massive particles, which tend to produce particles at high p_T and in the barrel region, from the vast majority of the background, which usually produces low p_T particles in the forward regions. Single- and double-object seeds in the menu cover about 75% of the available rate. Figure 2.9 shows the fraction of the maximum L1 trigger rate allocated to various object seeds in a typical Run 2 menu, where events triggered by n different seeds are weighted by $1/n$ [66].

The trigger rate from a given algorithm can be reduced by applying a *prescale* which imposes that only fraction of events selected by the seed will pass the trigger. A prescale is indicated by a positive integer N, which means that only one out of N of the events satisfying the seed condition will eventually be accepted. A "prescale column" is a set of prescale values applied to each of the seed of the L1 trigger meny. Prescale columns depends also on the instantaneous luminosity: at the end of an LHC fill multiple prescale columns with decreasing prescale values are used to maximize signal efficiency while keeping the L1 trigger rate under 100 kHz.

The L1 trigger algorithms are designed using either Monte Carlo simulated collision events or, if possible, previously collected data. The seed thresholds are adjusted to achieve a total menu rate that is less than 100 kHz. The trigger rate is usually studied as a function of the pileup: for most single-object seeds the trigger rate increases almost linearly with the pileup, while multi-object seeds might have a higher-order dependency. Rate and pileup are measured in a time interval of a "luminosity section", corresponding to 2^{18} LHC orbits or 23.3 seconds of data taking.

The HLT menu has a modular structure. It is divided in different *paths*, each path corresponds to a sequence of reconstruction and filtering modules, that are based on streamlined versions of the algorithms used for the offline reconstruction. The modules within a path, either object producers or filters, are arranged in blocks of increasing complexity so that faster algorithms are run first: if a filter within a path fails the rest of the path is skipped. To further reduce the CPU time consumption, the HLT can exploit features such as the "regionality", i.e. the object reconstruction is restricted to narrow regions around the L1 or higher-level candidates. The final trigger decision is the logical OR of the decisions of the single paths.

Events accepted by the trigger are then stored and processed for the physics object reconstruction (see chapter 3). The information on the luminosity recorded by CMS in the various data-taking periods is publicly available [53].

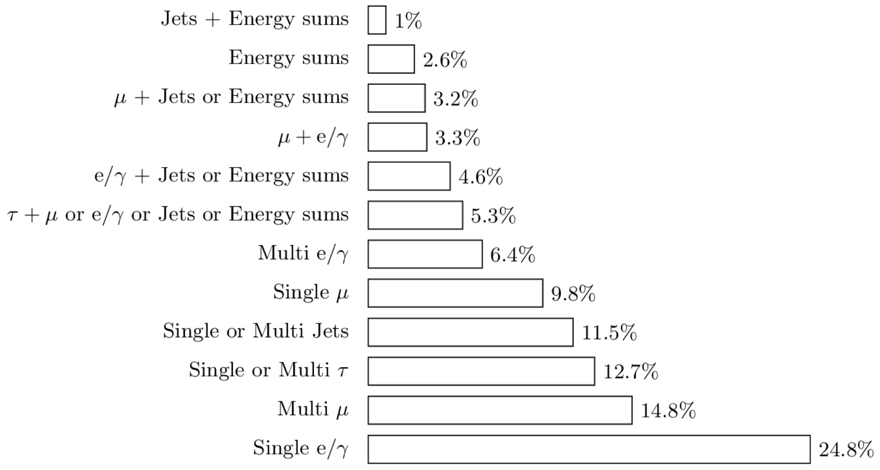

Figure 2.9: Fractions of the 100 kHz rate allocation for single- and multi-object triggers and cross triggers in a typical CMS physics menu during Run 2, from Ref. [66].

2.3.1 The Level-1 trigger

For each event the L1 trigger combines the information coming from muon and calorimeter detectors to build the physics objects that are eventually used to produce a trigger decision. A brief summary on how the calorimeter and muon detectors information is processed is reported below.

The calorimeter trigger consists of two layers: Layer-1 receives, calibrates (to take into account changing calorimeter response over time) and sorts the local energy deposits, i.e. the “trigger primitives”, which are sent to the trigger by the ECAL and HCAL; Layer-2 uses these calibrated trigger primitives to reconstruct and calibrate the physics objects such as electrons, tau leptons, jets and energy sums.

Electrons and photons are indistinguishable to the L1 trigger since the information on the tracks is not available. The electron/photon candidate is built by clustering the ECAL-plus-HCAL energy deposit around a “seed” trigger tower defined as the local energy maximum above $E_T = 2\,\mathrm{GeV}$. Clusters are then built dynamically by including surrounding towers having an energy deposit above 1 GeV.

Tau leptons with hadronic decay to pions tend to cause an energy deposit that is typically more spread out than that of an electron. However, the dynamic clustering developed for the electron/photon trigger has been adapted to reconstruct these individual clusters, which can subsequently be merged.

For jets the L1 trigger exploits a “sliding window” technique [67] based on 3×3 calorimeter regions (i.e. 144 trigger towers in the central region and up to 54 trigger towers in the forward region), spanning the full (η, ϕ) coverage. The L1 jet candidate is found if the energy deposit in the 3×3 window meets the following conditions: the central region of the 3×3 matrix must have an E_T higher than any of the eight neighbors, and exceed a specific threshold.

The muon trigger system includes three muon track finders (MTF) which reconstruct muons in the barrel (BMTF), DT-CSC overlap (OMTF), endcap (EMTF) regions of the detector and the global muon trigger (μGMT). Prior to the transmission to the respective track finders, the signals coming from the three muon subdetectors are combined into trigger primitives, which provide information about

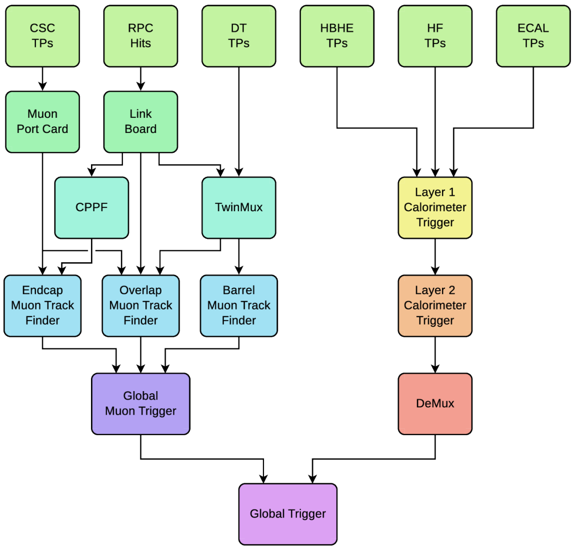

Figure 2.10: Diagram of the CMS L1 trigger system during Run 2, from Ref. [66].

coordinates, timing and quality from the detector hits. Track finders use muon detector trigger primitives to build muon track candidates, assign to them a quality flag and measure charge and p_T. Each track finder uses muon finding and p_T assignment logic optimized for its region, and assigns the track quality corresponding to the expected p_T resolution. Each track finder transmits up to 36 muons to the μGMT, which resolves duplicates from different boards, and sends the data for a maximum of eight muons of highest rank (a linear combination of p_T and a quality value) to the global trigger, where they are used in the final Level-1 trigger decision.

The global trigger finally collects muons and calorimeter objects and executes every algorithm in the menu in parallel for the final trigger decision. A diagram displaying the L1 trigger workflow is shown in Fig. 2.10.

2.3.2 The High Level Trigger

The event selection at the HLT level is performed in a similar way to that of offline processing. For each event objects as electrons, muons and jets are reconstructed and identification criteria are applied in order to select only those events which are of possible interest for data analysis. The HLT hardware consists of a farm of computers running Linux Operating System.

The data processing of the HLT is structured around the concept of an “HLT path”. Each path implements a sequence of reconstruction and filtering steps which use the software framework that is also used for offline reconstruction and analyses. If at least one path in the HLT menu has a positive outcome, the event is accepted and sent to another software process, the *storage manager*, for archival storage. The event data are stored locally on disk and eventually transferred to the CMS Tier-0 computing center for offline processing and permanent storage. The HLT rate of a given path can be reduced as well by applying a prescale with the same mechanism used for L1 seeds.

Track reconstruction and vertex finding are crucial for object reconstruction at the HLT. The knowledge of the position of the primary vertex is needed for many trigger paths. To reconstruct the primary vertex without having to run the full tracking algorithm, tracks are reconstructed requiring only the data from the pixel detector. With these tracks, a gap-clustering algorithm is used for vertex reconstruction [68]. The track reconstruction algorithm is a four-step combinatorial track finder algorithm implementing a Kalman filter technique [69]. Different objects use slightly different configurations for the track reconstruction, e.g. pixel-only tracking, iterative tracking, muon-seeded tracking, one-step-tracking for isolated leptons and for b-tagging.

The HLT electron and photon identifications begin with a regional reconstruction of the energy deposited in the ECAL crystals around the L1 electron/photon candidates. This is followed by the building of the so-called “supercluster” using offline reconstruction algorithms [70]. Electrons and photons are distinguished by requiring the presence of a reconstructed track compatible with the supercluster.

The muon HLT combines information from both the muon and the tracker subdetectors to identify muon candidates and measure their p_T. The HLT muon trigger algorithm consists of two steps [71]. The level-2 (L2) step performs a track reconstruction in the muon spectrometer using a Kalman filter technique, while at the level-3 (L3) muons are reconstructed using all available information regarding the trajectory of the muon from both the muon spectrometer and the inner tracking detector. The muon is reconstructed by either matching a track in the inner tracker with a L2 muon and performing a combined track fit using the information from both the muon system and tracking detectors, or by identifying an inner-detector track as a muon candidate by matching it to one or more DT-CSC track segments without performing a combined fit. In the former case, the final track parameters are obtained from a refit that is performed once all hits associated with the trajectory are known. Poor-quality tracks are rejected at the end of reconstruction.

Triggers based on jets and missing transverse energy play an important role for the search for new physics. At the HLT, jets are reconstructed using the anti-k_T clustering algorithm with cone size $\Delta R = 0.5$ [72, 73]. The HLT algorithms based on missing transverse energy use the same algorithms used offline to calculate E_T^{miss}. The τ-jet triggers identify and select events with hadronic decays of the τ leptons, while leptonic decays are selected as prompt electrons and muons.

2.4 The CMS computing model

To handle the processing and the storage of the large amount of data, CMS makes use of a distributed computing and analysis model [74, 75] which relies on the Worldwide LHC Computing Grid (WLCG) infrastructure. The computing resources are geographically distributed, interconnected via high-throughput networks and accessed by means of Grid techniques.

2.4.1 Computing centers

The CMS computing system is built using computing resources at a range of scales, provided by collaborating institutes around the world. The CMS uses a hierarchical architecture of tiered centers, with a single Tier-0 at CERN, few Tier-1 centers at national computing facilities, and several Tier-2 centers at institutes.

The Tier-0 primary functions are to:

- Accept data from the online system and copy it to permanent mass storage;
- Handle prompt reconstruction of RAW data (see Sec. 2.4.2);

- Export a copy of RAW and reconstructed data to Tier-1 centers. Data is not considered "safe" for deletion from Tier-0 buffers until it is stored on at least two independent sites (one of these is CERN computing centre, playing the role of a Tier-1).

The Tier-0 is a common CMS facility used only for well controlled batch work; it is not accessible for analysis use.

A few Tier-1 centers are hosted at collaborating national labs and computing centers around the world. Each site provides large batch CPU facilities, a mass storage system including a robotic tape archive, and very high speed international network links including a dedicated link to the LHC optical private network (LHC-OPN). The primary functions of a Tier-1 are to:

- Provide long-term safe storage of a second copy of RAW data from CMS;
- Store and transfer to Tier-2 centers simulated and derived data.
- Carry out new RAW data reconstruction using improved algorithms or calibrations; provide rapid access to very large data samples for skimming and data-intensive analysis which Tier-2 centers would not be able to handle.

Several Tier-2 centers of varying sizes are hosted at CMS institutes. Typically a Tier-2 center shares its resources between the local user community and the whole CMS community. The functions of a Tier-2 center may include:

- Support of analysis activities, such as local storage of data samples transferred from Tier-1, and access to a flexible CPU farm;
- Support of specialized activities such as offline calibration and alignment tasks, and detector studies;
- Production of Monte Carlo data, and its transfer to an associated Tier-1 center for long term storage.

2.4.2 The data flow

The CMS data acquisition (DAQ) system writes DAQ-RAW events (roughly 1.5 MB/event) to the HLT farm input buffer. The DAQ-RAW events contains data from the various subdetectors and the L1 trigger results. The HLT farm writes RAW events (roughly 1.5 MB/event) accepted by the trigger at a rate around 1 kHz. A RAW event contains the detector data, the L1 trigger result, the result of the HLT selections, and some of the higher-level objects created during the HLT processing. These events are classified to primary data sets depending on the trigger history. Primary data sets are grouped into online streams in order to optimize their transfer to the Tier-0 farm and the following reconstruction process.

The first event reconstruction is performed in a short time ($\sim 48\,\text{h}$) on the Tier-0 farm which writes RECO events (roughly 3 MB/event). A RECO event contains reconstructed physics objects and reconstructed hits and clusters. The RAW and RECO versions of each primary data set are stored on the Tier-0, while a copy is transferred to a Tier-1. Transfer to other Tier-1 centers is subject to additional bandwidth being available.

The analysis object data (AOD) are derived from RECO events and contain a copy of all high-level physics objects and a summary of other RECO information sufficient to support typical analysis actions. The AOD data have a 480 kB size per event and are produced in the Tier-0 reconstruction step and distributed to the Tier-1 centers. Whenever the AOD format is updated, the production is handled by Tier-1 centers which then distribute the new AOD version between themselves.

A further processing of the AOD data is finally performed to produce common light data skims which serve almost all physics analyses. Such data formats are derived from AOD and are called MINIAOD (35-60 kB/event) [76] and NANOAOD (1 kB/event) [77], the latter being derived from the MINIAOD.

All AOD of selected primary streams, as well as MINIAOD and NANOAOD, and a fraction of RECO and RAW events are transferred to Tier-2 centers which support interactive analysis of authorized groups of users.

2.4.3 The 2018 B-Parking data taking mode

During the 2018 data-taking period, a new trigger and data processing strategy was implemented for the first time, referred to as "B-Parking", which allowed to collect and reconstruct roughly 10^{10} unbiased b hadron decays.

Usually the core physics program of CMS exploits data streams that are promptly reconstructed at the CERN Tier-0 data center, and are generally available after 48 hours for physics analysis. The B-Parking data stream has a trigger rate of several kHz, which is beyond the standard processing capabilities of the Tier-0 center. However, the B-Parking data-taking was possible because the CMS trigger and DAQ system have the ability to record multiple data streams. B-Parking streams are defined by relaxed inclusive trigger requirements and are not processed immediately by the CMS reconstruction software. The collected data are stored unprocessed to permanent tape, and the reconstruction is performed only at later point, e.g. during an end-of-year or long shutdown of the LHC.

During the 2022 data-taking period of LHC Run 3, CMS implemented again a B-Parking data-taking strategy based on what was done in 2018. However, the trigger strategy in 2022 was different from the 2018 one (e.g. it includes also double-muon and double-electron triggers) and it will not be discussed in this section, which refers only to the 2018 B-Parking strategy.

The 2018 B-Parking trigger strategy

The trigger strategy for the B-Parking data samples was designed to record an unbiased sample of b hadron decays, exploiting the fact that in proton collisions at the LHC b hadrons are mainly produced in pairs. The trigger logic targets muons from the semi-leptonic decays of such b hadrons, which often produce low-p_T non-prompt muons: a b hadron usually travels $\sim$ 0.1-0.5 mm before decaying. This results in a "tag" b hadron that decays semileptonically and a "signal" b hadron that is unbiased by the trigger logic imposed on the "tag". The trigger logic based on the presence of a single muon accounts for approximately 20% of all b hadron decays.

The 2018 B-Parking data-taking program was able to record an unprecedented amount of b hadron decays while leaving the core CMS physics program unaffected. This is achieved by taking advantage of an increase of idle online computing resources as the instantaneous luminosity decreases at the end of an LHC fill. As the instantaneous luminosity decreases both the L1 and the HLT trigger rates decrease, and the event processing load decreases as well because of the reduced number of pileup events.

The L1 B-Parking trigger logic requires the trigger muon to have $|\eta| < 1.5$ and a p_T above a given threshold, that decreases as the instantaneous luminosity decreases. A summary of the L1 seeds requirement as a function of the instantaneous luminosity is shown in Tab. 2.3. Both the L1 and the HLT requirements are relaxed by changing the prescale-sets of the trigger menu causing the trigger rate to progressively increase while slightly reducing the b hadron purity, as summarized in Tab. 2.4. The HLT paths are seeded by the same set of L1 seeds, which are activated following the instantaneous luminosity decrease. The B-Parking samples are collected using a mixture of HLT paths, with different thresholds in p_T and on the muon impact parameter significance. Each path is identified with a HLT_MuX_IPY string, where X and Y indicates the threshold on p_T and impact parameter significance.

The b hadron purity is defined as the fraction of events triggered by muons which come from a B meson decay:

$$P = \frac{N\,(B \to X\mu)}{N(\mu)}\,. \tag{2.5}$$

The b hadron purity was found to range between 0.59 and 0.92 in simulation (depending on the trigger p_T and IPS thresholds), with an average value of approximately 0.75 that has been validated using data. The trigger purity has been measured in data exploiting the $B^0 \to D^{*+}\mu\nu$ decay because of its large branching fraction. The targeted D^{*+} decay is the $D^{*+} \to D^0\pi^+ \to K^-\pi^+\pi^+$ decay. The B^0 yield in data is measured by reconstructing $D^0 \to K^-\pi^+$ candidates, where the kaon must have the same sign of the triggering muon, and $D^{*+} \to D^0\pi^+$ candidates. The invariant mass difference between $K^-\pi^+\pi^+$, forming the D^{*+} candidate, and $K^-\pi^+$, forming the D^0 candidate, produces a very narrow peak around the mass of the pion, and it is shown in Fig. 2.11. The $B \to D^{*+}\mu\nu$ yield in data is derived by an unbinned maximum likelihood fit of this peak. The fit is performed with a second-order polynomial and Gaussian PDF for the background and signal, respectively. The number of D^{*+} candidates are determined from the fitted mass value $\pm 2\sigma$. The number of expected events triggered by a muon coming from a B meson decay, i.e. $N\,(B \to X\mu)$, is derived by correcting the $B \to D^{*+}\mu\nu$ yield for acceptance and reconstruction effects, as well as for the branching fraction of the decay channel. Further details on the trigger purity measurement can be found in Ref. [78].

The B-Parking L1 and HLT rates peak at approximately 50 kHz and 5.4 kHz, respectively. Because of the design of the B-Parking trigger logic, the highest rates are observed late in an LHC fill, which results in a pileup value of approximately 29 when averaged over the entire collected data set. This value is lower than that typically observed for the standard physics streams of CMS. The left and middle plots in Fig. 2.12 show the L1 rate during 2017 and 2018 data-taking periods, where prescale column changes are displayed as a vertical red dashed line. These plots illustrate how the B-Parking trigger strategy allowed to exploit spare L1 trigger rate bandwidth due to the decrease of instantaneous luminosity at the end of a physics fill: while in 2017 the L1 trigger rate decreased with the time, in 2018 the rate remained close to its optimum value, i.e. approximately 80 kHz. The right plot in Fig. 2.12 shows the trigger rates of the CMS HLT system as a function of time during an HLT fill in 2018 for both the standard physics and the B-Parking streams. When the B-Parking triggers are activated, starting from $1.7 \times 10^{34}\,\mathrm{cm^{-2}\,s^{-1}}$ instantaneous luminosity, the B-Parking stream shows sharp increases in the rate throughout the LHC fill, as the trigger requirements are loosened. On the other hand, the standard physics data streams decreases monotonically.

Peak $\mathcal{L}_{\text{inst}}$ ($10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$)	Loosest L1 p_T threshold (GeV)	Loosest L1 p_T threshold (GeV)	Loosest L1 p_T threshold (GeV)	Peak $\mathcal{L}_{\text{inst}}$ ($10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$)	Loosest L1 p_T threshold (GeV)
1.6	22	22	12	1.7	12
1.4	10	10	10	1.5	10
1.2	10	9	9	1.3	9
1.0	8	8	8	1.1	8
0.8	7	7	7	0.9	7
First fill	6672	6693	6763	First fill	7132

Table 2.3: Summary of the tag-side loosest p_T muon trigger L1 seeds threshold enabled for trigger menus designed to operate in different instantaneous luminosity ranges. All L1 seeds have the $|\eta| < 1.5$ requirement as well. Each column represents one of the different settings that were deployed throughout the year, together with the number of the first LHC Fill in which they were implemented.

Settings	Peak $\mathcal{L}_{\text{inst}}$ ($10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$)	L1 μ p_T threshold (GeV)	HLT μ p_T threshold (GeV)	HLT IPS threshold	Trigger purity (%)	Peak rate (kHz)
1	1.7	12	12	6	92	1.5
2	1.5	10	10	6	87	2.8
3	1.3	9	9	5	86	3.0
4	1.1	8	8	5	83	3.7
5	0.9	7	7	4	59	5.4

Table 2.4: The p_T and IPS thresholds on B-Parking trigger muon associated to each trigger menu designed to operate at different instantaneous luminosity ($\mathcal{L}_{\text{inst}}$) ranges. The last two columns show the trigger purity and the peak rate. The table refers to the trigger configuration that was online starting from LHC Fill 7132 of B-Parking era D.

The B-Parking sample has been collected during the June-November period in 2018. The sample recorded 12 billion events with a high trigger purity and contains approximately 10 billion unbiased b hadron decays. The B-Parking data set is divided in four eras, labeled as A, B, C and D, and each era is divided in different parts: eras A and B are divided in six parts, eras B and D are divided in five parts. If a given era is divided into N parts, N copies of the HLT path that is active at a given instantaneous luminosity are run, each prescaled by a factor N. To initiate the event counting of the N independent prescaled triggers, consecutive integer numbers are used as seeds. This partitioning strategy allows to separate a large amount of data into subsamples that can be stored and reconstructed independently, while collecting all the events selected by the B-Parking triggers. The whole B-Parking data samples correspond to a total integrated luminosity of $41.7\,\text{fb}^{-1}$.

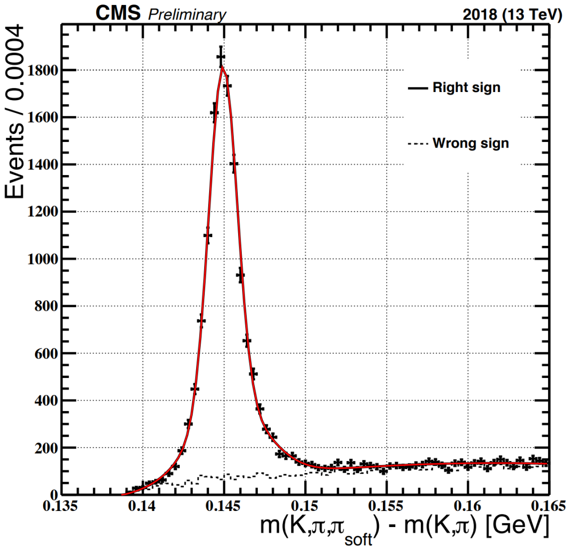

Figure 2.11: Invariant mass difference of the reconstructed D^{*+} and D^0 candidates. Points with error bars show events where the product of the K and the μ is +1 (right sign), the dashed line shows the events where this product is -1, and the red solid line is the fit [78].

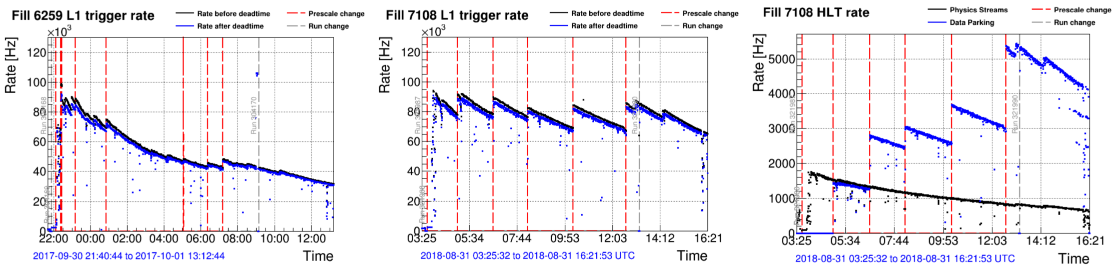

Figure 2.12: Rate of the CMS L1 trigger (blue data points), as a function of time, during an LHC fill in 2017 (left) and 2018 (center). The time intervals cover 13–15 hours. Changes in the run number and settings (prescale column) are indicated by vertical gray and red dashed lines. The right plot shows the trigger rates of the CMS HLT system as a function of time during an LHC fill in 2018, for the physics (black data points) and B-Parking (blue data points) streams.

Chapter 3

Event reconstruction in CMS

A particle traveling through the detector is expected to produce signals in the various CMS subdetectors, depending on the nature of the particle. Charged particles give rise to signals (*hits*) in the sensitive layers of the tracker, allowing to reconstruct their trajectory, or *track*, and their origin, or *vertex*. Electrons and photons are absorbed in the ECAL, where their energy is measured, and the deposited energy is being detected as *clusters*, which also allow to infer the particle direction. Charged and neutral hadrons may also deposit a fraction of their energy in the ECAL, but they are primarily absorbed in the HCAL. The deposited energy clusters in both ECAL and HCAL are used to detect the hadron energy and direction. At the energies involved at the LHC, muons are minimum ionizing particles and they loose only a small fraction of their energy when traversing the detector material. Muons are detected both in the tracker and in the muon system. Although muons are not absorbed in the calorimeters, they might deposit there a small fraction of their energy. Neutrinos pass through the CMS detector with almost no interaction, their presence is assessed by reconstructing the transverse momentum unbalance in collision events. The signals in the subdetectors are used to reconstruct various *physics objects* and measure their kinematics.

The software used by CMS to reconstruct such physics objects is based on the idea of correlating the basic elements from all detector layers (tracks and clusters) to identify each final-state particle, and reconstructing the particle properties by combining the corresponding measurements on the basis of this identification. This approach is called *particle-flow* (PF) *reconstruction* [79].

3.1 Particle flow elements

3.1.1 Charged-particle track reconstruction

Originally, charged particle tracks were reconstructed using combinatorial track finder algorithms based on the Kalman Filter technique. Such algorithms were composed of three stages: (i) initial seed generation with a few hits compatible with a charged-particle trajectory; (ii) pattern recognition step gathering hits from all the tracker layers along this charged-particle trajectory and (iii) fitting to determine the charged-particle properties, i.e. its origin, transverse momentum and direction. Reconstructed tracks were required to be seeded with at least two hits in consecutive pixel layers, to be reconstructed with at least eight total hits and at most one missing hit along the way. Furthermore, all tracks had to originate from within a cylinder of a few mm radius centered around the beam axis and to have a transverse momentum larger than 0.9 GeV. These reconstruction requirements were able to keep the misreconstructed track rate to few percent, but limited the reconstruction efficiency to only 70-80% for charged pions with $p_T > 1\,\mathrm{GeV}$, compared to 99% for isolated muons.

To increase the tracking efficiency while keeping the misreconstructed track rate at a similar level, the combinatorial track finder is applied in several successive iterations. At each step, the misreconstruction rate is reduced by applying quality criteria on the track seed, on the track fit χ^2, and on the tracks compatibility with originating from one of the reconstructed primary vertices. The hits associated with the selected tracks are masked for the following iterations. At each iteration, the remaining hits may be used to form new seeds and tracks with looser quality requirements, this allows to gain on the tracking efficiency side while limiting the degradation of the purity. The same operation is repeated with increasing complexity seeding, filtering and tracking algorithms. The complete list of seeding configurations of the ten tracking iteration can be found in Ref. [79].

The iterations addressing tracks seeded with at least one hit in the pixel detector recover about half of the tracks with p_T above 1 GeV missed by the original combinatorial track finder, with slightly smaller misreconstruction rate levels. These iterations allow to lower the p_T threshold to 0.2 GeV, typically below the calorimeter thresholds. In addition, after a loss of muon reconstruction efficiency that was observed in 2012 in the tracker, two more iterations have been implemented. An "inside-out" iteration was designed to re-reconstruct muon-tagged tracks with looser requirements to improve hit collection efficiency. An "outside-in" iteration seeded by tracks reconstructed in the muon system, was designed to recover the missing muon track in the tracker.

During reconstruction, tracks are separated in categories of expected purity based on a series of cuts on the normalized track fit χ^2, on longitudinal and transverse impact parameters, and their significance. Tracks failing the loosest selection are rejected, while those that pass the tighest selection are labeled as *highPurity*. A brief description of the selection cuts used to define *highPurity* tracks will be reported here, further details can be found in Ref. [80]. Selection cuts optimized using simulated samples lead to the following selection requirements for *highPurity* tracks:

- Normalized $\chi^2 < \alpha_0 \cdot n_{\text{layers}}$;
- $|d_{xy}| < (\alpha_1 \cdot n_{\text{layers}})^\beta \cdot \sigma_{d_{xy}}$;
- $|d_z| < (\alpha_2 \cdot n_{\text{layers}})^\beta \cdot \sigma_{d_z}$;
- $d_{xy}/\delta_{xy} < (\alpha_3 \cdot n_{\text{layers}})^\beta$;
- $d_z/\delta_z < (\alpha_4 \cdot n_{\text{layers}})^\beta$;

where d_{xy} (d_z) and δ_{xy} (δ_z) denote the transverse (longitudinal) impact parameter and their uncertainties from the track fit covaraint matrix, n_{layers} is the number of tracker layers with a hit belonging to the track, the resolutions $\sigma_{d_{xy}} = \sqrt{a^2 + (b/p_T)^2}$ and $\sigma_{d_z} = \cosh(\eta)\sigma_{d_{xy}}$ on d_{xy} and d_z are parameterized using $a = 30\,\mu\mathrm{m}$ and $b = 10\,\mu\mathrm{m}$ and the values of α_i and β have different values depending on the tracking iteration.

3.1.2 Primary vertex reconstruction

The primary vertex (PV) reconstruction aims at measuring the position, and the associated uncertainty, of all proton-proton interaction vertices in each event [81]. To do this, the primary vertex reconstruction algorithm use all available reconstructed tracks and it consists of three steps: (i) selection of the tracks, (ii) clustering of the tracks that originate from the same interaction vertex, and (iii) performing a fit to measure the position of each vertex using its associated tracks.

The tracks associated with a collision vertex are selected requiring them to be compatible with being produced promptly in the primary interaction region. This is achieved by imposing requirements on the maximum value of significance of the transverse impact parameter relative to the center of the beam spot, the number of strip and pixel hits associated with a track, and the normalized χ^2 from a fit to the trajectory. No requirements on the p_T of the track ensures a high reconstruction efficiency.

The selected tracks clustering is based on the track z coordinate at their point of closest approach to the beam spot. To have good primary vertex reconstruction performance at high pileup, track clustering is performed using a deterministic annealing algorithm [82]. The track clustering algorithm identifies primary vertex candidates. The candidates containing at least two tracks are then fitted using an adaptive vertex fitter [83]. In addition to the position on x, y and z, to each primary vertex it is associated also the fit covariance matrix, the numbers of the degrees of freedom of the fit and the weights of the tracks used in the vertex reconstruction. The more a track is compatible with the hypothesis of being originated from the vertex candidate the more is its weight closer to 1, otherwise it is closer to 0. The number of degrees of freedom (dof) in the fit is defined as

$$n_{\text{dof}} = -3 + 2 \sum_{i=1}^{N_{\text{tracks}}} w_1 \,, \tag{3.1}$$

where w_i is the weight of the ith track, and the sum is extended to the number of all tracks associated with the vertex (N_{tracks}). The value of n_{dof} is strongly correlated with the number of tracks compatible with arising from the interaction region, i.e. it can be used to discriminate true proton-proton interactions.

3.1.3 Tracking for electrons

Electrons (or positrons) are reconstructed exploiting the information coming both from the ECAL and the tracker. The electrons can be naturally seeded making use of energy clusters in the ECAL. The energy and the position of the ECAL clusters can be then used to infer the position of the hits expected in the tracker, given the assumption that the cluster has been produced by an electron or a positron. However this approach has some significant weaknesses when electrons have emitted a fraction of their energy before reaching the ECAL, which is caused by the significant tracker thickness. Furthermore, for electrons in jets the energy and position of the associated supercluster are often biased by the overlapping contributions from other particle deposits.

To reconstruct electrons missed by the ECAL-based approach, a tracker-based electron seeding method was developed in the context of PF reconstruction. The iterative tracking is designed to have a large efficiency for these electrons. All the tracks from the iterative tracking are therefore used as potential seeds for electrons, if their $p_T > 2\,\text{GeV}$. The tracks are preselected using the number of hits and the fit χ^2 and are fit again using a Gaussian-sum filter [84]. The GSF fitting is more suited for electrons than the Kalman filter used in the iterative tracking, as it allows for sudden and substantial energy losses along the trajectory, caused by bremsstrahlung.

3.1.4 Tracking for muons

When muons reach the muon system, most of the other particles (except neutrinos) have been absorbed in the previous detection layers. This allows to identify muons with high efficiency and purity over the full detector acceptance, while the inner tracker provides a precise measurement of the transverse momentum of these muons. The muon object can be reconstructed in different ways, yielding to three different muon *types*:

- *standalone muons* are reconstructed making use of muon subsystems only. The hits in DT and CSC subsystems are combined to form track segments, then used as seeds for the pattern recognition and fitting in the muon spectrometer, that combines all compatible DT, CSC and RPC hits along the muon trajectory. The result of the final fitting is called a *standalone muon track*;
- *tracker muons* are reconstructed starting from inner tracker tracks. Each inner track with a p_T larger than 0.5 GeV and a total momentum $p > 2.5\,\text{GeV}$ is extrapolated to the muon system. The track is identified as a *tracker muon track* if at least one DT or CSC segment matches the extrapolated track. The track-segment matching is performed in a local (x, y) coordinate system defined in a plane transverse to the beam axis. The extrapolated track and the segment are matched either if, the absolute value of the difference along local x is smaller than 3 cm, or if the ratio of this distance to its uncertainty is smaller than 4;
- *global muons* are reconstructed combining information from both the muon system and the inner tracker. Each standalone muon track is matched to a track in the inner tracker if the parameters of the two tracks propagated onto a common plane are compatible. If the two tracks are compatible, the hits from the inner track and from the standalone muon track are refitted to form a *global muon track*. Combined fits allow to improve the momentum resolution with respect to the tracker-only fit at transverse momenta above few hundreds of GeV or more.

Global muon reconstruction is efficient for muons passing through more than one muon detector station, because it requires segments or hits to be associated in at least two muon detector stations. For momenta below 10 GeV, the effect of multiple scattering is larger, and muons are less likely to reach the outermost stations, therefore the tracker muon reconstruction is more efficient, since it requires only one segment match in the muon system. About 99% of the muons produced within the geometrical acceptance of the muon system are reconstructed either as a global muon or a tracker muon, and very often as both.

3.1.5 Calorimeter clusters

The purpose of the clustering algorithm is to: (i) detect and measure the energy and direction of stable neutral particles such as photons and neutral hadrons; (ii) distinguish between neutral and charged hadron energy deposits; (iii) reconstruct and identify electrons and all emitted photons from bremsstrahlung processes; and (iv) help the energy measurement of charged hadrons for which the tracks parameters were not determined accurately.

The clustering algorithm has been designed to have a high detection efficiency even for low-energy particles and to be able to separate close energy deposits. The clustering is performed separately in each subdetector: ECAL barrel and endcaps, HCAL barrel and endcaps, and the two preshower layers.

In the first step of the clustering algorithm, *cluster seeds* are identified as cells with an energy larger than a given threshold, and larger than the neighboring cells energy. The cells considered as neighbors are either the four closest cells, sharing a side with the seed candidate, or the eight closest cells, including cells that share only a corner with the seed candidate. In the second step, *topological clusters* are grown from the seeds by aggregating cells with at least a corner in common with a cell already in the cluster and with an energy above twice the noise level. In the ECAL seeds are additionally required to satisfy a threshold requirement on E_T. Energy clusters are then extracted within the topological cluster using an expectation-maximization algorithm based on a Gaussian-mixture model. Further details on the clustering algorithm can be found in Ref. [79].

3.2 Particle-flow identification and reconstruction

Each particle is expected to give rise to several PF elements in the various CMS subdetectors. The reconstruction of a particle is achieved with a *link algorithm* that connects the PF elements from different subdetectors. Depending on the elements nature, the link algorithm can test and link any pair of elements in the event, which are restricted in a (η, ϕ) region of interest to reduce the computing time. When two elements are linked, an element-element distance is defined that quantifies the quality of the link. The link algorithm then produces *PF blocks* of elements associated either by a direct link or by an indirect link through common elements.

In each PF block, the identification and reconstruction sequence proceeds in the following order. The muon candidates are first identified and reconstructed, and the corresponding PF elements (tracks and clusters) are removed from the PF block. The electrons are then identified and reconstructed with the aim of collecting the energy of all bremsstrahlung photons. Energetic and isolated photons are identified in the same step. The corresponding tracks and ECAL or preshower clusters are excluded from further consideration. The remaining elements in the block are then subject to a cross-identification of charged hadrons, neutral hadrons and photons, arising from parton fragmentation, hadronization and decays in jets. When all PF blocks have been processed and all particles have been identified, the reconstructed event is revisited by a post-processing step aiming at reducing particle misidentification and misreconstruction.

3.2.1 Muons

For muons the PF algorithm applies a set of selection criteria to candidates reconstructed with the standalone, global, or tracker muon algorithms. These selection criteria are based on various quality parameters from the muon reconstruction, as well as information from other CMS subdetectors. Further to this, it can use some of the variables based on muon reconstruction such as the track fit χ^2, the number of tracker or muon system hits per track and the degree of matching between the tracker muon track and the standalone muon track (in case of global muons). For tracker muons, the muon segment compatibility is computed by propagating the tracker track into the muon system and by evaluating the number of matched segments and the distance of the matching in position and direction. The muon segment compatibility is represented by a number ranging between 0 and 1: the closer is the number to one, the higher is the degree of matching. A kink-finding algorithm splits the inner track into two separate tracks at several places along the trajectory. For each split the algorithm makes a comparison between the two separate tracks, with a large χ^2 indicating that the two tracks are incompatible with being a single track. Other variables such as the compatibility of the muon with the primary vertex exploit information outside the muon reconstruction. Different sets of cuts correspond to different identification types, often referred to as "muon ID" [85]. The ones most commonly used for physics analysis are:

- *Loose muon ID* has been developed to identify prompt muons originating at the primary vertex, and muons from light and heavy flavor decays while constraining the rate of the misidentification of charged hadrons as muons. A loose muon is a muon selected by the PF algorithm that is also either a tracker or a global muon.

