Search for new phenomena at CMS in the multilepton final state

A thesis
Submitted in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

by:

Shubhanshu Chauhan
ID No.: 20133284

Department of Physics
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH
PUNE-411008, INDIA

November 2019

Certificate

Certified that the work incorporated in the thesis entitled **"Search for new phenomena at CMS in the multilepton final states"** submitted by **Shubhanshu Chauhan** was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other university or institution.

Date: October 9, 2019 (Supervisor)

Declaration of Authorship

I declare that this written submission represents my ideas in my own words and where others ideas have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: January 28, 2019 (Shubhanshu Chauhan)

ID. No.: 20133284

Abstract

The standard model of particle physics has been extremely successful in describing the fundamental particles and their interactions via electroweak and strong forces. Despite this, there are various experimental observations that can't be explained using the standard model, such as, small non-zero neutrino mass and the presence of dark matter in the universe. Various extensions of SM are proposed to fix these inadequacies of the standard model. Searches for the standard model extensions, vectorlike leptons, and type-III seesaw fermions are carried out in multilepton final states using the 13 TeV proton-proton collision data collected by CMS experiment at LHC in 2016 and 2017.

The events are categorized based on the multiplicity of leptons (electrons, muons, and taus) and the invariant mass of the oppositely charged lepton pairs. The missing transverse momenta and the scalar sum of lepton transverse momentum are used to distinguish a signal from the standard model background. The observations are consistent with the standard model background. The results are used to place limits on the production of vectorlike leptons and heavy fermions of the type-III seesaw model. The vectorlike leptons with a mass between 130-690 GeV are excluded using 41.4 fb^{-1} of data collected in 2017, while the type-III seesaw fermions are excluded with a mass below 840 GeV using 35.9 fb^{-1} of data collected in 2016.

Acknowledgements

First and foremost I would like to express the deepest gratitude to my advisor, Prof. Sourabh Dube. He has been not only an excellent mentor but also a great friend. I feel very fortunate to have worked with him for my thesis. During my time as a graduate student, I could always discuss any issue with him, be it related to Physics for life in general.

A special thank to the members of the multilepton group at Rutgers University, Halil, Maxi, Peter, Matt, and Sunil. Prof. Sunil Somalwar has been very kind and helpful throughout the analysis. Halil has been a fantastic collaborator, his (virtual)company has made burning the midnight oil to meet the deadlines, a little bit easier.

I am also thankful to the HEP group members of IISER Pune, Kunal, Anshul, Angira, Arnab, Vinay, Seema, Aditee, Shubham, and Bhumika. Kunal, Anshul and I started our journey in HEP together, a bit later Angira joined the group. And it has been a great experience working with all of them since then. A special thanks to Angira who shared the office with me, she had enough patience to tolerate me. Thanks to Neha for also sharing the office and for all those the bike rides to fun places in Pune. I would also like to thank the members of my RAC committee, Prof. Seema Sharma and Prof. Sanjay Swain, who provided invaluable suggestions during the thesis. I would also like to thank Prof. Arun Thalapillil, who indulged in my all sort of theory questions.

In the end, I would like to thank my family, my parents, my brother, sister-in-law and my wife Himani for supporting me and believing in me. Thank you, Himani, for always being understanding, supportive and making my life beautiful.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Our understanding of the natural phenomena has evolved by trying to ask a few basic questions, such as 1) What is the universe made of? 2) How does it function? In ancient times in India, it was believed that everything around us is made out of five elements, namely, fire, air, earth, water, and space. This understanding of the fundamental constituents has since been updated multiple times as we better understand the natural phenomena. The main aim of particle physics is to find out what are the fundamental building blocks of nature, and how do they interact with each other? Answers to these questions have led to the development of the standard model (SM) of particle physics. The SM has been successful and has stood the test of time since the 1970s. It has also predicted the existence of top quark, tau neutrino, W, Z, and Higgs boson, all of which have been experimentally discovered, latest being the Higgs boson discovered in 2012 by CMS and ATLAS collaborations. Guided by a set of symmetry and written in the language of quantum field theory, it incorporates strong, weak, and electromagnetic forces in a single coherent framework.

Although SM has been hugely successful, there are a few shortcomings to the SM, such as the omission of the gravity, no explanation for the dark matter and dark energy. There are models beyond SM that attempt to address these shortcomings. The models such as supersymmetry, extra dimensions, seesaw, and additional vectorlike leptons are few such models. In this thesis, a search for such models (type-III seesaw and vectorlike taus) is performed. A brief description of the standard model and the phenomena beyond is given in Chapter 2.

One of the ways to examine the models beyond SM and measure the SM parameters is to collide high energy particles and then examine the debris coming out of the collision of these particles. This approach has been successful in discovering W, Z, Higgs boson, as well as top quark in the past 60 years. The large hadron collider (LHC) situated at CERN in Geneva, Switzerland, is currently the highest energy particle collider in the world with a center of mass energy of 13 TeV. At LHC, the proton beams traveling in opposite directions are made to collide at four interaction points. At each of these points, detectors are placed on examining the particles coming out of the collisions. Compact Muon Solenoid (CMS) is one of the multipurpose detectors at LHC (another being “ATLAS”). In this thesis, data collected by CMS at 13 TeV is used to search for the new phenomena. The description of the LHC and CMS detector is provided in Chapter 3.

The search is based on reconstructing at least three final state leptons (three electrons/muons, four or more electrons/muons and two electrons/muon + at least one hadronically decaying tau lepton) in each event. The description of the reconstruction of the objects used in the analysis is provided in the Chapter 4. There are many SM processes, such as WZ, ZZ, $\mathrm{t\bar{t}Z}$, which are the background of this search. The techniques used to estimate the SM background are described in the Chapter 5. The multileptonic events in each category are further divided into low and high missing transverse momentum ($\mathrm{p_T^{miss}}$) regions. In each of these regions scalar sum of the lepton transverse momentum is used to discriminate signal from the SM background. The search for vectorlike leptons is described in Chapter 6, and search for the evidence of the type-III seesaw is described in Chapter 7.

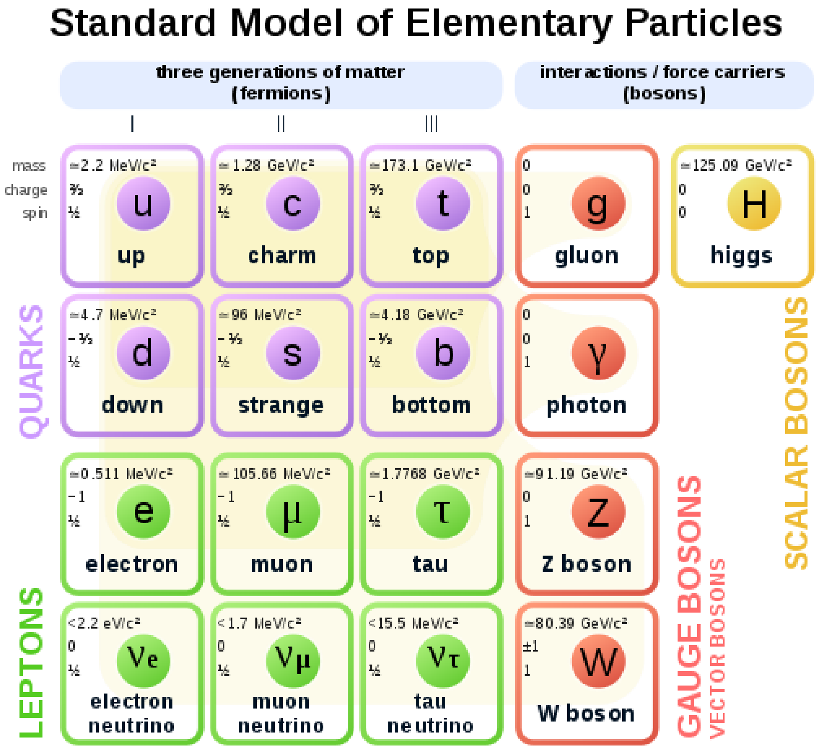
Chapter 2

The Standard Model and Beyond

In this chapter, a brief introduction to the Standard Model (SM) of particle physics is discussed. It is the most successful theory of fundamental particles and interactions. Various limitations of the SM are discussed, models such as vectorlike leptons and type-III seesaw mechanism, which attempt to resolve the limitations of the SM, are also summarized.

2.1 The Standard Model of Particle Physics

The standard model of particle physics (SM) is a quantum field theory that describes the fundamental particles and their interactions. It is a gauge theory based on $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ symmetry group, developed by Glashow, Weinberg, and Salam. The SM is developed from the work of many scientists over the years. Glashow in 1961, proposed a model that unified the electromagnetic and weak interactions [2]. Salam and Weinberg then integrated the Higgs mechanism into the electroweak framework [3, 4]. The quantum chromodynamics (QCD), a description of the strong interaction, was included by Gell-Mann, Gross and others, to complete the SM as we know it in 1973 [5, 6, 7]. The SM has been a very successful theory of the three fundamental interactions, namely, the strong, the weak, and the electromagnetic. Gravity, the fourth fundamental interaction is not described within SM, and including gravity in the same framework is a very active area of the particle physics research.



Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		
QUARKS	mass ≃2.2 MeV/c² charge ⅔ spin ½ u up	≃1.28 GeV/c² ⅔ ½ c charm	≃173.1 GeV/c² ⅔ ½ t top	0 0 1 g gluon	≃125.09 GeV/c² 0 0 H higgs
	≃4.7 MeV/c² -⅓ ½ d down	≃96 MeV/c² -⅓ ½ s strange	≃4.18 GeV/c² -⅓ ½ b bottom	0 0 1 γ photon	SCALAR BOSONS
LEPTONS	≃0.511 MeV/c² -1 ½ e electron	≃105.66 MeV/c² -1 ½ μ muon	≃1.7768 GeV/c² -1 ½ τ tau	≃91.19 GeV/c² 0 1 Z Z boson	
	<2.2 eV/c² 0 ½ ν_e electron neutrino	<1.7 MeV/c² 0 ½ ν_μ muon neutrino	<15.5 MeV/c² 0 ½ ν_τ tau neutrino	≃80.39 GeV/c² ±1 1 W W boson	GAUGE BOSONS VECTOR BOSONS

FIGURE 2.1: Figure shows the particle content of the standard mdoel of the particle phyiscs [1].