- *Medium muon ID* aims at prompt muons and muons from heavy flavor decay. A medium muon is a loose muon with the additional requirement that the inner track uses hits from more than 80% of the inner tracker layers it passes through. If the muon can be reconstructed as both a tracker and a global muon, it is required a muon segment compatibility greater than 0.303, a $\chi^2/n_{dof} < 3$, the inner track and standalone-muon matching $\chi^2 < 12$, and a maximum χ^2 computed by the kink-finding algorithm smaller than 20. For tracker-only muons the muon-segment compatibility is raised to 0.451 with no additional requirements.

- *Tight muon ID* is designed to suppress muons from decay in flight and from hadronic punch-through. A tight muon is a loose muon, which is reconstructed both as a tracker and a global muon, with a tracker track that uses hits from at least six layers of the inner tracker including at least one pixel hit. The global muon fit is required to have $\chi^2/n_{dof} < 10$ and include at least one hit from the two stations matched by the tracker muon. In addition it is required to be compatible with the primary vertex: the tracker track transverse (longitudinal) impact parameter significance with respect to the primary vertex must be $< 2(5)\,\mathrm{mm}$.

- *Soft muon ID* is optimized for low p_T muons and in general it is suited for flavor physics analyses. A soft muon is a tracker muon with a tracker track qualified as a *high-purity* track and using hits from at least six layers of the inner tracker including at least one pixel hit. The tracker muon track must have a tight muon-segment matching (i.e. having the ratio of the distance to its uncertainty in local x and in local y below 3) and be loosely compatible with the primary vertex: the tracker track transverse (longitudinal) impact parameter significance with respect to the primary vertex must be $< 0.3(10)\,\mathrm{cm}$.

- *High momentum muon ID* is optimized for muons with $p_T > 200\,\mathrm{GeV}$. A high momentum muon is required to be reconstructed as both tracker and global muon. The selection criteria are the same used for the tight ID, with the exception of the requirement on $\chi^2/n_{dof} < 10$ for the global track, which is removed to account for muons producing electromagnetic showers in the flux-return yoke hence producing additional hits in the muon chambers. In addition, the relative p_T uncertainty is required to be less than 30% to ensure a proper momentum measurement.

3.2.2 Electrons and isolated photons

Electron reconstruction combines information from the inner tracker and the calorimeters. Isolated photon reconstruction is conducted together with electron reconstruction because of the technical issue to be solved for the tracking, due to electromagnetic interaction within the tracker material, and because energy deposition patterns of electrons and photons are similar. An electron candidate is seeded from a GSF track, provided that the corresponding ECAL cluster is not linked to more than two additional tracks. A photon candidate is seeded from an ECAL supercluster with a transverse energy larger than 10 GeV and without any link to a GSF track.

Electron candidates must satisfy additional identification criteria. In particular, up to fourteen variables based on the GSF track and on the ECAL cluster information are combined in boosted-decision-tree (BDT) classifiers trained separately in ECAL barrel and endcaps acceptance, and for isolated and non-isolated electrons. Photon candidates must be isolated from other tracks and the associated clusters in the calorimeters must be compatible with a photon shower. The PF selection is looser than the typical requirements applied at the analysis level using isolated photons. All tracks and clusters used to reconstruct electrons and photons are masked against further processing, as well as tracks identified as originating from a photon conversion that have not been used for the electron reconstruction.

3.2.3 Hadrons and nonisolated photons

After muons, electrons and isolated photons are identified and removed from the PF blocks, the remaining particles to be identified are hadrons from jet fragmentation and hadronization. These particles may be detected as charged or neutral hadrons, as non-isolated photons, e.g. from π^0 decays, and more rarely as additional muons.

The ECAL and HCAL clusters not linked to any track give rise to photons and neutral hadrons. Within the tracker acceptance all ECAL clusters are turned into photon and all HCAL clusters are turned into neutral hadrons. Outside the tracker acceptance, ECAL clusters linked to an HCAL cluster are assumed to arise from the same hadron shower, while ECAL clusters without such a link are classified as photons. All tracks give rise to a charged hadron, the momentum and energy of which are directly taken from the corresponding track momentum, under the charged-pion mass hypothesis.

3.2.4 Jets

Jets are reconstructed with the anti-k_T algorithm [86, 87] with a distance parameter of 0.4. The algorithm clusters either all particles reconstructed by the PF algorithm, so-called *PF jets*, or the sum of the ECAL and HCAL energies deposited in the calorimeter towers, so-called *Calo jets*.

For PF jets, the momentum is determined as the vector sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5 to

10% of the true momentum over the whole p_T spectrum and detector acceptance. Particles from pileup collisions can contribute with additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions.

For Calo jets, the contribution from each calorimeter tower is assigned a momentum, the absolute value and the direction of which are given by the energy measured in the tower, and the coordinates of the tower. The raw jet energy is obtained from the sum of the tower energies, and the raw jet momentum by the vector sum of the tower momenta, which results in a nonzero jet mass. The raw jet energies are then corrected to establish a relative uniform response of the calorimeter in η and a calibrated absolute response in transverse momentum p_T.

3.2.5 Missing transverse momentum

The presence of particles that do not interact with the detector material, e.g. neutrinos, is indirectly detected by missing transverse momentum, or also referred to missing transverse energy [88]. CMS event reconstruction employs two distinct missing transverse energy reconstruction algorithms: the first one defines the missing transverse energy as the negative vector p_T sum of all PF candidates in the event; the second algorithm relies on the "pileup per particle identification" (PUPPI) [89, 90] method, and uses local shape information around each PF candidate in the event, pileup properties and tracking information to reduce the pileup dependence of jet and E_T^{miss} observables. The first algorithm is used in the majority of CMS analyses, since it provides a simple, robust and efficient estimate of the E_T^{miss} reconstruction.

3.3 CMS luminosity measurement

A precise measurement of the luminosity delivered to the CMS experiment by the LHC is crucial both during the data-taking operation and offline, when collected data are analyzed. The online luminosity measurement provides a real time feedback on the LHC performance and operation, as well as to CMS operations, e.g. if compared to trigger rate measurement. Offline, the luminosity measurement is a fundamental quantity to several physics analyses [91].

A total of seven systems are used for measuring the luminosity in CMS. The pixel luminosity telescope (PLT) [92] and the fast beam conditions monitor (BCM1F) [93] are dedicated for luminosity system. The hadronic forward calorimeter contribute as well to measuring the luminosity using a dedicated readout on an existing system. These three detectors use a separate data acquisition (DAQ) system, called BRIL-DAQ, which operates independently of the main CMS readout, so that it can provide luminosity measurements despite the ongoing CMS operating state. In addition, the following three methods use data from existing CMS detectors to perform a luminosity measurement using the main CMS DAQ system: the drift tube luminosity, the pixel cluster counting method (PCC) and the vertex counting method (VTX). Finally, the RAMSES (Radiation Monitoring System for the Environment and Safety) detector, even though its main function is to ensure the safety of personnel, also provides a luminosity measurement.

Each luminometer reads out a rate (R) of the specific quantities observed in the detector, which is expected to be proportional to the instantaneous luminosity $\mathcal{L}_{\mathrm{inst}}$

and the visible cross section σ_{vis}:

$$R = \mathcal{L}_{\text{inst}} \cdot \sigma_{\text{vis}} \, . \tag{3.2}$$

In reality, the luminometers show some non-linear dependence on the instantaneous luminosity due to pileup or to external factors such as the LHC filling scheme. In order to get an accurate luminosity measurement, these non-linear behavior is taken into account.

The measurement of the visible cross section is performed by van der Meer (VdM) scans performed with a dedicated LHC machine setup. The VdM scans consist of moving the two separate LHC beams across each other, and measuring the rate as a function of the beam separation. This measurement is used to derive the beam overlap width, thus the absolute luminosity value can be derived as a function of the machine parameters [94] and the measured rate can be used to derive σ_{vis}. This measurement can be used to determine the luminosity in regular physics collisions, during which the beam overlap width cannot be measured with the necessary accuracy.

3.4 Monte Carlo simulation

The production chain of a simulated sample in CMS is performed in the following steps:

1. event generation (“GEN” step);

2. simulation of the interaction between the final state particles and the CMS detector (“SIM” step);

3. the information on the interaction of the final state particles with the detector is used to simulate the response of the detector electronics (“DIGI” step), and the contribution from the expected pileup interactions is added (“PU mixing”);

4. the full CMS trigger chain is simulated using the simulated detector response;

5. the CMS reconstruction software is run on the simulated events and reconstructed events are saved in analysis data object (AOD) format that can be used for physics analysis (“RECO” step);

6. the data in AOD format are further refined and skimmed to produce much lighter data samples that, at the compromise of losing part of the information carried by the data, serve most of the CMS data analysis needs. These data formats are called MINIAOD and NANOAOD, the latter being the lightest data format.

Except for the event generation and the simulation of the particle interaction with the detector, all the production steps match exactly what happens during a real data-taking operation. Different MC sample production workflows, or “campaigns”, are available to adapt the simulations to the changes in data-taking conditions and CMS software updates.

Event generation

A Monte Carlo event generator is capable of simulating a wide range of the most interesting processes that are expected at the LHC [95]. Monte Carlo simulation of high energy physics processes and of particle interactions with the detector is an indispensable tool for the study of LHC events. Several Monte Carlo event generators are used to simulate the relevant physics processes expected to contribute to LHC collisions. Event generators are usually required to extract a signal of new physics from the background of SM processes or for comparisons with the data. They also provide realistic input for the design of new experiments, or for the selection of reconstruction procedures within an existing experiment.

For the simulation of a hadron-hadron collision, several basic steps of the process need to be simulated:

- the primary hard sub-process;
- parton showers associated with the incoming and outgoing colored participants in the sub-process;
- hadronization of the outgoing partons to final state hadrons;
- secondary interactions contributing to the underlying event;
- the decays of unstable particles that do not escape from the detector.

The result is a collection of generated final state particles. Given the significant event pileup obtained in LHC proton-proton collisions this is also a key element included in the simulations.

Detector response, trigger simulation and event reconstruction

After the generation of the event, the interaction of the particles with the detector material has to be simulated. This task is performed at CMS using GEANT4 [96], a C++ object-oriented toolkit that provide the simulation of the geometry of the detector and the description of the particle interactions in the detector material.

To reduce the computing time, the contribution from pileup interaction is added only at the DIGI step. This is done by overlaying, to the signal produced by the event of interest, the signal produced by inelastic non-diffractive QCD events (referred to as “minimum bias”), and reproducing the expected distribution of pileup events. The minimum bias events are generated only once and used for every sample of the same production campaign, resulting in a less CPU-expensive workflow. Only at this stage, first the L1 trigger emulation, afterwards the HLT software and finally the reconstruction software are run to perform the full event reconstruction.

Each MC sample is handled centrally by CMS and it undergoes a process of validation and approval before going to production. Every MC sample produced centrally is available to the whole collaboration.

Chapter 4

Search for heavy neutral leptons

As discussed in Chapter 1, among all the theories that could describe the existence of a neutrino mass term, the νMSM theory provides the simplest extension of the standard model that could explain the smallness of the neutrino masses, provide a dark matter candidate and give a possible explanation to the baryon asymmetry. The νMSM predicts the existence of three sterile right-handed neutrinos. While the lightest of the three could be a dark-matter candidate, the remaining two could be potentially detected in CMS. The smallness of the active neutrino masses can be explained through the see-saw mechanism. This means that the sterile right-handed neutrinos are expected to be much heavier than the active left-handed neutrinos and the matrix elements describing the active-sterile neutrino mixing are expected to be small.

Various experimental setups have provided stringent limits on the active-sterile neutrino mixing matrix elements in the hypothesis of a mixing to electron ($|V_e|$), muon ($|V_\mu|$) or tau ($|V_\tau|$) neutrinos, as a function of the sterile neutrino mass (see Sec. 1.5). The existence of sub-eV sterile neutrinos can be probed via neutrino oscillations measurement in the sub-eV range, while neutrinoless double-β decays can be used for masses below $\sim 1\,\text{MeV}$. The decay of light, strange or charmed mesons have been used especially in beam-dump facilities for heavy neutral lepton searches and provided the most stringent constraints below $\sim 1\,\text{GeV}$.

The LHC experiments are particularly suited for probing the higher part of the mass spectrum. The CMS [38] and ATLAS [39] collaborations recently published heavy neutral lepton searches from W decay that allowed to set limits up to $\sim 100\,\text{GeV}$. The LHCb collaboration has performed a heavy neutral lepton search targeting the lepton flavor violating $B^- \to \pi^+\mu^-\mu^-$ decay [12], which allowed to constrain the sterile neutrino for masses below the B meson mass. Also the Belle collaboration explored the same heavy neutral lepton mass range exploiting B meson decays [35]. No evidence of a sterile right-handed neutrino has been found so far.

While a search for heavy neutral leptons from B decays is currently ongoing in CMS, there are no public results on heavy neutral lepton searches from charmed meson decays at LHC experiments. As discussed in Sec. 1.6, among the charmed mesons, the two-body decays of D_s mesons are the most promising heavy neutral lepton (HNL) source and allow to explore the relatively widest mass range. Among the two-body D_s decay that could involve a heavy neutral lepton N, the $D_s^+ \to N\mu^+$ decay, and the subsequent $N \to \mu^+\pi^-$ decay, represents the cleanest experimental signature for the CMS experimental setup. This decay channel would be able to constrain the $|V_\mu|^2$ matrix element in the mass region below the D_s mass. Final states with electrons or τ leptons, although being of physics interest, would be

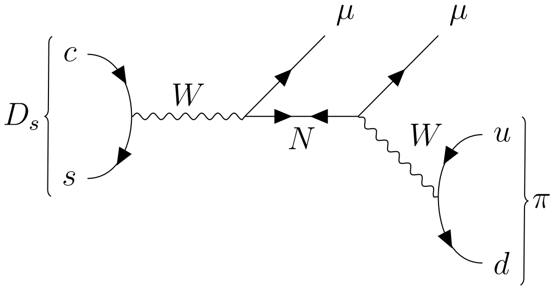

Figure 4.1: Feynman diagram for $D_s \to \pi\mu\mu$ decay via Majorana neutrino labeled N.

experimentally more challenging and would provide a smaller yield because of the presence of physics objects reconstructed with a lower efficiency than muons.

The most stringent limits on the branching fractions of the decays $D_s^\pm \to \pi^\pm\mu^\mp\mu^\pm$ and $D_s^\pm \to \pi^\mp\mu^\pm\mu^\pm$ have been set by LHCb Collaboration to 1.8×10^{-7} and 8.6×10^{-8}, at 90% confidence level (CL) [97]. However, the limits reported by LHCb do not consider the possibility for the decays to be mediated by an on-shell heavy neutral lepton, meaning that no information can be inferred on its mass and its coupling with SM particles.

In this work, the heavy neutral lepton is considered to be produced in the $D_s^+ \to N\mu^+$ decay, as shown in Fig. 4.1. The HNL later decays into a displaced $\mu\pi$ vertex, which represents the main signal signature. Background events coming from known D_s decays can be reduced by requiring the $\mu\pi$ pair to come from the same displaced vertex and requiring it to be well detached from the D_s vertex, making the combination of uncorrelated $\mu\pi$ pairs the main source of background events. If the decay happens inside the detector acceptance, both the D_s and the HNL can be fully reconstructed. Since the HNL can be a Majorana neutrino, the two muons from the D_s decay can have the same electric charge, allowing for a lepton flavor violating process. The analysis takes into account two possible neutrino masses, 1.5 and 1.0 GeV. The range of neutrino lifetimes considered, expressed as proper mean decay length $c\tau$, ranges from 1 cm to 1 m. Throughout the text, the proper mean decay length $c\tau$ will be referred to, as an abuse of language, as lifetime. Given that the neutrino lifetime depends on its mass m_N and the active-sterile neutrino mixing matrix element $|V_\mu|$ as $c\tau \sim m_N^{-5}|V_\mu|^{-2}$, larger neutrino lifetimes allow to explore smaller values of $|V_\mu|^2$. The sterile neutrino mass range taken into account in this analysis allows to probe the $|V_\mu|^2$ value down to $\sim 10^{-5}$. The search for HNL from D_s decay is performed by doing a bump hunt in the $\mu\pi$ invariant mass spectrum.

The analysis uses the B-Parking data set, that was recorded by the CMS detector during the 2018 data-taking period using proton-proton collisions at a center-of-mass energy of 13 TeV. The B-Parking data set (described in detail in Sec. 2.4.3) has been collected by triggering on low-p_T displaced muon. Although being designed to record a large unbiased b hadron sample, by targeting semileptonic b hadron decays, the B-Parking trigger is also well-suited for the search of displaced $N \to \mu^+\pi^-$ decays. The B-Parking data set corresponds to a total integrated luminosity of 41.7 fb^{-1}.

4.1 Search strategy

In this analysis the HNL search is performed exploiting the D_s decays. The choice of considering only the exclusive $D_s^+ \to N(\to \mu^\pm\pi^\mp)\mu^+$ decay has many advantages.

First of all, the low-p_T muons from the D_s decay can be selected by the B-Parking displaced muon triggers. This makes available a large event data set. A final state involving low momentum muons, rather than electrons, can profit from the excellent muon reconstruction and identification performance of the CMS detector. Furthermore, both the HNL and the D_s decays can be fully reconstructed allowing to put a cut on the reconstructed D_s mass and reject combinatorial background, and to search for the HNL performing a bump hunt over the $\mu\pi$ invariant mass spectrum. The mass of the HNL that can be probed using this decay is mostly limited by the mass of the D_s meson. The analysis is performed using a cut-based selection strategy, which is expected to have a good performance in the small mass range explored. While the analysis targets a specific D_s decay mechanism, the Ds mesons can be produced both promptly and as a result of a B meson decay. The scenario in which D_s are only produced promptly in the proton-proton collision has also been studied in Sec. 4.9 as a cross-check.

On-shell HNLs mediating the $D_s^+ \to N(\to \mu^\pm\pi^\mp)\mu^+$ decay are reconstructed by requiring the $\mu\pi$ pair to form a common vertex that can be combined with an additional muon into a good D_s candidate. The HNL decay is required to be within the tracker volume. This is the main limiting factor of the analysis which is less efficient as the HNL lifetime becomes larger, since more HNL are expected to decay outside the tracker volume. Requiring that the $N \to \mu^+\pi^-$ vertex is well detached from the $D_s^+ \to N\mu^+$ vertex allows to highly reduce background events from known SM processes, such as resonant and nonresonant $D_s \to h\mu\mu$ and $D_s \to hhh$ decays, where h is a pion or a kaon mis-reconstructed as a muon (the so-called *fake* muons). Signal events are expected to be mostly contaminated by combinatorial background, which can be modeled by simulated QCD events. Background MC samples can be used to adjust the analysis selection requirements and optimize the discovery sensitivity. On the other hand, a data-driven method is used to make an estimate of the background yield while avoiding MC mismodeling: for each HNL mass point the number of expected background events is estimated by fitting the $\mu\pi$ invariant mass distribution in the sidebands.

The number of produced D_s mesons can be measured on the data using the $D_s^+ \to \phi\pi^+$ decay channel, which has the same particles in the final state and a well-known branching ratio. The $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ decay channel will be referred to as "normalization channel". The expected signal yield associated with the HNL being produced in the $D_s^+ \to N\mu^+$ decay $N_{D_s^+\to N\mu^+}$ can be written as

$$N_{D_s^+\to N\mu^+} = \mathcal{L}\sigma(D_s)\mathcal{B}(D_s^+ \to N\mu^+)\mathcal{B}(N \to \mu^+\pi^-)\varepsilon_{D_s^+\to N\mu^+}\,, \tag{4.1}$$

while, the number of D_s events in the normalization channel $N_{D_s^+\to\phi\pi^+}$ is:

$$N_{D_s^+\to\phi\pi^+} = \mathcal{L}\sigma(D_s)\mathcal{B}(D_s^+ \to \phi\pi^+)\mathcal{B}(\phi \to \mu^+\mu^-)\varepsilon_{D_s^+\to\phi\pi^+}\,. \tag{4.2}$$

In the equations above, $\mathcal{L}$ is the integrated luminosity, $\sigma(D_s)$ is the D_s production cross section (including both prompt production and decays from heavier hadrons), $\mathcal{B}$ indicates the branching ratio of the decay under consideration, $\varepsilon_{D_s^+\to N\mu^+}$ and $\varepsilon_{D_s^+\to\phi\pi^+}$ indicate the total selection efficiency of the signal and the normalization channel, calculated as the fraction of generated events that survives all analysis selection cuts. By combining equations 4.1 and 4.2, the expected signal yield reads as:

$$N_{D_s^+\to N\mu^+} = N_{D_s^+\to\phi\pi^+}\frac{\mathcal{B}(D_s^+ \to N\mu^+)\mathcal{B}(N \to \mu^+\pi^-)}{\mathcal{B}(D_s^+ \to \phi\pi^+)\mathcal{B}(\phi \to \mu^+\mu^-)}\frac{\varepsilon_{D_s^+\to N\mu^+}}{\varepsilon_{D_s^+\to\phi\pi^+}}\,. \tag{4.3}$$

A key advantage of measuring the D_s yield with respect to other decays is that it allows the cancellation of systematic uncertainties on the luminosity, the production cross section and, at least partially, on the selection and reconstruction efficiencies of the signal and normalization channels. The trigger and reconstruction efficiencies may be different in data and simulation, meaning that the total selection efficiency has to be scaled by a factor that takes into account the difference between data and simulation.

The selection efficiency is evaluated by means of simulated samples: it is the ratio between the number of events that survive all analysis requirements and the number of generated MC events. The D_s yield in the normalization channel can be measured in data (see. Sec. 4.4.1), while the expected branching fractions for the signal and normalization channels are external inputs from theoretical predictions (see Sec. 1.7) and measurements, respectively. The signal selection optimization as well as the signal efficiency measurements in simulation is presented in Sec. 4.5.5.

For any of the HNL mass points, the number of signal events is extracted from the data using an unbinned maximum likelihood (UML) fit of the HNL invariant mass. A limit on the value of $|V_\mu|^2$ is eventually set using asymptotic approximations [98].

To avoid unconscious biases, the analysis used a “data blinding” technique. For each HNL mass point taken into consideration, all optimization studies were performed on data excluding events compatible with the HNL mass. Once the selection criteria and measurement procedure will be settled, the data will be unblinded.

4.2 Data and Monte Carlo samples

The analysis presented in this thesis uses proton-proton collision data collected at $\sqrt{s} = 13\,\text{TeV}$ during 2018 data-taking period with the B-Parking trigger. The structure of the B-Parking data set will be presented in Sec. 4.2.1, while the simulated samples used for the analysis will be discussed in Sec. 4.2.2.

4.2.1 Data

Events are selected with a trigger logic that requires a single displaced muon, the so-called “B-Parking” trigger [78]. The overall trigger logic is designed to exploit the last part of an LHC fill, when the instantaneous luminosity decreases and the B-Parking trigger can exploit the spare L1 bandwidth as the rate of triggers designed for high-p_T objects decreases. The L1 trigger logic requires either a higher p_T muon ($p_T > 22\,\text{GeV}$) or muon with lower p_T thresholds restricted in pseudorapidity to $|\eta| < 1.5$. Lower p_T threshold evolves with the instantaneous luminosity. At the HLT level, the muon is then required to have an impact parameter significance (IPS) above a value that changes as the instantaneous luminosity decreases. The B-Parking data-taking mode has been discussed in detail in Sec. 2.4.3.

Several HLT paths are defined with different p_T and IPS thresholds, respectively ranging from 7 to 12 GeV and from 3.5 to 6. A summary of all HLT paths is reported in Tab. 4.1, along with the integrated luminosity collected by each path for each data-taking era, labeled as A, B, C, and D. The HLT paths are indicated as HLT_MuX_IPY, where X refers to the muon p_T threshold and Y to the threshold on its impact parameter significance. Each HLT path is divided into six (five) parts in eras A and B (C and D), and each path has a prescale value identical in all parts and such that the sum of all parts results in a prescale that is either 0 or 1.

HLT path	**Integrated Luminosity ($\mathbf{fb^{-1}}$)**				
	era A	era B	era C	era D	Total
HLT_Mu7_IP4	0.0	0.365	0.157	6.418	6.940
HLT_Mu8_IP3	0.442	0.718	0.017	0.406	1.583
HLT_Mu8_IP5	0.0	0.0	0.006	8.254	8.259
HLT_Mu8_IP6	0.0	0.0	0.006	8.254	8.259
HLT_Mu8p5_IP3p5	0.320	0.0	0.0	0.0	0.320
HLT_Mu9_IP4	0.0	0.0	0.006	0.004	0.010
HLT_Mu9_IP5	0.0	1.367	2.698	16.885	20.89
HLT_Mu9_IP6	4.642	4.323	3.838	20.867	33.669
HLT_Mu10p5_IP3p5	0.320	0.0	0.0	0.0	0.320
HLT_Mu12_IP6	0.0	2.673	5.515	26.603	34.791

Table 4.1: Total integrated luminosity for each B-Parking trigger path divided by data-taking era.

(p_T,IPS)	(7,4)	(8,6)	(8,5)	(8,3)	(8.5,3.5)	(9,6)	(9,5)	(9,4)	(10.5,3.5)	(12,6)
(7,4)	6.940	4.931	4.931	0.788	-	6.940	6.940	0.010	-	6.940
(8,6)	4.931	8.259	8.259	0.258	-	8.259	8.259	0.010	-	8.259
(8,5)	4.931	8.259	8.259	0.258	-	8.259	8.259	0.010	-	8.259
(8,3)	0.788	0.258	0.258	1.583	-	1.583	0.788	0.010	-	0.788
(8.5,3.5)	-	-	-	-	0.320	0.320	-	-	0.320	-
(9,6)	6.940	8.259	8.259	1.583	0.320	33.669	20.950	0.010	0.320	26.768
(9,5)	6.940	8.259	8.259	0.788	-	20.950	20.89	0.010	-	20.950
(9,4)	0.010	0.010	0.010	0.010	-	0.010	0.010	0.010	-	0.010
(10.5,3.5)	-	-	-	-	0.320	0.320	-	-	0.320	-
(12,6)	6.940	8.259	8.259	0.788	-	26.768	20.950	0.010	-	34.791

Table 4.2: Integrated luminosity (fb^{-1}) shared between every pair-wise combination of the B-Parking trigger paths. The HLT path p_T and IPS cuts are specified in the (p_T,IPS) parentheses.

To compute the total integrated luminosity of the B-Parking data set, the overlap between different HLT paths has to be considered. The total integrated luminosity shared between each B-Parking HLT path pair is shown in Tab. 4.2, where only the p_T and the IPS of the HLT paths are indicated. The total integrated luminosity collected using the B-Parking trigger is $41.7\mathrm{fb}^{-1}$, and it has been computed after removing the overlaps between HLT paths: all luminosity sections for which one of the B-Parking trigger paths is switched on are taken, and their luminosity is summed, but whenever two or more trigger paths share the same luminosity section, the luminosity is taken only once. A summary of the integrated luminosity collected by the B-Parking trigger is shown in Tab. 4.3, where the integrated luminosity is shown separately for each era and part.

4.2.2 Simulated samples

Simulated samples are crucial in many analysis steps: background shape modeling and signal selection optimization, acceptance and efficiency measurement, systematic uncertainties estimation. Six signal samples have been used, corresponding to two possible neutrino masses, $m_N = 1.0, 1.5\,\mathrm{GeV}$, and three possible neutrino lifetimes, $c\tau_N = 10, 100, 1000\,\mathrm{mm}$. Two sets of signal samples have been produced to explore both the scenario where the D_s is produced promptly and the inclusive D_s

		B-Parking era				All eras
		A	B	C	D	
$\mathcal{L}_{\text{int}}$ (fb^{-1})	part 0	0.774	0.911	1.103	5.321	8.108
	part 1	0.774	0.911	1.103	5.321	8.108
	part 2	0.774	0.911	1.103	5.321	8.108
	part 3	0.774	0.911	1.103	5.321	8.108
	part 4	0.774	0.911	1.103	5.321	8.108
	part 5	0.774	0.377	-	-	1.150
	Total	4.644	4.932	5.515	26.605	41.7

Table 4.3: Integrated luminosity (fb^{-1}) of the B-Parking data set in each era and part, as well as the sum among all eras of the same part (last column) and the sum on the single era (last row).

production scenario. The $D_s^+ \to \phi\pi^+$ decay, with ϕ decaying to $\mu^+\mu^-$, is used to measure the acceptance and efficiency of the normalization channel. The $D_s^+ \to \phi\pi^+$ generated sample takes into account D_s mesons from any production source, i.e. both prompt and from decays-in flight. Also a $D_s^+ \to \phi(\to K^+K^-)\pi^+$ sample, where the D_s comes from the B_s decay, has been used to estimate the possible contribution from K misidentified as muons. The MC samples mentioned above have been produced using PYTHIA8.240 [99] with the CP5 tune [100], which generates the hard event and handles both the hadronization of the partons and the particles decay, except of the decay of interest, which is handled by the EVTGEN1.6.0 software [101]. The PYTHIA CP5 tune corresponds to a proper choice of values of the strong coupling used in the modeling of the initial- and final-state radiation, hard scattering, and multiparton interactions, as well as the order of its evolution as a function of the four-momentum squared. The $D_s^+ \to N\mu^+$ decay and the subsequent $N \to \mu^+\pi^-$ decay are generated using the EVTGEN phase space model, which assumes no spin correlation between the particles involved.

The background from SM is simulated by generating events with jets in the final state, here referred to as quantum chromodynamics (QCD) dijet events. The QCD dijet events have been used to model the distribution of the background events for the variables that are expected to discriminate signal from background, and optimize the signal selection. For this purpose, QCD dijet events allow to have a good description of the combinatorial background, which is expected to be the main source of background. On the other hand, the background estimation for testing the existence of a HNL is data-driven, and does not rely on the MC modeling of the background. In simulated QCD dijet events, the range of transverse momenta in the rest frame of the $2 \to 2$ process, here denoted as $\hat{p}_T$, varies from 20 to 300 GeV. Since the production cross section depends on $\hat{p}_T$, the event generation is more efficient at low $\hat{p}_T$. For generation efficiency reasons, at the production stage, the QCD dijet sample has been divided into different samples depending on $\hat{p}_T$. The range 20 to 300 GeV is covered by six different samples, each corresponding to a $\hat{p}_T$ sub-range. All background samples have been generated using PYTHIA with the CP5 tune.

After event generation, additional conditions are required for the event to be finally stored. These requirements depend on the generated samples and generally are requirements on p_T and η of the final-state particles to make them fall into detector acceptance and to meet the trigger conditions, or requirements for final-state particles to be generated from the decay of interest. These requirements are often referred to as “generator filters”, and the fraction of events that survive such requirements is called filter efficiency. The QCD samples are "muon enriched", meaning that events are filtered at generator level based on the requirement of at least one muon with $p_T > 5\,\text{GeV}$ and $|\eta| < 2.5$. The simulated samples used in the analysis are summarized in Tab. 4.4. For each sample, the corresponding number of total events produced, its cross section, and the production filter efficiency are reported. The reported cross section was also calculated using the PYTHIA software, and it refers to the production process only. In Tab. 4.4, whenever the cross section reported is identical to that of the previous row, it is indicated with the "-" symbol. A summary of the generator settings and filters used for each of the simulated processes can be found in Appendix A.

The simulation of pileup interactions in the heavy neutrino signal samples have been generated so that its distribution is closer to the one observed in the B-Parking trigger data set. As a matter of fact, the observed mean number of pileup interactions in the B-Parking data set is smaller than the one observed in standard CMS physics streams. For this reason, CMS has developed a dedicated pileup distribution profile to produce simulated samples with an appropriate pileup distribution for B-Parking data set analysis. Since only the heavy neutrino signal samples were produced for the purpose of this analysis, the other MC samples were generated with the standard data-taking condition pileup profile, because they have a broader use within the CMS Collaboration.

4.3 Objects Definition

The D_s candidates, both in the signal channel and in the normalization channel, are built using three objects in the final state: two muons and a pion. After the reconstruction process, the data are further refined and skimmed to produce the MINIAOD data format, which is used for this work. In this section, the objects used to reconstruct the D_s candidates are defined.

4.3.1 Muons

Reconstructed muon objects are stored at the MINIAOD level if they satisfy at least one of the following conditions: (i) muon $p_T > 5\,\text{GeV}$; (ii) muon $p_T > 3\,\text{GeV}$ and pass any of the following requirements: is a *global* muon, is a *tracker* muon, is a *standalone* muon, pass *RPCMuLoose* condition (i.e. requiring the muon to match at least two hits in the RPCs); (iii) the muon passes the *PF* ID [102].

For the purpose of this work, the following requirements are added to the reconstructed muons:

- $p_T > 3.5\,(2.0)\,\text{GeV}$ if $|\eta| < 1.2$ $(1.2 < |\eta| < 2.4)$: such cuts are meant to allow the muon to reach the muon chambers (also in the barrel);
- pass *Soft* ID [85]: this muon identification is typically used in B-physics analyses in CMS (short description is given in Sec. 3.2.1);

4.3.2 Pions

Among all candidates reconstructed by the PF algorithm, only the objects fulfilling the following requirements are used as pion candidates:

- the track must satisfy the *highPurity* requirement;

Process	$\boldsymbol{N_{\mathrm{TOT}}}$	$\boldsymbol{\sigma}$ **(pb)**	$\boldsymbol{\varepsilon_{\mathrm{filter}}}$ **(%)**
$D_s^{\pm} \to N(\to \mu\pi)\mu^{\pm}$ (prompt)			
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 10\,\mathrm{mm}$	126247	2.4×10^9	0.05
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 100\,\mathrm{mm}$	239566	-	0.045
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 1000\,\mathrm{mm}$	219247	-	0.053
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 10\,\mathrm{mm}$	268863	-	0.051
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 100\,\mathrm{mm}$	269755	-	0.054
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 1000\,\mathrm{mm}$	270859	-	0.041
$D_s^{\pm} \to N(\to \mu\pi)\mu^{\pm}$ (inclusive)			
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 10\,\mathrm{mm}$	215946	55.6×10^9	0.005
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 100\,\mathrm{mm}$	213340	-	0.005
$m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 1000\,\mathrm{mm}$	214422	-	0.003
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 10\,\mathrm{mm}$	277288	-	0.004
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 100\,\mathrm{mm}$	255234	-	0.005
$m_{\mathrm{HNL}} = 1.0\,\mathrm{GeV}; c\tau_{\mathrm{HNL}} = 1000\,\mathrm{mm}$	259243	-	0.003
$D_s^{\pm} \to \phi(\mu\mu)\pi^{\pm}$ (inclusive)	2964234	78.050×10^9	0.0160
$B_s \to D_s^{\pm} \to \phi(KK)\pi^{\pm}$	2989965	0.454×10^9	0.23
Muon enriched QCD dijet			
$20 < \hat{p}_T < 30\,\mathrm{GeV}$	60640516	3.977×10^8	0.65
$30 < \hat{p}_T < 50\,\mathrm{GeV}$	58627984	1.070×10^8	1.25
$50 < \hat{p}_T < 80\,\mathrm{GeV}$	40022458	1.573×10^7	2.35
$80 < \hat{p}_T < 120\,\mathrm{GeV}$	45566475	2.341×10^6	3.81
$120 < \hat{p}_T < 170\,\mathrm{GeV}$	39114558	4.086×10^5	5.17
$170 < \hat{p}_T < 300\,\mathrm{GeV}$	71926577	1.036×10^5	6.83

Table 4.4: Total events (N_{TOT}), cross section (σ) and filter efficiency ($\varepsilon_{\mathrm{filter}}$) of simulated samples. When present, the labels *prompt* and *inclusive* indicate the D_s production from prompt $pp \to D_s X$ interactions or if it includes also the D_s production from decays in flight of heavier hadrons. Whenever the cross section reported is identical to the one of the previous row it is indicated with the - symbol.

- the PDG ID assigned by the reconstruction algorithm must be that of a charged hadron;
- $p_T > 0.5\,\text{GeV}$ and $|\eta| < 2.4$.

4.3.3 Trigger matching

In both data and MC, at least one of the muons that form the D_s candidate must be the muon that fired the B-Parking trigger. A muon is said to be matched to a trigger object if the following conditions are met:

- The trigger object is used in the final filter of the HLT path;
- The trigger object is used in the L3 filter of the HLT path;
- The muon and the trigger object $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ must be smaller than 0.005;
- The muon to be matched with the trigger object is required to have η and p_T compatible with the cuts applied at the L1 trigger and HLT level, respectively.

4.4 Signal yield normalization

The number of expected signal events is evaluated from the data by means of the normalization channel as described in Sec. 4.1. The D_s yield in the normalization channel is measured using data collected with the same trigger logic as in the signal channel.

4.4.1 Measurement of $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ yield

The $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ candidates are selected by requiring the presence of at least two oppositely charged muons compatible with the $\phi \to \mu^+\mu^-$ decay, meaning the dimuon invariant mass must be within a 0.05 GeV window centered on the known ϕ mass and the vertex fit probability must be greater than 0.01. The invariant mass distribution of the reconstructed ϕ mesons is shown in Fig. 4.2, and a clear peak is visible at 1.02 GeV. For each muon pair forming a good $\phi \to \mu^+\mu^-$ candidate, the two muons are combined with a π candidate to form a D_s candidate, requiring the muon pair to be constrained to the ϕ mass and having a $\mu\mu\pi$ vertex fit probability above 0.01. Both D_s and ϕ candidates are reconstructed using a kinematic fitter described in Ref. [103]. For each event, all $\mu\mu\pi$ combinations forming a good D_s candidate are selected if the fitted D_s candidate has a mass between 1.75 and 2.15 GeV.

The D_s decay vertex is expected to be relatively well separated from the PV, thus it can usually be fully reconstructed. Many useful variables are built by combining the position of the PV with the D_s decay vertex and are used throughout the analysis. Among all PVs reconstructed in an event, the PV associated with the D_s production is selected as the one that has the smallest distance along z with respect to the leading muon track, i.e. the one having the largest p_T.

After applying the selection requirements, it can happen that more than one D_s candidate is selected in a single event. Among all selected $\mu\mu\pi$ combinations forming a good D_s candidate, different criteria have been tested to select the optimal one. The different criteria consisted in selecting the best D_s candidate as the one having:

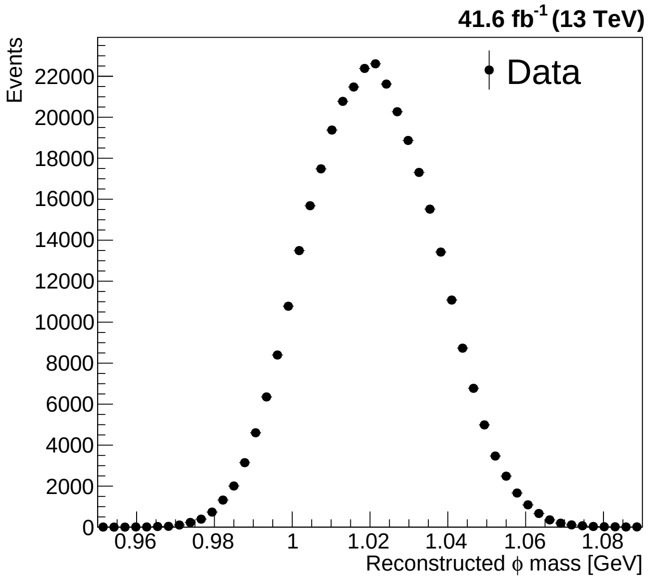

Figure 4.2: Reconstructed mass of the ϕ meson candidates used to build the $D_s^+ \to \phi\pi^+$ candidates.

- the largest p_T:
- the largest fit probability of the D_s vertex;
- the cosine of the “back-pointing” angle ($\cos(\alpha_{D_s})$), i.e. the projection on the xy plane of the angle between the D_s momentum and the PV-D_s vertex vector, closest to 1;
- the largest significance of the D_s vertex in the xy plane, i.e. the ratio of its decay length (L_{xy}) and the associated uncertainty ($\sigma_{L_{xy}}$) on the transverse plane.

The performance of these selection requirements has been tested on simulation using as a figure of merit the fraction of events with a selected D_s candidate matched with generator-level information. All variables taken into account could select a fraction of generator-level matched candidates above 98% and have difference in the performance below few 0.1%. Given the similar performances, the optimal D_s candidate is selected as the one having the highest p_T.

The D_s yield is extracted in the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ channel by performing an unbinned maximum likelihood fit of the reconstructed $\mu\mu\pi$ invariant mass, in the 1.75 to 2.15 GeV region. In this mass region, both the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ and the $D^+ \to \phi(\to \mu^+\mu^-)\pi^+$ signal peaks are present. The D and D_s signal probability density functions (PDFs) are modeled using Voigtian functions, i.e. the convolution of a Gaussian distribution, indicated as $G(x;\mu,\sigma)$ and a Lorenz, or Breit-Wigner, distribution, indicated as $L(x;\mu,\Gamma)$:

$$V(x;\mu,\sigma,\Gamma) = \int_{-\infty}^{+\infty} G(x';\mu,\sigma)L(x-x';\mu,\Gamma)dx' \,, \tag{4.4}$$

while the background shape is modeled using an exponential function, namely

$$B(x;\lambda) = \lambda e^{-\lambda x} \,. \tag{4.5}$$

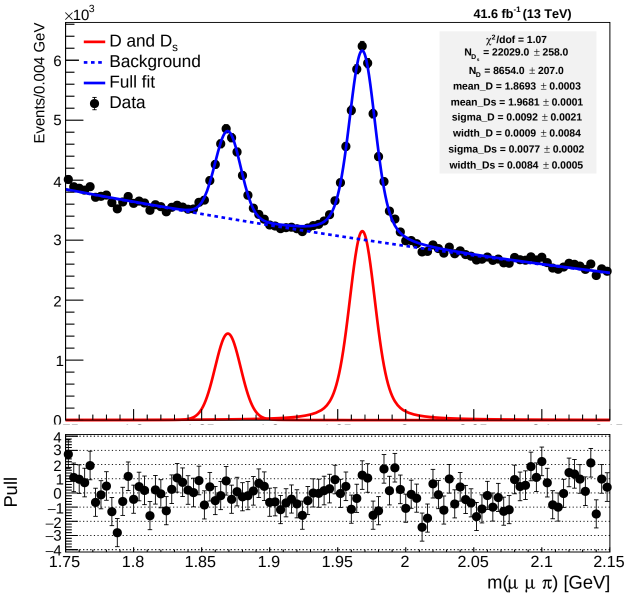

Figure 4.3: The $\mu\mu\pi$ invariant mass distribution with fits to the D and D_s peaks and the background. The bottom panel shows the pulls of the data with respect to the fit curve.

The μ and σ parameter of the Gaussian PDF indicates its mean and standard deviation, the Γ parameter of the Voigtian indicates its width while the λ parameter of the exponential is related to its steepness.

For simplicity, the signal and background PDFs parameters can be indicated as $\vec{\theta} = (\mu_D, \mu_{D_s}, \sigma_D, \sigma_{D_s}, \Gamma_D, \Gamma_{D_s}, \lambda)$, and the N mass measurements as m_i, where $i = 1, \ldots, N$. Then the extended unbinned likelihood function used to fit the data can be written as:

$$L(m_i; \vec{\theta}, N_D, N_{D_s}, N_{\mathrm{b}}) = \\ \frac{e^{N_D + N_{D_s} + N_{\mathrm{b}}}}{N!} \prod_{i=1}^{N} [N_D V(m_i; \mu_D, \sigma_D, \Gamma_D) + N_{D_s} V(m_i; \mu_{D_s}, \sigma_{D_s}, \Gamma_{D_s}) + N_{\mathrm{b}} B(m_i; \lambda)]. \tag{4.6}$$

In the above equation, the D, D_s and background yields are indicated with N_D, N_{D_s} and N_b, respectively.

The fit of the reconstructed $\mu\mu\pi$ invariant mass has been performed by leaving all the parameters free to float. The $\mu\mu\pi$ invariant mass distribution as well as the result of the fit is shown in Fig. 4.3. The final fit χ^2 normalized to the number of degrees of freedom is 1.07. The best estimate of the parameters of the PDFs obtained after the fit is summarized in Tab. 4.5, where the D and D_s subscripts refer to the PDF of the D and D_s mesons, respectively.

4.4.2 Contribution from mis-identified muons

The measurement of the D_s yield using the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ decay channel could be affected by contamination by other D_s decays, such as $D_s^+ \to \phi(\to K^+K^-)\pi^+$, where both K from ϕ decay could be misidentified as muons. The contamination from this source has been investigated using simulated $D_s^+ \to \phi(\to K^+K^-)\pi^+$ decays.

Parameter	Best estimate
μ_D	1.8693 ± 0.0003
σ_D	0.009 ± 0.003
Γ_D	0.001 ± 0.009
μ_{D_s}	1.9681 ± 0.0001
σ_{D_s}	0.0077 ± 0.0002
Γ_{D_s}	0.0084 ± 0.0005
λ	-1.12 ± 0.02
N_D	8654 ± 207
N_{D_s}	22029 ± 258
N_{b}	309762 ± 1347

Table 4.5: Fitted parameters as result of the $\mu\mu\pi$ invariant mass fit in the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ channel.

The simulated D_s candidates are reconstructed as for the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ channel, with the same preselection requirements on objects kinematics, but with few differences: *highPurity* reconstructed tracks have been used instead of reconstructed muons to build ϕ candidates; no requirements on fitted ϕ mass are applied; the D_s vertex is fitted using a kinematic vertex fitter without constraints on the ϕ mass and no requirements are applied on the fitted D_s mass. Furthermore, the mass of the objects used to build ϕ candidates is assumed to be equal to the known muon mass. The kinematic requirements applied to the reconstructed tracks used to build ϕ candidates are the same as those used for muons in the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ channel. Among all $\mu\mu\pi$ combinations, only the one that has the highest p_T is kept for each event, as for the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ channel.

The reconstructed D_s candidate invariant mass distribution is shown in Fig. 4.4, where the objects used to build the ϕ candidate are assumed to have either the muon or the kaon mass. When the mass is correctly assigned, the D_s invariant mass distribution clearly shows a peak corresponding to the known D_s mass. If the muon mass is assigned to kaons, the D_s invariant mass does not show any structure at the known D_s mass. The background of misidentified kaons coming from the $D_s^+ \to \phi(\to K^+K^-)\pi^+$ decay is expected to be well described by the background model used so far and not contaminate the extraction of the D_s yield.

4.5 Signal channel

4.5.1 Signal preselection

The HNL candidates are selected by first applying a loose set of preselection requirements, on top of which the final selection strategy is optimized. The $N \to \mu^+\pi^-$ candidates are selected by combining a muon and a pion candidate, requiring them to form a common vertex and have a $p_T > 1.0$ GeV. A requirement on the $\mu\pi$ invariant mass ($m_{\mu\pi}$) is also applied: $0.2 < m_{\mu\pi} < 2.0$ GeV. In addition, the μ and the π forming the HNL candidate are required to have an opposite electric charge. The fit of the $\mu\pi$ into a common vertex is performed using a kinematic vertex fitter, and the fit probability is required to be greater than 0.01. At this stage no constraint on the $\mu\pi$ invariant mass, the various HNL mass hypotheses will be tested at the statistical data analysis level (see Sec. 4.6.2). The fitted HNL candidate is then combined with an additional muon and fitted to a common vertex to form a D_s candidate. Only

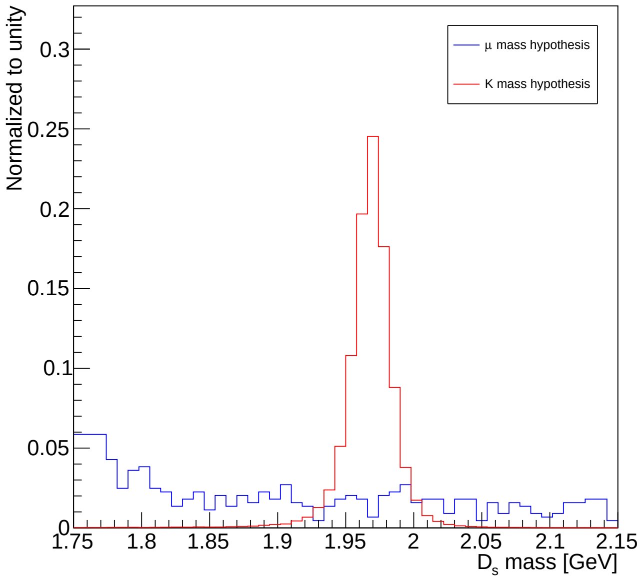

Figure 4.4: Reconstructed D_s mass in simulated $D_s^+ \to \phi(\to K^+K^-)\pi^+$ sample assuming either the muon (blue) or the kaon (red) mass hypothesis for the tracks used to build the ϕ mass.

combinations of $\mu\mu\pi$ that have an invariant mass between 1.5 and 2.5 GeV are taken into account. The vertex fit probability of the D_s candidate must also be greater than 0.01.

After the application of these set of preselection criteria, most of the selected events end up with one or two $\mu\mu\pi$ candidates. The number of $\mu\mu\pi$ candidates selected in each event was determined using the simulated HNL samples. This number depends on the generated HNL lifetime and the fraction of events with a single preselected candidate is higher for HNL generated with a smaller $c\tau$. However, despite the generated HNL mass and $c\tau$, the fraction of events ending up with one or two $\mu\mu\pi$ candidates is above 80%. Several selection criteria have been tested to select the best $\mu\mu\pi$ combination for each event. The efficiency of a selection criterion is evaluated using simulated signal samples and it is defined as the fraction of events where the correct $\mu\mu\pi$ combination has been selected, among all preselected events that contain at least a signal candidate matched with generator-level information.

Particles that form a D_s signal candidate are expected to have a greater significance of the transverse impact parameter than random combinations $\mu\mu\pi$, and the decay length of the $\mu\pi$ pair that forms the HNL candidate is expected to always be greater than the D_s candidate decay length in signal events. Under these considerations, two variables have been used to build selection criteria to choose the best $\mu\mu\pi$ combination: the sum of the significance of the transverse impact parameter, calculated with respect to the position of the beam spot, of the two muons and pion forming the D_s candidate (denoted as $\sum_{i=\mu,\pi} \mathrm{IPS}_i$) and the ratio between the 3D decay length of the HNL and the D_s, both calculated with respect to the primary vertex (denoted as L_{HNL}/L_{D_s}). In addition, the cosine of the HNL back-pointing angle $\cos(\alpha_{\mathrm{HNL}})$ is expected to be closer to unity in signal events than in background events.

All selection criteria taken into account are listed in Tab. 4.6: for each listed variable, only the candidate having the highest value of this variable is kept. The

Variable	Additional condition	Efficiency (%)
L_{HNL}/L_{D_s}	-	93.0 ± 2.0
L_{HNL}/L_{D_s}	$\cos\alpha_{\mathrm{HNL}} > 0$	95.0 ± 2.0
$\sum_{i=\mu,\pi} \mathrm{IPS}_i$	-	94.0 ± 2.0
$\sum_{i=\mu,\pi} \mathrm{IPS}_i$	$\cos\alpha_{\mathrm{HNL}} > 0$	95.0 ± 2.0
$(\sum_{i=\mu,\pi} \mathrm{IPS}_i) \times (L_{\mathrm{HNL}}/L_{D_s})$	$\cos\alpha_{\mathrm{HNL}} > 0$	95.0 ± 2.0

Table 4.6: Efficiency of the criteria used to select the best $\mu\mu\pi$ combination. The best combination is selected choosing the candidate having the highest value of the variable in the first column. In cases where another condition is added, it is indicated in the second column. The efficiency has been evaluated using MC samples.

uncertainty reported is only statistical. The criterion used for the analysis is the one selecting the $\mu\mu\pi$ combination based on the L_{HNL}/L_{D_s} ratio, with the additional condition that the candidate $\cos(\alpha_{\mathrm{HNL}})$ must be nonnegative. This is one of the criteria having the best efficiency in selecting the correct $\mu\mu\pi$ candidate among those that have been tested.

4.5.2 Pileup reweighting

Monte Carlo events must include the simulation of pileup collisions. The observed pileup distribution in data is different from the one obtained from the MC simulation. A good description of LHC events is fundamental for studying the selection efficiency in MC. For this reason, each MC sample is reweighted to better match the pileup distribution observed in data.

For standard physics analyses, the *Luminosity Physics Object Group* (Lumi POG) provides run-by-run pileup information. The average pileup distribution for a particular data-taking period can be retrieved for application to data analysis. Information on the mean number of interactions per crossing is already stored in simulated samples. The ratio between the average pileup distribution in data and MC represents the weight that has to be applied to MC events.

As mentioned in Sec. 4.2.2, pileup distributions in signal and background samples have been generated using different shapes. While signal samples have a pileup distribution closer to the one observed in the B-Parking data set, in QCD background samples the pileup interactions have been simulated so that their distribution is closer to the one observed during the standard data-taking operations, which is shifted to higher values. For this reason, two sets of pileup weights have been derived, the first dedicated to signal samples and the second for the remaining samples.

By construction, after the application of the aforementioned pileup weights, the distribution of the pileup interactions in simulated samples is identical to the one observed in data. The result of the application of the pileup weights is shown in Fig. 4.5, where the distribution of the number of pileup interactions is shown in data and QCD (left plot) and signal (right plot) MC samples. As expected the distribution of pileup interactions in QCD MC is shifted to larger values, whereas in signal the distribution before the application of pileup weights (red) is closer to data than the QCD one.

Since the number of pileup interactions is expected to affect the primary vertex reconstruction efficiency, the application of the pileup weights to MC samples should result in a better agreement between data and MC on the number of reconstructed primary vertices. Figure 4.6 shows the normalized distribution of the number of

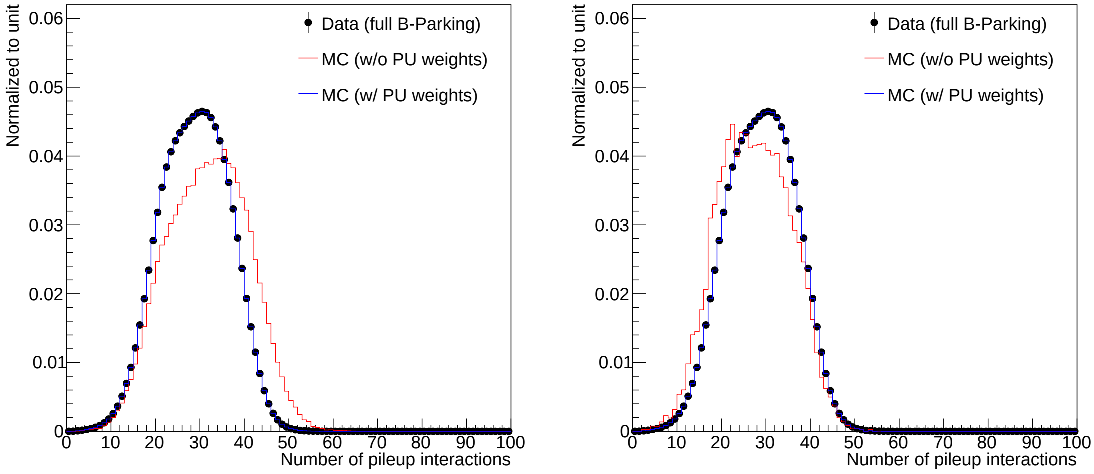

Figure 4.5: Normalized distribution of the number of pileup interactions in data and MC. The left plot shows the effect of pileup weights on QCD MC samples, while the right plot shows the pile up reweighting on signal samples.

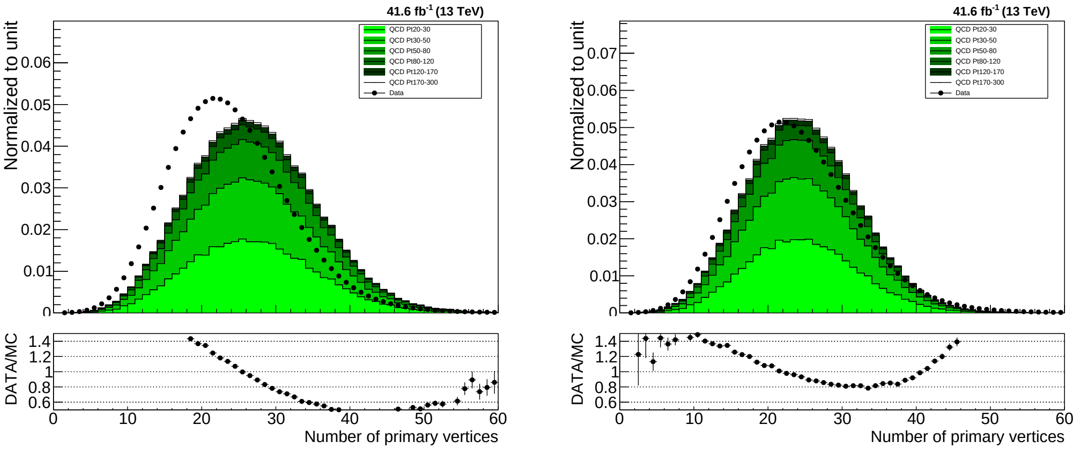

Figure 4.6: Normalized distribution of the number of primary vertices in data and MC: simulated events in the right plot have been reweighted to match the pileup distribution in data.

primary vertices of events that pass the HNL preselection in data and in MC. When comparing the two distributions before (left) and after (right) pileup weights are applied, the PV distribution for MC events shows a better agreement with data when the pileup weights are applied. However, data and MC shows still a quite large disagreement. Although after pileup reweighing the data/MC agreement for the number of reconstructed vertices is expected to improve, in general a perfect closure is not expected. The reason lies on the fact that the vertex reconstruction depends on some subtleties that may differ between data and MC, so the efficiency to reconstruct a vertex is in general different between them. Nevertheless, throughout the whole analysis, MC events will implicitly include the pileup reweighting.

4.5.3 Muon reconstruction and identification scale factors

Since the muon identification and reconstruction efficiencies might be different in data and simulation, a set of identification and reconstruction scale factors have been taken into account to correct for any eventual efficiency difference. These scale factors are provided by the CMS *Muon physics object group* (Muon POG) to the whole CMS collaboration. The muon identification and reconstruction efficiency is computed in both data and MC using the tag-and-probe technique. This technique exploits well-known resonances decaying to a pair of muons and it is often used to measure the reconstruction, identification and trigger efficiency as a function of muon kinematics. In the tag-and-probe method a mass resonance is reconstructed often exploiting its decay to two leptons. One of the two objects in which the resonance decays into must be a “tag”, i.e. it is required to pass a set of stringent selection criteria designed to isolate the required particle type. The tag is paired with a “probe” which is required to pass looser selection criteria. The efficiency is measured as the fraction of probes that pass the selection criteria under investigation.

The set of identification scale factors used in this work have been calculated specifically for low-p_T muons exploiting the $J/\psi \to \mu\mu$ decay. For the tag-and-probe-based efficiency measurement, the tag muon is required to have $p_T > 8\,\mathrm{GeV}$, $|\eta| < 2.4$ and pass the Tight identification criteria (see Sec. 3.2.1). In addition, the tag muon must be matched to a trigger object from a prescaled single muon trigger requiring $p_T > 8$, 17 or 20 GeV. The probe is defined as a tracker muon track. The tag-and-probe muon pair is formed if its invariant mass is within the 2.9 to 3.3 GeV mass window, the ΔR between the two muons is larger than 0.1 and the two tracks vertices have a distance smaller than 0.5 cm along z. After the tag-and-probe muon pair selection, the number of passing and failing probes is determined by fitting the J/ψ candidates invariant mass where the probe muon either passes or fails the Soft identification criteria. The J/ψ signal peak and the continuous background are fitted using a Crystal ball and an exponential shapes. Since the probes used to measure the muon identification efficiency were required to be *tracker* muons, the efficiency of reconstructing a muon as a tracker muon track has also been assessed. The tracker muon track reconstruction efficiency measurement is also provided by the Muon POG. The Muon POG realized a muon reconstruction efficiency measurement in data and simulation in the same way it has been done for the muon identification efficiency.

The efficiency measurement is repeated in bins of the probe muon p_T and η using data and a $J/\psi \to \mu\mu$ simulated sample. The ratio of the efficiencies in bins of muon p_T and η results in a set of scale factors. The muon reconstruction and identification scale factors are shown in Fig. 4.7. The muon reconstruction scale factors are close to 1 within 1% for almost all bins. Few bins show a deviation within 10% and just a single bin has data/MC reconstruction efficiency difference of about 23%. Also the muon ID scale factors obtained are all close to 1, with the largest deviations being within 3%. Since the final state involve two muons that are required to pass the Soft ID criteria, the final weight applied to each event is the product of the scale factors of the single muons. Throughout the whole analysis, MC events will implicitly include the muon ID and reconstruction scale factors.

4.5.4 Trigger scale factors

The B-Parking data set has been collected using many different HLT paths. During the data-taking, not all HLT paths were active, but only few were operated at the same time. This was possible by changing the trigger prescales depending on the instantaneous luminosity during a physics run, as described in Sec. 2.4.3. Moreover, only a single path was always collecting data throughout the whole data taking, i.e.

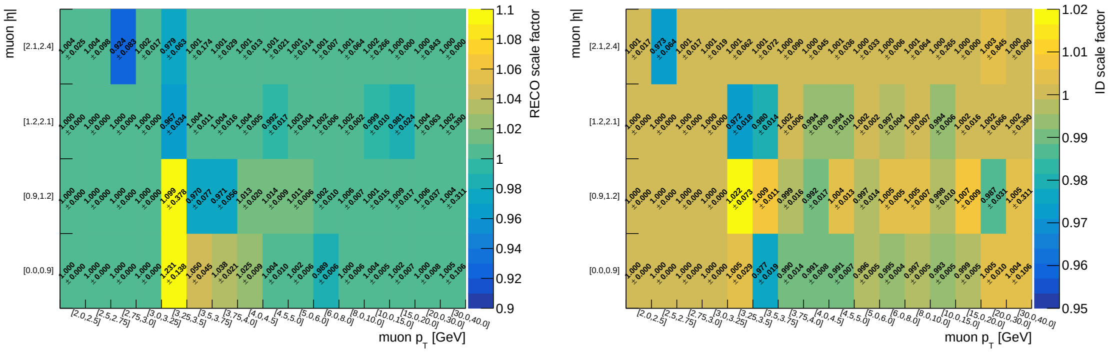

Figure 4.7: Muon reconstruction (left) and identification (right) scale factors in bins of p_T and $|\eta|$.

the HLT_Mu9_IP6 path. Usually the trigger efficiency in data is not modeled well by MC simulations. In addition, the peculiar activity of the B-Parking trigger on HLT side cannot be reproduced at the simulation level. For this reason the trigger efficiency in MC is expected to show a large difference with respect to data. This difference has been taken into account by including a correction to MC events which corresponds to the ratio of the trigger efficiency in data and simulation, the so-called trigger scale factors. The trigger efficiencies have been computed in the context of a CMS analysis searching for heavy neutral leptons in B meson decays [104], which uses the B-Parking data set, as well as all B-Parking HLT paths. However, the scale factor calculation has been performed using only the part 0 of era D of the B-Parking data set, which corresponds to a total integrated luminosity of $5.3\,\text{fb}^{-1}$. The scale factor computation procedure will be briefly described below.

The trigger efficiency is computed in both data and MC using the tag-and-probe technique. One of the muons, the *tag* muon is required to satisfy tight selection cuts to purify the sample, specifically the Soft muon identification, which is a standard method to select low-p_T muons. In addition, the tag muon has to match with either the HLT_Mu9_IP6 path or the HLT_Mu12_IP6 path, whose logical OR covers the full B-Parking data set. This request is applied to avoid biasing the probe muon. The other muon, the *probe* muon, is left unconstrained and it is used to study the trigger efficiency. The trigger efficiency is measured as the fraction of probes that satisfy the requirement on trigger.

The trigger efficiency in MC is measured using a $B \to J/\psi K^*$ sample, where the J/ψ decays to $\mu\mu$. The muons forming the J/ψ decay are used to measure the trigger efficiency. The J/ψ events are reconstructed by fitting two muons to a common vertex and requiring: the two muons to be oppositely charged, to have $p_T > 1\,\text{GeV}$, $|\eta| < 2.3$ and an invariant mass in the 2.9-3.3 GeV window; the fitted B vertex probability to be greater than 10^{-5} and its back-pointing angle $\cos\alpha_B > 0.9$. The J/ψ candidate coming from a B meson decay is then used as a tag-and-probe muon pair. The efficiency is computed by fitting the following two distributions:

- Invariant $\mu\mu$ mass where the probe muon is matched to at least one B-Parking HLT path (so-called *passing* probe muon);
- Invariant $\mu\mu$ mass without any trigger matching requirement on the probe muon.

The efficiency is computed as the ratio between the yield measured using only passing probe muons and the yield measured using all probe muons.

Furthermore, the efficiency is computed in bins of probe muon p_T and its impact parameter significance (IPS) on xy:

- probe p_T: $[6.0, 7.0, 8.0, 8.5, 9.0, 10.0, 10.5, 11.0, 12.0, 20.0, 100.0]$ GeV;
- probe xy IPS: $[0.0, 3.0, 3.5, 4.0, 5.0, 6.0, 8.0, 10.0, 20.0, 500.0]$.

The p_T and IPS binning is chosen following the different conditions of the B-Parking HLT triggers. The efficiency is computed in data and simulation, and from them a trigger scale factor is determined, to be used as a reweighting factor applied to the MC events, to account for the different trigger efficiency in real data. The trigger efficiency in data and simulation in the various p_T and IPS bins is shown in Fig. 4.8. The trigger efficiency in simulation shows a sharp turn-on as a function of the muon p_T at 7 GeV, which corresponds to the lowest p_T threshold value among all B-Parking trigger paths. In the 6 to 7 GeV bin the efficiency is almost 0% regardless of the IPS bin. There is also a quite sharp turn-on as a function of the IPS, and the trigger efficiency becomes almost flat around approximately 85% for the region $p_T > 7\,\text{GeV}$ and IPS > 4. On the other hand, the trigger efficiency measured in data shows a very slow increase as a function of the muon p_T and IPS with respect to the MC simulation. In data, the trigger efficiency reaches a stable value around $\approx 80\%$ only for $p_T > 12\,\text{GeV}$ and IPS > 6. The significant difference between data and MC can be explained by the fact that the MC does not reproduce the B-Parking trigger setting used during data-taking. As explained in Sec. 2.4.3, different L1 trigger seeds were activated as the instantaneous luminosity decreased at the end on an LHC fill, and not all HLT paths where active during the whole B-Parking data-taking. The trigger conditions have not been simulated in a time-dependent manner, and as a consequence the discrepancy between trigger efficiency in data and MC is relevant.

The trigger scale factors are derived from trigger efficiency measurements by defining the per-event trigger efficiency. Since both muons can trigger an event, the per-event trigger efficiency (ε_{evt}) is a combination of the trigger efficiency of the two separate muons:

$$\varepsilon_{evt} = \varepsilon_1 + \varepsilon_2 - \varepsilon_1 \varepsilon_2 \,, \tag{4.7}$$

where ε_1 and ε_2 indicate the trigger efficiency of the two triggering muons. The trigger scale factor used to reweight simulated events is then the ratio of the per-event efficiency in data and simulation.

The resulting scale factors have been validated by comparing data and simulation distributions for the muon p_T, η and impact parameter significance on the xy plane. Since the scale factors have been computed using only a subset of the B-Parking data set, the effect of the application of the trigger scale factors on MC has been validated by looking only at the agreement with the same data subset. The comparison between data and simulation is shown in Figs. 4.9 and 4.10 for the muon coming from the candidate $D_s^+ \to N\mu^+$ and $N \to \mu^+\pi^-$ decays, respectively. The simulated distributions on the right column of Figs. 4.9 and 4.10 have been reweighted by the trigger scale factor. Only events passing the preselection requirements are taken into account. The pileup weights, the muon reconstruction and ID scale factors derived in the previous sections have been taken into account as well.

The muon p_T distributions show a drastic improvement when trigger scale factors are applied. On one hand, the muon p_T distribution reflects the selection cut requirements based on the muon η, i.e. muon $p_T > 3.5\,\text{GeV}$ ($p_T > 2\,\text{GeV}$) for $|\eta| < 1.2$ ($1.2 < |\eta| < 2.4$), on the other hand the structures starting at $p_T > 7\,\text{GeV}$, and a small bump around 12 GeV, reflect the different trigger thresholds on the muon

p_T. The modeling of the muon η distribution shows a slight improvement as well. Although the modeling of the muon IPS also shows a slight improvement, data and simulation still show a disagreement of about 20%. The residual disagreement between data and simulation might be caused by a mismodeling of the transverse IPS in the MC, which has been observed by other CMS analyses.

Since the analysis is based on the full B-Parking data set, the agreement between data and simulation must be assessed using the entire B-Parking data set. Figure 4.11 show the comparison between data and MC, where the data corresponds to the full B-Parking data set. The level of data/MC agreement is similar to that shown for the B-Parking data subset above. However, the simulation shows a worse agreement with data for muon p_T values around the 7 to 9 GeV, where most of the HLT p_T thresholds are. Moreover, the discrepancy between full B-Parking data and MC reaches higher values, beyond 20%, at high muon p_T values. No particular improvement nor worsening are observed in the agreement between data and simulation for muon η and IPS. The bottom row in Fig. 4.11 show the data/MC agreement for pion p_T, η and IPS along xy. While the pion p_T is in agreement with data within 10%, its η and IPS show a similar level of agreement with data to the one observed for muons. This might suggest an underlying simulation mismodeling of data that cannot be accounted for in trigger scale factors. The residual difference between data and simulation has been taken into account as a systematic uncertainty, and it will be described in detail in Sec. 4.7.

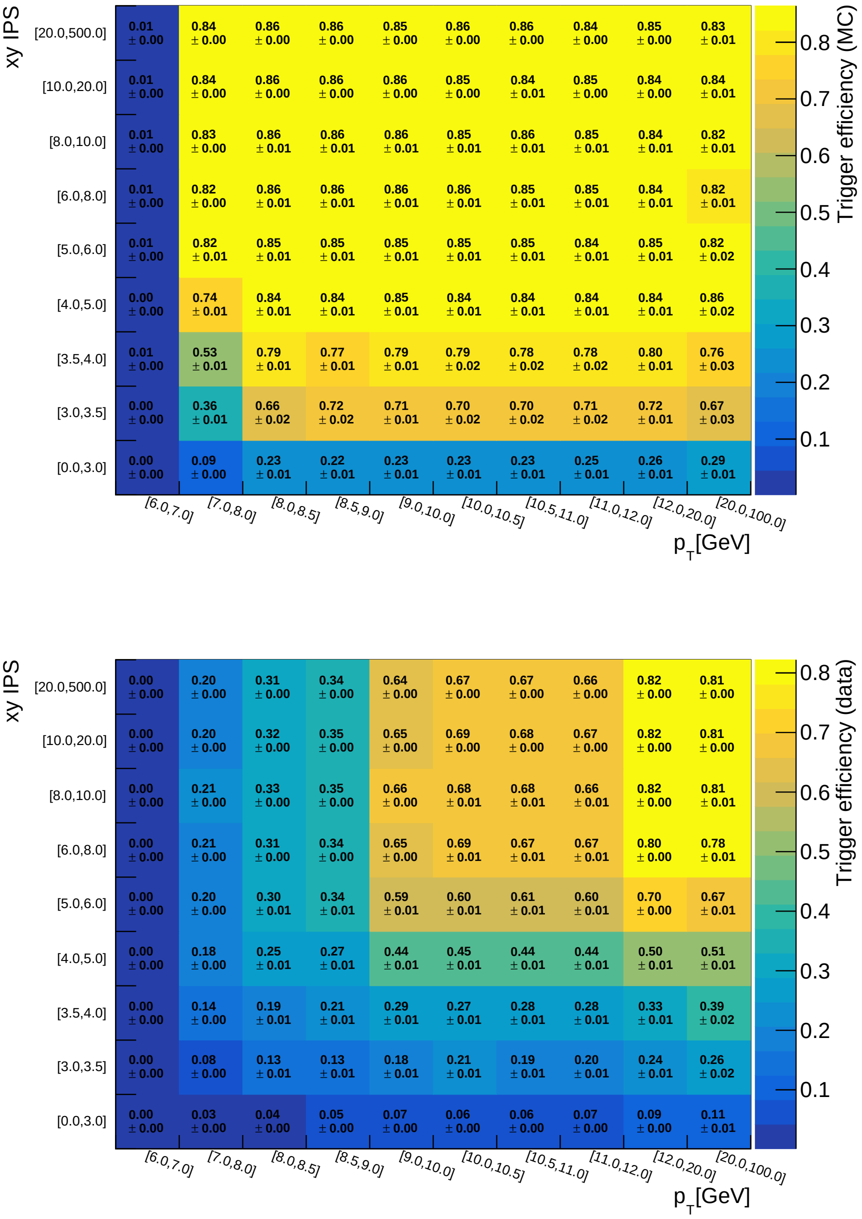

Figure 4.8: B-Parking trigger efficiency in simulation (top) and data (bottom) in bins of the muon p_T and impact parameter significance (IPS).

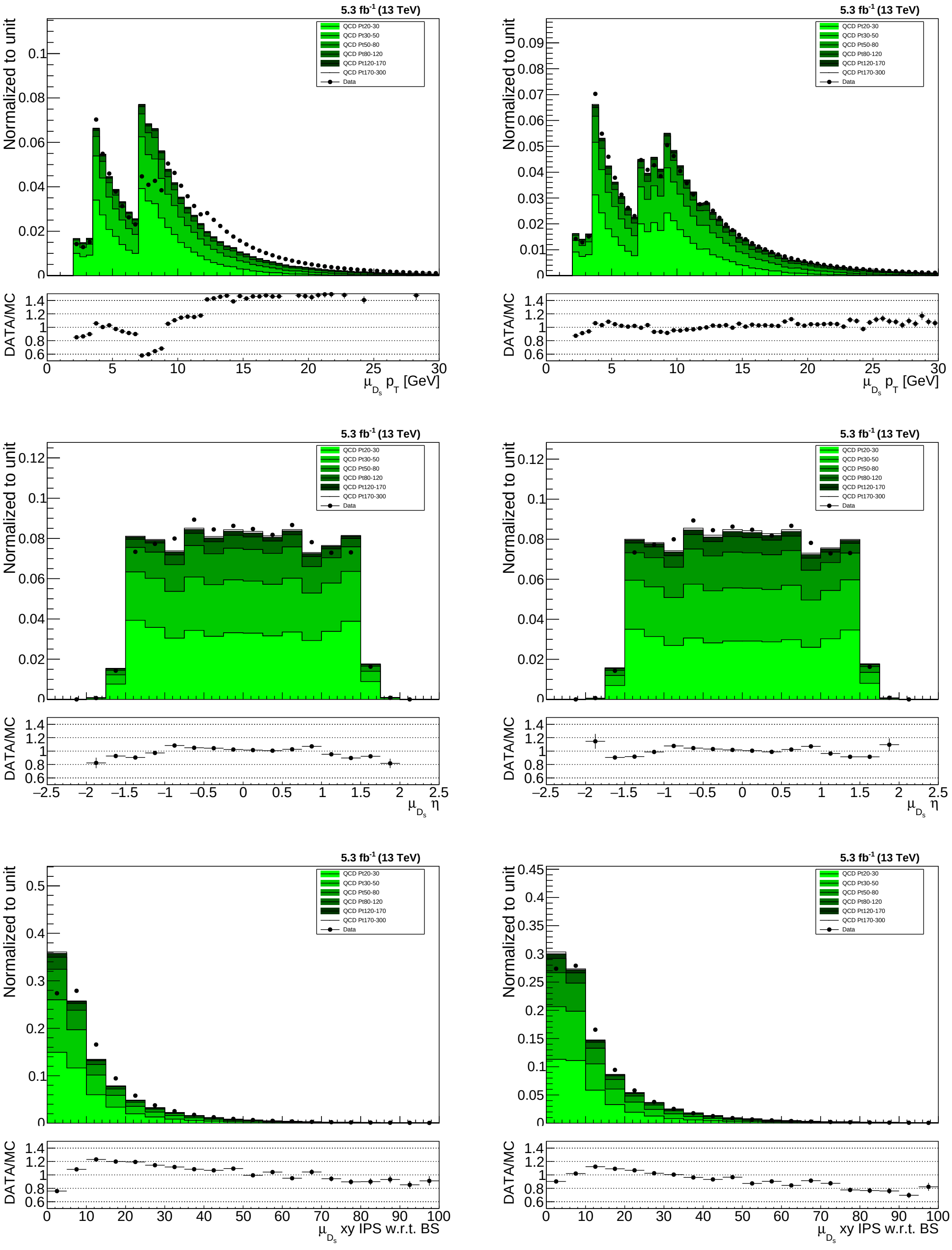

Figure 4.9: Data/simulation comparison for p_T (top), η (middle) and impact parameter significance (IPS) on the xy plane (bottom) for the muon forming the $D_s \to N\mu$ candidate. The simulated events in the right column distributions have been reweighted by the B-Parking trigger scale factor. Only the b-Parking data subset corresponding to era D and part 0 has been used.