SM includes all the fundamental particles and also describes the interactions among them. Particles in SM could broadly be split into two categories, the fermions and the bosons. Fermions in the SM form all the matter while the bosons govern the interactions among SM particles. The fermions that do not take part in strong interactions and only interact via electroweak force are called leptons. On the other hand, the fermions that interact via electroweak, as well as the strong force are called quarks. In SM, the matter fields are spin 1/2 particles that come in 3 generations. The lepton family is composed of the electron (e), the muon (μ), and the tau (τ), which have 1 unit of a negative charge, and correspondingly three neutrinos (ν_e, ν_μ, ν_τ) which are electrically neutral. The Quark family is composed of the "up type" up (u), charm (c) and top (t) having +2/3 electric charge, and "down type" down (d), strange (s) and bottom (b) having -1/3 of electric charge. Each fermion has a corresponding antiparticle in the SM. The vector bosons in SM are the mediators of the interactions. The electroweak interaction is mediated via photon, $W^{\pm}$, and Z bosons, while the strong interaction is mediated via gluons. The Higgs boson, a scalar boson, arises due to the spontaneous symmetry breaking of the electroweak theory. All the massive SM particles in the SM acquire their mass upon interacting with the Higgs field. A summary of all the particles in SM with corresponding electric charge and mass is shown in Fig. 2.1.

In the next few sections, fundamental interactions incorporated in the SM are discussed briefly. The description follows Ref. [8, 9].

2.1.1 Electromagnetic Interaction

The particles carrying the electric charge have electromagnetic interaction via an exchange of the photon and is described by a quantum field theory called quantum electrodynamics (QED). The Lagrangian density for a free fermionic field, ψ, with mass m can be written as:

$$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi \tag{2.1}$$

where the field, $\bar{\psi}$, is the conjugate field obtained by $\bar{\psi} = \psi^{\dagger}\gamma^{0}$. This Lagrangian remains invariant under the global gauge transformation of the form $\psi \rightarrow \psi' \equiv e^{iq\theta}\psi$, where θ and q are real constants. This Lagrangian leads to the equation of motion for the spin-$\frac{1}{2}$ fermions, called the Dirac equation. Global gauge invariance implies that the addition of any phase to the wavefunction, ψ, does not change the equation of motion as long as the transformation parameter θ remains independent of the local space-time coordinates. The local gauge invariance is broken by the action of the partial derivative on the field. For the theory to describe the nature, the Lagrangian should also have local gauge invariance, since the physics should not change by choosing a different reference point for the coordinate system. The theory could be made local gauge-invariant if we can make the derivative transform like fields. To accomplish this, a new covariant derivative is defined as:

$$\mathcal{D}_{\mu} = \partial_{\mu} - iqA_{\mu} \tag{2.2}$$

where A_μ is a gauge field with spin-1, which transform in the following way:

$$A_\mu \rightarrow A'_\mu = A_\mu + \partial_\mu \theta \tag{2.3}$$

Using this definition of the covariant derivative, $\mathcal{D}_\mu$, its transformation is expressed as :

$$\mathcal{D}_\mu \psi \rightarrow (\mathcal{D}_\mu \psi)' = e^{iq\theta} \mathcal{D}_\mu \psi \tag{2.4}$$

Now the Lagrangian for the theory can be written as:

$$\mathcal{L} = i\bar{\psi}(\gamma^\mu(\partial_\mu - iqA_\mu) - m)\psi \tag{2.5}$$

By requiring the local gauge invariance of the Lagrangian has led to a new term $\bar{\psi}A_\mu\psi$ in the Eq. 2.5. This term represents the interaction between spin-1 gauge field A_μ with the fermion field ψ. By defining $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ the Lagrangian for a spin-1 field could be written as :

$$\mathcal{L}_A = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \frac{1}{2} m_A^2 A_\mu A^\mu \tag{2.6}$$

m_A is the mass of the boson A. If we add the mass term from the Eq. 2.6 to the Eq. 2.5 the Lagrangian will no longer be invariant under the U(1) symmetry transformation. Hence if this theory is to describe the electromagnetic interaction, then the gauge field A_μ has to be massless. We can identify the gauge field A_μ as the photon, q as the charge for the electromagnetic interaction, and $F_{\mu\nu}$ as the field strength tensor. We can now write down the full Lagrangian for QED as:

$$\mathcal{L}_{QED} = i\bar{\psi}(\gamma^\mu \mathcal{D}_\mu - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{2.7}$$

From this Lagrangian Dirac equation for the electromagnetic field could be obtained, the equation comes out as:

$$i\gamma_\mu \partial^\mu \psi - m\psi = q\gamma_\mu A^\mu \psi \tag{2.8}$$

The range of interaction for the electromagnetic force can be obtained using the Heisenberg uncertainty principle, $\Delta x \Delta p \geq \hbar/2$. The force carrier for the electromagnetic interaction, photon, is massless, this implies that electromagnetic interaction has an infinite range, while its strength varies inversely with the square of the distance. The strength of the electromagnetic interaction is given by a constant α, also known as the fine structure constant. It could be obtained by a comparison between the electrostatic energy of the two electrons kept at one unit natural length to the rest mass of the electron.

$$\alpha = \frac{e^2}{4\pi\epsilon_0 \hbar c} \approx \frac{1}{137} \tag{2.9}$$

where e is the unit charge, m_e is the mass of an electron, ϵ_0 is the permittivity of the free space, c is the speed of light and $\hbar$ is reduced Plank constant. $\hbar$ is defined as $\hbar = h/2\pi$. The fine structure constant depends on the amount of energy transfer during the interaction [10]. The value of the α increases with the increasing energy of the interaction. All the particles with non-zero electric charge participate in the electromagnetic interaction, α being the strength of that interaction.

2.1.2 Electroweak Unification and Symmetry Breaking

The $W^\pm$ and Z bosons are the force carriers of the weak interaction. They were discovered in 1983 by UA1 and UA2 collaborations [11, 12, 13, 14]. Masses of these bosons have been measured very precisely. The W boson has a mass of 81.4 GeV, while the Z boson has a mass of 91.2 GeV [15]. In electrodynamics, the gauge boson, photon, is massless, and its easier to write down the QFT for electromagnetic interaction. The massive gauge bosons in the weak theory are problematic since the mass term in the Lagrangian will break the symmetry. Moreover, the interaction of W bosons with the

SM fermions depends on their chirality. The W bosons only couple to left-handed particles, having negative chirality and right-handed antiparticles, having positive chirality. On the other hand, the Z boson couples to both right-, left-handed particles and their antiparticles. This behavior breaks the charge and parity symmetry maximally.

The presence of the three force carriers of weak interaction hints towards $SU(2)_L$ symmetry group in theory. The $SU(2)_L$ symmetry group would lead to all three gauge bosons having the same couplings to the SM fermions. We know from the experimental observations that the couplings of W and Z bosons to fermions are different; moreover, W bosons are electrically charged. To account for these observations, we can try to write down a theory having $SU(2)_L \otimes U(1)_Y$ symmetry group. Following the above argument, Lagrangian for SM fermions of a given family having a negative chirality (left-handed) doublet (ψ_L) and the positive chirality (right-handed) singlet (ψ_R) could be written as:

$$\mathcal{L}_{free} = \bar{\psi}_L \gamma_\mu \partial^\mu \psi_L + \bar{\psi}_R \gamma_\mu \partial^\mu \psi_R \tag{2.10}$$

The transformation of the left-handed and right-handed component of the fermion fields under local $SU(2)_L \otimes U(1)_Y$ could be written as:

$$\begin{aligned} \psi_L \to \psi'_L &= e^{(\frac{1}{2} i \vec{\xi}.\vec{\sigma} + iy\beta)} \psi_L \\ \psi_R \to \psi'_R &= e^{(iy\beta)} \psi_R \end{aligned} \tag{2.11}$$

Where $\vec{\xi}$ and β are function of spacetime coordinates, y is hypercharge and $\vec{\sigma}$ are the Pauli matrices.

The fermion mass term for these theories would be of the form $-m(\bar{\psi}_L \psi_R)$. If we add fermion mass term to the Lagrangian, the $SU(2)_L$ symmetry will break. To avoid this, let's consider the fermions to be massless for the time being and develop the theory. Just like QED, we can define covariant derivatives:

$$\begin{aligned} \mathcal{D}_\mu \psi_L &= (\partial_\mu - \frac{1}{2} i g \vec{W}_\mu . \vec{\sigma} + \frac{1}{2} i g' B_\mu) \psi_L \\ \mathcal{D}_\mu \psi_R &= (\partial_\mu + \frac{1}{2} i g' B_\mu) \psi_R \end{aligned} \tag{2.12}$$

The gauge boson fields ($\vec{W}$, B) transform as:

$$\begin{aligned} W^i_\mu \rightarrow W^i_\mu + \frac{1}{g}\partial_\mu \xi^i - \epsilon^{ijk}\xi^j G^k_\mu \\ B_\mu \rightarrow B_\mu + \frac{1}{g'}\partial_\mu \beta \end{aligned} \qquad (2.13)$$

Both $\vec{W}$ and B are spin one gauge bosons, the mass term from these fields in the Lagrangian would break the symmetry of the theory. By defining $\vec{W^{\mu\nu}}, B^{\mu\nu}$ as the field strength tensors for the gauge fields, the Lagrangian for the electroweak theory without the mass term can be expressed as:

$$\mathcal{L}_{EWK} = -\frac{1}{4}\vec{W}^{\mu\nu}.\vec{W}_{\mu\nu} - \frac{1}{4}B^{\mu\nu}B_{\mu\nu} + i\bar{\psi}_L\gamma_\mu\mathcal{D}^\mu\psi_L + i\bar{\psi}_R\gamma_\mu\mathcal{D}^\mu\psi_R \qquad (2.14)$$

Two components of the $\vec{W}$ boson, namely (W^1_μ, W^2_μ), participate in the charged current interaction, such as the interaction between an up-type quark with a down-type antiquark. We can identify the $W^\pm$ bosons which participate in the charged current interactions, by making a linear combination of the (W^1_μ, W^2_μ):

$$W^\pm_\mu \equiv \frac{1}{\sqrt{2}}(W^1_\mu \mp W^2_\mu) \qquad (2.15)$$

Wienberg, Salam and Glashow [2, 4] also showed that the third component of the $\vec{W}_\mu$, W^3_μ can mix with B_μ, giving rise to the photon field A_μ and a new neutral gauge field, later discovered as Z boson. The mixing strength between W^3_μ, B_μ is given by an angle θ_w called Weinberg angle. Hence the photon field A_μ and neutral boson field Z_μ could be written in the following way:

$$\begin{aligned} A_\mu &= B_\mu cos\theta_w + W^3_\mu sin\theta_w \\ Z_\mu &= -B_\mu sin\theta_w + W^3_\mu cos\theta_w \end{aligned} \qquad (2.16)$$

This theory provides a unification between the electromagnetic and the weak force. Particles in this theory are still massless and give rise to a long-range interaction. We know from the experimental observations that the weak interaction has massive gauge bosons, resulting in a short-range interaction. In the next section, we shall discuss a mechanism that can provide mass to these particles in the electroweak theory.