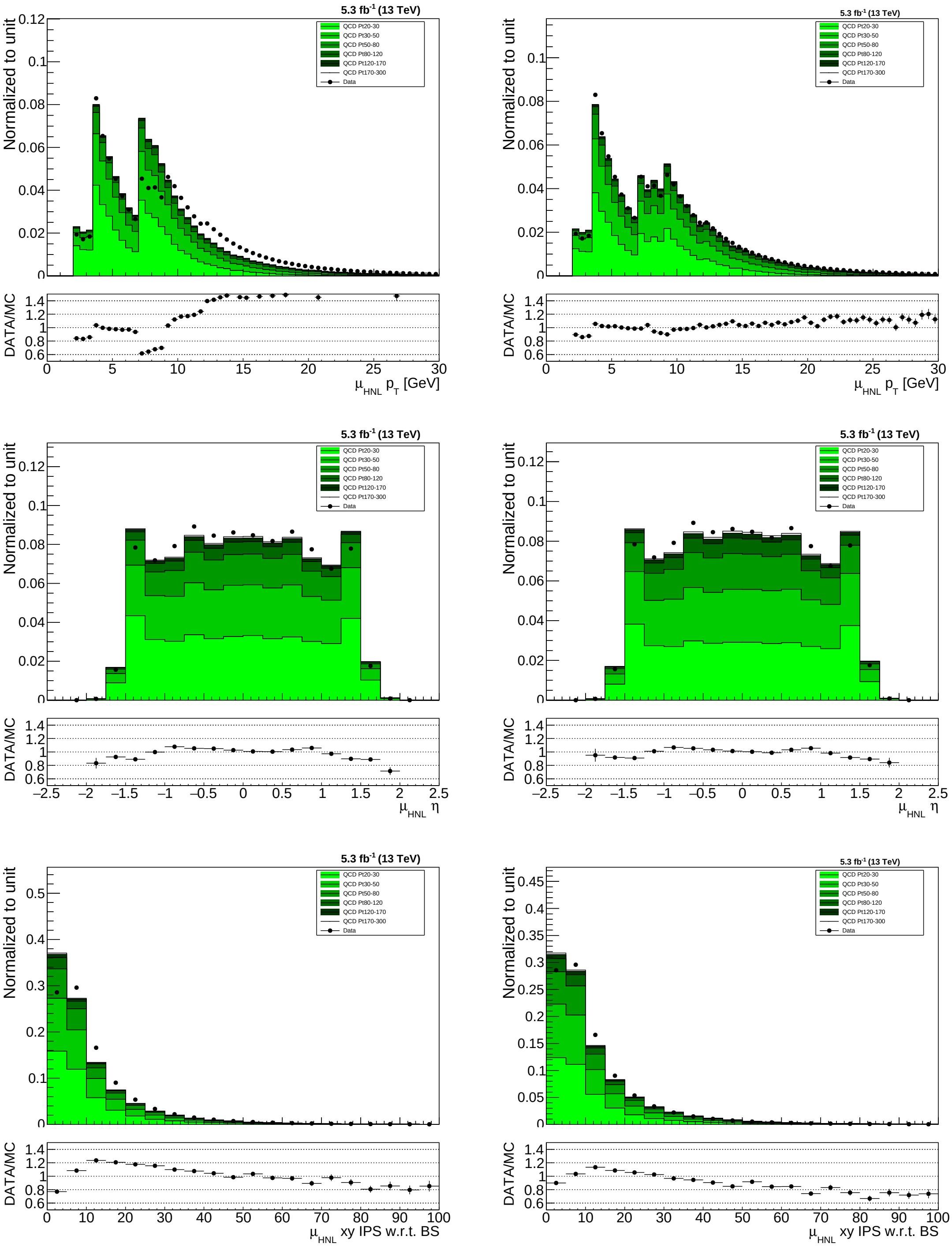

Figure 4.10: Data/simulation comparison for p_T (top), η (middle) and impact parameter significance (IPS) on the xy plane (bottom) for the muon forming the $N \rightarrow \mu^+\pi^-$ candidate. The simulated events in the right column distributions have been reweighted by the B-Parking trigger scale factor. Only the b-Parking data subset corresponding to era D and part 0 has been used.

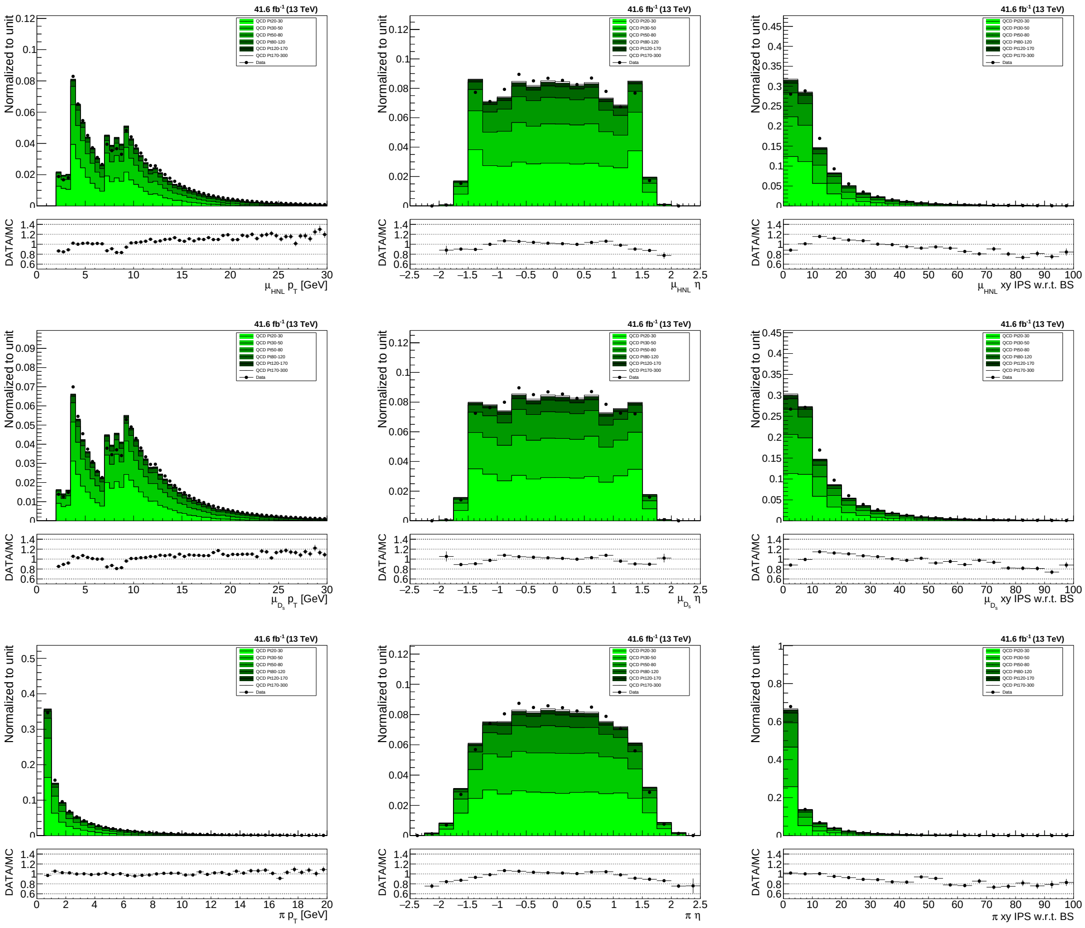

Figure 4.11: Data/simulation comparison for p_T (left column), η (middle column) and impact parameter significance (IPS) on the xy plane (right column) for the muon forming the $D_s^+ \to N\mu^+$ (top row) and the $N \to \mu^+\pi^-$ (middle row) candidate, and for the pion (bottom row). The data corresponds to the full B-Parking data set.

4.5.5 Selection optimization

After the selection of the best candidate has been made, the selection requirements are optimized to improve the discovery significance. In addition, the events are categorized based on:

- relative muon sign: same sign muons (SS), opposite sign muons (OS);
- HNL decay length on the transverse plane (L_{xy}): $0 < L_{xy} < 1\,\mathrm{cm}$ (lxy0to1), $1 < L_{xy} < 5\,\mathrm{cm}$ (lxy1to5) and $L_{xy} > 5\,\mathrm{cm}$ (lxy5toInf).

Six categories are formed, labeled as SSlxy0to1, SSlxy1to5, SSlxy5toInf, OSlxy0to1, OSlxy1to5, OSlxy5toInf. The categorization based on the charge of the muons is meant to reflect the possibility of a lepton flavor violating process, i.e. when the final state muons have the same sign. The final state with two same-sign muons can happen only if the heavy neutral lepton is a Majorana neutrino, while in the case where the two muons have an opposite charge the heavy neutral lepton could be either a Majorana or a Dirac particle. The categorization on the transverse decay length is based on the fact that the reconstructed HNL vertex is expected to be produced with a significantly greater transverse displacement with respect to reconstructed vertices from combinatorial background. Moreover, the categories having a larger displacement are expected to be more efficient for HNL with a longer lifetime. This behavior can be observed in Fig. 4.12, where the selection efficiency of preselected signal events is displayed as a function of the cut on the HNL transverse displacement, for the various hypotheses of HNL mass and lifetime (expressed as $c\tau$). The categories with large L_{xy} are expected to be populated by long-lived neutrinos, while those with small L_{xy} would be populated by neutrinos with short lifetime. This is verified in Fig. 4.13, that shows the L_{xy} distributions of background and signal for 10 and 1000 mm lifetime hypotheses, along with the L_{xy} cuts used for the categorization.

The selection is then optimized for each category using simulated samples, by testing selected variables to discriminate signal from background. For each variable, the final selection requirement is fixed by maximizing the expected significance of the discovery in the asymptotic approximation of large statistics. The signal yield is estimated using Eq. 4.3 assuming the HNL mass is 1.5 GeV and the lifetime is 10 mm. The background yield is estimated using the HNL mass sidebands, where no HNL events are expected. The mass sidebands are defined as the HNL invariant mass range outside the signal region, that is a $\pm 3\sigma_{\mathrm{HNL}}$ region centered at the HNL mass hypothesis where σ_{HNL} is estimated by fitting the signal peak in simulated events using a Voigtian function. The background yield is estimated by fitting the HNL candidate invariant mass sidebands in data, after the preselection, with an exponential function in a $\pm 6\sigma_{\mathrm{HNL}}$ window around the HNL mass hypothesis. The background yield is then estimated as the integral of the fitted function in the signal region.

Assuming a signal strength $\mu = 1$, the discovery significance for a given expected signal (s) and background (b) yield can be written as:

$$Z_0 = \sqrt{2\left[(s+b)\ln\left(1+\frac{s}{b}\right) - s\right]}\,, \tag{4.8}$$

and it reduces to $s/\sqrt{b}$ in the $s \ll b$ limit. For the selection optimization, the value for s (b) is estimated as the product of the expected signal (background) yield

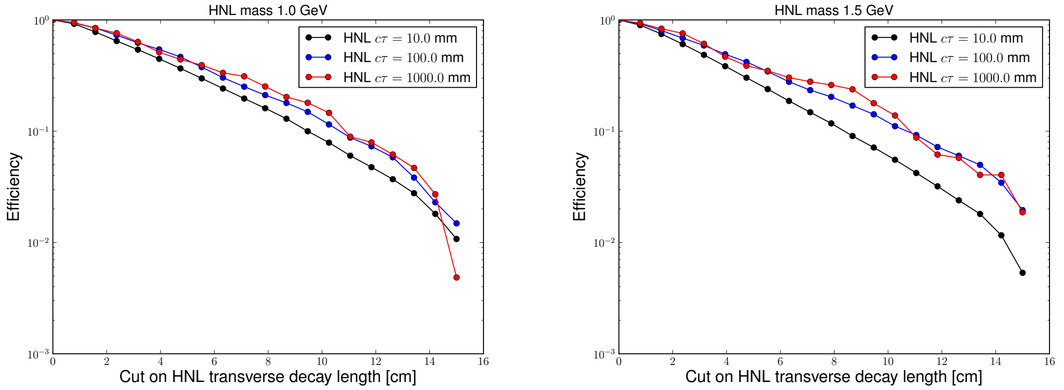

Figure 4.12: Selection efficiency of preselected signal events as a function of the requirement on the HNL transverse decay length, for 1.0 GeV (left) and 1.5 GeV (right) HNL mass hypothesis. Black, blue and red markers refers to the 10, 100 and 1000 mm HNL $c\tau$ hypothesis.

after the preselection and the efficiency of the selection requirement. The selection efficiency is estimated using simulated HNL events and simulated QCD dijet events for signal and background respectively, and is taken as the fraction of generated events that survives all selection cuts.

To separate signal from background, the fact that the HNL vertex is expected to decay well detached both from PV and from D_s decay vertex is being exploited. The following variables are expected to have a significantly different distribution in signal and background: the significance of the HNL decay length on the transverse plane, that is, the ratio of the HNL decay length and its uncertainty, measured with respect to the beam spot position; the impact parameter (IP) and its significance (IPS) on the transverse plane for the muon and the pion that form the HNL candidate, measured with respect to the beam spot position. The normalized distribution for each of the above variables is shown in Fig. 4.14, for QCD dijet simulation and HNL signal simulation after the preselection. All the variables taken into account show a quite different distribution with respect to the combinatorial background from QCD dijet events: signal events tend to have a distribution with significantly longer tail compared to background events.

In order to consider the QCD dijet MC sample as a good proxy for the expected combinatorial background, the agreement between data and MC has been tested. In Fig. 4.15 the distribution of data and MC for the above variables is displayed after preselection: the distribution for data and simulation show an overall reasonable agreement. However, even after the reweighting of the signal using the trigger scale factors there is still some disagreement between data and simulation ranging from 10 to 20%, especially for variables describing the muon and pion impact parameter and impact parameter significance. The residual disagreement might be caused by a mismodeling of these variables in the Monte Carlo simulation, as it was observed in previous CMS analysis.

The discovery significance defined in Eq. 4.8 is evaluated for the aforementioned variables, in each category. The final set of selection requirements optimized for each category is obtained via the following algorithm:

1. the significance is tested separately for each variable and category by varying the value of the selection cut;

2. the variable giving the largest significance gain is selected and the optimized

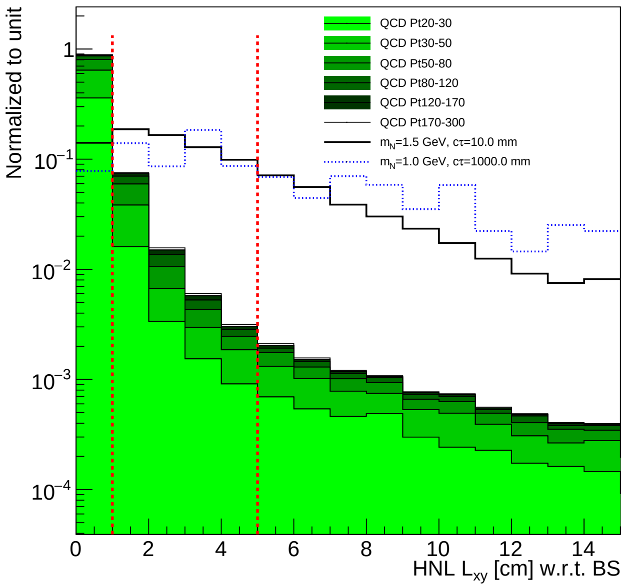

Figure 4.13: QCD background and HNL signal distribution of the transverse HNL decay length with respect to the beam spot. The signal is shown only for the $m_{\text{HNL}} = 1.5\,\text{GeV}$ and $c\tau_{\text{HNL}} = 10\,\text{mm}$, and for $m_{\text{HNL}} = 1.0\,\text{GeV}$ and $c\tau_{\text{HNL}} = 1000\,\text{mm}$. The vertical dashed red lines indicate the categorization cuts on L_{xy}.

   selection cuts are determined for each category as the ones maximizing the discovery significance;

3. the variable selected in the previous step is excluded and the optimized selection cuts applied;

4. the procedure starts over.

The procedure is repeated until no relevant gain is shown in the discovery significance scan of the remaining variables. The optimized selection requirements for each category obtained with the above method are summarized in Tab. 4.7. The variables reported in Tab. 4.7 follow the order in which they have been selected in the various iterations of the above algorithm. The first is the displacement significance of the HNL vertex in the xy plane, denoted as $L_{xy}/\sigma(L_{xy})$, then the pion IP on xy and finally the muon IPS on xy. After applying the optimized selection cuts on these variables, no significant gain in the discovery significance has been observed if further requirements where applied on the pion IPS or muon IP on the xy plane.

Figures 4.16 to 4.21 show the discovery significance as a function of the cut on the discriminating variables. The discovery significance scans that are shown contain the optimized selection requirement from the previous iteration, as reported in Tab. 4.7. Namely, the scan over the HNL $L_{xy}/\sigma(L_{xy})$ is performed only after the preselection requirements, the pion IP scan is performed after the application of the optimized cuts on the HNL $L_{xy}/\sigma(L_{xy})$ and the muon IPS contains in addition also the selection requirements on the pion IP. From figs. 4.16 to 4.21 it can be observed that, as more selection requirements are added the discovery significance gain decreases. The discovery significance scan on the pion IPS or muon IP are performed after the application of the HNL $L_{xy}/\sigma(L_{xy})$, pion IP and muon IPS optimized selection cuts. Following these requirements, any further cut on either

Variable	**Categories**					
	SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
$L_{xy}/\sigma(L_{xy})$	> 35.3	> 107.0	> 191.8	> 38.2	> 120.5	> 255.0
π xy IP (cm)	> 0.03	> 0.03	> 0.03	> 0.05	> 0.03	> 0.03
μ xy IPS	> 2.8	> 2.6	> 3.0	> 3.5	> 2.2	> 3.4

Table 4.7: Selection requirements optimized for each category.

m_{HNL} (GeV)	$c\tau_{\text{HNL}}$ (mm)	ε_{pre} (%)	Selection efficiency (%) per category					
			SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
1.5	10	6.0	34.2	70.6	71.9	30.0	67.2	65.3
	100	0.9	31.6	70.3	65.1	15.5	65.5	63.4
	1000	0.1	8.3	60.0	72.7	3.0	48.4	65.4
1.0	10	4.5	21.7	57.2	63.0	14.6	50.8	56.4
	100	0.6	9.6	57.5	64.6	5.1	53.2	59.8
	1000	0.08	2.9	43.5	61.9	1.1	30.2	55.9

Table 4.8: Summary of the preselection efficiency (ε_{pre}) and the selection efficiency for each category after the preselection. All the quantities are displayed for each HNL mass (m_{HNL}) and lifetime ($c\tau_{\text{HNL}}$) hypothesis.

the pion IPS or muon IP on does not show any potential gain on the discovery significance. For this reason, no requirements are applied on these variables.

In principle one could expect a higher background in the opposite-sign categories since they are expected to be populated by known standard model processes in addition to combinatorial events. However, the discovery significance shape is found to be similar in the categories having the same L_{xy} displacement. This might indicate that the main contribution to background is combinatorial. With the exception of the large displacement categories SSlxy5toInf and OSlxy5toInf, similar selection requirements have been found for categories sharing the same L_{xy} displacement.

The efficiency values of the whole selection are summarized in Tab. 4.8. As expected the preselection requirements are more efficient for simulated HNL decays with a smaller $c\tau$, since HNL decays with a larger lifetime are more likely to fall outside the detector acceptance. The preselection efficiencies for HNLs having a different mass but the same lifetime are comparable. The selection efficiency of Tab. 4.8 refers only to the optimized selection cuts summarized in Tab. 4.7 and it is computed considering at the denominator the events that pass the preselection cuts.

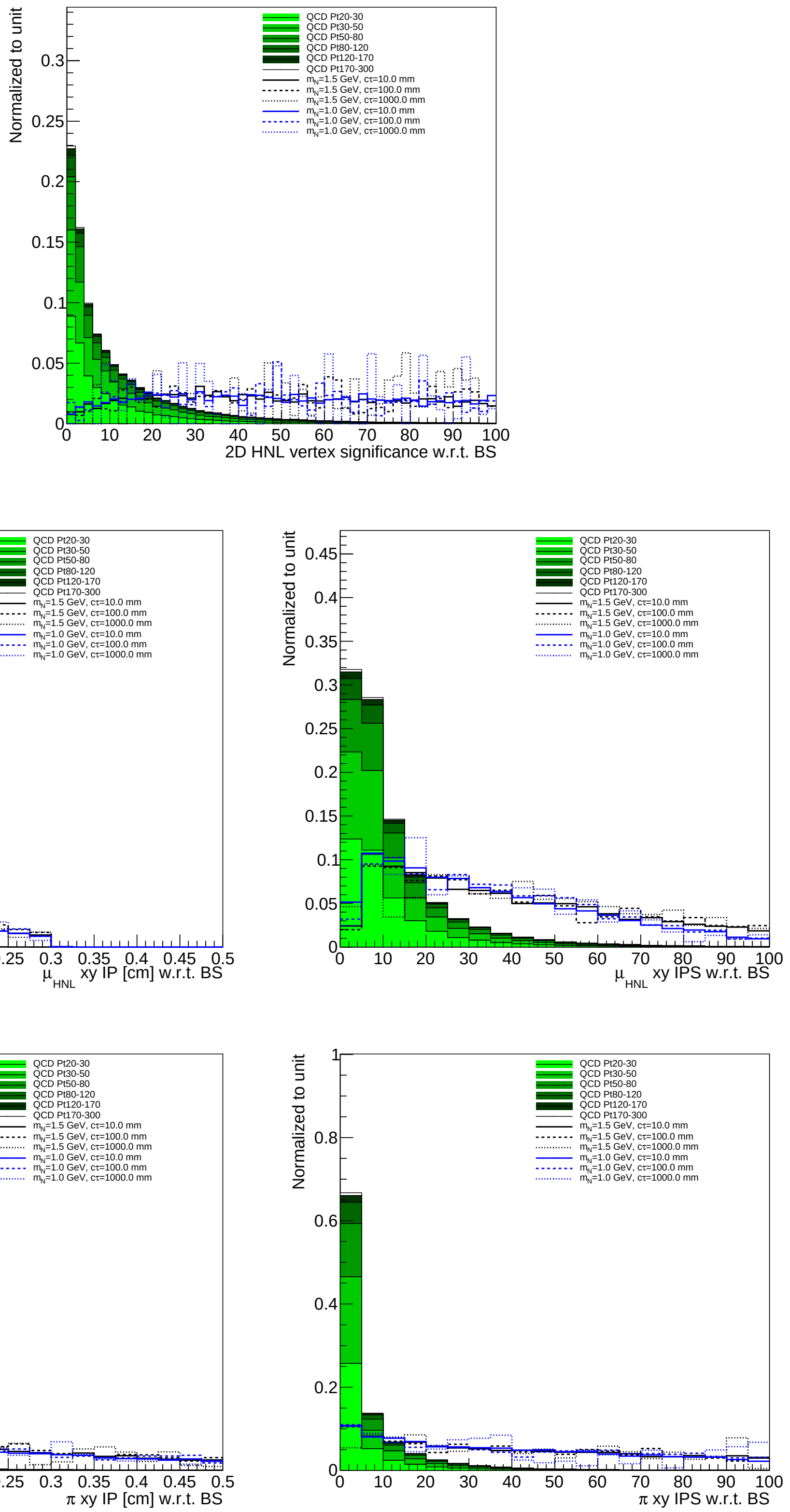

Figure 4.14: QCD background and HNL signal distributions for various observables. Top: HNL vertex significance on the transverse plane measured with respect to the beam spot. Middle: Impact parameter (IP, left) and its significance (IPS, right), for muons forming the HNL candidate, measured in the transverse plane with respect to the beam spot. Bottom: IP (left), and IPS (right), for pions forming the HNL candidate, measured on the transverse plane with respect to the beam spot.

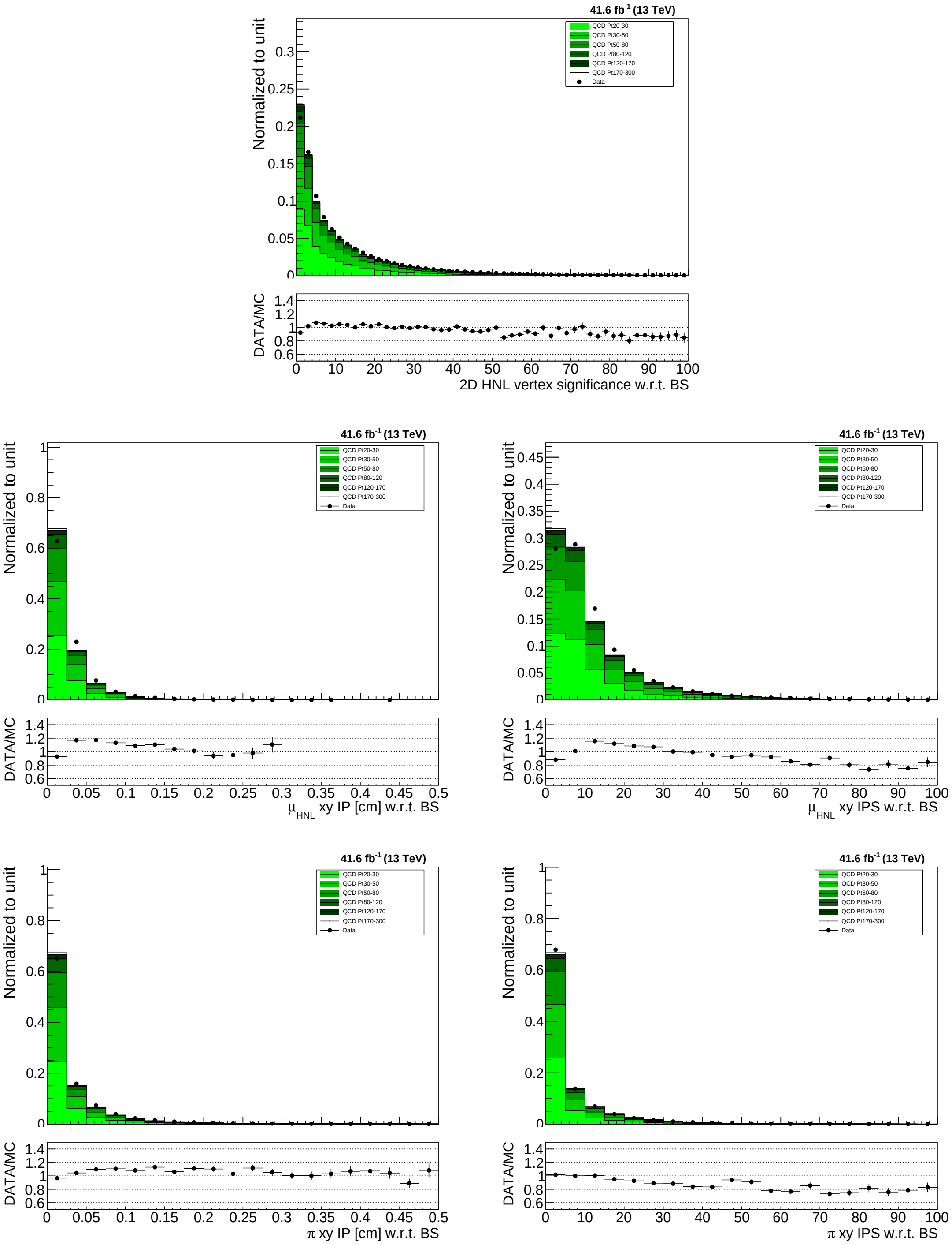

Figure 4.15: Comparison of data and QCD background simulation for various observables. Top: HNL vertex significance on the transverse plane measured with respect to the beam spot. Middle: Impact parameter (IP, left) and its significance (IPS, right) for muons forming the HNL candidate, measured in the transverse plane with respect to the beam spot. Bottom: IP (left) and IPS (right) for pions forming the HNL candidate, measured on the transverse plane with respect to the beam spot.

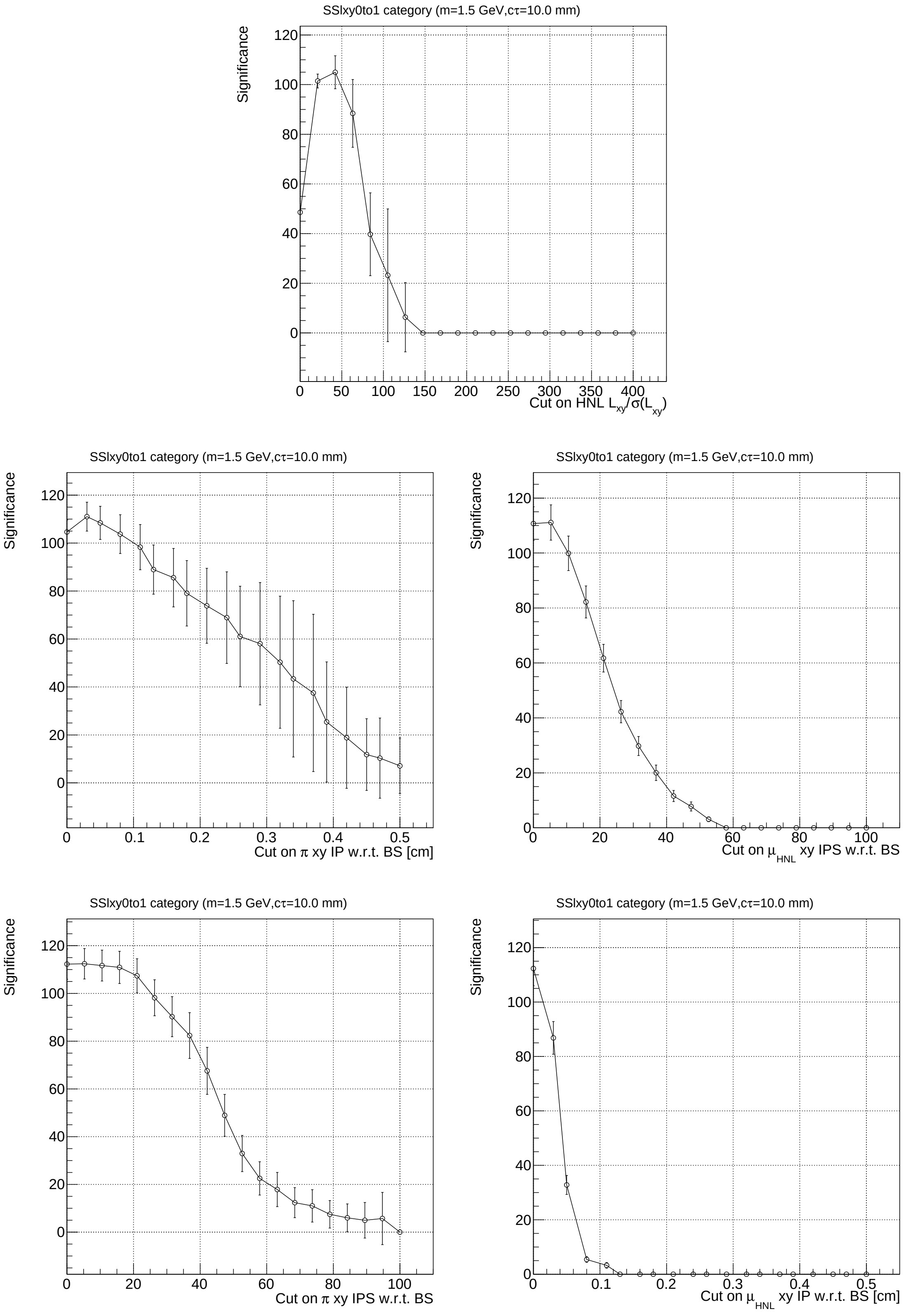

Figure 4.16: Expected discovery significance for SSlxy0to1 category.

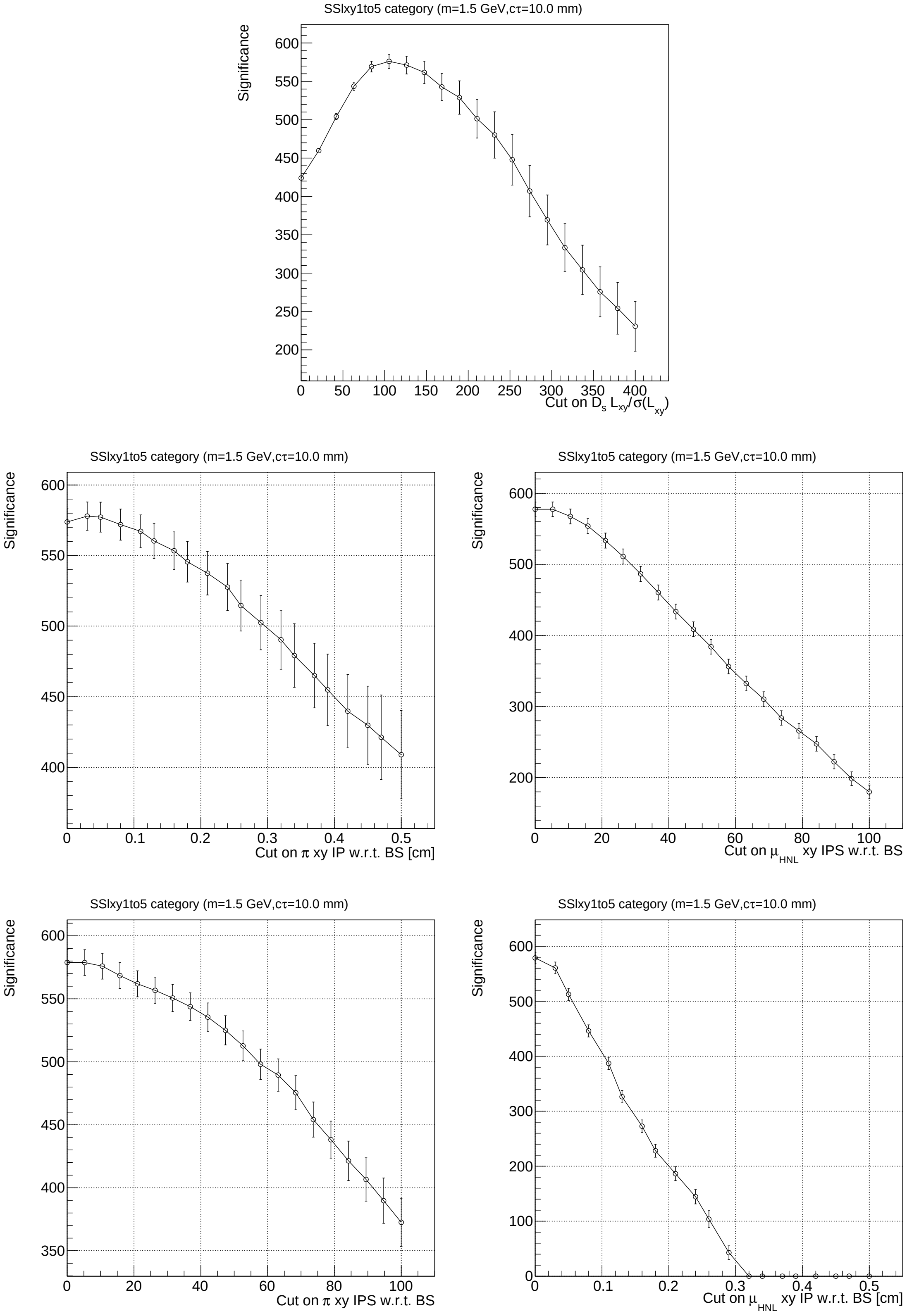

Figure 4.17: Expected discovery significance for SSlxy1to5 category.

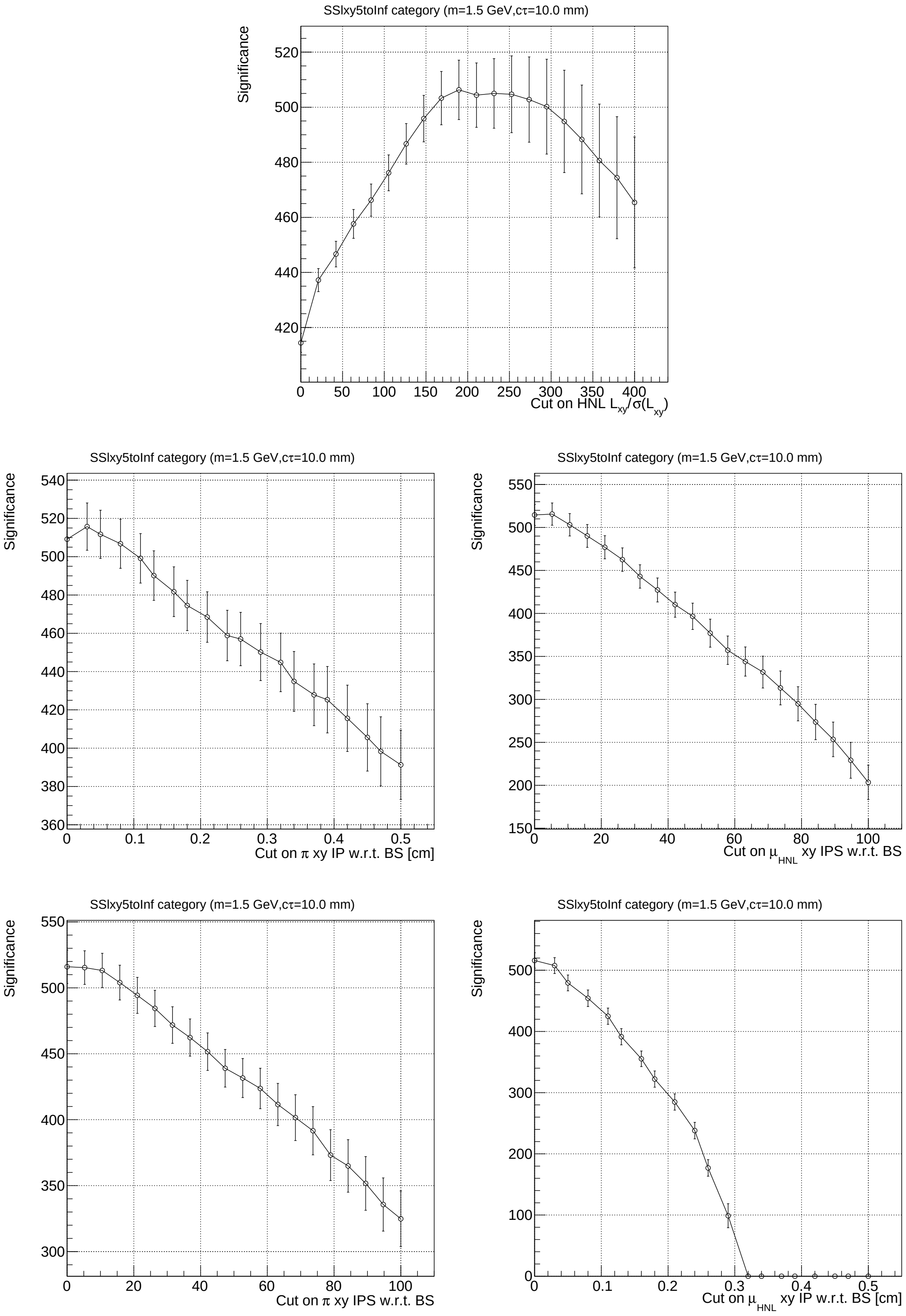

Figure 4.18: Expected discovery significance for SSlxy5toInf category.