The masses to the particles in the electroweak theory are provided by the use of Goldstone's theorem [16, 17, 18], It proposes the existence of a massless field (Goldstone boson) for every broken symmetry. Higgs, Brout, and Englert used Goldstone's theorem to propose a mechanism which breaks the electroweak symmetry spontaneously and makes the electroweak bosons massive [3, 19]. A brief description of this procedure is given here. Any complex scalar field can be written as a two-component field as follows:

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \tag{2.17}$$

Using the field, ϕ, we can write a generic Lagrangian of the form:

$$\mathcal{L} = (\mathcal{D}_\mu \phi)^\dagger (\mathcal{D}^\mu \phi) - V(\phi, \phi^\dagger) \tag{2.18}$$

In this Lagrangian, we have used the covariant derivative, $\mathcal{D}_\mu$, for this complex scalar field defined as:

$$\mathcal{D}_\mu = (\partial_\mu - \frac{1}{2} ig \vec{W}_\mu . \vec{\sigma} + \frac{1}{2} ig' B_\mu)\phi \tag{2.19}$$

We can now write down a generic potential, $V(\phi, \phi^\dagger)$ of the form:

$$V(\phi, \phi^\dagger) = \mu^2 \phi\phi^\dagger + h(\phi\phi^\dagger)^2 \tag{2.20}$$

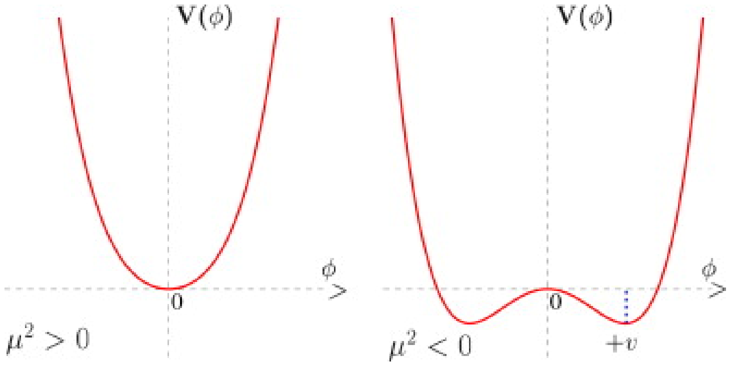


FIGURE 2.2: Figure shows sketch of a potential V as defined in Eq. 2.20.The sketch is taken from the reference [20]

Figure 2.2 shows a sketch of potential V for the case $\mu^2 > 0$ and $\mu^2 < 0$ [20]. It is evident from the potential that, it has a trivial minimum at $\phi = 0$ for $\mu^2 > 0$, while for $\mu^2 < 0$ the minimum of the potential occurs at:

$$|\phi_{min}| = \sqrt{\frac{-\mu^2}{2h}} \tag{2.21}$$

Now we can see that we started with a Lagrangian, which preserved the symmetry, but for a certain choice of the quadratic term, the ground state of the potential breaks the symmetry. This kind of phenomenon is known as the spontaneous breaking of symmetry. Now the initial Lagrangian could be perturbed around the new minima at ϕ_{min} by defining a new field H, called Higgs field. We can write

$$\phi = \phi_{min} + H \tag{2.22}$$

By doing an expansion around ϕ_{min}, the Lagrangian could be written using the electroweak boson as follows:

$$\mathcal{L} = \frac{1}{2}\partial_\mu H \partial^\mu H + \mu^2 H^2 + \frac{1}{4}g^2 v^2 W_\mu^+ W^{\mu-} + \frac{1}{8}v^2(g^2 + g'^2) Z_\mu Z^\mu \tag{2.23}$$

Using the Lagrangian in Eq. 2.23, we can identify the mass terms for electroweak bosons $(W^\pm, Z)$ as well as for the new Higgs field H as:

$$\begin{aligned} M_{W^\pm} &= \frac{1}{2} g v \\ M_Z &= \frac{1}{2} v \sqrt{g^2 + g'^2} \\ M_H &= \sqrt{-2\mu^2} \end{aligned} \tag{2.24}$$

With the help of the scalar Higgs field, we can write gauge-invariant couplings of Higgs to the fermions of the SM. These coupling terms are known as Yukawa terms, $\bar{\psi}_L \phi \psi_R$. The mass of the fermions in terms of the Higgs vacuum expectation value, v, and Yukawa coupling y_f could be expressed as $M_f = y_f \frac{v}{\sqrt{2}}$.

2.1.3 Strong Interaction

The strong force governs the interaction between the quarks and the gluons; the corresponding QFT that describes these interactions is called the quantum chromodynamics (QCD). QCD is a gauge theory, similar to QED but follows a non-abelian symmetry group, SU(3). The SU(3) symmetry of the theory means that the gauge bosons of the theory interact with themselves as a direct consequence of the non-commuting nature of the generators of the symmetry group SU(3). Each quark has a color charge under SU(3), labeled as red, green, and blue. Lagrangian for the theory without interaction for a given quark flavor can be written as follows:

$$\mathcal{L}_{free} = \bar{q}_f (i\gamma^\mu \partial_\mu - m_f) q_f \tag{2.25}$$

Where q is a three-component vector referring to three colors. The Lagrangian for the free theory is invariant under global gauge transformations of SU(3), while the local gauge invariance is not present. To make the Lagrangian invariant under local gauge transformations, the partial derivatives can be promoted to covariant derivatives. Since there are eight generators of SU(3), we need eight additional gauge fields to write a covariant derivative. The covariant derivative, in this case, could be defined as:

$$\mathcal{D}_\mu = \partial_\mu - \frac{1}{2} i g_s G^a_\mu \lambda_a \tag{2.26}$$

Where a takes integer values from 1 to 8, λ_a are the Gell-Mann matrices, and G_a are eight gauge fields. These gauge fields correspond to 8 gluons, which are carriers of the strong force. Using this definition of the covariant derivative complete QCD Lagrangian having global as well as local gauge invariance can be written as:

$$\mathcal{L}_{QCD} = \sum_f \bar{q}_f (i\gamma_\mu \mathcal{D}^\mu - m_f) q_f - \frac{1}{4} G^a_{\mu\nu} G^{a\mu\nu} \tag{2.27}$$

where $G^{a\mu\nu}$ is defined as, $G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu + g_s f^{abc} G^b_\mu G^c_\nu$ and is called gluon field strength tensor.

The standard model has been a hugely successful theory in describing the observed phenomenon in nature. It has been extensively tested at the electroweak scale at various experiments. There are many measurements, such as electron g-factor, where the SM prediction and the measured experimental value agree to unprecedented precision, having a relative uncertainty of the order of 10^{-13} [21]. However, the standard model is not complete, and various observed phenomena cannot be explained by the standard model. In the next section, some inadequacies of the standard model are discussed.

2.1.4 Inadequacies of the Standard Model

SM is the most consistent theory of fundamental interactions. Although SM has been very successful in describing the natural phenomenon, there are various experimental observations that SM fails to explain. Some of these inadequacies of the SM are discussed here.

In the SM, the neutrinos are massless. The Super-Kamiokande experiment [22] and the Sudbury Neutrino Observatory (SNO) [23] observed neutrino oscillations where the neutrinos oscillate among the flavor eigenstates. The SNO experiment observed conversion of neutrinos from the sun (ν_e) into ν_μ, ν_τ [24]. This is only possible if the flavor eigenstates are a linear superposition of the different mass eigenstates. These observations imply that at least two generations of the neutrinos have finite nonzero mass [25]. In the SM, there is no explanation for these finite neutrino mass.

In the standard model, there is no candidate for the dark matter. The presence of dark matter could be inferred by observing galaxy rotation curves [26, 27]. The velocity of stars in a galaxy as we move away from the center of the galaxy depends on the amount of matter enclosed. The visible matter content of the galaxy cannot account for the observed rotation curve. These observations could be explained if the entire galaxy is sitting in a dark matter halo. The word "dark" refers to the property that this matter does not interact via electromagnetic interactions and only interacts gravitationally (or weakly). There are other observations such as gravitational lensing, which also point to the same conclusions. The standard model does not have any explanation for these observations.

The observed universe to a large extent is made out of matter particles (not antimatter), which is only possible if there was a huge violation of Charge and Parity (CP) symmetries in the early universe. In the SM, there is only minimal CP violation, which is not enough to account for the observed universe today. In the evolution of the universe, there must have been substantial CP violation, which caused this large asymmetry between the matter and the antimatter.

The CMS and the ATLAS collaborations have observed the mass of the Higgs boson to be around 125 GeV [28, 29]. In the SM, the Higgs mass has terms contributions from quantum corrections of the order of (Planck scale)^2. This means the quantum corrections will have to be extremely fine-tuned to give rise to 125 GeV, which is many many orders of magnitude smaller than the corrections. In the standard model, there is no explanation for such cancellations.

2.2 Beyond Standard Model

There are various models/theories beyond the standard model (BSM) that attempt to address the issues with the SM. Some of the most common models are Supersymmetry

(SUSY), seesaw, extra dimensions, additional vectorlike fermions, and compositeness models. Supersymmetry is a proposed symmetry between the fermions and bosons, and it predicts the existence of a superpartner for each SM particle. For all fermions, there are bosonic superpartners, and for each boson, there are fermionic superpartners. SUSY can explain the fine-tuning of Higgs mass by canceling the large quadratic divergence terms. If a certain quantum number in SUSY is taken to be conserved, then it also provides a dark matter candidate who can explain the relic density of the dark matter observed today. The extra-dimensional models try to address the fact that gravity is much much weaker than any of the other fundamental forces and predict the existence of massive gravitons. The compositeness models, on the other hand, explore the possibility that the fundamental particles we see today might be made out of some other particles.

Many of above-mentioned BSM models, such as extra dimensions, compositeness models, and SUSY, often have additional vectorlike leptons (VLLs) in them. Moreover, one most straightforward extension to SM is the addition of VLL doublet. In this thesis, a search for the VLLs that couple to the third generation of the standard model leptons is presented. In Chapter 6, more details about the VLLs, as well as the strategy and results of the search are provided.

One of the most glaring deficiencies of the SM is the finite nonzero mass of the neutrinos. The seesaw mechanism is one technique that extends the SM and attempts to explain this discrepancy. In this thesis, we also performed a search for evidence of the type-III seesaw mechanism. Chapter 7 discusses the details of the type-III model as well as the results of the search.

Chapter 3

The Experimental Apparatus

The Large Hadron Collider is a superconducting hadron accelerator located at CERN near Geneva, Switzerland [30]. The Compact Muon Solenoid (CMS) experiment is a multipurpose detector at LHC [31]. It measures the properties of the particles produced in collisions at LHC. A description of the LHC and CMS detector is provided in this chapter.

3.1 The Large Hadron Collider

The LHC is a two-ring superconducting hadron accelerator, and it is designed to collide counter-rotating proton beams with a center-of-mass energy of 14 TeV (7 TeV per beam) and luminosity of $10^{34}\ cm^{-2}s^{-1}$. It can also collide heavy (Pb) ions with an energy of 2.8 TeV per nucleon and peak luminosity of $10^{27} cm^{-2}s^{-1}$. Figure 3.1 shows a view of the accelerator complex of CERN [30].

The number of events per second generated in the LHC collisions could be written as:

$$N_{event} = L\sigma_{event} \tag{3.1}$$

where σ_{event} is the cross section for the event under study and L the machine luminosity. The machine luminosity depends only on the beam parameters and can be written for a Gaussian beam distribution as:

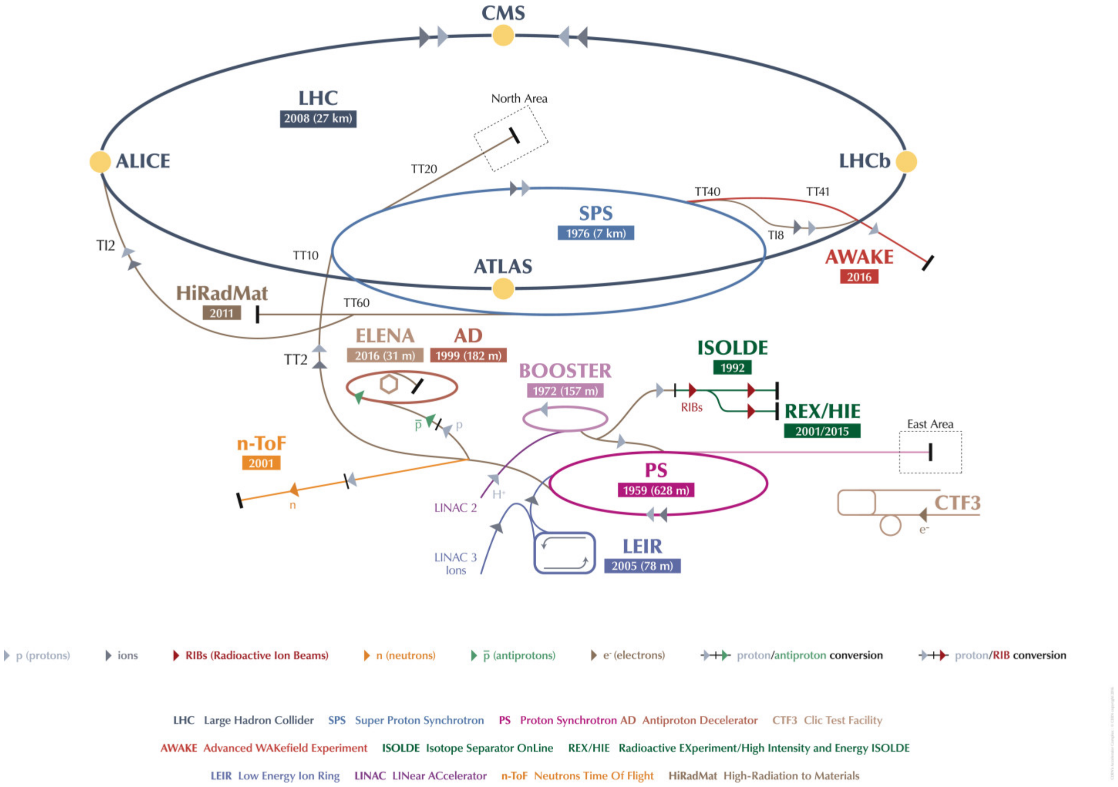


FIGURE 3.1: Figure shows a schematic diagram of the CERN accelerator complex showing various stages of the accelaration for proton beam and the collision points on the LHC ring where detectors are placed to study the collisions. [30]

$$L = \frac{N_b^2 n_b f_{rev} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{3.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP):

$$F = (1 + (\frac{\theta_c \sigma_z}{2\sigma^*})^2)^{-1/2} \tag{3.3}$$

θ_c is the full crossing angle at the IP, σ_z the RMS bunch length, and σ^* the transverse RMS beam size at the IP.

The amount of proton-proton collision data delivered by LHC to CMS in year 2016 and 2017 at $\sqrt{s}$=13 TeV is shown in Figure 3.2. In 2016 CMS recorded 37.76 fb^{-1} of data while in 2017 CMS recorded 44.98 fb^{-1} of data. In the analysis, we use the data after it has passed certain filters to make sure we do not use the events where the important detector was not operating properly for the beam conditions were favorable. After allying these event filters, 41.4 fb^{-1} and 35.9 fb^{-1} data is used in the analysis from the years 2017 and 2016, respectively.

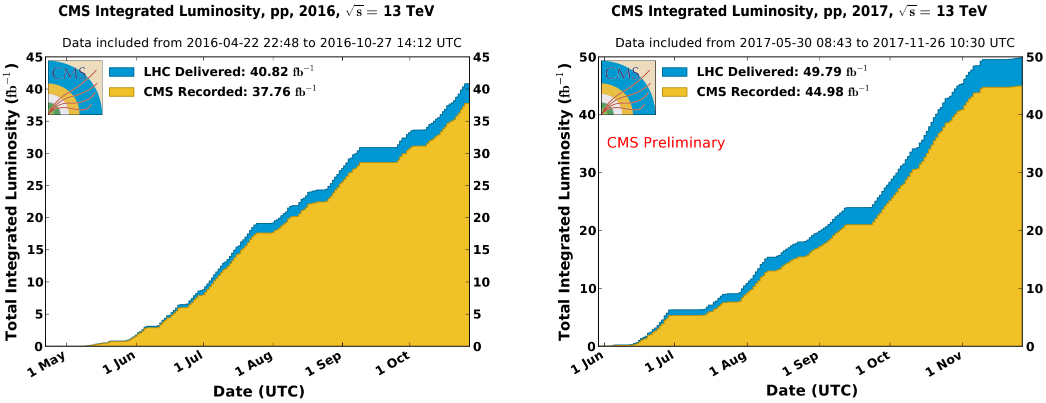


FIGURE 3.2: Figure shows the total amount of proton proton collision data collected by CMS detector in 2016 (left), and in 2017 (right). In 2016 CMS recorded 37.76 fb^{-1} of data while in 2017 CMS recorded 44.98 fb^{-1} of data.

3.2 The CMS Detector

The CMS detector is located 100 meters underground close to the French village of Cessy, at point 5 of the LHC ring. It is one of the two (another being ATLAS) multipurpose detectors at LHC. CMS was designed to discover the Higgs boson and explore BSM physics at the TeV scale. Due to its design goals, CMS has an excellent silicon tracker, calorimeters, and outer muon tracker [32]. By combining the information from various parts of the detector, CMS provides an excellent reconstruction efficiency and energy resolution for electrons, photons, muons, and hadrons.

The CMS detector gets its name from its large superconducting solenoid. It has a 6-meter internal diameter and provides a uniform 3.8-T magnetic field. The inner silicon tracker, electromagnetic calorimeter and the hadron calorimeter are inside the superconducting solenoid. The presence of a high magnetic field in the tracker volume provides an excellent measurement of the charged particle momentum and also separates the energy deposits of charged and neutral particles in the calorimeters. The magnetic field is returned via steel yoke, which also houses the muon chambers and provides structural integrity to the structure of the CMS detector. Figure 3.3 shows the basic design of the CMS detector.

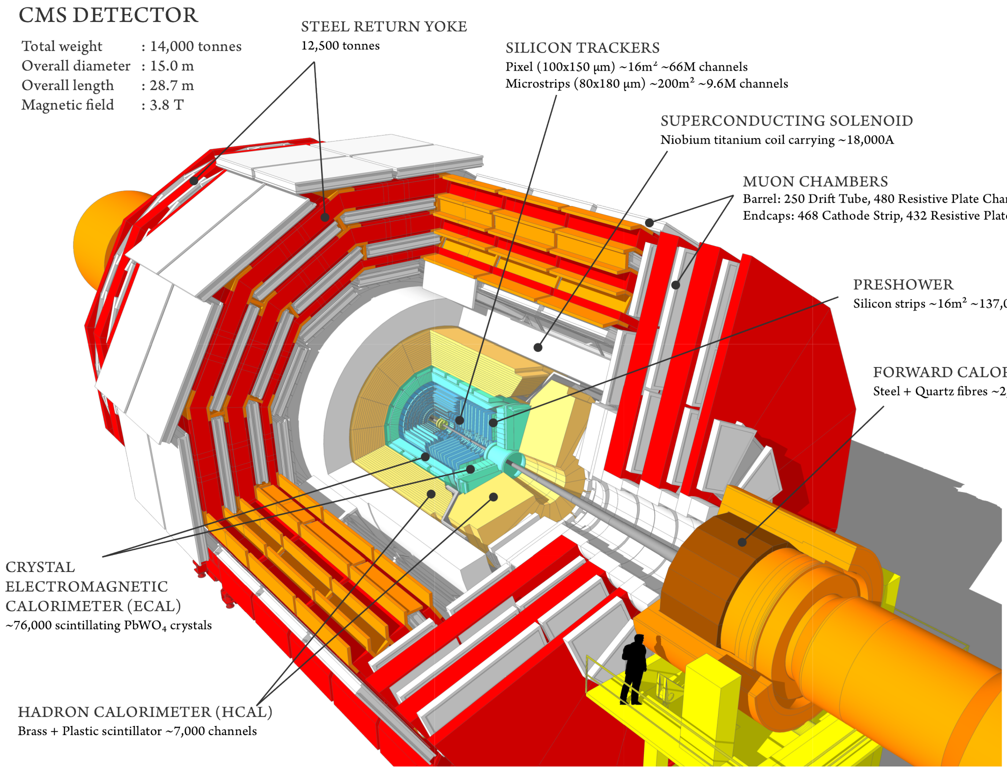


FIGURE 3.3: A view of the CMS detector showing its cylindrical structure, with endcaps layers on either of the cylinderical layers.