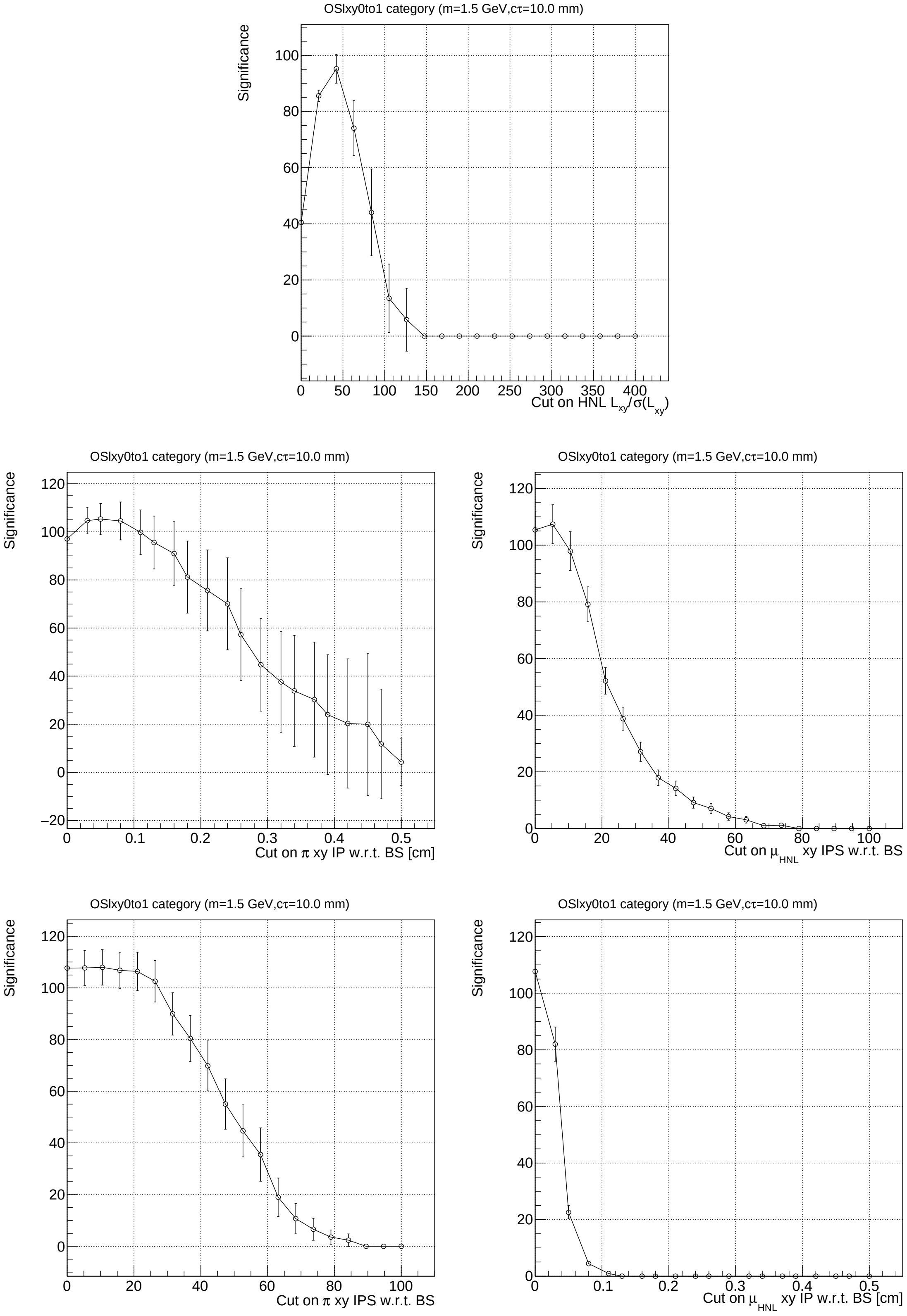

Figure 4.19: Expected discovery significance for OSlxy0to1 category.

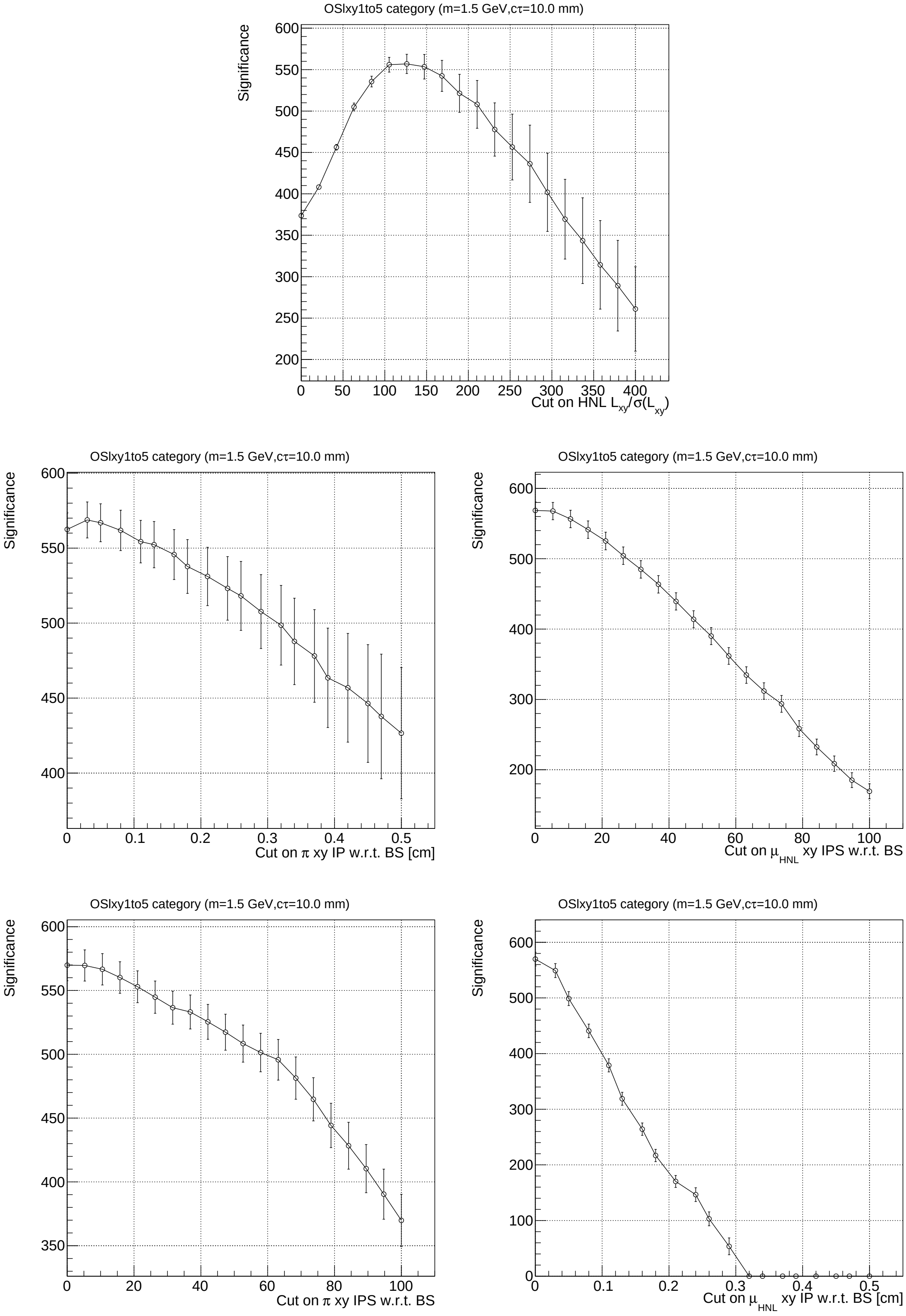

Figure 4.20: Expected discovery significance for OSlxy1to5 category.

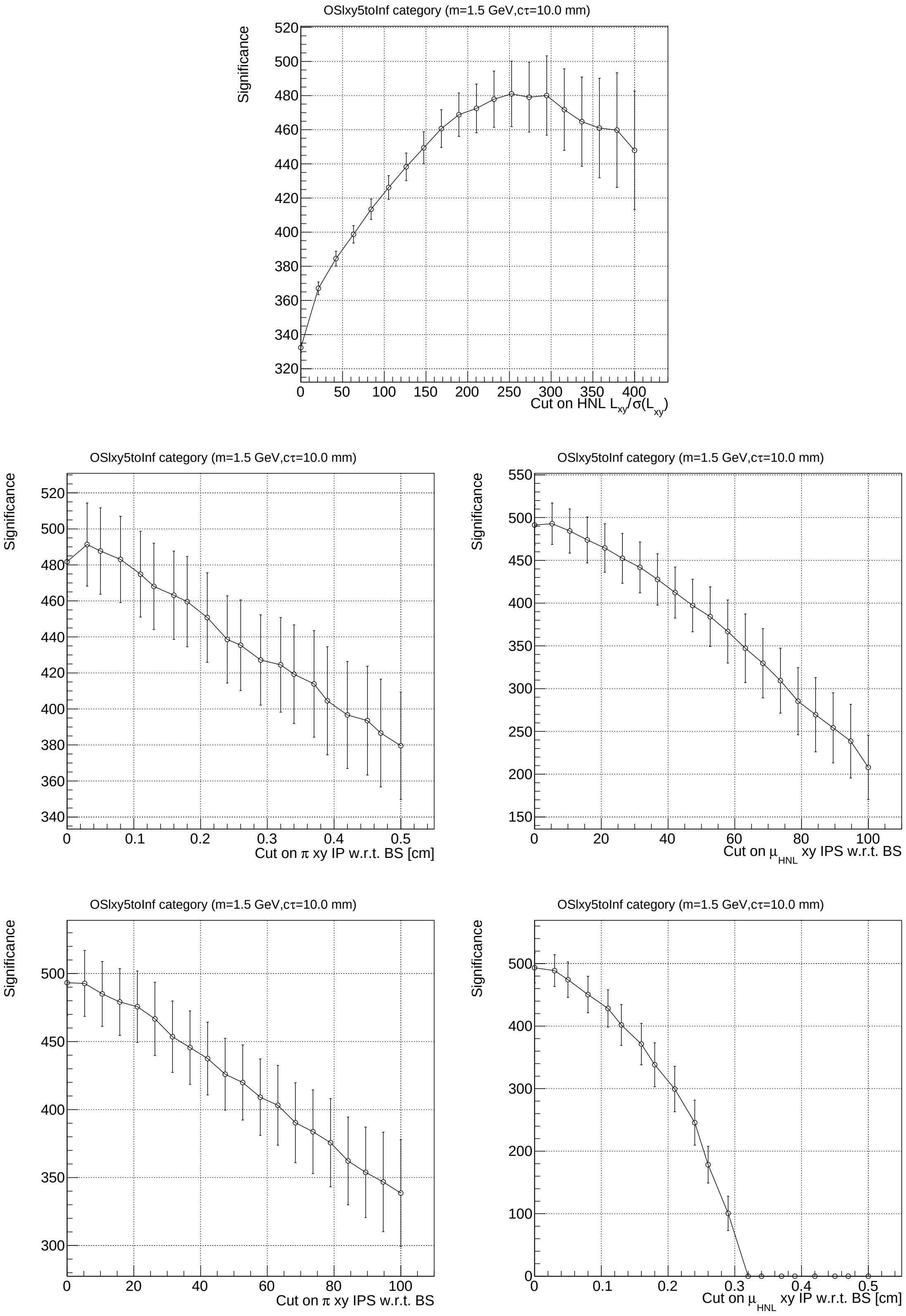

Figure 4.21: Expected discovery significance for OSlxy5toInf category.

4.6 Limits extraction

4.6.1 Lifetime reweighting

Simulated signal samples have been generated only for fixed values of $c\tau$ and m_{HNL}. To obtain a fine scan of the lifetime of the HNL, simulated samples are reweighted to cover the $c\tau$ spectrum ranging from 10 to 1000 mm. To reweight a simulated sample with lifetime $c\tau_0$ to a target lifetime $c\tau_1$, the events must be multiplied by the following weight:

$$w(ct, \tau_0 \to \tau_1) = \frac{\frac{1}{c\tau_1} \exp\left(-\frac{ct}{c\tau_1}\right)}{\frac{1}{c\tau_0} \exp\left(-\frac{ct}{c\tau_0}\right)}, \tag{4.9}$$

where ct is the generator-level proper lifetime of the HNL in the event, calculated as:

$$ct = \frac{LM}{p} \tag{4.10}$$

with L, M and p being the HNL decay length, mass and momentum at generator level. To validate the procedure, a study was performed at the reconstruction level. The reconstructed HNL decay length is measured in simulation for two samples generated for the mass value $m_{\mathrm{HNL}} = 1.5\,\mathrm{GeV}$: the first having $c\tau_0 = 10\,\mathrm{mm}$ and the second having $c\tau_0 = 100\,\mathrm{mm}$. The HNL decay length distribution is then studied for HNL candidates passing the preselection requirements. Figure 4.22 shows the reconstructed HNL decay length for both simulated samples. After applying lifetime weights to the sample with $c\tau_0 = 10\,\mathrm{mm}$, a good agreement was obtained with the sample with $c\tau_0 = 100\,\mathrm{mm}$.

4.6.2 Signal and background fit

The analysis consists of a search for heavy neutral leptons of unknown mass. To search for different mass hypotheses, a scan in the displaced lepton and pion invariant mass spectrum has been performed. To extract the results, a simultaneous fit of signal and background is performed to extract the signal strength. The fit is performed, for each category, using $\mu\mu\pi$ candidates that pass the optimized selection requirements, defined in Sec. 4.5.5, and that have a mass compatible with the known D_s mass within 3σ. The resolution σ on the D_s mass is extracted by fitting the $\mu\mu\pi$ invariant mass, in simulated HNL events, using a Voigtian function with the mean fixed to the known D_s mass, and it is found to be 14.3 MeV.

For each signal mass hypothesis, and in each category, a signal region and a sideband region around the HNL mass hypothesis are defined, as described in Sec. 4.5.5. A window of size $6\sigma_{\mathrm{HNL}}$ is defined around the given mass, where σ_{HNL} is the signal peak resolution. The signal resolution is determined by a fit of the $\mu\pi$ invariant mass in HNL simulated events using a Voigtian function with the mean fixed at the generated HNL mass value. The value of σ_{HNL} is averaged between the different generated heavy neutrino lifetime, since it is expected to depend only on the mass of the heavy neutral lepton. This means that the mass window of the signal region depends only on the heavy neutrino mass hypothesis. The value of σ_{HNL} that has been used is 12.8 MeV for the 1.5 GeV HNL and 7.7 MeV for the 1.0 GeV HNL. The resulting PDF fitting the signal in simulation is then used to fit the contribution from signal in data. The background yield is extracted from the data

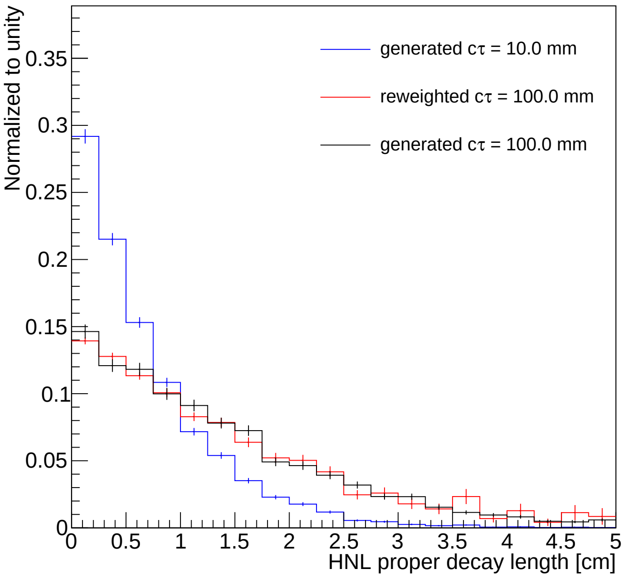

Figure 4.22: HNL reconstructed proper decay length normalized distribution for $c\tau = 10$ (blue) and 100 mm (black) simulated samples. The 10 mm reweighted distribution to match the 100 mm simulation is shown in red.

as described in Sec. 4.5.5. The HNL mass sidebands are fitted with an exponential function. The fitted function is then used to extract the background yield in the signal region. This procedure is then repeated in all the mass windows.

The result of the fit of the mass sidebands is shown in Fig. 4.23, where the top two rows show the fit in the case of a 1.0 GeV HNL and the bottom two rows show the case of a 1.5 GeV HNL. On the other hand, Figs. 4.24, 4.25 and 4.26 show the fit of the HNL simulated samples in the various category for a generated lifetime of 10, 100 and 1000 mm, respectively. The top (bottom) two rows show the fit of a 1.0 (1.5) GeV HNL. As expected, the categories having a small transverse displacement are more efficient for generated signal having a small $c\tau$. Also, the background populates more the categories having a small transverse displacement, since combinatorial background is expected to have a small displacement with respect to the beam spot. The background yield is found to be similar for the SS and OS categories corresponding to a given L_{xy} range. This is expected when the main source of the background is combinatorial.

4.6.3 Expected limits

Given that the analysis is still blinded, only expected exclusion limits can be extracted at this stage. Limits are computed using the CLs criterion with a binned profile likelihood ratio in the asymptotic approximation as the test statistic [98]. The CLs criterion consists in excluding the signal based on the combination of the p-value computed under the signal+background hypothesis (p_μ), and the one computed in the background only hypothesis (p_b), namely:

$$\mathrm{CL}_s = \frac{p_\mu}{1 - p_b}\,. \tag{4.11}$$

The p-values are computed using $q_\mu = -2\ln\lambda(\mu)$ as test statistic, where $\lambda(\mu)$ is the profile likelihood ratio which depends on the signal strength modifier μ. Denoting

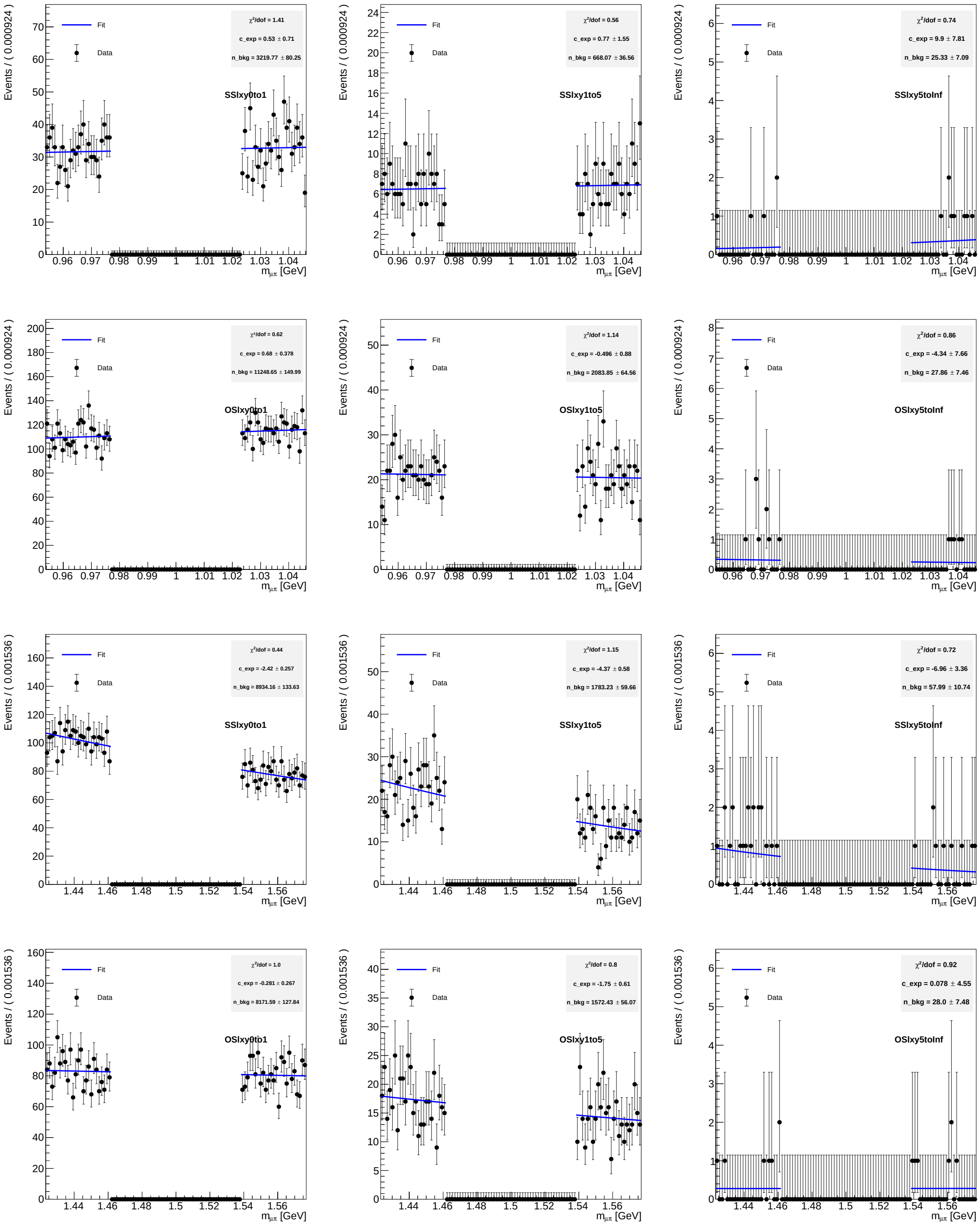

Figure 4.23: Fit of the $\mu\pi$ invariant mass sidebands in data in the case of 1.0 GeV mass (top two rows) and 1.5 GeV mass (bottom two rows) HNL hypothesis. First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

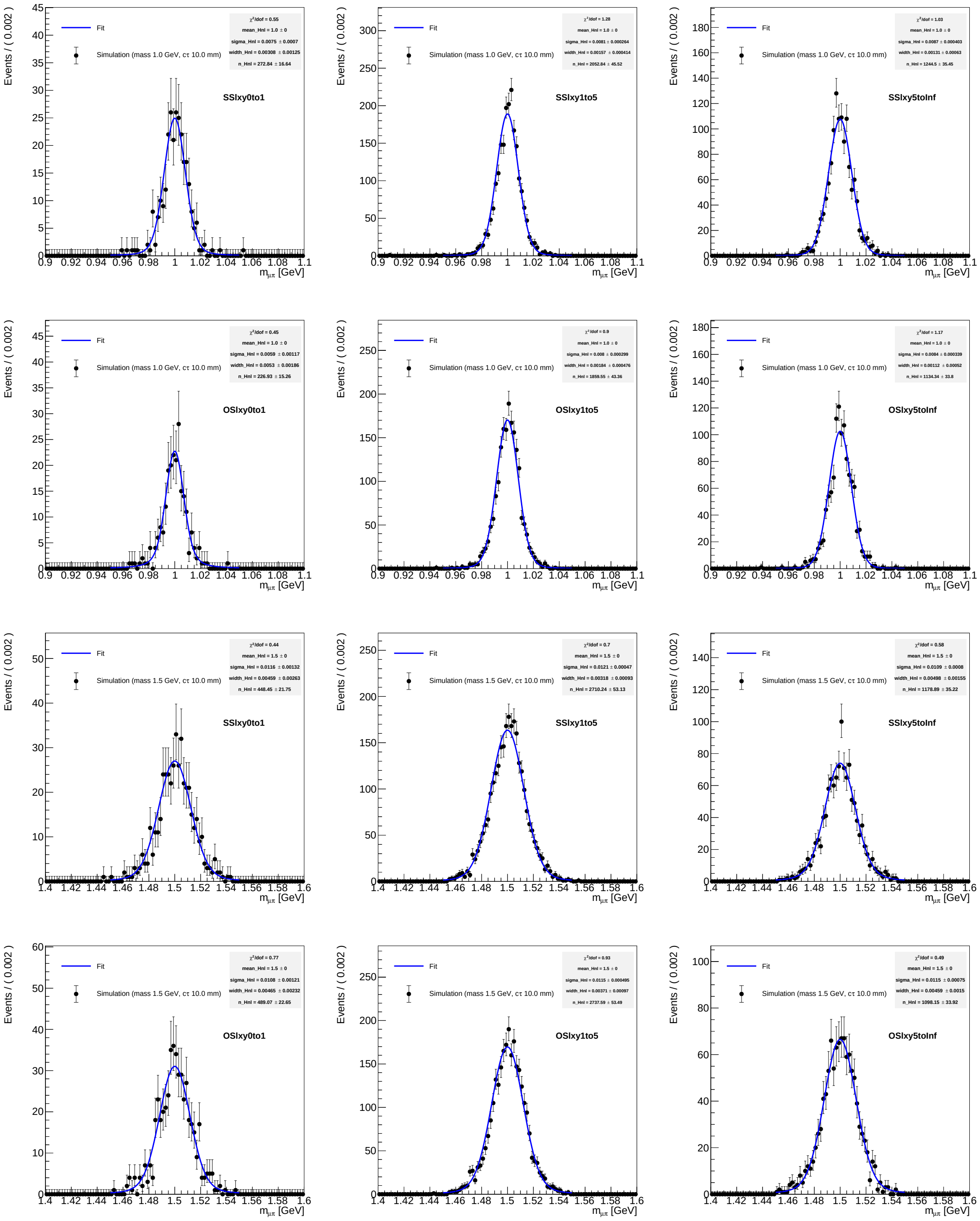

Figure 4.24: Fit of the $\mu\pi$ invariant mass of a $c\tau = 10\,\mathrm{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

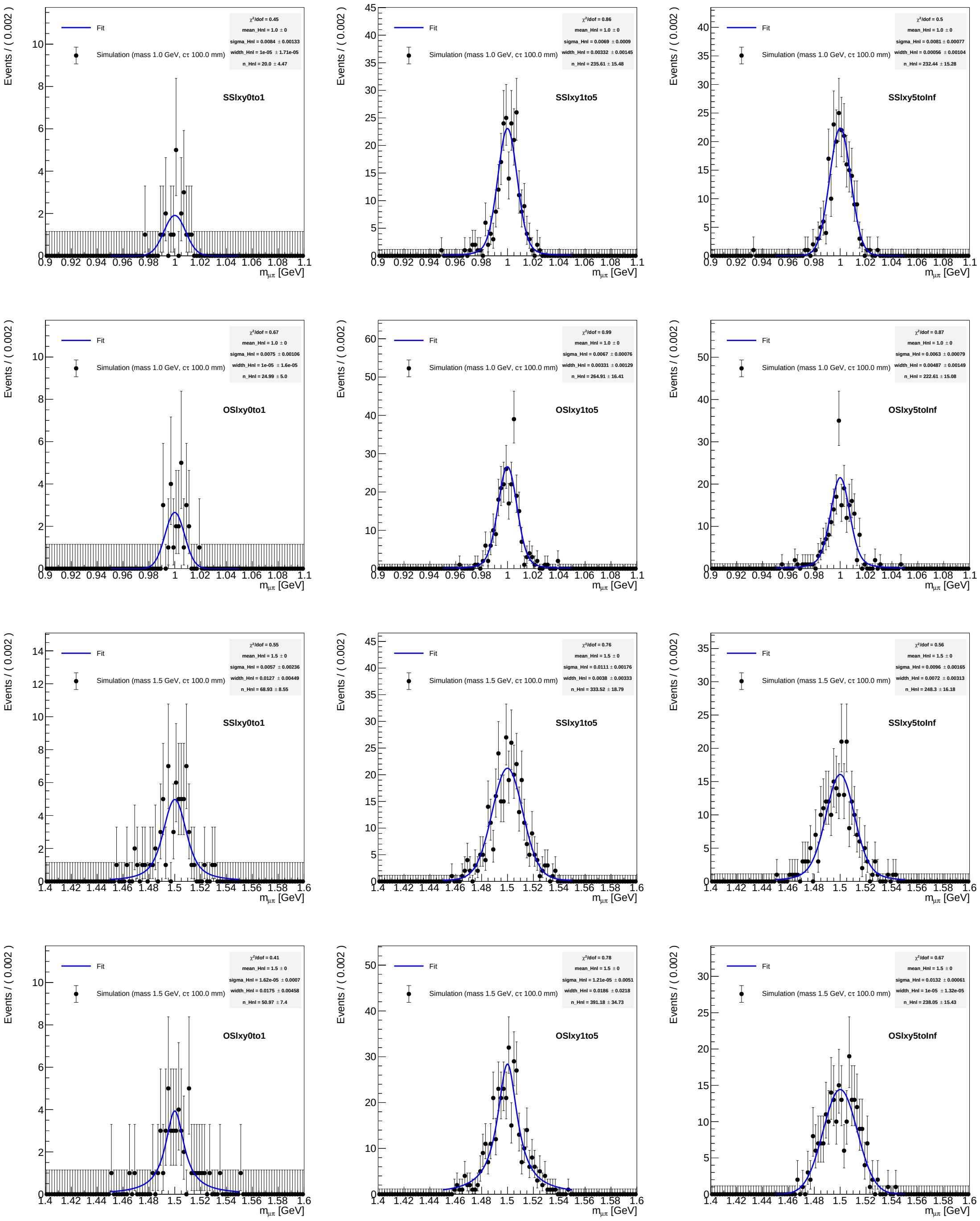

Figure 4.25: Fit of the $\mu\pi$ invariant mass of a $c\tau = 100\,\text{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

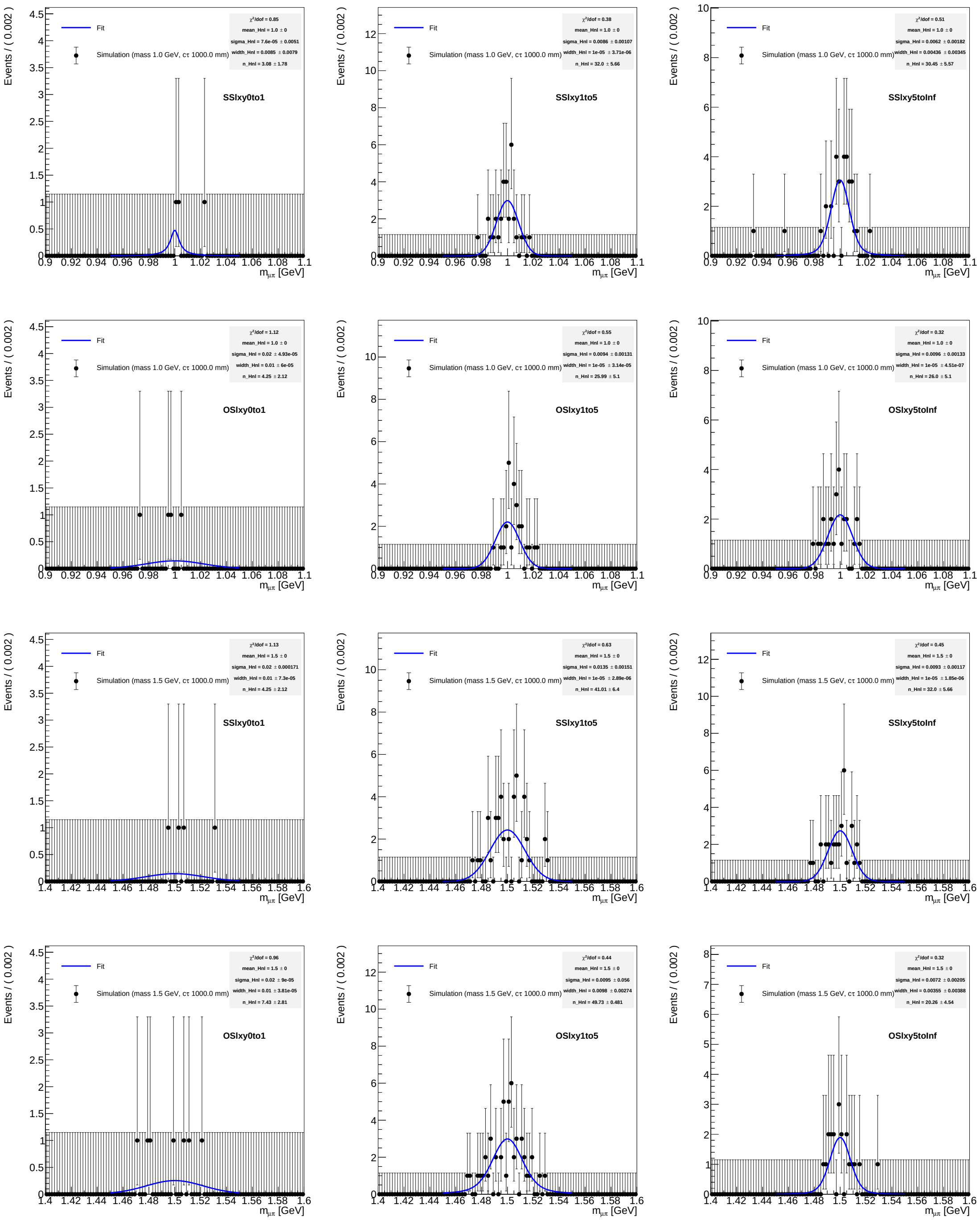

Figure 4.26: Fit of the $\mu\pi$ invariant mass of a $c\tau = 1000\,\text{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

the N mass measurements as m_i, where $i = 1, \ldots, N$, the likelihood function which describes signal and background can be written as

$$L(m_i; \mu, \vec{\theta}) = \frac{e^{\mu N_\mathrm{s} + N_\mathrm{b}}}{N!} \prod_{i=1}^{N} [\mu N_\mathrm{s} V(m_i; \mu_\mathrm{HNL}, \sigma_\mathrm{HNL}, \Gamma_\mathrm{HNL}) + N_\mathrm{b} B(m_i; \lambda)] , \tag{4.12}$$

where $V(m_i; \mu_\mathrm{HNL}, \sigma_\mathrm{HNL}, \Gamma_\mathrm{HNL})$ (see Eq. 4.4) and $B(m_i; \lambda)$ (see Eq. 4.5) are the Voigtian and exponential PDFs describing signal and background, $\vec{\theta}$ is the set of nuisance parameters, and N_s (N_b) is the number of signal (background) events.

The parameter of interest, onto which the limits are set, is the signal strength modifier μ. The limits are derived for a confidence level (CL) of 95%. The expected background yield is extracted from data fitting HNL mass sidebands, as described in Sec. 4.6.2, while the number of expected signal events is extracted using Eq. 4.3, where the signal yield is reweighted as described in Sec. 4.6.1 to span over the entire 10 to 1000 mm lifetime spectrum. The expected exclusion limit on the signal strength at 95% confidence level is derived for the two HNL mass scenarios as a function of the $|V_\mu|^2$ mixing parameter. Before showing the final expected exclusion limits, the derivation of the systematic uncertainties will be discussed in the next section.

4.7 Systematic uncertainties

The sources of systematic uncertainties might affect the background yield or the signal yield estimation. The systematic uncertainties affecting the signal yield come from any eventual bias that is introduced in each term of the calculation of Eq. 4.3. Any systematic uncertainty that affects the selection efficiency measurement of the signal and normalization channels may partially cancel out in the $\varepsilon_{D_s^+ \to N\mu^+} / \varepsilon_{D_s^+ \to \phi\pi^+}$ ratio. However, the signal and normalization channels are expected to have different kinematics. This can lead the systematic uncertainties to not cancel out completely in the efficiency ratio. On the other hand, since the background yield is estimated using real data, the main contribution to systematic uncertainties comes from the modeling of the background shape, which in principle could lead to a different yield. In this section the sources of systematic uncertainties considered so far are presented, divided on whether they affect the signal or the background yield estimation. A summary of the systematic uncertainties that will be discussed in the following sections is showed in Tab. 4.9 where to each source it is associated the systematic uncertainty, that is expressed as a range if it varies among the different categories.

4.7.1 Number of expected D_s mesons

The number of reconstructed D_s mesons in the normalization channel enters Eq. 4.3 and it is crucial to have an estimate of the expected number of D_s mesons. This number depends on the PDF used to model the D_s signal peak. To account for systematic uncertainties affecting the D_s yield measured in the normalization channel, the D and D_s signal peaks have been fitted with two alternative functions: a Gaussian PDF, which depends on two parameters, μ and σ, and a Crystal ball function [105], which depends on four parameters, m_0, σ, α and n. In both cases, the background has been fitted using an exponential function, as in Sec. 4.4.1, Eq. 4.5. The result of the $\mu\mu\pi$ invariant mass fit in the normalization channel is shown in Fig. 4.27, where the D and D_s signal peaks have been fitted using Crystal ball

Source	Systematic uncertainty (%)	Affected yield
$D_s^+ \to \phi\pi^+$ PDF	10	signal
$D_s^+ \to \phi\pi^+$ fixed/floating mass	0.2	signal
Data/MC agreement	1-15	signal
Vertex reconstruction efficiency	7	signal
Muon ID/reconstruction scale factor	0.1-1	signal
Background PDF	0.1-7	background

Table 4.9: Summary of the systematic uncertainties affecting signal and background yield estimation. First and second columns indicate the source of the systematic uncertainty and the associated value. The third column indicates if the systematic uncertainty affects the signal or background yield.

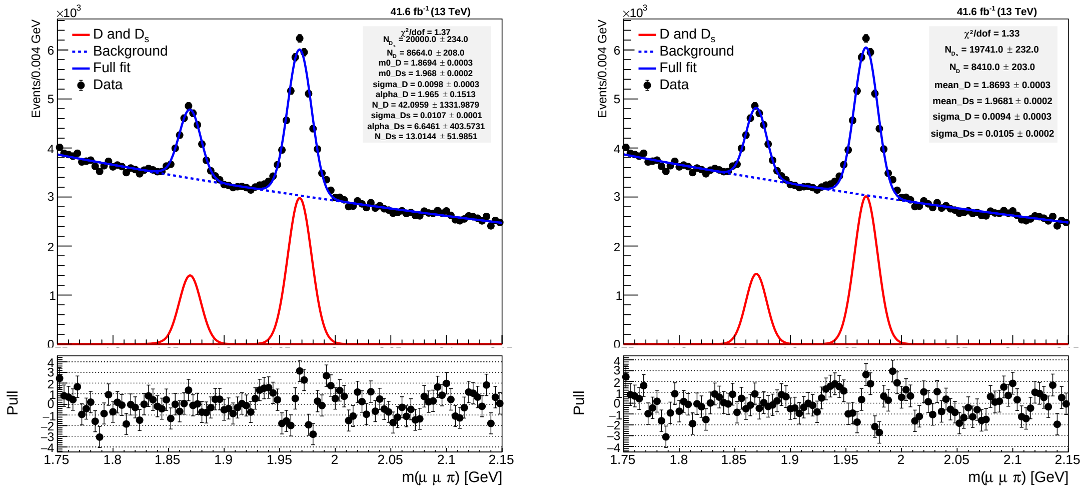

Figure 4.27: Fit of the D and D_s candidates invariant mass in the $D_{(s)} \to \phi\pi$ normalization channel using Double Crystal ball functions (left) and Gaussian functions (right).

PDFs (left plot) and Gaussian PDFs (right plot). The parameters obtained in the fit are summarized in Tab. 4.10, where the parameters labeled with the D and D_s subscripts refer to the PDF of the D and D_s mesons, respectively.

The maximum value of the relative difference between the D_s yield measured using a Voigtian function and each of the two alternative function results in a systematic uncertainty of 10%. In addition, the effect of fixing the means of the PDFs used to fit the D and D_s peaks to the known values of the masses of the D and D_s mesons has been tested. When using the reference model (Eq. 4.6) to fit the $\mu\mu\pi$ invariant in the normalization channel, the effect of leaving the mean of the Voigtian PDF fixed or free to float resulted in a relative D_s yield difference of 0.2%, which is considered as a sytematic uncertainty.