3.2.1 Inner Tracking System

The inner tracking system of the CMS detector is designed to provide a precise and efficient measurement of the momentum of the charged particles arising from the collision point of the LHC. It also provides a precise reconstruction of the secondary vertices, which helps mitigate the effect of multiple interactions on particle reconstruction and is essential in efficiently identifying the b-quark jets.

The inner tracking system of CMS is made out of silicon detectors. It has a cylindrical shape with 5.8 m length and 2.5 m diameter. The tracker small pixel detector close to the interaction point. A large silicon strip detector surrounds the pixel detector. Both of these subdetectors are submerged in a homogeneous magnetic field provided by the superconducting solenoid. A schematic diagram of the CMS tracker is provided in Figure 3.4.

The pixel tracker has three cylindrical barrel layers with two endcap layers on either side. It provides a 3-dimensional positional measurement of the charged particle passing through it. The transverse and longitudinal hit resolution in the pixel tracker is of the order of 10 μm and 20-40 μm, respectively. The position of the sensor provides the third coordinate of the hit. It consists of 66 million pixel sensors and covers an area of about one square meter.

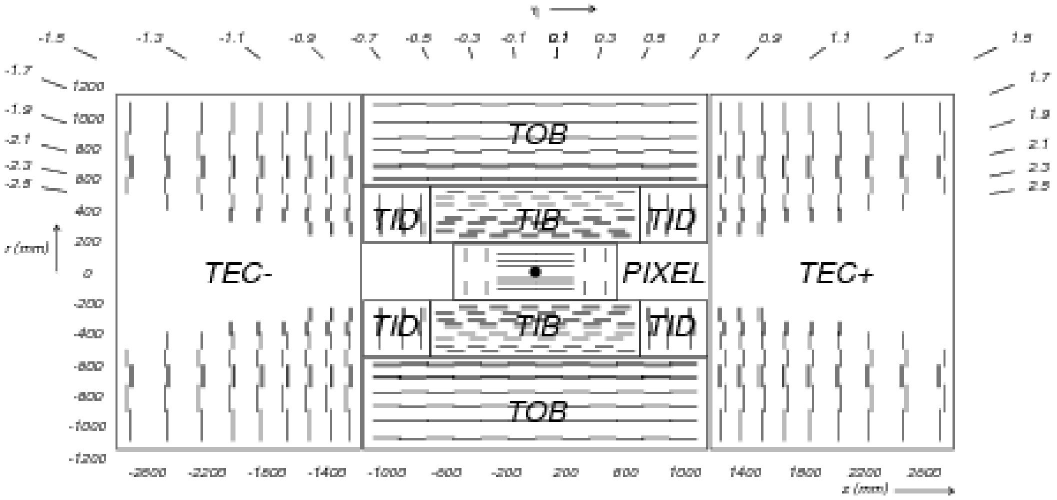


FIGURE 3.4: A schematic diagram of the CMS inner tracking system [32].

The strip tracker consists of four subsystems, the Tracker Inner Barrel (TIB) and Disks (TID), and Tracker Outer Barrel (TOB) and Tracker EndCaps (TEC). In total, the strip detector has 10 barrel layers and 12 endcap layers on either side. The inner layers TIB, TID provides a resolution in the range of 13-38 μm in the r-ϕ plane, while the outer layers, TOB, TEC delivers a resolution of 18-47 μm. The strip tracker has 15148 silicon modules that cover an area of around 198 m^2 and has 9.3 million channels. Figure 3.5 shows the total thickness of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity. More details about the CMS tracking system are provided in reference [33].

The inner pixel detector was upgraded in early 2017 to account for radiation damage to the pixels and to cope with the increased instantaneous luminosity [34]. The upgrade is referred to as the Phase-1 upgrade of the pixel detector. The upgraded detector has one extra layer of pixels in the barrel and an extra disk in the forward endcaps. The new barrel layers are placed at 2.9 cm, 6.8 cm, 10.9 cm, and 16.0 cm from the beam axis. The

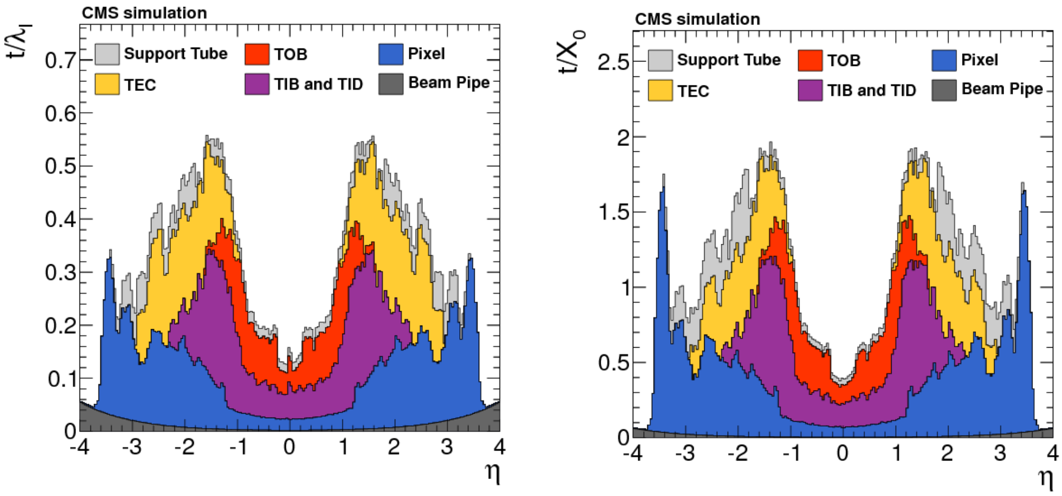


FIGURE 3.5: Total thickness of the tracker material traversed by a particle produced at the nominal interaction point, as a function of pseudorapidity, expressed in units of radiation length (left) and nuclear interaction length(right) [32].

endcap disks are placed at 3.2 cm, 3.9 cm, and 4.8 cm on both z directions. The upgraded pixel detector was used to collect the data in 2017. The updated detector helps increase the tracking efficiency, reduce the fake rate, and extend the lifetime of the detector. This leads to a better performance in reconstructing the leptons (electrons, muon, and taus), photons, b-jets, and better resolution of the missing transverse momentum. Figure 3.6 shows the efficiency and fake rates for the upgraded pixel detector.

3.2.2 The Electromagnetic Calorimeter

The CMS Electromagnetic Calorimeter (ECAL) is designed to measure the energy of particles predominantly deposited via electromagnetic interactions. The design goal of the CMS ECAL was to provide fast measurements of the energy deposits with fine granularity while also being radiation-resistant. The lead tungstate ($PbWO_4$) crystals with high density (8.28 g/cm^3), short radiation length (0.89 cm), and small Moliere radius (2.2 cm) used in ECAL satisfy all the design goals. ECAL has 61200 trapezoidal lead tungstate crystals mounted in the barrel, and 7324 crystals in each of the endcaps. The layout of the CMS ECAL is shown in the Figuer 3.7. ECAL is constructed out of single lead tungstate crystals, which provides an excellent energy resolution.

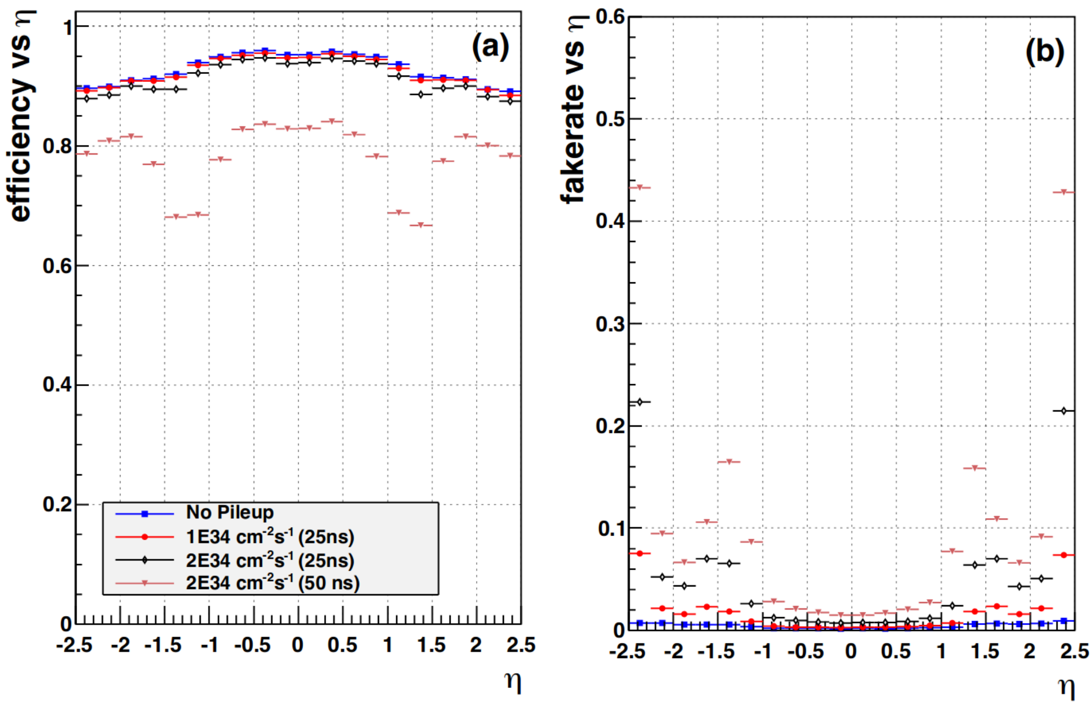


FIGURE 3.6: The efficiency (left) and the fake rates (right) for the upgraded pixel detector for various pileup scenario calculated in the simulated $t\bar{t}$ events. [34].