4.7.2 Data/MC discrepancies

The residual discrepancies between data and simulation after the application of the trigger scale factors might result in a bad estimate of the selection efficiencies, i.e. $\varepsilon_{D_s^+ \to N\mu^+}$ and $\varepsilon_{D_s^+ \to \phi\pi^+}$ of Eq. 4.3. To take into account for this effect, another source of systematic uncertainty affecting the signal yield has been considered. Although

Fitted parameters			
Crystal ball + exponential		Gaussian + exponential	
λ	-1.12 ± 0.02	λ	-1.11 ± 0.02
N_D	8717 ± 199	N_D	8401 ± 293
N_{D_s}	20015 ± 226	N_{D_s}	19741 ± 283
N_b	311936 ± 576	N_b	312301 ± 686
$m_{0,D}$	1.8694 ± 0.0003	μ_D	1.8693 ± 0.0003
σ_D	0.0098 ± 0.0003	σ_D	0.0094 ± 0.0004
α_D	2.0 ± 0.2	μ_{D_s}	1.9681 ± 0.0002
n_D	42 ± 1330	σ_{D_s}	0.0104 ± 0.0002
m_{0,D_s}	1.9680 ± 0.0002		
σ_{D_s}	0.0107 ± 0.0001		
α_{D_s}	6 ± 404		
n_{D_s}	13 ± 52		

Table 4.10: Fitted parameters of the alternative PDFs used to fit the $\mu\mu\pi$ invariant mass in the normalization channel.

the correction of the trigger efficiency is responsible for a significant improvement of the data/MC agreement, especially for the muon p_T, a discrepancy of 10-15% is still visible between data and simulation for the p_T and the impact parameter significance of the muons (see figs. 4.9 and 4.10).

In order to quantify these systematic uncertainties, the residual data/MC disagreement has been used to reweight both signal and $D_s^+ \to \phi\pi^+$ simulated samples. The variables taken into account are the same used for the trigger scale factor calculation, i.e. the p_T of the muon from D_s and HNL decay as well as their impact parameter significance. For the simulated $D_s^+ \to \phi\pi^+$ sample the p_T, or IPS, of the muons from the ϕ decay have been used. The simulated samples have been reweighted separately for each of these variables. The systematic uncertainty has been evaluated by taking the relative difference of the efficiency ratio in Eq. 4.3 before and after reweighting the simulated samples for the residual data/MC disagreement. The procedure has been repeated for each aforementioned kinematic variable, category, HNL mass and lifetime point, and the resulting uncertainties have been included in the limit calculation. A summary of these systematic uncertainties are shown in Tab. 4.11. The uncertainties vary depending on the category and on the reweighting variable and range between few 1% to approximately 15%.

4.7.3 Vertex reconstrucion efficiency

Since the most important signature of the production of a heavy neutral lepton from the D_s decay is its displaced $\mu\pi$ vertex, the reconstruction efficiency as a function of the heavy neutral lepton displacement is a key feature of this analysis. Any discrepancy in the vertex reconstruction efficiency between data and MC affects the selection efficiency estimate in Eq. 4.3 and could introduce a bias in the signal yield estimate.

The vertex reconstruction efficiency has been measured in a CMS search for long-lived heavy neutrinos originating from the W boson decay, and decaying leptonically [106]. The vertex reconstruction efficiency has been measured in data and MC exploiting the K_S^0 meson decays to a pair of charged particles. The measurement has been done in bins of radial distance from the primary vertex and transverse

m_{HNL} (GeV)	$c\tau_{\text{HNL}}$ (mm)	Systematic uncertainty (%)						
		SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf	Variable
1.0	10	4	4	4	4	4	4	μ_{D_s} p_T
		0.3	0.4	0.3	0.4	0.5	0.1	μ_{HNL} p_T
		6	6	7	7	6	7	μ_{D_s} IPS
		9	2	5	9	2	5	μ_{HNL} IPS
	100	5	3	3	7	3	4	μ_{D_s} p_T
		1	1	1	3	1	0	μ_{HNL} p_T
		8	6	7	9	5	7	μ_{D_s} IPS
		9	2	4	9	3	5	μ_{HNL} IPS
	1000	4	1	4	6	4	5	μ_{D_s} p_T
		0	6	0	2	0	1	μ_{HNL} p_T
		6	5	10	15	2	7	μ_{D_s} IPS
		12	0	2	11	3	3	μ_{HNL} IPS
1.5	10	7	6	6	7	6	6	μ_{D_s} p_T
		3	2	2	3	2	2	μ_{HNL} p_T
		13	12	12	13	11	11	μ_{D_s} IPS
		9	5	9	8	4	8	μ_{HNL} IPS
	100	7	6	7	6	7	6	μ_{D_s} p_T
		3	2	3	2	2	2	μ_{HNL} p_T
		13	11	10	12	11	11	μ_{D_s} IPS
		8	5	7	8	6	8	μ_{HNL} IPS
	1000	4	4	7	10	6	7	μ_{D_s} p_T
		0	0	3	6	2	3	μ_{HNL} p_T
		8	10	13	16	11	14	μ_{D_s} IPS
		10	5	2	2	9	3	μ_{HNL} IPS

Table 4.11: Systematic uncertainties (%) due to the residual p_T and IPS mismodeling in simulations. The first two columns indicate the mass and lifetime of the simulated samples. The last column shows the variable used to reweight the MC: μ_{D_s} (μ_{HNL}) p_T indicates the transverse momentum of the muon forming the $D_s^+ \to N\mu^+$ ($N \to \mu^+\pi^-$) candidate, and the same convention has been used to name the impact parameter significance (IPS) variable. The uncertainties are reported for each category and variable separately.

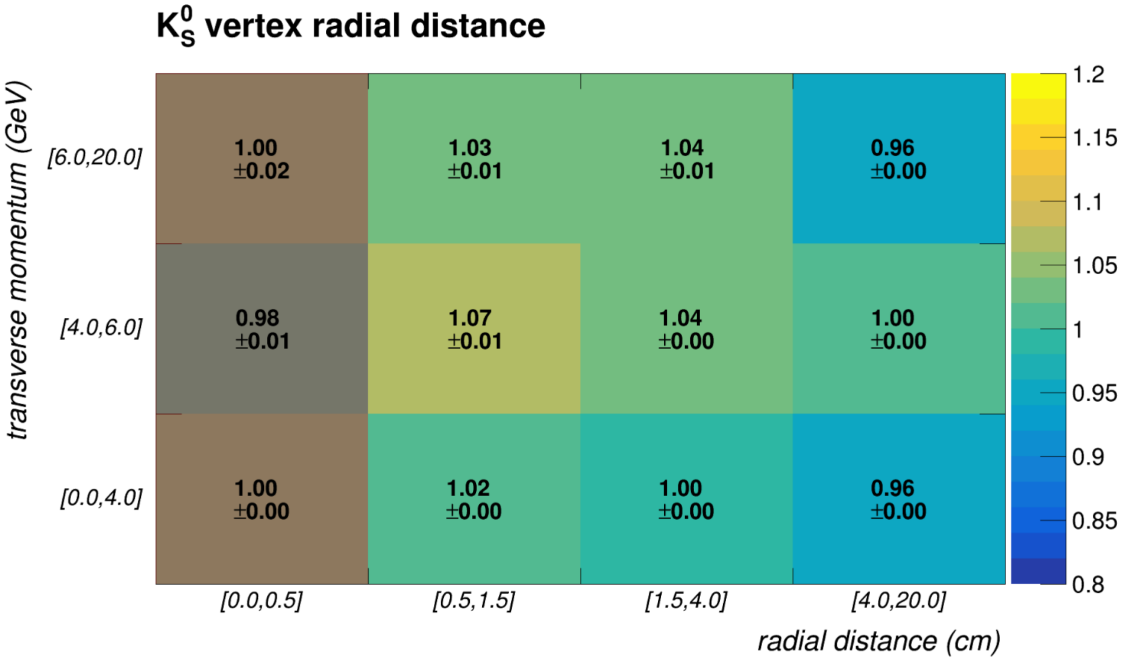

Figure 4.28: Ratio of the K_s^0 vertex reconstruction efficiency determined in data and in MC simulation in bins of radial distance and transverse momentum [106].

momentum of the K_S^0 meson. A set of scale factors was derived by dividing the efficiency in data and MC. The vertex reconstruction efficiency scale factors are shown in Fig. 4.28. The scale factors are all close to 1 with 7% being the largest discrepancy between data and MC. Since the difference between data and simulation is small, no correction is applied and rather a systematic uncertainty that account for this effect is applied. A conservative flat 7% systematic uncertainty in the signal selection efficiency $\varepsilon_{D_s^+ \to \phi\pi^+}$ (Eq.4.3) is considered.

4.7.4 Muon ID and reconstruction scale factors

In order to estimate the systematic uncertainties related to the muon identification and reconstruction, the expected signal yield has been calculated after varying the scale factors by their associated uncertainty. Since these scale factors are applied to MC simulations, they affect the selection efficiency estimate and thus could result in a biased HNL signal yield estimate.

Namely, for the identification scale factors $w_{\text{ID}}(p_T, \eta)$, the expected HNL signal yield has been re-calculated by using a new set of muon ID scale factors $w_{\text{ID}}^{\text{up}}(p_T, \eta) = w_{\text{ID}}(p_T, \eta) \pm \delta_{\text{ID}}(p_T, \eta)$ which have been obtained by varying the nominal scale factor by the related uncertainty $\delta_{\text{ID}}(p_T, \eta)$. When the + (−) sign is considered in the variation it will be referred to as up (down) variation. The relative difference with respect to the signal yield calculated using the scale factors without any variation is considered as systematic uncertainty. This procedure has been repeated for each generated HNL mass and lifetime, and for the muon reconstruction scale factors as well. Thus, the obtained systematic uncertainties are summarized in Tab. 4.12. The uncertainties are almost always the same for an up or down variation of the scale factor and, most importantly, always below 0.5% with few exceptions. In summary, two sources of systematic uncertainty are considered, originating from the muon ID and reconstruction scale factors application. Since the difference between the

m_N [GeV]	$c\tau$ [mm]	type	Systematic uncertainty after up/down scale factors variation (%)					
			SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
1.5	10	ID	0.4	0.3	0.3	0.5	0.3	0.1
		RECO	0.3	0.2	0.2	0.3	0.2	< 0.1
	100	ID	0.5	0.2	0.3	0.5	0.4	0.4
		RECO	0.4	< 0.1/0.1	0.1	0.4	0.2	0.4
	1000	ID	0.5	0.4	0.4	0.9	0.6	0.4
		RECO	0.3	0.3	0.4	1.2	0.7	0.3
1.0	10	ID	0.4/0.3	0.3	0.4	0.4	0.3	0.4
		RECO	0.2	0.3	0.4	0.2	0.3	0.3
	100	ID	0.4	0.4	0.3	0.4	0.4	0.4
		RECO	< 0.1/0.1	0.4	0.3	0.3	0.3	0.3
	1000	ID	0.2	0.6	0.5	0.9	0.5	0.5
		RECO	0.2	0.6	0.5	1.0	0.6	0.8

Table 4.12: Systematic uncertainties derived by adding or subtracting to the identification (ID) and reconstruction (RECO) muon scale factors their associated error. These variations are referred to as up/down variations. The procedure has been repeated for each HNL mass and $c\tau$ generated point. When a single value is displayed, the systematic uncertainty derived for up and down variation are identical (up to the first decimal digit).

systematic uncertainty coming from an up or down variation of the scale factors is almost always below 0.1%, for each of the sources the largest systematic uncertainty between up and down variation is considered if it is not below 0.1%.

4.7.5 Background yield estimation

As described in Sec. 4.6.2, the expected number of background events in the signal region is estimated by fitting the HNL candidate mass sidebands. Thus, an alternative PDF to the exponential function might result in a different background yield. In order to take into account this effect the HNL candidate mass sidebands have been fitted using two alternative functions: a power law and a first order polynomial. This study has been repeated in each category for the two mass points taken into account.

The fit of the $\mu\pi$ invariant mass using the alternative PDFs in each category is shown in Figs. 4.29 and 4.30 for the 1 and 1.5 GeV mass points. It can be observed that all the alternative shapes are almost perfectly overlapped and close to a straight line shape. Small differences can be observed in large displacement categories, where the number of event is small. For each of the alternative PDFs the background yield in the signal region is extracted, and the largest difference with respect to the yield extracted with the exponential shape is taken as systematic uncertainty. The procedure is repeated for each category, and the resulting systematic uncertainties are reported in Tab. 4.13. The resulting systematic uncertainties due to the different parametrization of the background shape is around few percent for almost all categories with many values smaller than 0.1%.

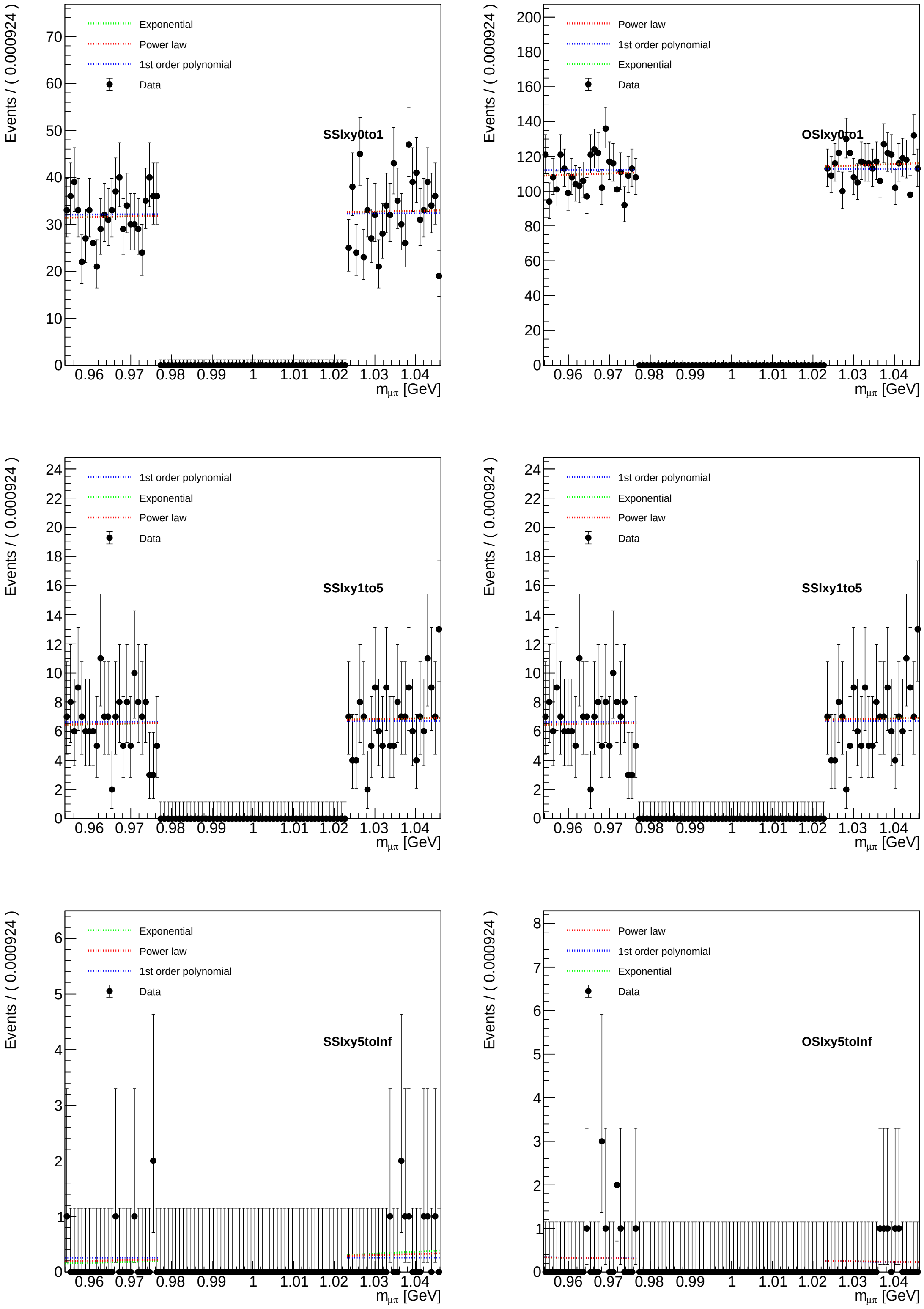

Figure 4.29: Fit of the $\mu\pi$ invariant mass spectrum in data with different background hypotheses for a 1 GeV heavy neutral lepton.

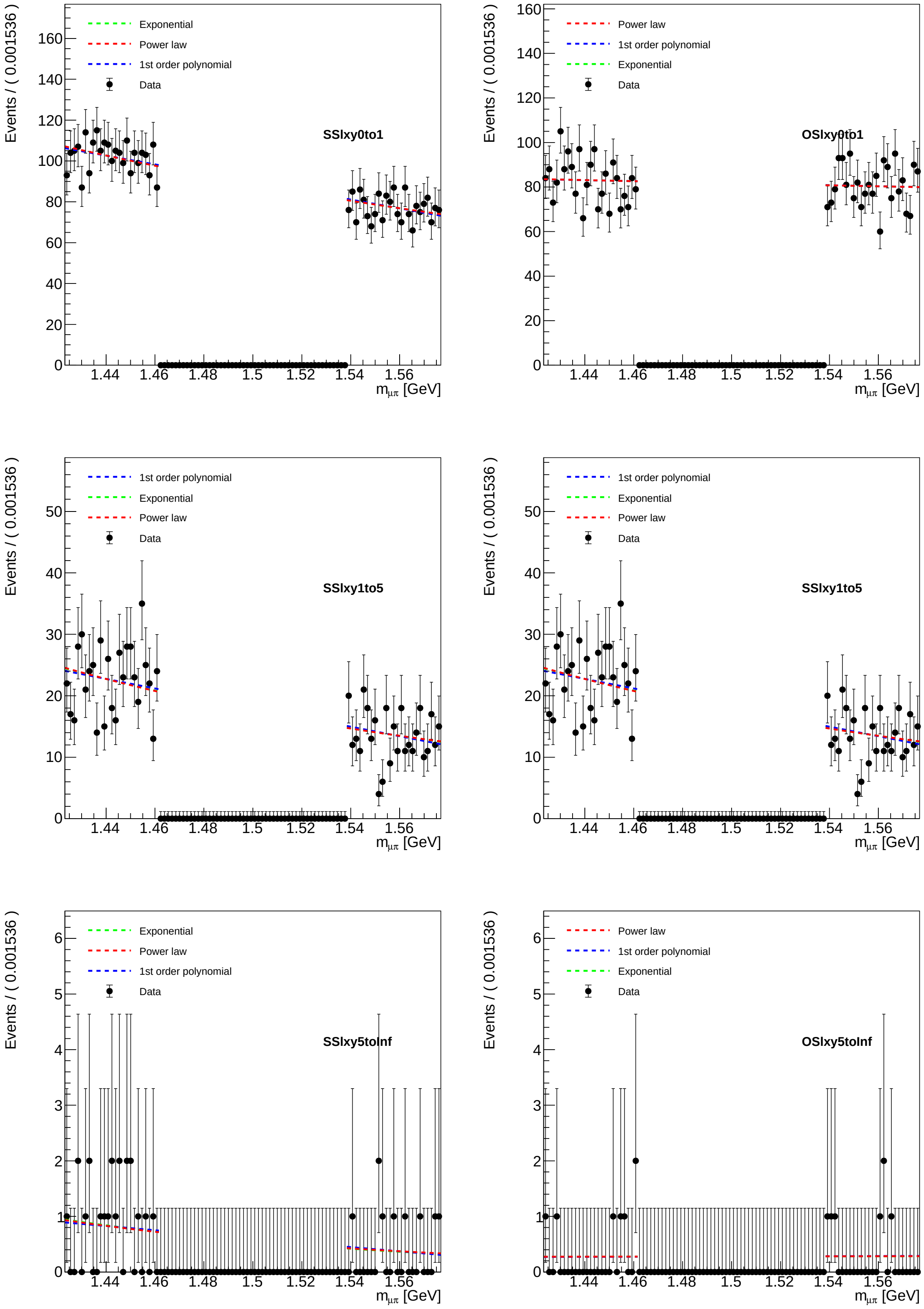

Figure 4.30: Fit of the $\mu\pi$ invariant mass spectrum in data with different background hypotheses for a 1.5 GeV heavy neutral lepton.

HNL mass (GeV)	Background yield systematic uncertainty (%)					
	SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
1.0	0.8	2.8	7.1	< 0.1	0.4	0.1
1.5	< 0.1	< 0.1	5.3	< 0.1	< 0.1	1.1

Table 4.13: Systematic uncertainties on background yield estimation for each category derived using alternative PDFs to fit the HNL candidate mass sidebands.

4.8 Results

The expected exclusion limit on the signal strength at 95% confidence level is derived for the two HNL mass scenarios as a function of the $|V_\mu|^2$ mixing parameter, and it is shown in Fig. 4.31, where all the systematic uncertainties listed in Sec. 4.7 have been taken into account. All systematic uncertainties are considered to be uncorrelated and to follow a log-normal distribution. The usage of a log-normal distribution is preferred to limit the distribution to positive values and avoid negative yields. The expected exclusion limits on $|V_\mu|^2$ for the 1 and 1.5 GeV mass hypothesis are derived by intercepting the expected exclusion limit as a function of $|V_\mu|^2$ with the hypothesis of unit signal strength, shown in 4.31 as the black solid horizontal line. Figure 4.32 shows the expected excluded $|V_\mu|^2$ region as a function of the signal mass, where the two expected limits at 1 and 1.5 GeV mass points are interpolated using a straight line. The sensitivity of the analysis ranges between approximately 10^{-4}, for the 1 GeV mass point, and 7×10^{-5} for the 1.5 GeV mass point.

In the context of the existing limits on the active-sterile neutrino mixing matrix parameters, this result can be compared with the $|V_\mu|^2$ regions excluded by existing results, which are summarized in Fig. 4.33. The left plot contains all the limits obtained so far and the potential excluded regions at future or planned experiments [11]. The right plot contains only the most recent limits on $|V_\mu|^2$ at low masses [107]. The expected limits from the HNL search from D_s decay presented in this work are not competitive with the results obtained by beam dump experiments such as CHARM [108] and NuTeV [109], which excluded $|V_\mu|^2$ values down to $10^{-6} - 10^{-7}$ in the region below the D_s mass. However, looking at results from collider experiments, the Belle [35] and LHCb [12] collaborations were able to set limits down to about 10^{-4}-10^{-5} on $|V_\mu|^2$. In particular, in the 1.0 to 1.5 GeV mass range, the Belle collaboration was able to set an observed limit on $|V_\mu|^2$ between $\sim 0.7 \times 10^{-4}$ and $\sim 0.5 \times 10^{-4}$ while the LHCb collaboration was able to set a limit between $\sim 5 \times 10^{-3}$ and $\sim 2 \times 10^{-3}$, and both analyses targeted B meson decays. The results obtained by Belle are comparable with the ones expected from the HNL search presented in this work.

4.9 The prompt D_s production scenario

The results presented so far have been obtained without targeting any specific D_s meson production mechanism. Events where the D_s mesons are produced as a decay product of a heavier hadron include additional final state particles with respect to events where the D_s is produced promptly in the primary interaction. The presence of more tracks in the final state might result in less clean final state and a worse signal-to-background ratio could be expected. The scenario where only prompt D_s production processes are taken into account, has also been explored. To obtain new

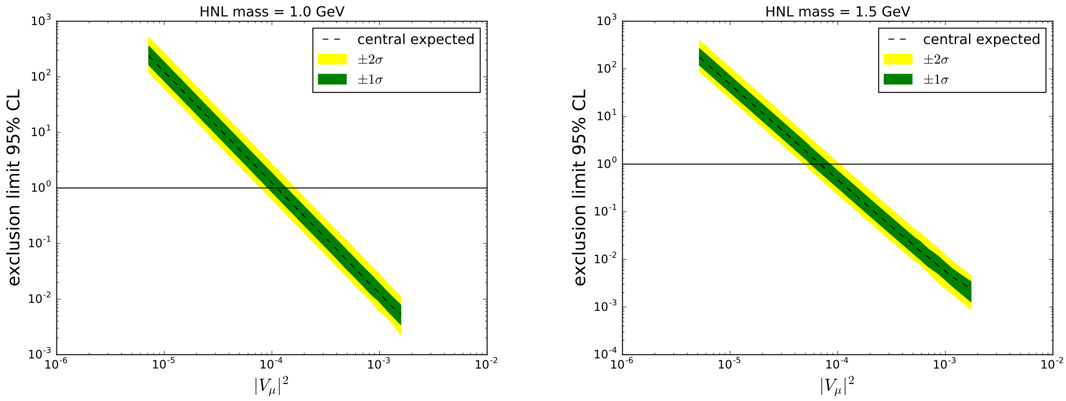

Figure 4.31: Expected exclusion limit at 95% CL on the signal strength μ as a function of $|V|^2_\mu$ for a 1.0 GeV HNL (left) and a 1.5 GeV HNL (right).

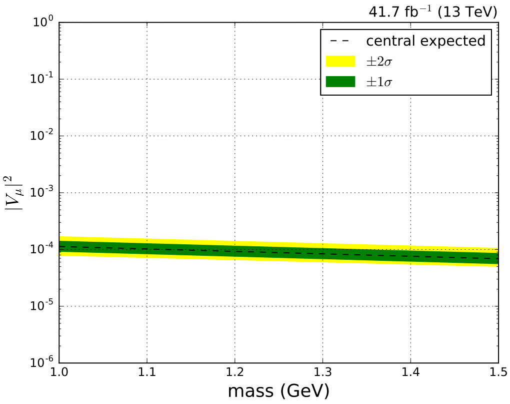

Figure 4.32: Minimum excluded $|V_\mu|^2$ mixing term as a function of the signal mass.

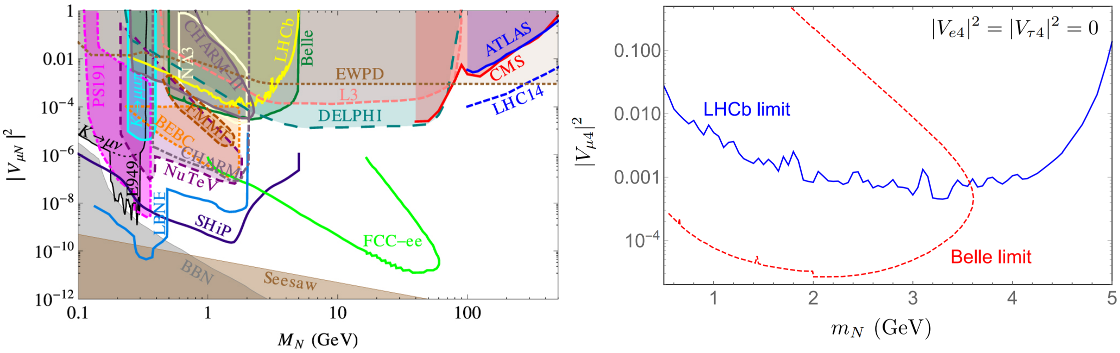

Figure 4.33: Left: existing limits on HNL masses and their couplings to muons (filled areas), and projected sensitivity achievable in planned future facilities (solid lines) [11]. Right: only the most recent limits on $|V_\mu|^2$ obtained by Belle and LHCb are shown [107].

Variable	**Categories**					
	SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
$L_{xy}/\sigma(L_{xy})$	> 33.5	> 114.5	> 192.3	> 38.7	> 123.8	> 255.5
π xy IP (cm)	> 0.04	> 0.03	> 0.03	> 0.05	> 0.05	> 0.05
μ xy IPS	> 2.8	> 2.5	> 3.3	> 3.4	> 3.7	> 3.8

Table 4.14: Selection requirements optimized for each category, using simulated HNL events from prompt D_s.

results with the prompt D_s production the analysis has been repeated by using a HNL Monte Carlo sample where the D_s was forced to be produced promptly. The expected signal yield can be estimated following the same procedure used in Sec. 4.1 to derive Eq. 4.3. However, since in the reconstruction of the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ decay, the contribution from prompt D_s cannot be decoupled from the non-prompt D_s contribution, the fraction of expected prompt D_s mesons (f_{prompt}) must be taken into account to get an estimate of the prompt D_s yield. As it will be shown in Sec. 4.9.1 f_{prompt} can be estimated from the data. By combining equations 4.1 and 4.2, and including the f_{prompt} fraction, the expected signal yield from prompt D_s reads:

$$N_{D_s^+\to N\mu^+} = N_{D_s^+\to\phi\pi^+} f_{\text{prompt}} \frac{\mathcal{B}(D_s^+ \to N\mu^+)\mathcal{B}(N \to \mu^+\pi^-)}{\mathcal{B}(D_s^+ \to \phi\pi^+)\mathcal{B}(\phi \to \mu^+\mu^-)} \frac{\varepsilon_{D_s^+\to N\mu^+}}{\varepsilon_{D_s^+\to\phi\pi^+}} . \tag{4.13}$$

Apart from the fact that to estimate the signal yield one has to take into account the f_{prompt} factor, and the different simulated signal sample, to probe the prompt D_s production scenario, the analysis has been repeated in the exact same way as for the inclusive D_s production scenario. The analysis selection requirements have been optimized in the same way as it is described in Sec. 4.5.5, i.e. by maximizing the discovery significance for each category using the simulated HNL samples from prompt D_s mesons. Moreover, the variables used to separate data from background are the same used for the inclusive D_s production scenario, since they seem to have a similar distribution in both prompt and inclusive simulations. The distributions of the HNL transverse decay length significance, the muon and pion transverse impact paramenter and impact parameter significance are shown in Fig. B.2 of Appendix B for the various HNL mass and $c\tau$ points and compared to the dijet QCD events used to model the background contribution. The discovery significance as a function of the selection cut on each of the variable used for the selection optimization is shown in figs. B.3 to B.8 of Appendix B. The new optimized cuts are listed in Tab. 4.14 while a summary of the efficiency of the whole analysis is summarized in Tab. 4.15. As observed in the inclusive channel (see Sec. 4.5.5), the preselection requirements are more efficient for simulated HNL decays with a smaller $c\tau$. The preselection efficiencies for HNLs having a different mass but the same lifetime are comparable. As for the prompt D_s channel, the selection efficiency is similar for categories sharing the same L_{xy} range.

4.9.1 Prompt D_s yield extraction

The selection of D_s candidates, as described in Sec. 4.4.1, is expected to be contaminated by D_s originating from decays of B mesons. To estimate the number of expected HNL events from prompt D_s decays, it is crucial to have an estimate of the yield of prompt and non-prompt D_s in the normalization channel.

The prompt and non-prompt D_s components can be separated on the basis of the proper distance between the D_s decay vertex and the PV: prompt D_s are ex-

m_{HNL} (GeV)	$c\tau_{\mathrm{HNL}}$ (mm)	$\varepsilon_{\mathrm{pre}}$ (%)	Selection efficiency (%) per category					
			SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf
1.5	10	3.1	38.3	70.9	74.0	29.5	65.5	68.9
	100	1.1	38.4	70.5	74.1	19.5	67.9	68.9
	1000	0.1	21.5	62.4	66.7	4.8	63.8	72.5
1.0	10	4.5	24.8	55.4	64.1	16.3	49.0	56.3
	100	0.7	14.3	55.1	65.8	11.4	53.4	56.0
	1000	0.1	6.0	63.2	55.6	3.3	39.7	52.9

Table 4.15: Summary of the preselection efficiency ($\varepsilon_{\mathrm{pre}}$) and the selection efficiency for each category after the preselection. All the quantities are displayed for each HNL mass (m_{HNL}) and lifetime ($c\tau_{\mathrm{HNL}}$) hypothesis. The values have been computed considering HNLs events from prompt D_s mesons.

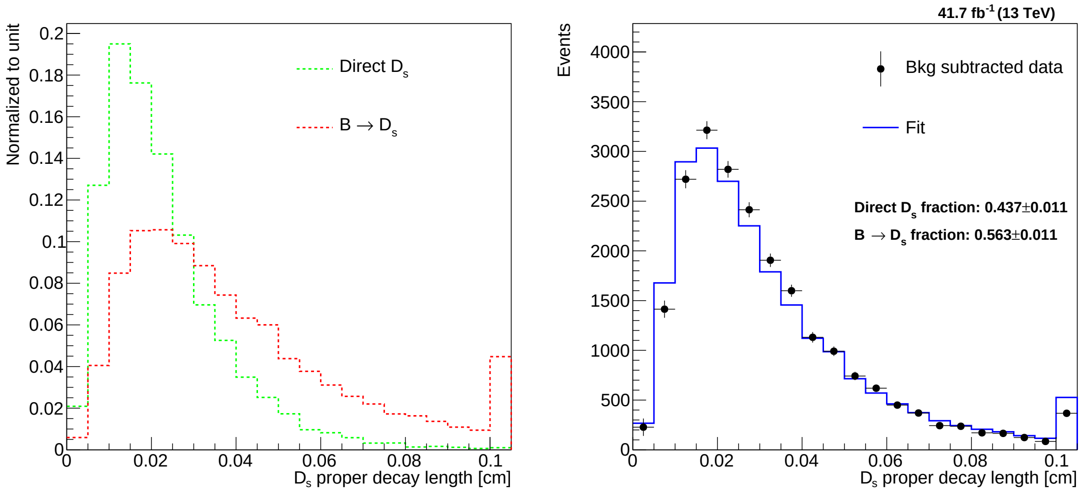

Figure 4.34: Left: the D_s reconstructed proper decay length distribution in $D_s \to \phi(\mu\mu)\pi$ MC simulation where the direct D_s and $B \to D_s$ contributions are shown in green and red. Right: the two shapes in the left plot are used to fit the background-subtracted data.

pected to show a distribution more populated at lower values with respect to D_s originating from B decays. Figure 4.34 shows the proper decay length distribution of D_s mesons in data (right) and simulation (left), where the two contributions from direct D_s production and from B decays are displayed separately. As expected, the distribution of the prompt D_s (green) populates the region with lower values of the proper decay length, while the distribution $B \to D_s$ (red) is more uniform throughout the spectrum. Data events have been reweighted using the *sPlot* technique [110] applied to the UML fit of the $D_s^+ \to \phi\pi^+$ invariant mass to subtract the background. The value of f_{prompt} is then extracted by fitting the relative prompt and non-prompt rates to the data using a maximum likelihood technique [111]. As a result of the fit, the expected fraction of prompt D_s is $f_{\mathrm{prompt}} = 0.437 \pm 0.011$, where the quoted uncertainty is statistical only.

4.9.2 Results with prompt D_s production

Prior to the extraction of the limits in the prompt D_s scenario, the systematic uncertainties calculation associated with the remaining disagreement between data and MC after the trigger scale factor reweighting has been repeated with the prompt D_s signal MC. The procedure used to extract these systematic uncertainties is identical

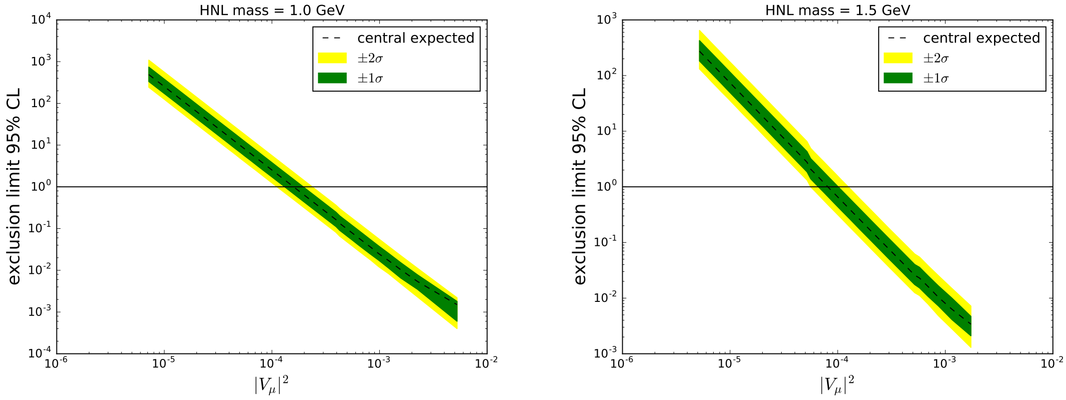

Figure 4.35: Expected exclusion limit at 95% CL on the signal strength μ as a function of $|V|^2_\mu$ for a 1.0 GeV HNL (left) and a 1.5 GeV HNL (right). To extract the limit, the expected signal yield has been estimated considering only the prompt D_s mesons.

to the one described in Sec. 4.7. The resulting systematic uncertainties are summarized in Tab. B.1 of Appendix B. For the prompt D_s channel, an additional source of systematic uncertainty must be considered. The f_{prompt} computation is affected by the choice of the PDF used to derive the signal yield in the normalization channel. To estimate the relative systematic uncertainty, the f_{prompt} measurement is repeated using the different fit functions mentioned in Sec. 4.7. A systematic uncertainty of 1.5% associated with the f_{prompt} extraction has been derived. The systematic uncertainty associated with the D_s yield is identical to the inclusive D_s production channel.

The expected exclusion limit on the signal strength at 95% confidence level is derived for the two HNL mass scenarios as a function of the $|V_\mu|^2$ mixing parameter, for the prompt D_s channel. The expected exclusion limits as a function of $|V_\mu|^2$ are shown in Fig. 4.35, where all the systematic uncertainties are considered to be uncorrelated and to follow a log-normal distribution. The expected exclusion limits on $|V_\mu|^2$ for the 1 and 1.5 GeV mass hypothesis are derived by intercepting the expected exclusion limit as a function of $|V_\mu|^2$ with the hypothesis of signal strength equal to one, shown in 4.35 as the black solid horizontal line. Figure 4.36 shows the expected excluded $|V_\mu|^2$ region as a function of the signal mass. The expected limits extracted for the 1.0 and 1.5 GeV mass points are interpolated with a straight line. The sensitivity of the prompt D_s channel ranges between approximately 2×10^{-4}, for the 1 GeV mass point, and 9×10^{-5} for the 1.5 GeV mass point.

The expected limit in the prompt D_s production channel is less stringent than the inclusive D_s case for both 1.0 and 1.5 GeV mass points. The similar distribution of the discriminating variables in inclusive and prompt channels results in similar set of optimized selection requirements. This results in similar background level in all categories and mass points. Therefore the better expected limits found for the inclusive D_s production can be attributed to the higher statistics.

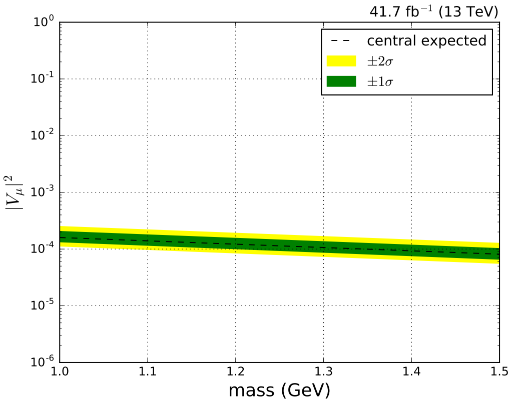

Figure 4.36: Minimum excluded $|V_\mu|^2$ mixing term as a function of the signal mass. To extract the limit, the expected signal yield has been estimated considering only prompt D_s mesons.