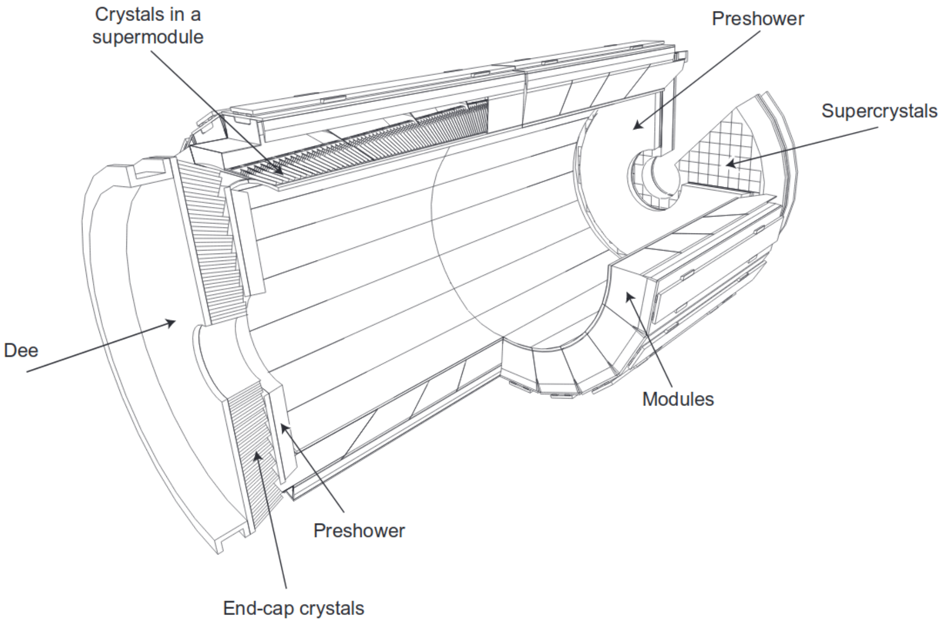


FIGURE 3.7: Layout of the CMS electromagnetic calorimeter showing the arrangement of crystal modules, supermodules and endcaps, with the preshower in front [32].

There is a pre-shower detector installed in the endcaps within the fiducial region $1.653 < |\eta| < 2.6$. In the endcaps, the radiation is very high, and ECAL is not efficient at distinguishing neutral pions and photons. The primary aim of the CMS pre-shower detector is to identify neutral pions in the high occupancy endcap regions. The preshower is a sampling calorimeter with two layers and a total thickness of 20 cm. The layers consist of lead radiators that initiate the electromagnetic showers and silicon strip sensors, which measure the energy and the shower profile of incoming particles.

3.2.3 The Hadron Calorimeter

Precise measurements of the hadronic jet properties are essential in fulfilling the goals of the CMS physics program. Additionally, precise measurement of missing transverse energy [31] arising from the neutrinos or other exotic particles is also equally vital. The CMS Hadron Calorimeter (HCAL) is designed to measure the energies and directions of quark, gluon, and neutrino by measuring the energy and direction of hadron jets [35]. The HCAL is composed of three sub-detectors, HCAL Barrel (HB), the HCAL endcaps (HE), and the HCAL forward detector (HF). These sub-detectors cover region up to $|\eta| < 1.3$, $1.3 < |\eta| < 3.0$, and $3.0 < |\eta| < 5.2$ respectively. In order to have a larger amount of material needed to contain the hadron shower than the available volume in the barrel region, an additional outer HCAL (HO) is placed outside of the solenoid behind the magnet. A schematic drawing of the HCAL can be seen in Figure 3.8

Copper alloy and stainless steel are used to build the sampling calorimeters, HB and HE. They are placed in the 3.8 Tesla magnetic field of the CMS magnet. Plastic scintillators of the thickness of a few mm are placed between the copper. The plastic acts as an active material, while the copper plates act as the absorber. The barrel calorimeter has a depth of 79 cm. The innermost and outermost plates are 7cm thick and are made of stainless steel for structural integrity. The granularity of the HB and HE calorimeters depending on the η coverage are $\Delta\eta\times\Delta\phi = 0.087\times0.087$ for for $|\eta| < 1.6$ and $\Delta\eta\times\Delta\phi = 0.17\times0.15$ for $|\eta| \geq 1.6$.

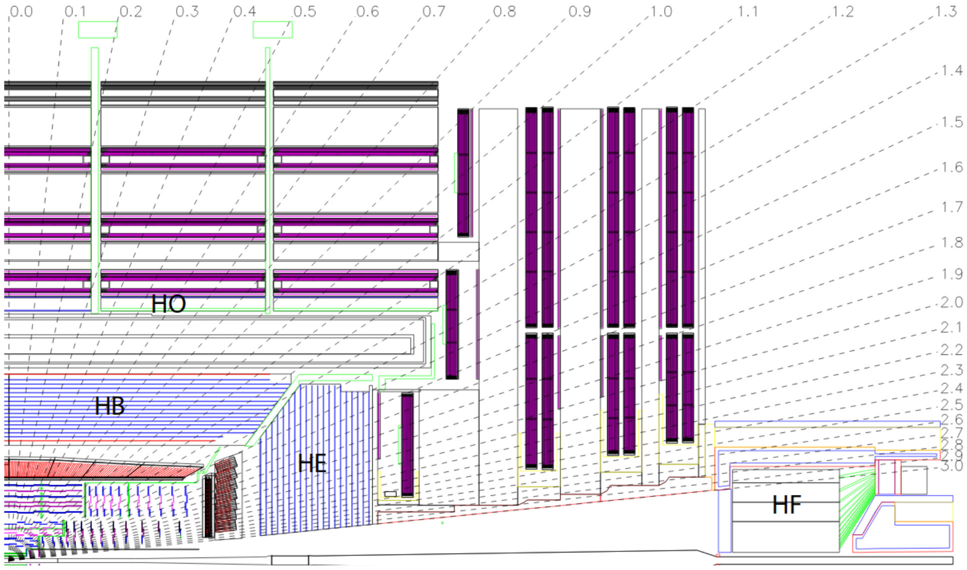


FIGURE 3.8: A schematic drawing of the CMS HCAL in the r-z plane. The dashed lines denote different values of pseudorapidity [32].

3.2.4 Muon Systems

To fulfill the discovery potential of the CMS detector, robust, precise, and efficient detection of muons is critical [36]. The muon system has a hermetic coverage over the pseudorapidity range up to ±2.4. It consists of three different technologies to detect and measure the trajectories of muons. The drift tubes (DT) are in the barrel region and extend up to 1.3 in $|\eta|$, cathode strip chambers (CSC) are in the endcap region cover from 0.9 to 2.4 in $|\eta|$, and resistive-plate chambers (RPC) are in both the barrel and endcap and cover the region from 0.0 to 2.1 in $|\eta|$. An overview of the muon system can be seen in Figure 3.9.

There are four stations in the barrel of the muon systems. The inner and the outer faces of the yoke of the magnet house two of these systems, while the other two stations are inside the steel yoke. The coordinates of the muon candidates in the $r - \phi$ plane and z-direction are provided by the first three stations consisting of eight chambers each. The endcap muon system houses four muon stations (ME1, ME2, ME3, ME4) of CSCs. These are designed to operate at very high speeds in the presence of a large magnetic field. A continuous ϕ coverage is achieved, arranging CSCs with a trapezoidal shape in an overlapping manner. The anode wires placed radially in each CSC measure the radial

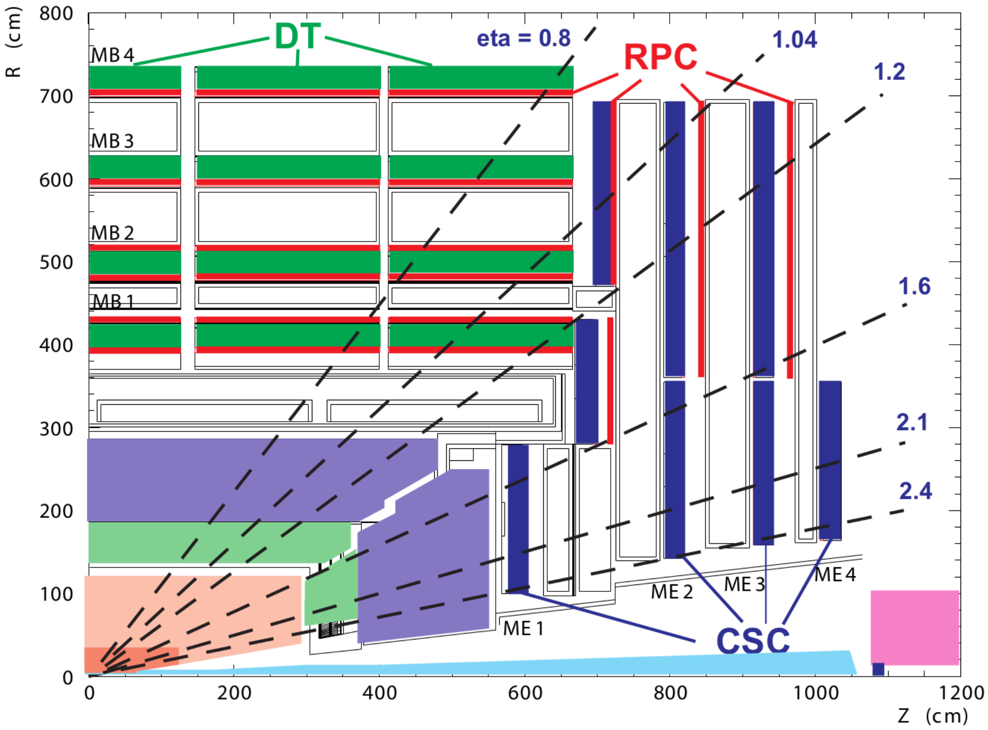


FIGURE 3.9: A view of the Muon systems of the CMS detector [32].

coordinate for tracks. The cathodes run lengthwise at a constant $\Delta\phi$ width. Additional RPC detectors are installed in both barrel and endcap muon systems. The RPCs are parallel-plate chambers filled with gas and are capable of good spatial resolution with excellent time resolution. The excellent time resolution of the RPC detector is utilized in quickly identifying the signal and to have complementary sensitivity to the cosmics.