4.10 Outlook

The search presented in this work sets an expected limit on the active-sterile mixing matrix element $|V_\mu|^2$ between $\sim 10^{-4}$ and 7×10^{-5} in the 1.0 to 1.5 GeV mass range. Although this result is not competitive with beam-dump experiments, these limits are better than the one obtained by LHCb and competitive with results from Belle (see Sec. 4.8) for masses below the D_s mass. However, since the analysis is still ongoing it might benefit from further improvements.

To further improve the background description, the possibility that a $D_s^+ \to \pi^+\pi^+\pi^-$ decay could fake the $D_s^+ \to N(\to \mu^\pm\pi^\mp)\mu^+$ should be taken into account. Since the difference between the pion and the muon masses is small, a peaking structure might appear in the background shape, in the invariant mass spectrum of both HNL and D_s candidates. The probability that a pion is misreconstructed as a muon has been studied in the context of the $B_s^0 \to \mu^+\mu^-$ analysis in CMS [112], and, although it depends on the p_T and the position where the pion originated, one could consider as a conservative value of 0.16%. To estimate the expected yield due to misreconstructed pions the $D_s^+ \to \pi^+\pi^+\pi^-$ selection efficiency needs to be measured. This can be done using simulated events, similarly to what has been done for the $D_s^+ \to \phi(\to K^+K^-)\pi^+$ decay in Sec. 4.4.2. Moreover, the $D_s^+ \to \pi^+\pi^+\pi^-$ decay could fake the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ decay used for the normalization channel as well, when the $\pi^+\pi^-$ pair comes from a ϕ meson or a $f(980)$ resonance, since it has a mass close to the ϕ meson mass. Such possibility should also be taken into account.

Another improvement that should be considered concerns the B-Parking trigger scale factor calculation. As shown in Sec. 4.5.4 the peculiar configuration of the B-Parking triggers results in a large difference of the trigger efficiency between data and simulation. Although the application of the trigger scale factors shows a significant improvement of the agreement between data and MC, a residual disagreement is still present and it reflects on the estimated selection efficiency, which is determined using simulated events. The residual disagreement between data and MC has been taken into account as a source of systematic uncertainty. The trigger scale factors have been calculated using only a subset of the B-Parking data set: the final version of the analysis should include the measurement performed on the whole data set. This could result in a different selection efficiency measurement.

This analysis is planned to undergo the analysis approval procedure of CMS, which means that the analysis will be reviewed internally by the CMS B-Physics working group until the analysis procedure is finalized. Until then the analysis could benefit from improvements in the background description and in the MC modeling of the data and of the expected signal. This will allow to reach a more efficient and solid event selection procedure. Only at that point the data will be unblinded and the hypothesis of the existence of a sterile right-handed neutrino can be tested.

Conclusion

The work presented in this thesis describes a search for sterile right-handed neutrinos in D_s meson decays, in the hypothesis of a sterile neutrino mixing only with the active muon neutrino. In the hypothesis that the see-saw mechanism is responsible for the smallness of the active neutrino masses, the active-sterile neutrino mixing parameter $|V_\mu|^2$ is expected to be small, and the sterile neutrino to have a large mass compared to the active one. The analysis uses proton-proton collisions collected by the CMS experiment at a center-of-mass energy of 13 TeV. The data set used for the analysis, the so-called B-Parking data set, has been collected during the 2018 data-taking period by triggering on low-p_T single non-prompt muons, and it corresponds to an integrated luminosity of $41.7\,\mathrm{fb}^{-1}$. Since the analysis is still blinded, only an expected limit on the active-sterile neutrino mixing matrix parameter $|V_\mu|^2$ has been determined as a function of the sterile neutrino mass hypothesis.

The analysis targeted the $D_s^+ \to N(\to \mu^\pm\pi^\mp)\mu^+$ decays, where the $N \to \mu^+\pi^-$ decay is expected to happen at a relatively large distance from the primary hard collision. Moreover, since the sterile neutrino could be a Majorana neutrino, the D_s decay could also violate the lepton flavor number conservation, allowing for the presence of two same-sign muons in the final state. To consider both the HNL displacement and the possibility of a lepton flavor violating decay, the events are categorized into six classes depending on the relative sign of the muons (same-sign or opposite-sign) and on the HNL transverse decay distance. A cut-based event selection has been optimized using simulated HNL events and considering the HNL vertex displacement significance, muon and pion impact parameter and impact parameter significance as key variables. The expected limit is then extracted by performing a fit to the $\mu\pi$ invariant mass spectrum for two HNL mass hypotheses, 1.0 and 1.5 GeV. For each mass hypothesis, the signal yield has been estimated performing a fine scan over the HNL lifetime spectrum, allowing to probe $|V_\mu|^2$ values down to $\sim 10^{-5}$. To estimate the expected D_s yield the $D_s^+ \to \phi(\to \mu^+\mu^-)\pi^+$ decay has been used as a normalization channel, while the background yield is estimated using mass sidebands in data, defined depending on the HNL mass hypothesis.

The analysis obtained an expected limit at the 95% C.L. on $|V_\mu|^2$ as a function of the neutrino mass. The limit ranges between $\sim 10^{-4}$ for a 1 GeV neutrino to $\sim 7 \times 10^{-5}$ for a 1.5 GeV neutrino. Although previous beam dump experiments allowed to put more stringent constraints for this mass region, the results obtained in this thesis are comparable with the ones obtained by the Belle collaboration and better than the ones obtained by LHCb, over the same mass range. However, the analysis is still blinded and it might still benefit from further improvements.

The searches for heavy neutral leptons are nowadays among the most important physics topics as the mechanism generating the neutrino masses is still unknown, despite the observation of the neutrino flavor oscillations. Since the standard model predicts massless active left-handed neutrinos, the neutrino mass generation mechanism must involve physics beyond the standard model.

Bibliography

[1] ATLAS Collaboration. “A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery”. In: *Nature* 607.7917 (2022), pp. 52–59.

[2] CMS Collaboration. “A portrait of the Higgs boson by the CMS experiment ten years after the discovery”. In: *Nature* 607.7917 (2022), pp. 60–68.

[3] Workman R. L. *et al.* “Review of Particle Physics”. In: *PTEP* 2022 (2022), p. 083C01. DOI: `10.1093/ptep/ptac097`.

[4] Pontecorvo B. “Inverse beta processes and nonconservation of lepton charge”. In: *Sov. Phys. JETP* 7 (1958), p. 172.

[5] Pontecorvo B. “Mesonium and anti-mesonium”. In: *Sov. Phys. JETP* 6 (1957), p. 429.

[6] Maki Z., Nakagawa M. and Sakata S. “Remarks on the Unified Model of Elementary Particles”. In: *Progress of Theoretical Physics* 28.5 (1962), pp. 870–880. ISSN: 0033-068X. DOI: `10.1143/PTP.28.870`.

[7] Appelquist T. and Shrock R. “Neutrino masses in theories with dynamical electroweak symmetry breaking”. In: *Physics Letters B* 548.3-4 (2002), pp. 204–214. DOI: `https://doi.org/10.1016/S0370-2693(02)02854-X`.

[8] Appelquist T. and Shrock R. “Dynamical symmetry breaking of extended gauge symmetries”. In: *Physical review letters* 90.20 (2003), p. 201801.

[9] Asaka T., Blanchet S. and Shaposhnikov M. “The νMSM, dark matter and neutrino masses”. In: *Physics Letters B* 631.4 (2005), pp. 151–156.

[10] Asaka T. and Shaposhnikov M. “The νMSM, dark matter and baryon asymmetry of the universe”. In: *Physics Letters B* 620.1-2 (2005), pp. 17–26.

[11] Deppisch F. F., Dev P. S. B. and Pilaftsis A. “Neutrinos and collider physics”. In: *New Journal of Physics* 17.7 (2015), p. 075019. DOI: `10.1088/1367-2630/17/7/075019`.

[12] LHCb Collaboration. “Search for Majorana Neutrinos in $B^- \to \pi^+\mu^-\mu^-$ Decays”. In: *Phys. Rev. Lett.* 112 (13 2014), p. 131802. DOI: `10.1103/PhysRevLett.112.131802`.

[13] Glashow S. L., Iliopoulos J. and Maiani L. “Weak Interactions with Lepton-Hadron Symmetry”. In: *Phys. Rev. D* 2 (7 1970), pp. 1285–1292. DOI: `10.1103/PhysRevD.2.1285`.

[14] Salam A. “Elementary particle theory”. In: *Prog. Of the Nobel Symposium, 1968, Stockholm, Sweden.* Vol. 367. 1968.

[15] Weinberg S. “A Model of Leptons”. In: *Phys. Rev. Lett.* 19 (21 1967), pp. 1264–1266. DOI: `10.1103/PhysRevLett.19.1264`.

[16] Glashow S. L. “Partial-symmetries of weak interactions”. In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2.

[17] Halzen F. and Martin A. D. *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. Wiley, 1984. ISBN: 0471887412.

[18] Peskin M. E. *An introduction to quantum field theory*. CRC press, 2018.

[19] Fukugita M. and Yanagida T. “Barygenesis without grand unification”. In: *Physics Letters B* 174.1 (1986), pp. 45–47.

[20] Akhmedov E. Kh., Rubakov V.A. and Smirnov A. Yu. “Baryogenesis via neutrino oscillations”. In: *Physical review letters* 81.7 (1998), p. 1359.

[21] Canetti L. *et al.* “Dark matter, baryogenesis and neutrino oscillations from right-handed neutrinos”. In: *Physical Review D* 87.9 (2013), p. 093006.

[22] Canetti L., Drewes M. and Garbrecht B. “Probing leptogenesis with GeV-scale sterile neutrinos at LHCb and Belle II”. In: *Physical Review D* 90.12 (2014), p. 125005.

[23] Antusch S. *et al.* “Probing leptogenesis at future colliders”. In: *Journal of High Energy Physics* 2018.9 (2018), pp. 1–53.

[24] Cvetič G., Das A. and Zamora-Saá J. “Probing heavy neutrino oscillations in rare W boson decays”. In: *Journal of Physics G: Nuclear and Particle Physics* 46.7 (2019), p. 075002.

[25] Anamiati G., Fonseca R. M. and Hirsch M. “Quasi-Dirac neutrino oscillations”. In: *Physical Review D* 97.9 (2018), p. 095008.

[26] Antusch S., Cazzato E. and Fischer O. “Resolvable heavy neutrino–antineutrino oscillations at colliders”. In: *Modern Physics Letters A* 34.07n08 (2019), p. 1950061.

[27] Bondarenko K. *et al.* “Phenomenology of GeV-scale heavy neutral leptons”. In: *Journal of High Energy Physics* 2018.11 (2018), pp. 1–48.

[28] De Gouvêa A. “Seesaw energy scale and the LSND anomaly”. In: *Physical Review D* 72.3 (2005), p. 033005.

[29] Cirelli M. *et al.* “Probing oscillations into sterile neutrinos with cosmology, astrophysics and experiments”. In: *Nuclear Physics B* 708.1-3 (2005), pp. 215–267.

[30] De Gouvêa A., Huang W. and Jenkins J. “Pseudo-Dirac neutrinos in the new standard model”. In: *Physical Review D* 80.7 (2009), p. 073007.

[31] Donini A. *et al.* “The minimal 3+ 2 neutrino model versus oscillation anomalies”. In: *Journal of High Energy Physics* 2012.7 (2012), pp. 1–20.

[32] Gorbunov D. and Shaposhnikov M. “How to find neutral leptons of the νMSM?” In: *Journal of High Energy Physics* 2007.10 (2007), p. 015. DOI: 10.1088/1126-6708/2007/10/015.

[33] Boyarsky A., Ruchayskiy O. and Shaposhnikov M. “The role of sterile neutrinos in cosmology and astrophysics”. In: *Annual Review of Nuclear and Particle Science* 59 (2009), pp. 191–214.

[34] Ruchayskiy O. and Ivashko A. “Restrictions on the lifetime of sterile neutrinos from primordial nucleosynthesis”. In: 2012.10 (2012), p. 014. DOI: 10.1088/1475-7516/2012/10/014.

[35] Belle Collaboration. "Search for heavy neutrinos at Belle". In: *Phys. Rev. D* 87 (7 2013), p. 071102. DOI: 10.1103/PhysRevD.87.071102.

[36] Kobach A. and Dobbs S. "Heavy neutrinos and the kinematics of tau decays". In: *Physical Review D* 91.5 (2015), p. 053006.

[37] Atre A. *et al.* "The search for heavy Majorana neutrinos". In: *Journal of High Energy Physics* 2009.05 (2009), p. 030.

[38] CMS Collaboration. "Search for heavy Majorana neutrinos in same-sign dilepton channels in proton-proton collisions at $\sqrt{s} = 13$ TeV ". In: *Journal of High Energy Physics* 2019.1 (2019), pp. 1–55.

[39] ATLAS Collaboration. "Search for heavy neutral leptons in decays of W bosons using a dilepton displaced vertex in $\sqrt{s} = 13$ TeV pp collisions with the ATLAS detector". In: (Apr. 2022).

[40] L3 Collaboration. "Search for isosinglet neutral heavy leptons in Z^0 decays". In: *Physics Letters B* 295.3-4 (1992), pp. 371–382.

[41] DELPHI Collaboration. "Search for neutral heavy leptons produced in Z decays". In: *Zeitschrift für Physik C Particles and Fields* 74 (1997), pp. 57–71.

[42] Blondel A. *et al.* "Search for heavy right handed neutrinos at the FCC-ee". In: *Nuclear and particle physics proceedings* 273 (2016), pp. 1883–1890.

[43] Abada A. *et al.* "Indirect searches for sterile neutrinos at a high-luminosity Z-factory". In: *Journal of High Energy Physics* 2015.4 (2015), pp. 1–33.

[44] Benedikt M. *et al. Future Circular Collider - European Strategy Update Documents.* Tech. rep. Geneva: CERN, 2019.

[45] Graverini E., Van Herwijnen E. and Ruf T. "Mass dependence of branching ratios into HNL for FairShip". In: (2016).

[46] ALICE Collaboration. "Charm-quark fragmentation fractions and production cross section at midrapidity in pp collisions at the LHC". In: *Physical Review D* 105.1 (2022), p. L011103.

[47] Källén G. *Elementary Particle Physics.* A-W series in advanced physics. Addison-Wesley Publishing Company, 1964.

[48] Gorishnii S. G., Kataev A. L. and Larin S. A. "The $O(\alpha_s^3)$-corrections to $\sigma_{tot}(e^+e^- \to \text{hadrons})$ and $\Gamma(\tau^- \to \nu_\tau + \text{hadrons})$ in QCD". In: *Phys. Lett. B* 259 (1991), pp. 144–150. DOI: 10.1016/0370-2693(91)90149-K.

[49] Shuve B. and Peskin M. E. "Revision of the LHCb limit on Majorana neutrinos". In: *Phys. Rev. D* 94 (11 2016), p. 113007. DOI: 10.1103/PhysRevD.94.113007.

[50] Jonathan L Rosner, Sheldon Stone, and Ruth S Van de Water. "Leptonic decays of charged pseudoscalar mesons-2015". In: *arXiv preprint arXiv:1509.02220* (2015).

[51] Evans L. and Bryant P. "LHC machine". In: *Journal of instrumentation* 3.08 (2008), S08001.

[52] Brüning O., Burkhardt H. and Myers S. "The Large Hadron Collider". In: *Progress in Particle and Nuclear Physics* 67.3 (2012), pp. 705–734. ISSN: 0146-6410. DOI: https://doi.org/10.1016/j.ppnp.2012.03.001.

[53] *CMS Luminosity - Public Results*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults. Accessed: 2022-11-17.

[54] Chatrchyan S. *et al.* “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[55] CMS Collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. Geneva: CERN, 2017.

[56] Colaleo A. *et al. CMS Technical Design Report for the Muon Endcap GEM Upgrade*. Tech. rep. 2015.

[57] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[58] The Tracker Group of the CMS collaboration. “The CMS Phase-1 Pixel Detector Upgrade”. In: *JINST* 16.02 (2021), P02027. DOI: 10.1088/1748-0221/16/02/P02027.

[59] CMS Collaboration. “Track impact parameter resolution for the full pseudo rapidity coverage in the 2017 dataset with the CMS Phase-1 Pixel detector”. In: (2020).

[60] CMS Collaboration. “Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC”. In: *Journal of Instrumentation* 16.05 (2021), P05014. DOI: 10.1088/1748-0221/16/05/P05014.

[61] CMS Collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997.

[62] CMS Collaboration. “Calibration of the CMS hadron calorimeters using proton-proton collision data at $\sqrt{s} = 13, \mathrm{TeV}$”. In: *Journal of Instrumentation* 15.05 (2020), P05002. DOI: 10.1088/1748-0221/15/05/P05002.

[63] ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020.

[64] CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021.

[65] CMS Collaboration. “The CMS trigger system”. In: *Journal of Instrumentation* 12.01 (2017), P01020. DOI: 10.1088/1748-0221/12/01/P01020.

[66] CMS Collaboration. “Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV”. In: *JINST* 15.10 (2020), P10017. DOI: 10.1088/1748-0221/15/10/P10017.

[67] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS.

[68] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009.

[69] Frühwirth R. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/0168-9002(87)90887-4`.

[70] CMS Collaboration. “Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\mathrm{TeV}$”. In: *Journal of Instrumentation* 10.06 (2015), P06005. DOI: `10.1088/1748-0221/10/06/P06005`.

[71] CMS collaboration. “Performance of the CMS muon trigger system in proton-proton collisions at $\sqrt{s} = 13, \mathrm{TeV}$”. In: *Journal of Instrumentation* 16.07 (2021), P07001. DOI: `10.1088/1748-0221/16/07/P07001`.

[72] Cacciari M., Salam G. P. and Soyez G. “FastJet user manual”. In: *The European Physical Journal C* 72.3 (2012), pp. 1–54.

[73] Cacciari M., Salam G. P. and Soyez G. “The anti-kt jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (2008), p. 063. DOI: `10.1088/1126-6708/2008/04/063`.

[74] Bayatyan G. L. *et al.* “CMS: The computing project. Technical design report”. In: (June 2005).

[75] Grandi C. *et al. The CMS computing model.* Tech. rep. CERN-LHCC-2004-035/G-083, 2004.

[76] Petrucciani G. *et al.* “Mini-AOD: A new analysis data format for CMS”. In: *Journal of Physics: Conference Series.* Vol. 664. 7. IOP Publishing. 2015, p. 072052.

[77] Rizzi A., Petrucciani G. and Peruzzi M. “A further reduction in CMS event data for analysis: the NANOAOD format”. In: *EPJ Web of Conferences.* Vol. 214. EDP Sciences. 2019, p. 06021.

[78] CMS Collaboration. “Recording and reconstructing 10 billion unbiased b hadron decays in CMS”. In: (2019). CMS Note CERN-CMS-DP-2019-043.

[79] CMS Collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *Journal of Instrumentation* 12.10 (2017), P10003. DOI: `10.1088/1748-0221/12/10/P10003`.

[80] CMS Collaboration. *Tracking and Vertexing Results from First Collisions.* Tech. rep. Geneva: CERN, 2010.

[81] CMS Collaboration. “Description and performance of track and primary-vertex reconstruction with the CMS tracker”. In: *Journal of Instrumentation* 9.10 (2014), P10009. DOI: `10.1088/1748-0221/9/10/P10009`.

[82] Rose K. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2210–2239. DOI: `10.1109/5.726788`.

[83] Frühwirth R., Waltenberger W. and Vanlaer P. *Adaptive Vertex Fitting.* Tech. rep. Geneva: CERN, 2007.

[84] Adam W. *et al.* “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”. In: *Journal of Physics G: Nuclear and Particle Physics* 31.9 (2005), N9. DOI: `10.1088/0954-3899/31/9/N01`.

[85] CMS Collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$”. In: *Journal of Instrumentation* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015.

[86] Cacciari M., Salam G. P. and Soyez G. “The anti-k_t jet clustering algorithm”. In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063.

[87] Cacciari M., Salam G. P. and Soyez G. “FastJet user manual”. In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2.

[88] Sirunyan A. M. *et al.* “Performance of missing transverse momentum reconstruction in proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ using the CMS detector”. In: *JINST* 14 (2019), P07004. DOI: 10.1088/1748-0221/14/07/P07004.

[89] Bertolini D. *et al.* “Pileup Per Particle Identification”. In: *JHEP* 10 (2014), p. 059. DOI: 10.1007/JHEP10(2014)059.

[90] Sirunyan A. M. *et al.* “Pileup mitigation at CMS in 13 TeV data”. In: *JINST* 15 (2020), P09018. DOI: 10.1088/1748-0221/15/09/p09018.

[91] CMS Collaboration. *CMS luminosity measurement for the 2018 data-taking period at* $\sqrt{s} = 13$ TeV. Tech. rep. Geneva: CERN, 2019.

[92] Lujan P. “Performance of the Pixel Luminosity Telescope for Luminosity Measurement at CMS during Run2”. In: *PoS* EPS-HEP2017 (2017), p. 504. DOI: 10.22323/1.314.0504.

[93] Hempel M. “Development of a Novel Diamond Based Detector for Machine Induced Background and Luminosity Measurements”. Dissertation, Brandenburg University of Technology Cottbus-Senftenberg, 2017. Dissertation. Hamburg: Brandenburg University of Technology Cottbus-Senftenberg, 2017, p. 151. DOI: 10.3204/PUBDB-2017-06875.

[94] van der Meer S. *Calibration of the effective beam height in the ISR*. Tech. rep. Geneva: CERN, 1968.

[95] Andy B. *et al.* “General-purpose event generators for LHC physics”. In: *Physics Reports* 504.5 (2011), pp. 145–233. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2011.03.005.

[96] Agostinelli S. *et al.* “Geant4–a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(03)01368-8.

[97] LHCb Collaboration. “Searches for 25 rare and forbidden decays of D^+ and D_s^+ mesons”. In: *Journal of High Energy Physics* 2021.6 (2021), p. 44. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2021)044.

[98] Cowan G. *et al.* “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (2011), p. 1554. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0.

[99] Sjöstrand T. *et al.* “An introduction to PYTHIA 8.2”. In: *Computer Physics Communications* 191 (2015), pp. 159–177. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2015.01.024.

[100] CMS Collaboration. “Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements”. In: *The European Physical Journal C* 80.1 (2020), p. 4. ISSN: 1434-6052. DOI: `10.1140/epjc/s10052-019-7499-4`.

[101] Lange D. J. “The EvtGen particle decay simulation package”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: `https://doi.org/10.1016/S0168-9002(01)00089-4`.

[102] CMS Collaboration. “Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7\,\mathrm{TeV}$”. In: *Journal of Instrumentation* 7.10 (2012), P10002. DOI: `10.1088/1748-0221/7/10/P10002`.

[103] Prokofiev K. and Speer T. “A kinematic and a decay chain reconstruction library”. In: (2005).

[104] Dissertori G. *et al.* “Search for long-lived heavy neutrinos in B meson decays using the B-parking dataset”. In: (2022). CMS Analysis Note AN-2022/138 of EXO-22-019.

[105] Oreglia M. “A Study of the Reactions $\psi' \to \gamma\gamma\psi$”. Other thesis. Dec. 1980.

[106] CMS Collaboration. “Search for long-lived heavy neutral leptons with displaced vertices in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$”. In: *Journal of High Energy Physics* 2022.7 (2022), pp. 1–56.

[107] Cai Y. *et al.* “Lepton number violation: seesaw models and their collider tests”. In: *Frontiers in Physics* 6 (2018), p. 40.

[108] Dorenbosch J. *et al.* “A search for decays of heavy neutrinos in the mass range 0.5–2.8 GeV”. In: *Physics Letters B* 166.4 (1986), pp. 473–478. ISSN: 0370-2693. DOI: `https://doi.org/10.1016/0370-2693(86)91601-1`.

[109] Vaitaitis A. *et al.* “Search for Neutral Heavy Leptons in a High-Energy Neutrino Beam”. In: *Phys. Rev. Lett.* 83 (24 1999), pp. 4943–4946. DOI: `10.1103/PhysRevLett.83.4943`.

[110] Pivk M. and Le Diberder F. R. “Plots: A statistical tool to unfold data distributions”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 555.1-2 (2005), pp. 356–369.

[111] Barlow R. and Beeston C. “Fitting using finite Monte Carlo samples”. In: *Computer Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/0010-4655(93)90005-W`.

[112] Chen K. *et al.* “Soft muon fakes in rare B decays”. In: (2022). CMS Analysis Note CMS AN-2021/141 of BPH-21-006.

[113] Ryd A. *et al.* “EvtGen: a Monte Carlo generator for B-physics”. In: *BAD* 522 (2005), p. v6.

[114] Scora D. and Isgur N. “Semileptonic meson decays in the quark model: An update”. In: *Phys. Rev. D* 52 (5 1995), pp. 2783–2812. DOI: `10.1103/PhysRevD.52.2783`.

Appendix A

Monte Carlo samples generation

As mentioned in the Sec. 3.4, the first step of a Monte Carlo sample production chain consists in the simulation of the main collision event. The Monte Carlo event generators allow the users to set many of the physics parameters describing the hard event, to choose the physics processes of interest and to force the decay of some of the final state particles. Depending on the Monte Carlo program used for the generation, it might be needed to further refine the selection of the generated particles by adding some requirements on the generated event. The settings used for the production of the simulated samples used in this work are summarized in Tab. A.1.

For each simulated process in Tab. A.1 the simulation of the hard event has been performed using the PYTHIA MC generator. Among the processes used for the hard event generation the `SoftQCD:nonDiffractive` setting allows to select the inelastic nondiffractive part of the total proton-proton cross section, the `SoftQCD:inelastic` setting excludes only the elastic processes, and the `HardQCD:all` setting generates QCD jet production within a certain $\hat{p}_T$ range that can be specified with the `PhaseSpace:pTHatMin` and `PhaseSpace:pTHatMax` settings. The `PTFilter` settings allow to filter the event at the generation level by choosing the quark involved in the process, and allowing to set a scale at which the parton-level downwards evolution in p_T is temporarily interrupted, so the event can be vetoed.

On the third column of Tab. A.1 it is specified which decays have been forced to be handled by the EVTGEN software, their branching ratio and the model used to simulate the decay. A description of the decay models available in EVTGEN can be found in the EVTGEN documentation [113]. The phase space (PHSP) model is used to model the decays involving the sterile neutrino N and it assumes no spin correlation between the particles involved. Other decay models such as SVS, VLL or VSS take into account the spin of the particles involved in case they are a spin-0 scalar (S), half-integer spin lepton (L) or a spin-1 vector (V) particle. Finally, the ISGW2 model is used to describe B, B_s, D and D_s mesons and implements the form factors defined by Scora and Isgur [114]. The last column in Tab. A.1 contains all the conditions used to accept or reject the generated event.

Process	**PYTHIA settings**	**EVTGEN settings**	**Generator filters**

Prompt
$D_s \to N(\to \mu\pi)\mu$
	`SoftQCD:nonDiffractive = on`
	`PTFilter:filter = on`
	`PTFilter:quarkToFilter = 4`
	`PTFilter:scaleToFilter = 5.0`
		$\mathcal{B}(D_s \xrightarrow{\text{PHSP}} N\mu) = 1.0$
		$\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^+\pi^-) = 0.5$
		$\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^-\pi^+) = 0.5$
			Final state μ having:
			• $p_T^\mu > 6.8$;
			• $\lvert\eta^\mu\rvert < 1.55$.
			At least a $N \to \mu^+\pi^-$ decay having:
			• D_s mother;
			• $p_T^{\mu,\pi} > 0.5\,\text{GeV}$;
			• $\lvert\eta^{\mu,\pi}\rvert < 2.5$.

Inclusive $D_s \to N(\to \mu\pi)\mu$	`SoftQCD:inelastic = on`	$\mathcal{B}(D_s \xrightarrow{\text{PHSP}} N\mu) = 1.0$ $\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^+\pi^-) = 0.5$ $\mathcal{B}(N \xrightarrow{\text{PHSP}} \mu^-\pi^+) = 0.5$	Final state μ having: • $p_T^\mu > 6.8$; • $\lvert\eta^\mu\rvert < 1.55$. At least a $N \to \mu^+\pi^-$ decay having: • D_s mother; • $p_T^{\mu,\pi} > 0.5\,\text{GeV}$; • $\lvert\eta^{\mu,\pi}\rvert < 2.5$.
$D_s \to \phi(\to \mu\mu)\pi$	`SoftQCD:inelastic = on`	$\mathcal{B}(D_s \xrightarrow{\text{SVS}} \phi\pi) = 1.0$ $\mathcal{B}(\phi \xrightarrow{\text{VLL}} \mu\mu) = 1.0$	At least a D_s. Final state μ having: • $p_T > 2.7\,\text{GeV}$; • $\lvert\eta\rvert < 2.45$.

$D_s \to \phi(\to K^+K^-)\pi$	`SoftQCD:nonDiffractive = on`
	`PTFilter:filter = on`
	`PTFilter:quarkToFilter = 5`
	`PTFilter:scaleToFilter = 5.0`
		$\mathcal{B}(B_s \xrightarrow{\text{ISGW2}} D_s\mu\nu_\mu) = 1.0$
		$\mathcal{B}(D_s \xrightarrow{\text{SVS}} \phi\pi) = 1.0$
		$\mathcal{B}(\phi \xrightarrow{\text{VSS}} KK) = 1.0$
			Final state μ having:
			• $p_T^\mu > 6.8$;
			• $\lvert\eta^\mu\rvert < 1.55$;
			At least a $\phi \to K^+K^-$ decay having:
			• D_s mother;
			• $p_T^K > 0.8\,\text{GeV}$;
			• $\lvert\eta^K\rvert < 2.5$.

QCD dijet	`ParticleDecays:limitTau0 = off`
	`ParticleDecays:limitCylinder = on`
	`ParticleDecays:xyMax = 2000`
	`ParticleDecays:zMax = 4000`
	`HardQCD:all = on`
	`PhaseSpace:pTHatMin =` p_T^{min}
	`PhaseSpace:pTHatMax =` p_T^{max}
	`130:mayDecay = on`
	`211:mayDecay = on`
	`321:mayDecay = on`
		-	Final state μ having:
			• $p_T^\mu > 5.$;
			• $\lvert\eta^\mu\rvert < 2.5$;
			• decay length $< 2\,\text{m}$ on xy;
			• decay length $< 4\,\text{m}$ on $\lvert z\rvert$.

Table A.1: Summary of the generator parameters and generator filter settings used for the simulation of each of the processes.

Appendix B

Prompt D_s production: extra plots

The HNL search has been performed considering either no constraints on the D_s production mechanism or requiring only the prompt D_s production. The inclusive D_s production channel results in more stringent expected limits (see Sec. 4.8) with respect to the prompt D_s channel (see Sec. 4.9.2). This appendix contains extra plots and tables concerning the prompt D_s production channel results that were not reported in the main body of the text.

Figure B.1 shows the selection efficiency of preselected signal events as a function of the cut on the HNL transverse displacement, for the various hypothesis of HNL mass and lifetime (expressed as $c\tau$).

Figure B.2 the distributions of the HNL transverse decay length significance, the muon and pion transverse impact paramenter and impact parameter significance for the various HNL mass and $c\tau$ points and compared to the dijet QCD events used to model the background contribution.

Figures B.3 to B.8 show the discovery significance as a function of the selection cut on each of the variable used to perform the selection optimization.

The top two and bottom two rows in Fig. B.9 show the background fit in the various categories for the 1 GeV and 1.5 GeV heavy neutrino mass hypothesis, respectively. The top two and bottom two rows in Figs. B.10, B.11 and B.12 show the result of signal fit in each category for the 1 GeV and 1.5 GeV heavy neutrino mass hypothesis, respectively. Each of figs. B.10 to B.12 refer to a different generated $c\tau$ hypothesis.

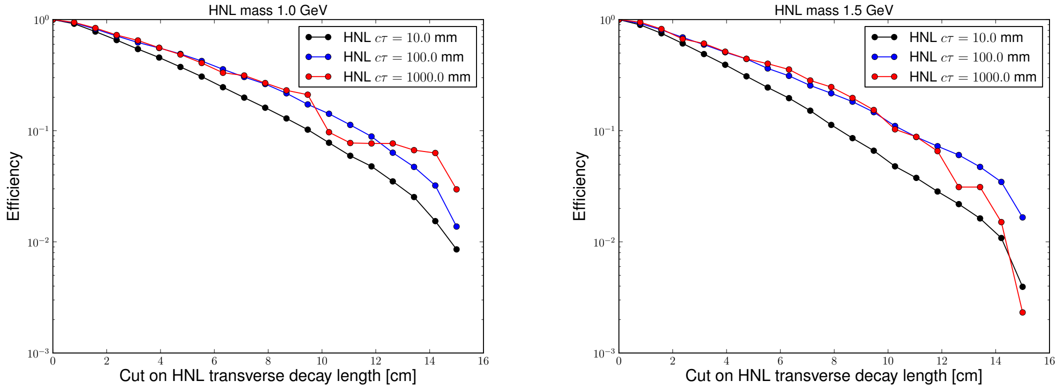

Figure B.1: Selection efficiency of preselected signal events as a function on the requirement on the HNL transverse decay length, for 1.0 (left) and 1.5 GeV (right) HNL mass hypothesis. Black, blue and red markers refers to the 10, 100 and 1000 mm HNL $c\tau$ hypothesis. The HNL simulated events are from prompt D_s decays only.

m_{HNL} (GeV)	$c\tau_{HNL}$ (mm)	Systematic uncertainty (%)						
		SSlxy0to1	SSlxy1to5	SSlxy5toInf	OSlxy0to1	OSlxy1to5	OSlxy5toInf	Variable
1.0	10	5	4	4	4	4	4	μ_{D_s} p_T
		2	4	7	2	3	7	μ_{HNL} p_T
		10	8	9	8	8	8	μ_{D_s} IPS
		8	1	6	9	1	6	μ_{HNL} IPS
	100	4	4	4	5	5	4	μ_{D_s} p_T
		2	4	6	5	3	6	μ_{HNL} p_T
		9	8	8	9	7	8	μ_{D_s} IPS
		8	0	5	11	3	5	μ_{HNL} IPS
	1000	2	6	4	7	3	2	μ_{D_s} p_T
		1	4	6	2	2	3	μ_{HNL} p_T
		5	10	9	14	7	6	μ_{D_s} IPS
		5	0	3	12	2	3	μ_{HNL} IPS
1.5	10	6	6	6	7	6	6	μ_{D_s} p_T
		6	7	10	6	7	10	μ_{HNL} p_T
		14	13	13	14	13	12	μ_{D_s} IPS
		9	5	9	8	5	8	μ_{HNL} IPS
	100	7	6	6	6	6	7	μ_{D_s} p_T
		6	7	9	5	7	9	μ_{HNL} p_T
		14	13	12	13	13	12	μ_{D_s} IPS
		8	3	6	10	5	5	μ_{HNL} IPS
	1000	7	6	7	7	6	6	μ_{D_s} p_T
		7	5	7	0	7	10	μ_{HNL} p_T
		16	11	10	12	13	14	μ_{D_s} IPS
		5	5	1	6	1	8	μ_{HNL} IPS

Table B.1: Systematic uncertainties (%) due to the residual p_T and IPS mis-modeling in simulations. The first two columns indicate the mass and lifetime point of the simulated samples. The last column shows the variable used to reweight the MC: μ_{D_s} (μ_{HNL}) p_T indicates the transverse momentum of the muon forming the $D_s^+ \to N\mu^+$ ($N \to \mu^+\pi^-$) candidate, and the same convention has been used to name the impact parameter significance (IPS) variable. The uncertainties are reported for each category and variable separately. The systematic uncertainties are computed using HNL simulated events from prompt D_s decays only.

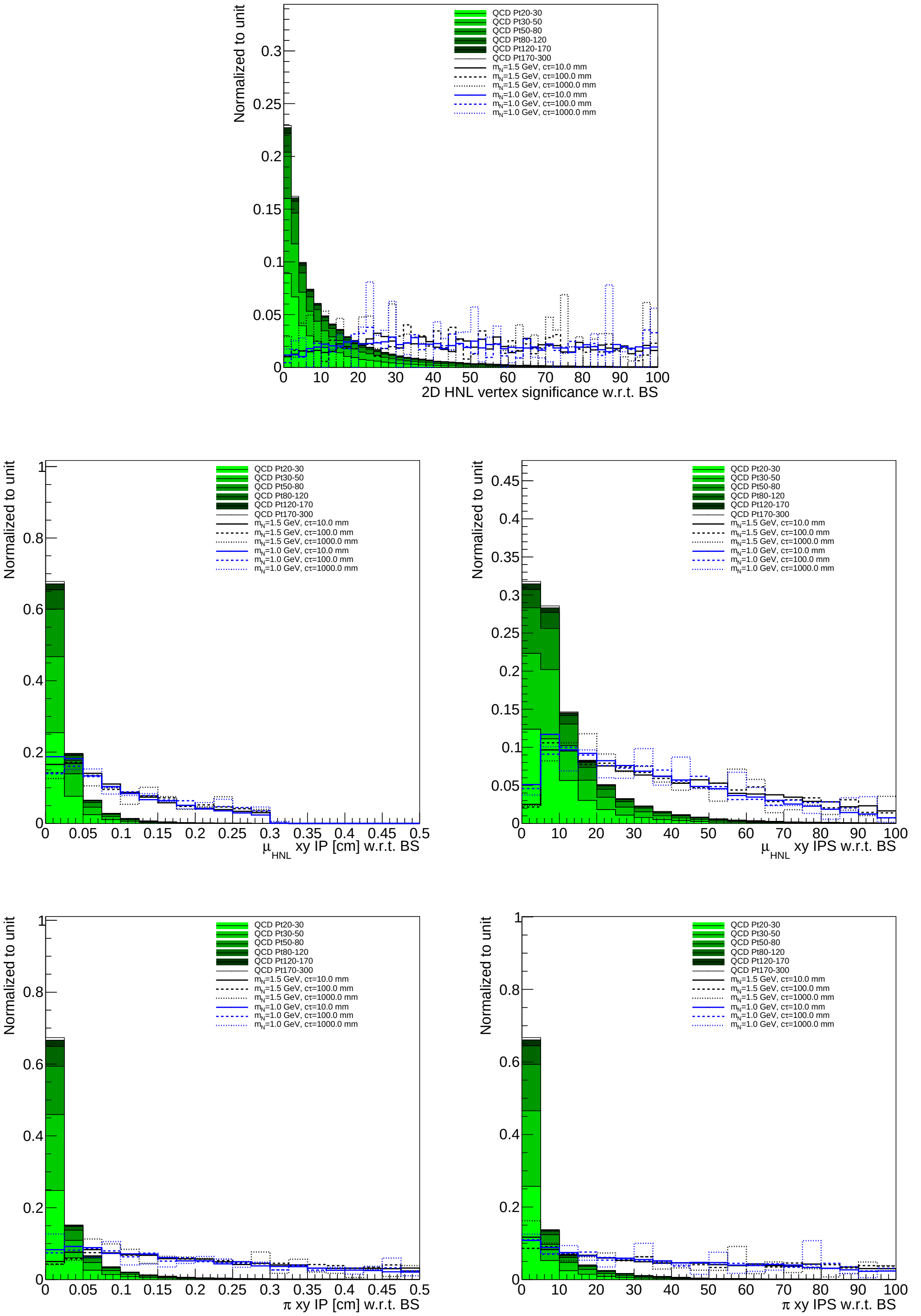

Figure B.2: QCD dijet and HNL signal distribution for various observables. Top: HNL vertex significance on the transverse plane measured with respect to the beam spot. Middle: Impact parameter (IP) left and its significance (IPS), right, for muons forming the HNL candidate, measured in the transverse plane with respect to the beam spot. Bottom: IP, left, and IPS, right, for pions forming the HNL candidate, measured on the transverse plane with respect to the beam spot. Signal events are from prompt D_s production MC.