3.2.5 Trigger and Data Acquisition System

The LHC collides proton bunches every 25 ns (40 MHz collision frequency), corresponding to $\sim 10^9$ interactions per second. Every second 40TB of storage required to store all the interaction events. It is impossible to record events at such a high rate with the existing technology. Moreover, most of these interactions are of very low energy and are unlikely to have interesting physics. CMS employs a two-level trigger system to filter out the interesting events and reduce the number of events to be stored to $\sim$100 interaction every second. The two-step are called Level-1 (L1) [37, 38] and High-Level Trigger (HLT) [39].

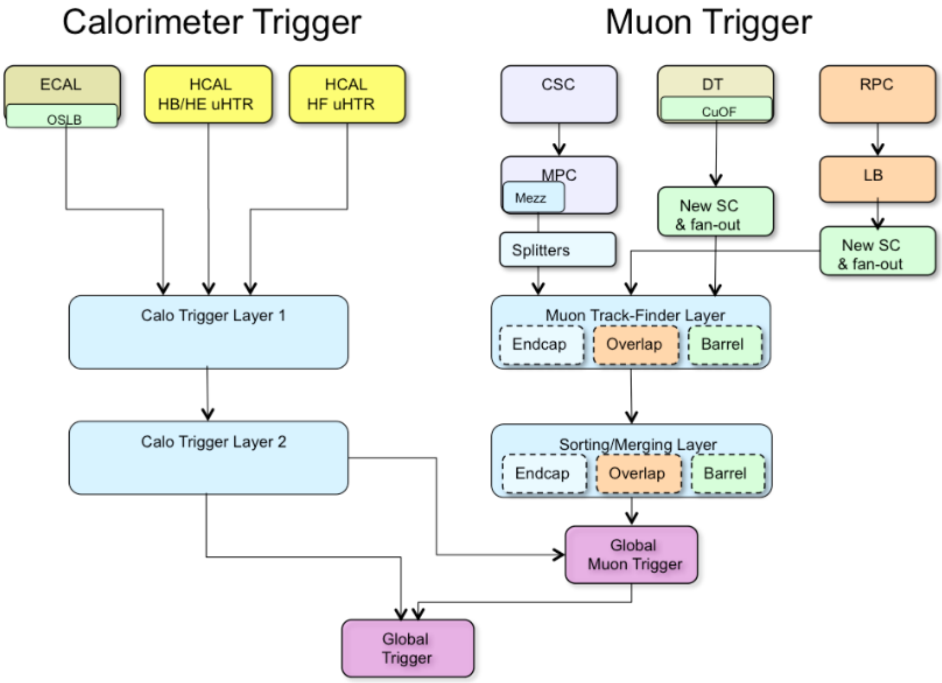


FIGURE 3.10: A schematic diagram of the L1 trigger logic is shown.

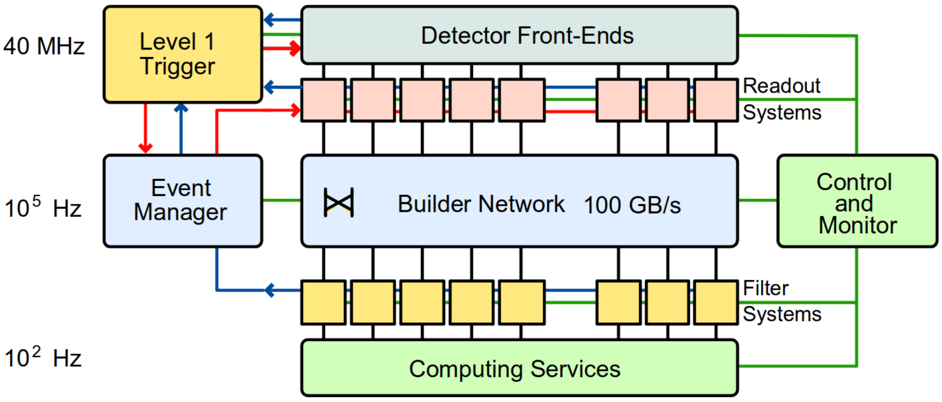


FIGURE 3.11: Architecture of the DAQ system used in the CMS experiment is shown [32].

The Level-1 Trigger consists of custom-designed, largely programmable electronics, whereas the HLT is a software system implemented in a filter farm of about one thousand commercial processors. The rate reduction capability is designed to be at least a factor of 10^6 for the combined L1 Trigger and HLT. The design output rate limit of the L1 Trigger is 100 kHz. The L1 Trigger uses coarsely segmented data from the calorimeters and the muon system to construct objects such as electrons, muons, jets, and missing energy. These reconstruct objects are referred to as trigger primitives (TPs). During L1 decision making buffer memories in the front-end electronics hold high-resolution data. A regional trigger system uses the TPs to form trigger objects on which the global trigger (GT) decision is based. Trigger objects are then sorted based on the quality of reconstruction, energy, or momentum of TPs. The highest-ranked trigger objects are passed on to the GT. GT uses a total of 128 programmable algorithms to evaluate these trigger objects. If any of the algorithms accept the event, then the event is passed on to HLT via CMS Data Acquisition (DAQ) system. A schematic of the L1 trigger system is provided in Figure 3.10. A schematic of the DAQ system is shown in Figure 3.11 [40]. The event data is extracted from the buffers in the front-end electronics. Front-End Drivers (FEDs) send this data to the DAQ system. The event builder takes the accepted L1 information and combines the different fragments into an event. The event filter, a computing farm of about 1000 nodes, takes the event and performs the Data Quality Monitoring (DQM). The event is then sent to the HLT system. HLT has access to the complete read-out data and has a processing delay time of about 100 ms. HLT can perform complex calculations and construct more realistic objects, similar to those made in the off-line analysis software if required for especially interesting events. Where possible, the L1 Trigger hardware is implemented in a field programmable gate array (FPGA) technology, allowing much flexibility in implementation. In certain cases where speed, density, and radiation resistance requirements are important, programmable memory lookup tables (LUT) are also commonly used.

Chapter 4

Monte Carlo Simulation and Identification of Physics Objects

The CMS detector is designed to identify various particles passing through it. In this chapter, the techniques used in reconstructing and identifying these particles are described. All the particles with electric charge passing through the silicon strip and pixels produce a signal in these detectors. Electrons and photons deposit most of their energy via electromagnetic interactions in the ECAL. Charged and the neutral hadrons deposit most of their energy via strong interaction in the HCAL.

Various subdetectors of the CMS experiment can independently reconstruct the physics objects, such as the muon system could deliver the muon candidates, ECAL can provide electrons/photon candidates. CMS also employes an integrated approach, called Particle Flow (PF), of combing the information from various subdetectors to coherently reconstruct and identify all the individual particles. Composite objects such as hadronic taus, jets, MET are then reconstructed using PF objects. Before discussing the object reconstruction, lets first discuss how the collision events are simulated in CMS. Once the simulation and digitization process is complete, the reconstruction of the objects is the same for both data and simulated samples.

In the next few sections, some details of the CMS simulation are discussed. After the simulation, object identification and reconstruction techniques are described.

4.1 Monte Carlo Simulation

Simulation techniques are used to mimic the detector environment to understand the experimental conditions and performance of the detector. Backgrounds from the SM processes for a BSM search could also be modeled using simulation techniques. We can interpret the results of the proton-proton collisions by comparing the simulation to the experimental measurements. Since we want to replicate the randomness of nature in the simulations, Monte Carlo (MC) simulation techniques are employed. Various MC techniques are used to model many processes, such as the production of the hard interaction and pileup, and transport of the particles produced through the detector and response of the detector for different particles having different energy and momentum.

The simulation chain starts from a theorist writing down the Lagrangian for a particular theory under consideration. This Lagrangian is used to derive various couplings decay widths of that model. A universal file format called UFO is used to store the model files. We use the UFO files in the event generators to produce the events. There are four main parts of connecting the simulation to the experimental measurements, event generation, simulation, digitization, and reconstruction. The simulation process is described using the production of the vectorlike lepton signal samples as an example below.

4.1.1 Event generation

The process of producing the signal samples starts from the model files provided by the authors of a given model. These files are used in an event generator like MadGraph 5_aMC@NLO v5.2.2 to produce the grid packs for further processing [41, 42]. The grid pack production is especially important in the case of large simulation samples, use of the grid packs allows us to utilize the grid computing resources fully. Madgraph uses the parton distribution functions (PDF) to pick a parton for interaction from each proton beam. These partons then interact according to the couplings and interaction strength provided in the model to produce the hard interaction vertex. This interaction yields a set of final state particles, and these particles may not be color singlets under QCD interaction. PYTHIA 8 takes the particles provided by madgraph and does fragmentation and hadronization to make color singlet states [43]. It also performs the initial state radiation (ISR) and final state radiation (FSR) to mimic the actual parton interaction.

4.1.2 Simulation

The particles produced at the interaction point travel through the detector and deposit their energy in various detector modules. In the simulation, the particles provided from the event generation step are transported through the detector. The interaction of the particles with the detector material is modeled using GEANT4 toolkit [44]. GEANT4 describes the energy loss of various particles as they interact with the material in the detector.

The GEANT4 based simulation (Full Simulation) of the CMS detector is used for the backgrounds as well as most signal samples. However, for some huge samples, a faster version of the simulation is desired. In the next section, a faster simulation technique “Fast Simulation” is discussed.

4.1.3 Fast Simulation

The GEANT4 based simulation is generally quite slow for complicated processes. It takes about 100 seconds to process one $t\bar{t}$ like event. Models beyond SM generally have some free parameters; in a search for BSM physics, a range of these parameters needs to be tested against the data. This means that a large number of simulation samples are needed to make a grid of samples corresponding to variation in the free parameters. To produce these large simulation samples, huge computing resources are needed. This problem is mitigated by developing a speedy model to do detector simulation by making some simplifying assumptions. This fast model of the detector simulation is referred to as Fast Simulation. Below are some of the key differences/similarities of Fast Simulation:

- Complete detector geometry.
- Simplified analytic detector responses.
- Infinitesmall thin detector layers.
- Complete emulation of trigger and electronics.