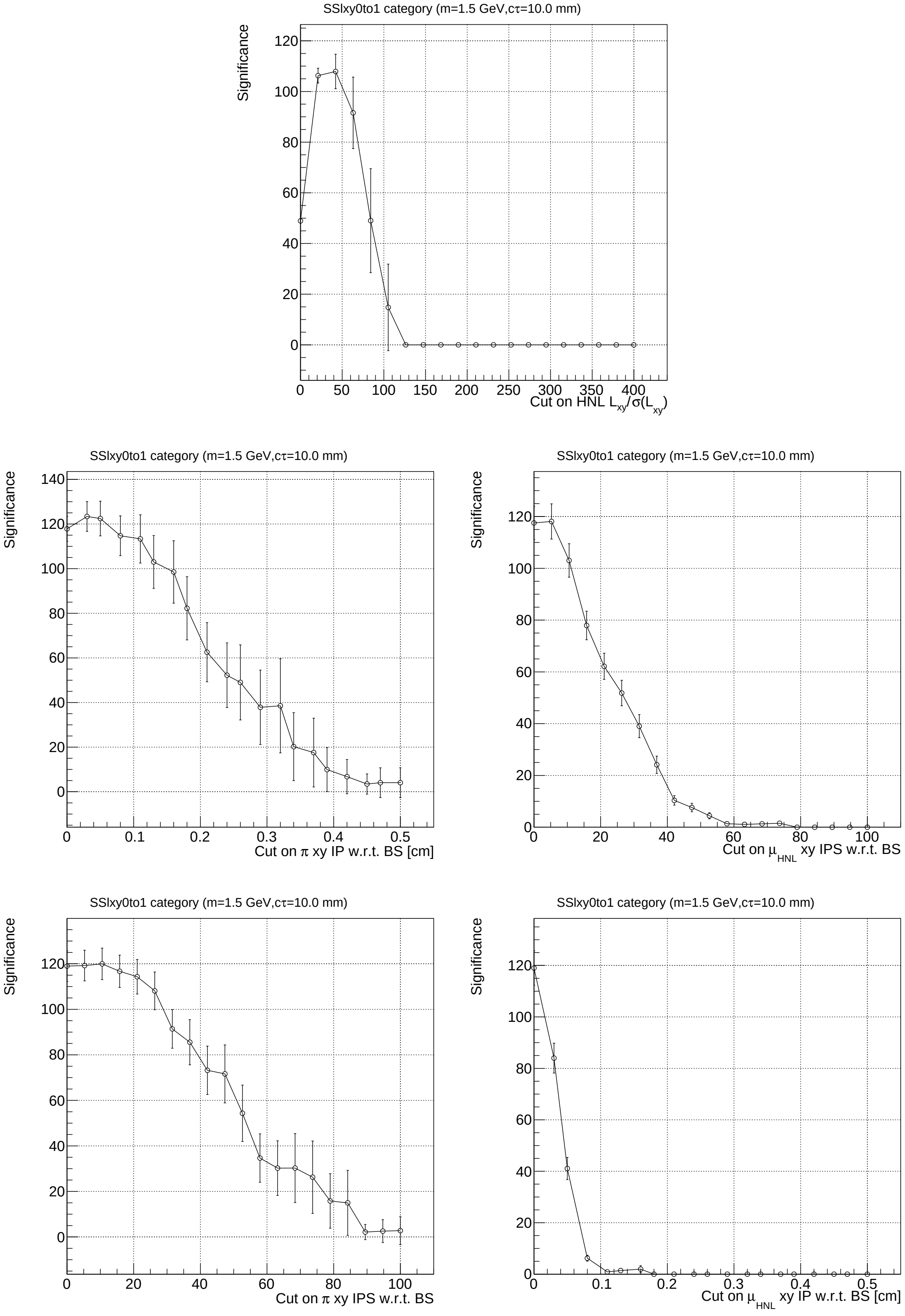

Figure B.3: Expected discovery significance for SSlxy0to1 category.

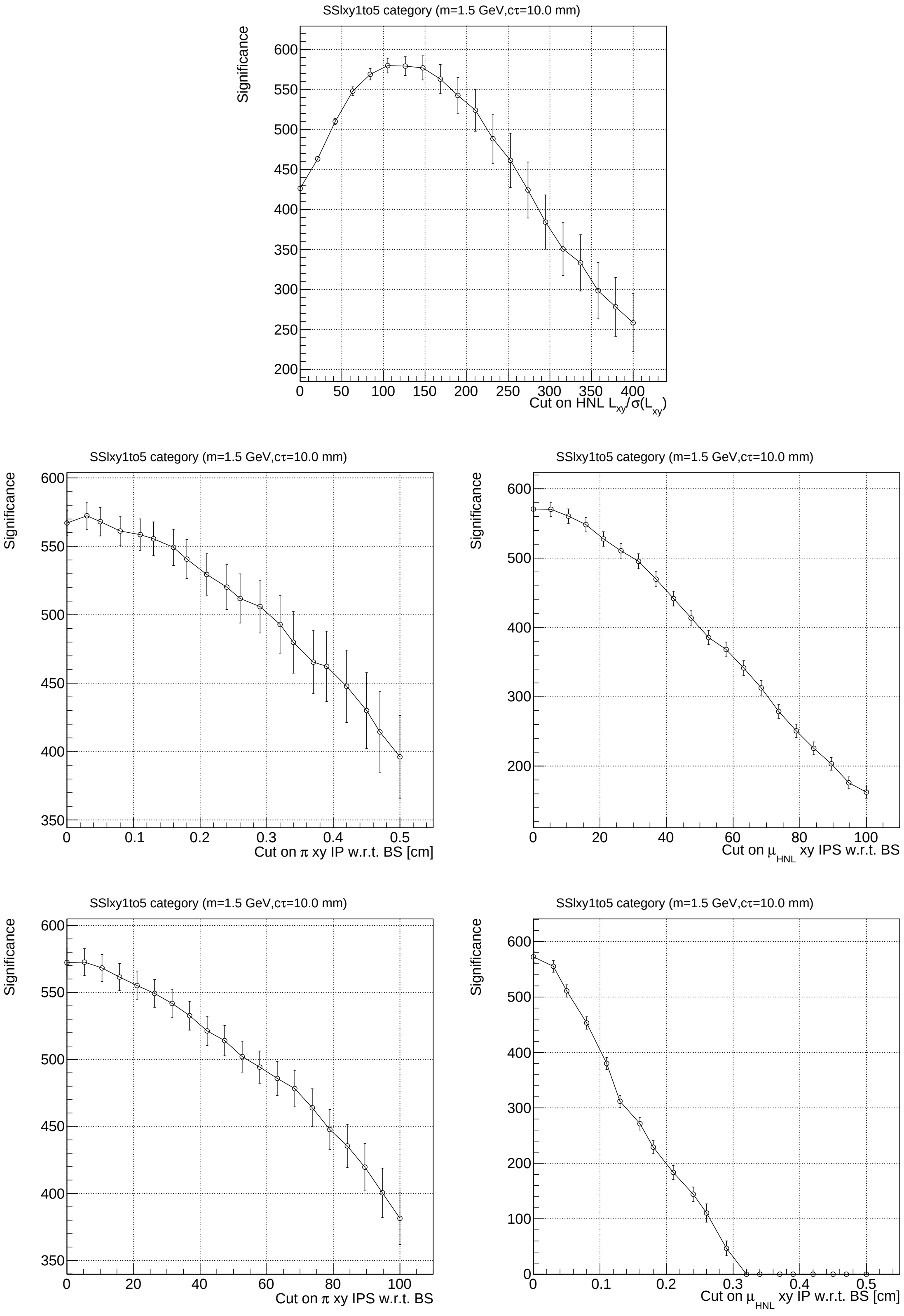

Figure B.4: Expected discovery significance for SSlxy1to5 category.

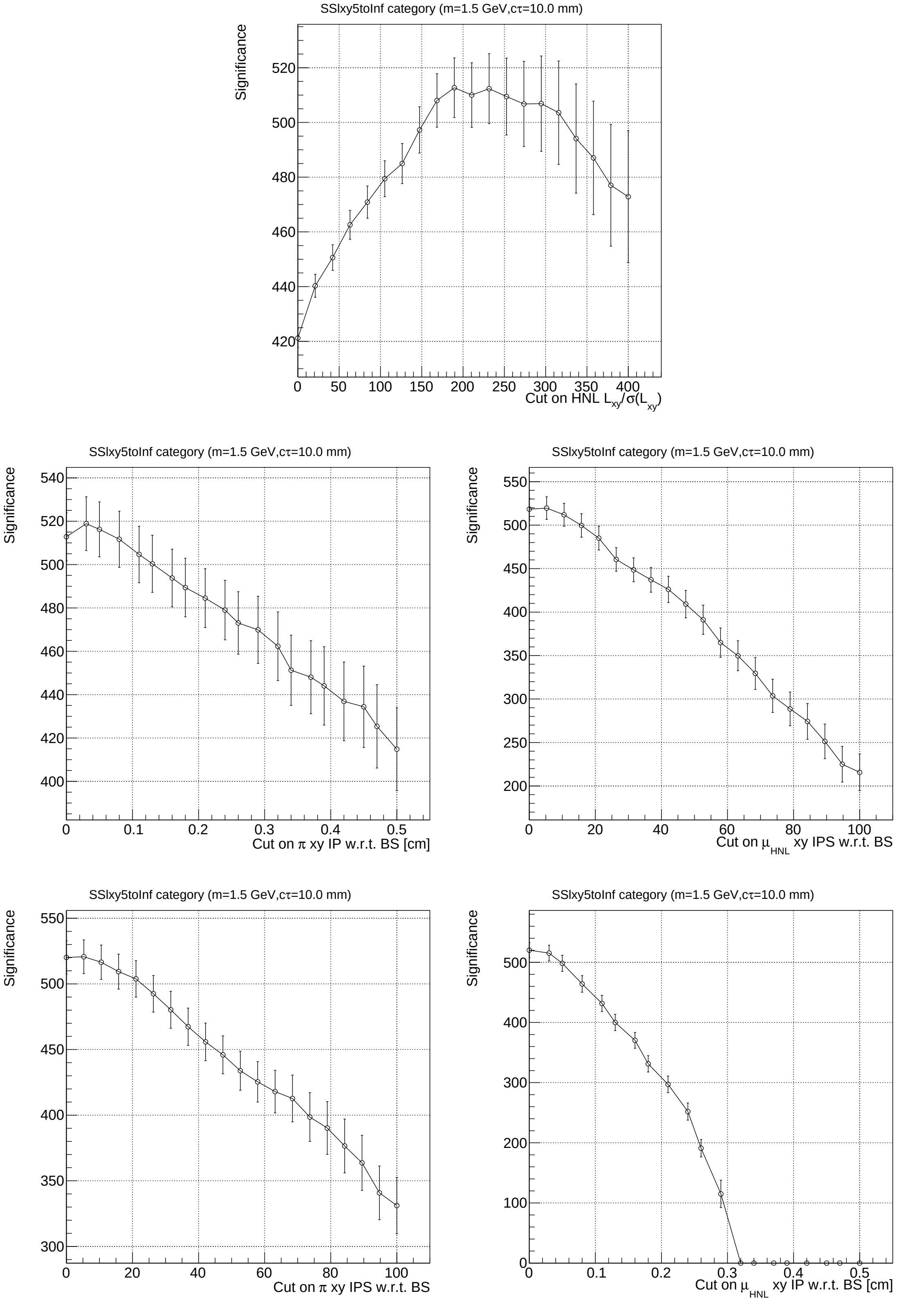

Figure B.5: Expected discovery significance for SSlxy5toInf category.

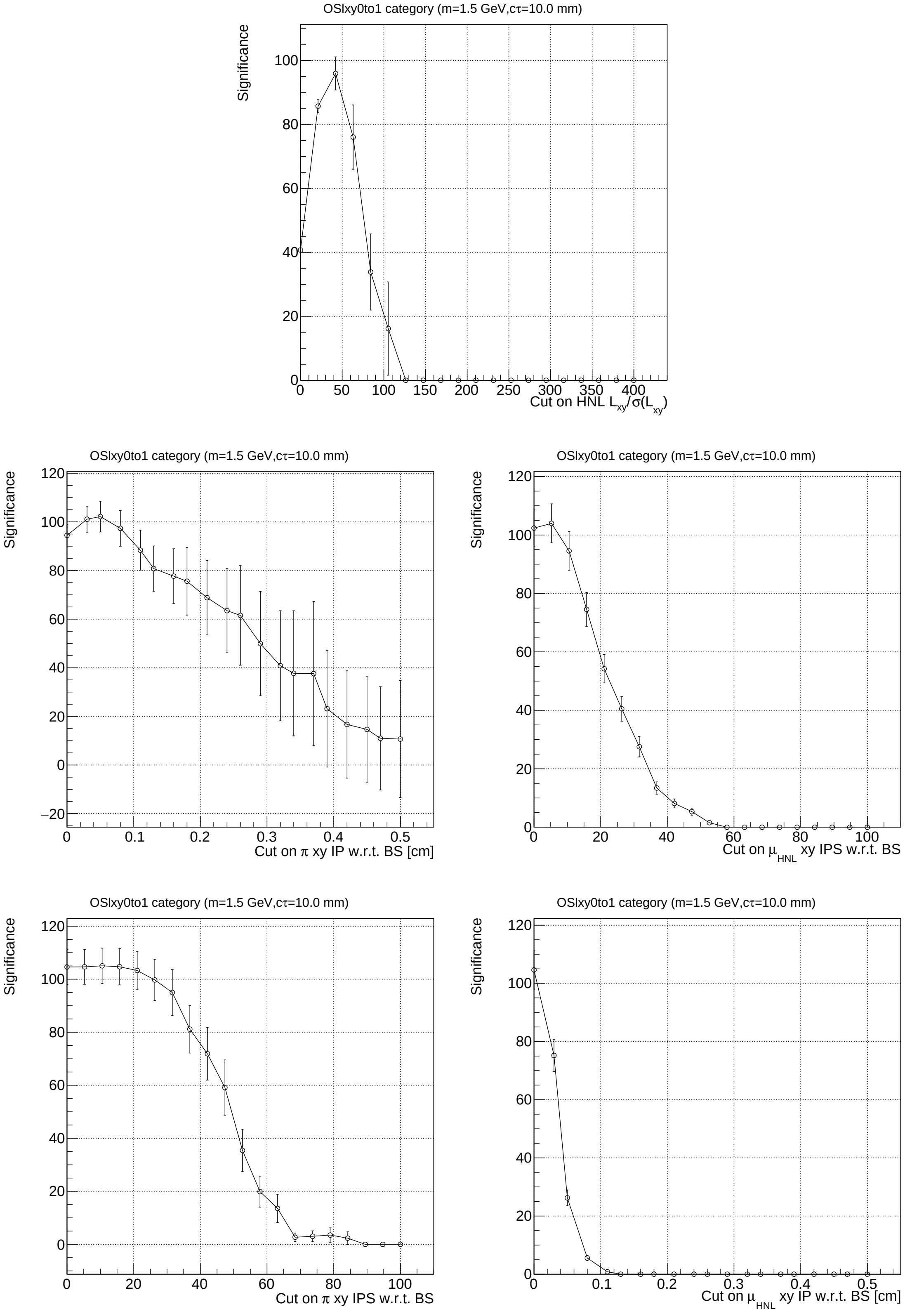

Figure B.6: Expected discovery significance for OSlxy0to1 category.

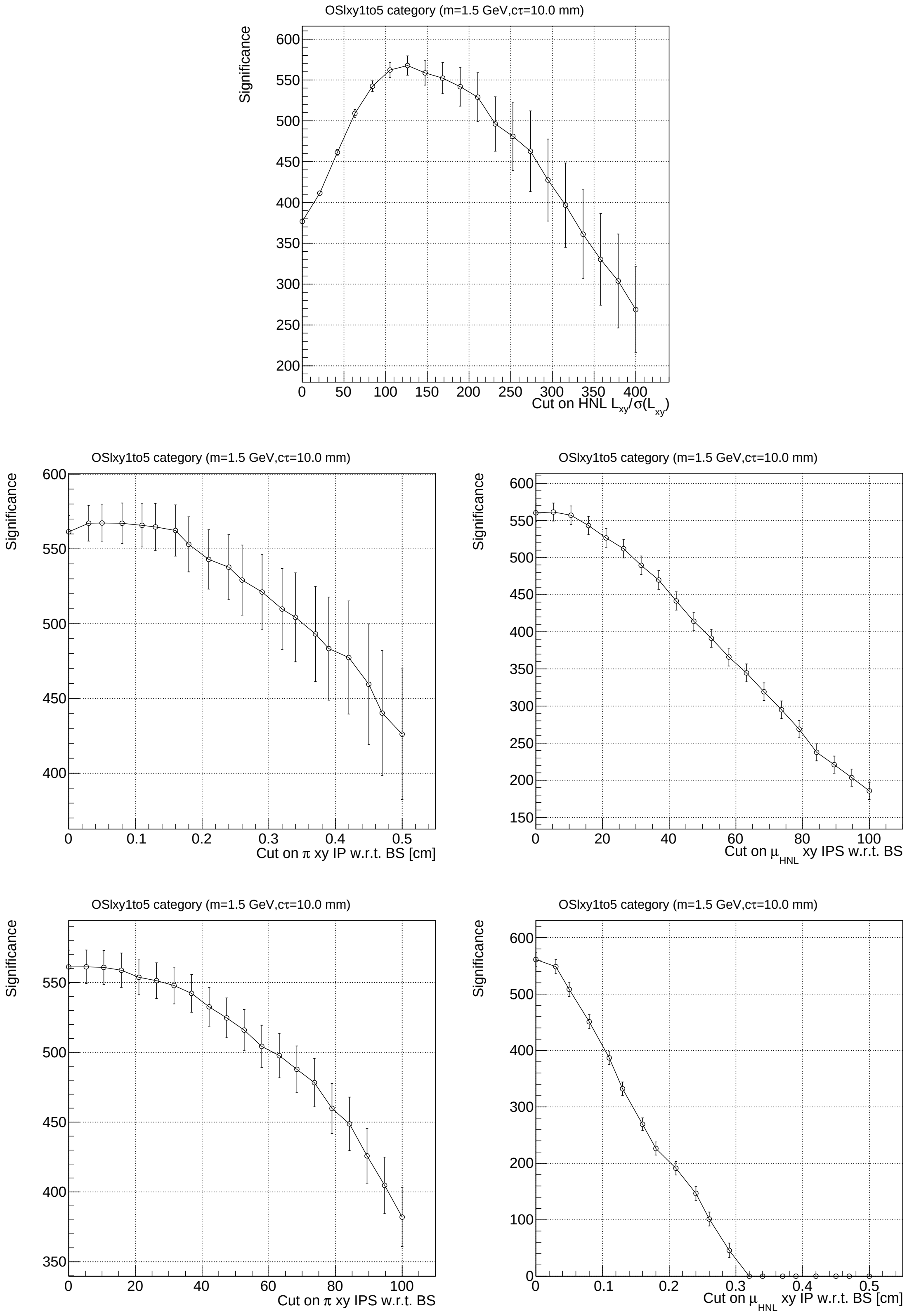

Figure B.7: Expected discovery significance for OSlxy1to5 category.

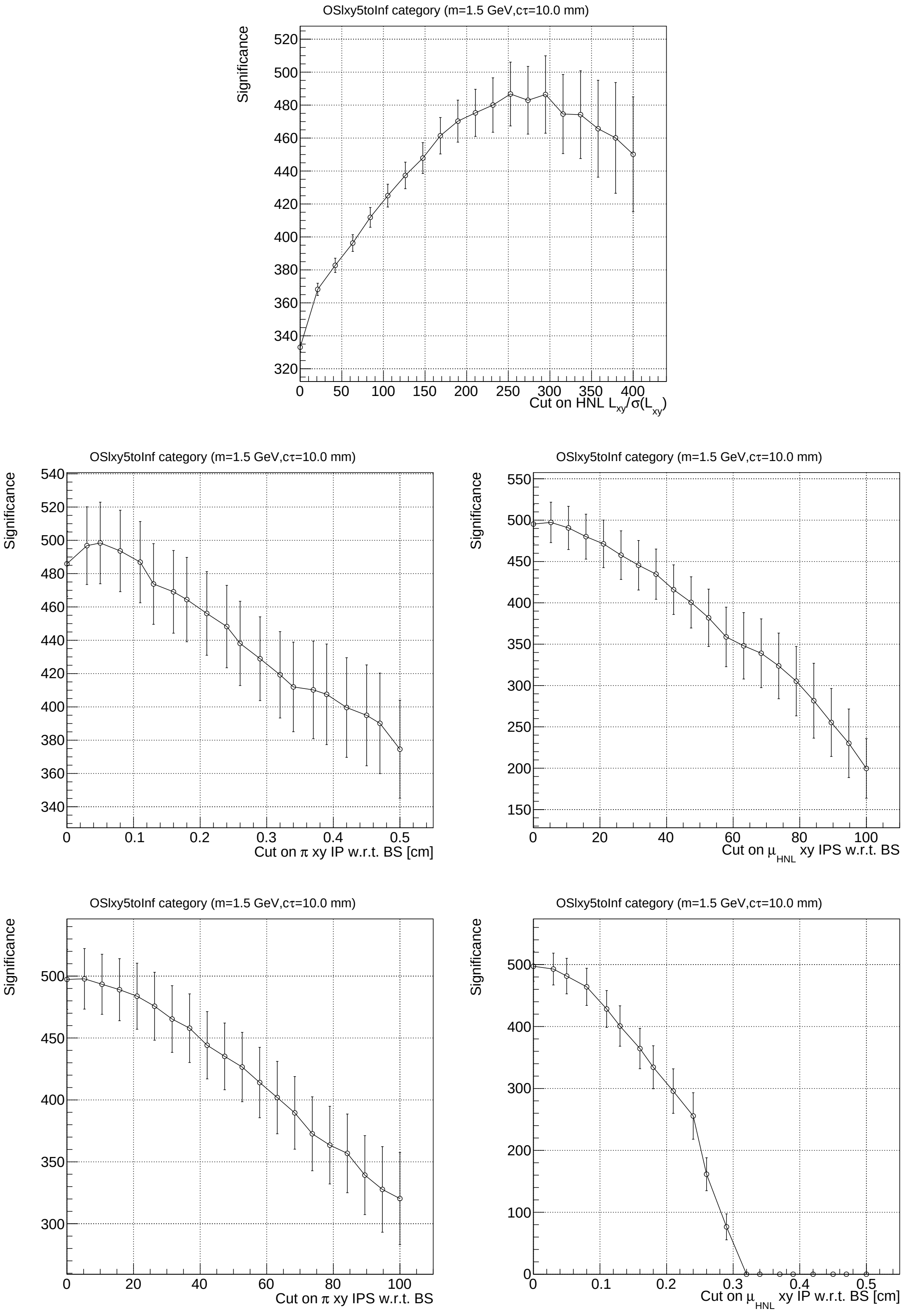

Figure B.8: Expected discovery significance for OSlxy5toInf category.

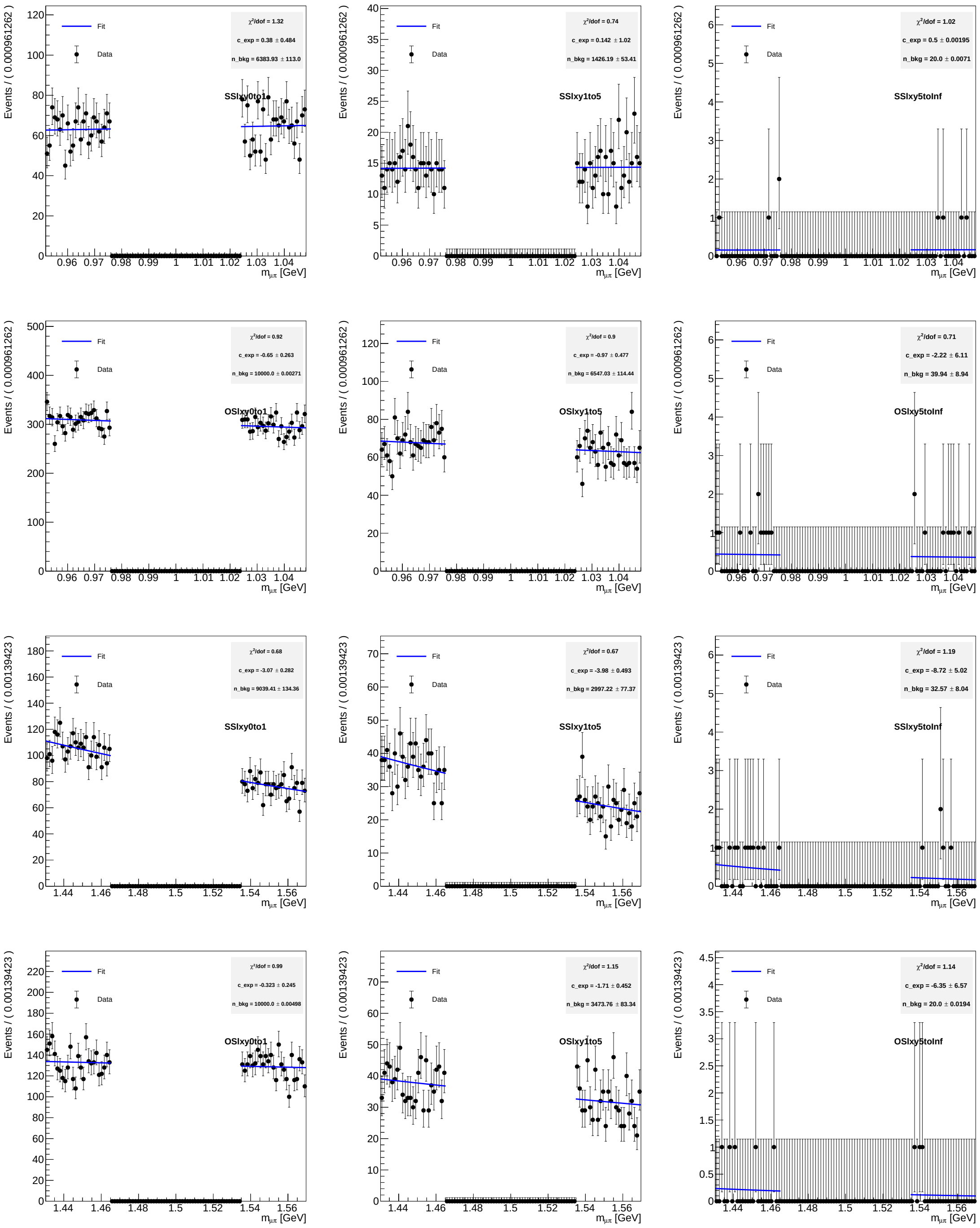

Figure B.9: Fit of the $\mu\pi$ invariant mass sidebands in data in the case of 1.0 GeV mass (top two rows) and 1.5 GeV mass (bottom two rows) HNL hypothesis. First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

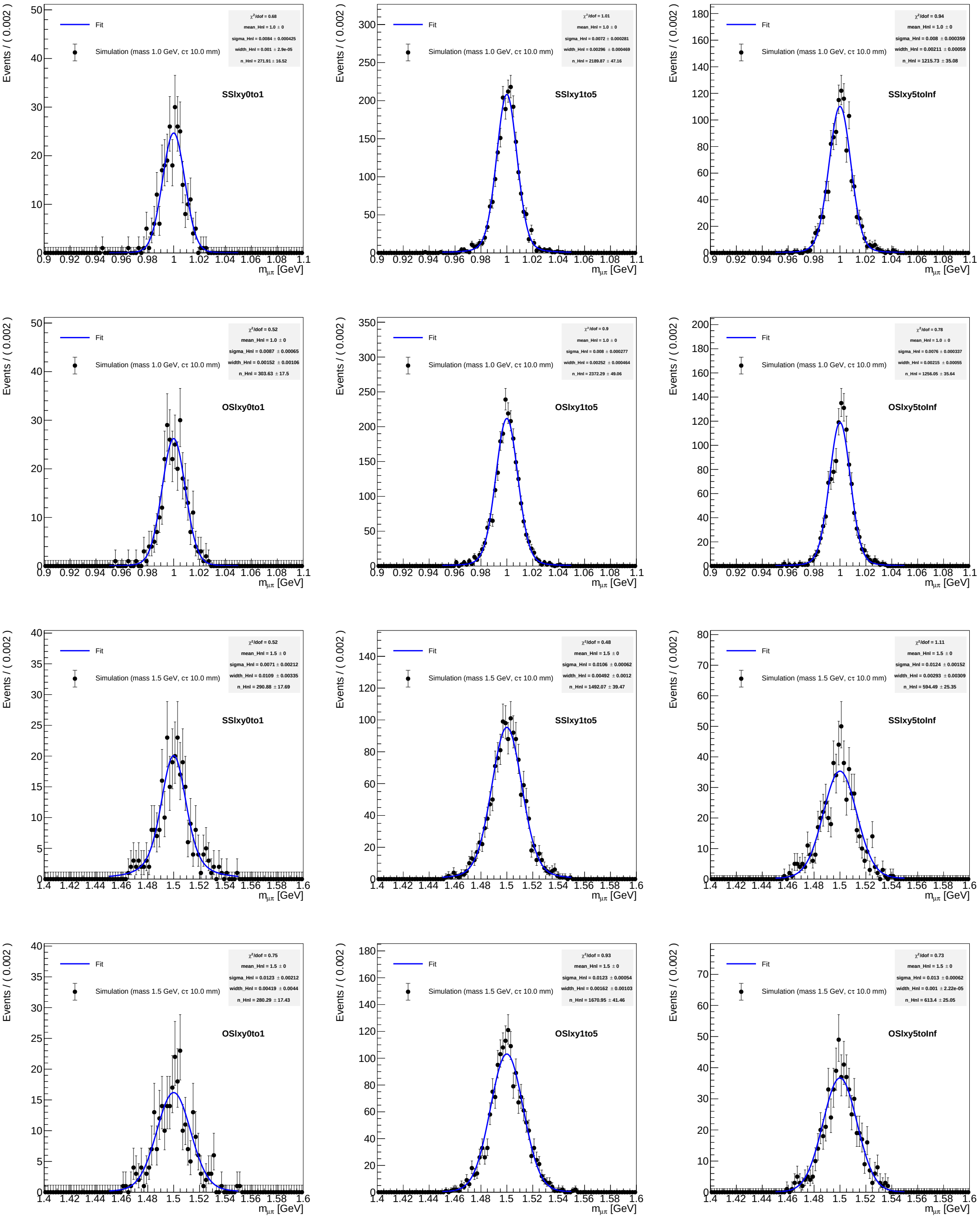

Figure B.10: Fit of the $\mu\pi$ invariant mass of a $c\tau = 10\,\text{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

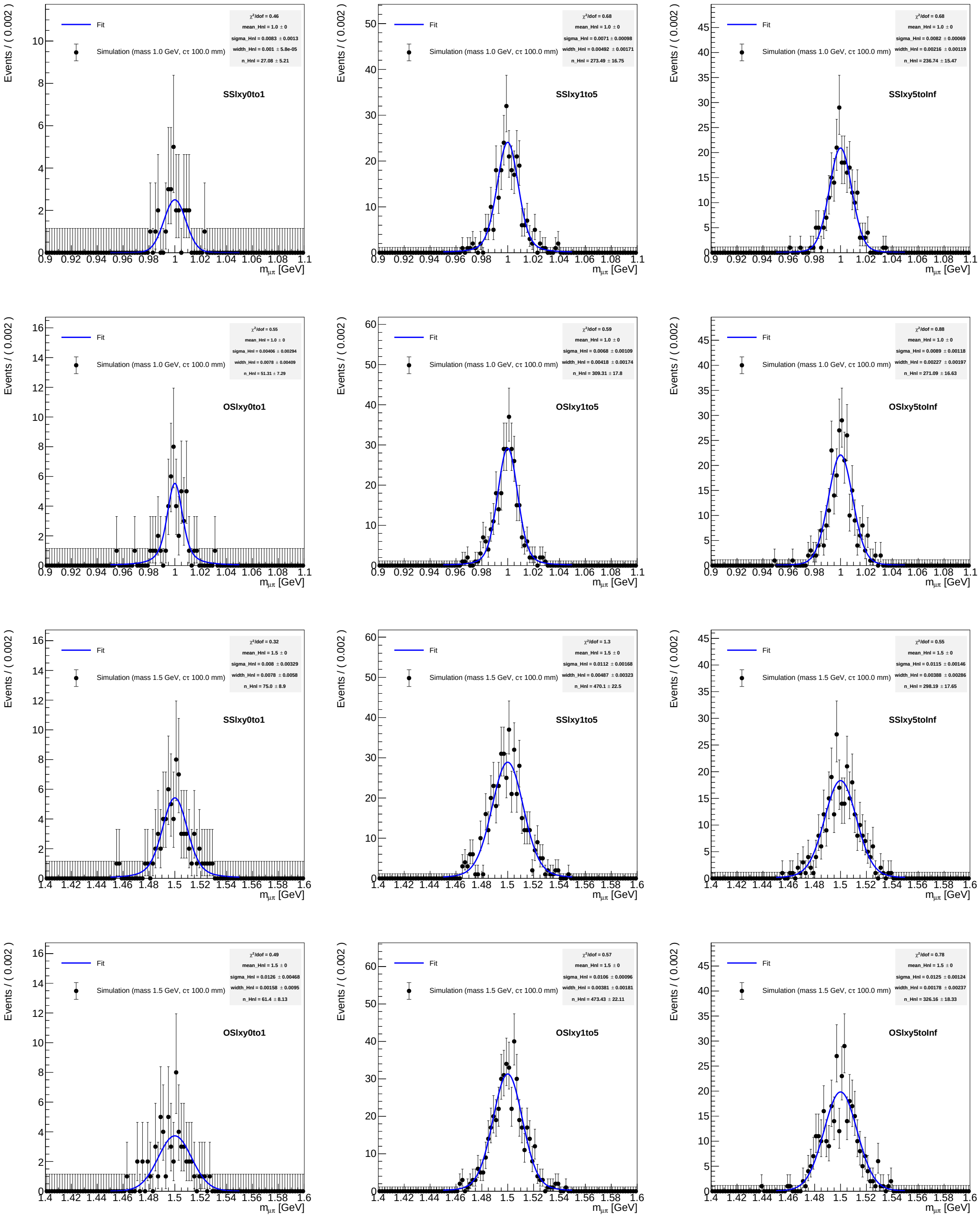

Figure B.11: Fit of the $\mu\pi$ invariant mass of a $c\tau = 100\,\text{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.

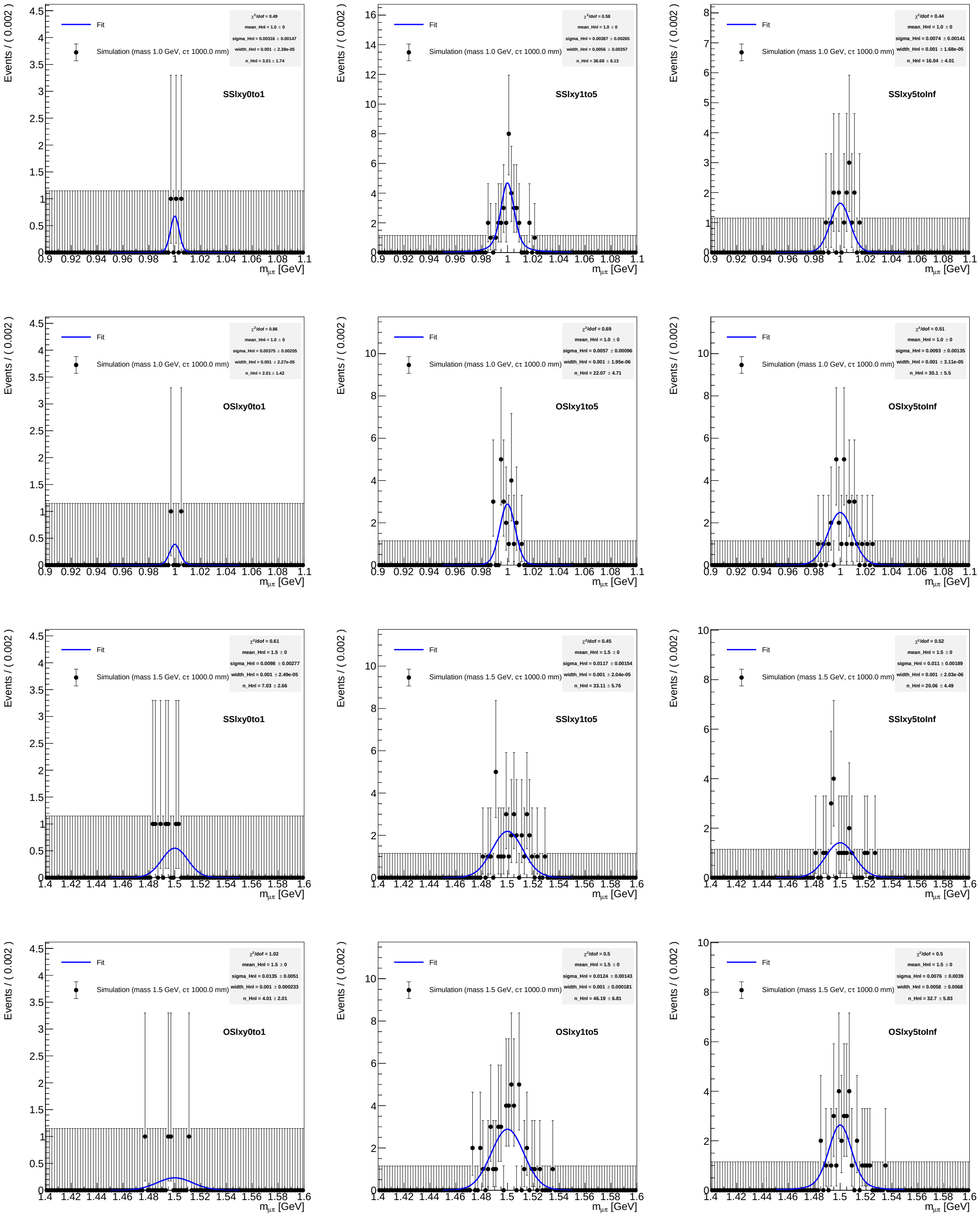

Figure B.12: Fit of the $\mu\pi$ invariant mass of a $c\tau = 1000\,\mathrm{mm}$ HNL of 1.5 GeV mass (top two rows) and 1.0 GeV mass (bottom two rows). First and third rows refer to same-sign categories, second and fourth rows to the opposite-sign category.