CMS Fast simulation is built around these assumptions, and the resultant software is speedy compared to the Full Simulation. The fast simulation takes about 5 seconds to process one $t\bar{t}$ like event. Many of the signal processes, especially in for the SUSY models are generated using Fast Simulation. Extensive validation tests of Fast Simulation are performed against Full Simulation to judge its accuracy in simulating the different processes. Figure 4.1 shows the efficiency of reconstructing tracks and muons in Fast and Full Simulation for $t\bar{t}$ events [45]. Fast simulation is adequate to model high level quanties such as, lepton p_T, p_T^{miss} but there are notable difference in lower level quantities like track reconstruction efficiency. An effort to correct for the discrepancies of the fast simulation is ongoing in CMS.

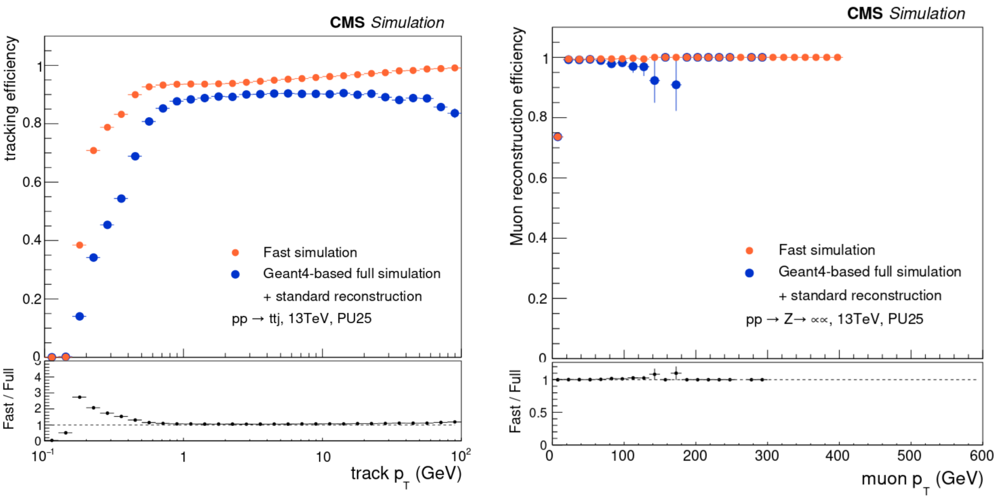


FIGURE 4.1: Comparison of reconstruction efficiencies for tracks (left) and muons (right) in Fastsim and Fullsim.

4.1.4 Digitization and Reconstruction

After having the energy deposits in various detector components, the raw signals are digitized by backend electronics (data)/digitizer software(MC), and hits are constructed. A hit refers to the energy deposit above some predefined threshold in a given detector module. These hits are then used by the reconstruction algorithms to reconstruct the physics objects.

Once the simulation chain is finished, the next step is to reconstruct and identify the physical objects of interest. In the following few sections, reconstruction and identification of physics objects are discussed.

4.2 Tracks and Vertices

As the charged particles traverse the tracker volume, they leave an electromagnetic signature in the detector. These electronic signals are used to reconstruct the tracks of the charged particles. The charged particle tracks are used in the reconstruction of advanced objects such as electrons, muons, charged hadrons, and also used in reconstructing interaction vertices. CMS tracker and the algorithms are designed to efficiently reconstruct the tracks of charged particles in the p_T range of 0.1-1000 GeV. The tracker volume near the interaction point has a very high occupancy of hits; the goal of the track reconstruction is to correctly identify the tracks of charged particles with a low misidentification rate.

The tracking in CMS could be divided into 5 logical steps: hit reconstruction, seed generation, pattern recognition, ambiguity resolution, and final track fit [46]. The signals from pixel and strip are used to build hits. The raw signals are zero-suppressed and are clustered in a local reconstruction. The clusters satisfying a predefined signal-to-noise ratio threshold in pixel and strip detectors are used to determine the initial hit position. Tracks using pixel hits are reconstructed and are used by beam spot fitter [47] to determine approximate location and size of the beam spot.

The magnetic field inside the detector volume is quite uniform; using a quasi-uniform magnetic field it is possible to estimate the initial track trajectory and the uncertainties assuming particles traverse helical paths in the detector. Seeds for the track reconstruction are generated using this approach with utilizing up to 3 hits and beam spot constraints. After this step, an iterative tracking technique with multiple levels is used to reconstruct tracks from the seeds. In the iterative tracking, high momentum tracks originating from interaction point are reconstructed first. Then the hits used in this step are not considered for further trajectory building. A Combinatorial Track Finder (CTF) algorithm processes the hits starting from the seed layer searching outwards in an attempt to improve the trajectory and momentum measurements by adding more hits to the reconstructed tracks. Particle trajectories are corrected for possible inhomogeneities in the magnetic field as well as the effects of energy loss due to ionization and multiple Coulomb scattering [33, 48].

After the iterative pattern recognition step, any hit or seed ambiguities the track candidates have are resolved using the information on shared hit fraction to prevent double counting. The resulting track candidates are fitted again using a least-squares approach with relaxed seed constraints to eliminate any biases that might arise from the initial coarse trajectory estimates. The reconstructed tracks, together with the pixel-based beam spot measurement, are used to reconstruct the primary vertices in the event, including those due to pileup collisions. For vertex reconstruction, tracks are required to satisfy specific quality requirements on their transverse impact parameter, the number of the strip and pixel hits, and the normalized χ^2 of their trajectory fit, and then clustered along the beam axis and fitted with the adaptive vertex fitter [49].

4.3 Muons

The CMS detector is very efficient in reconstructing the muons coming out of the proton collisions. Muons produce tracks in the inner detector and the muon detectors. Both tracker and the muon detectors are capable of reconstructing the muons and giving a precise measurement of the muon momentum. Depending on the detector system used in the reconstruction, three types of muon candidates have been defined: the standalone muons, the tracker muons, and the global muons.

The standalone techniques solely rely on the information provided by DT and CSC detectors. RPC detectors give additional information in the barrel endcap overlap region. The reconstruction starts from the track segments produced by local reconstruction in the DT and CSC chambers. In each DT chamber, triplets of hits in the SLs are used as seeds and are grown into linear segments with additional hits, consistent with the beam spot location and size. The information from strips and wire-bundles in the CSS is combined to construct 2-dimensional hits in each chamber layer. A series of at least four hits consistent with a linear fit with a maximum $r - \phi$ spread of 1 cm are retained as segments. Unlike the DT and CSC systems, local RPC reconstruction yields hit locations instead of segments. These are calculated as the geometric centers of the signal carrying strip clusters in the RPC detector plane. Following the local reconstruction step, the innermost DT and CSC segments are treated as seeds for the standalone muon trajectory reconstruction. This step is performed with the Kalman filter technique taking into account the effects of energy losses in the material, multiple scattering, and non-homogenous magnetic fields. The reconstruction is first performed inside-out, and then outside-in with the added inclusion of the beam spot constraint, yielding the final standalone muon candidates.

Tracks in the tracker with $p_T > 0.5$ GeV and $p_T < 2.5$ GeV are projected to the muon systems, if the track matches to a DT or a CSC segment then the track is considered as a tracker muon candidate.

Global muons are constructed by the combination of tracker tracks with the standalone muon tracks obtained in the muon systems. The matching is performed by projecting the standalone muon trajectories to the outer tracker surface, and the global muon track is accomplished via a re-fit of these hits with the Kalman filter technique. This is especially important for TeV range muons. The global muon reconstruction achieves a better p_T resolution in comparison to tracker-only measurements for muons with $p_T > 200$ GeV. Figure 4.2 shows the RMS of relative q/p_T as a function of cosmic rays p_T, q is charge of the muon [50].

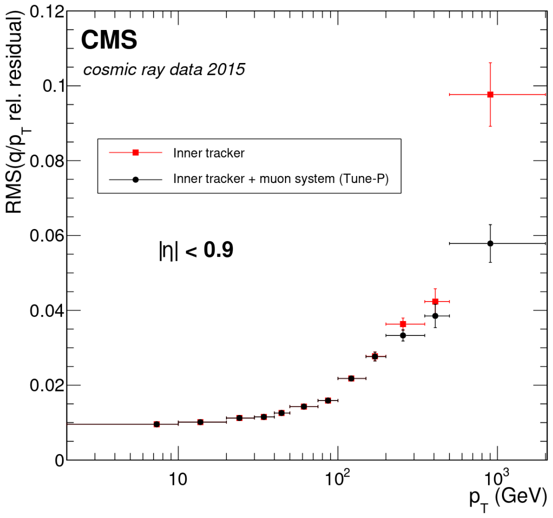


FIGURE 4.2: Figure shows the RMS of relative q/p_T as a function of cosmic rays p_T, q is charge of the muon. Inner track only fit is shown by squars and by clircles including the muon systems. [50].

4.4 Electrons

The electrons being electrically charged produce a track in the inner silicon tracker. The electron identification algorithm takes this information from the tracker and matches the energy deposits in the ECAL. Electrons deposit most of their energy in the ECAL, but the bremsstrahlung radiation in the tracker volume plays a significant role in the electron reconstruction [51]. As the electrons travel through tracker in the magnetic field, their trajectory curve in the ϕ direction. Due to this curvature, the bremsstrahlung photons deposits are spread in ECAL according to their p_T. Usually, the electromagnetic energy deposits in ECAL for a single electron/photon are contained in a 5×5 lattices of ECAL crystals (97% of times). Due to this spread of energy in the ECAL, a combination of clusters in local $\eta - \phi$ coordinates is needed to measure the total energy of the incident electron. In the CMS, two techniques are used in ECAL to form the cluster of energy deposits for electrons. In the endcap region, clusters of individual showers are produced using a modified island (5×5) algorithm, which is then further combined along ϕ-segments to produce super-clusters. In the barrel region, a hybrid algorithm, which dynamically seeks bremsstrahlung deposits along the ϕ axis, is used. In each event, the ECAL crystal channels are read out using a Selective Readout (SR) zero-suppression algorithm, and crystals with energy deposits above a predetermined energy threshold and corresponding to local maxima are used as seeds in both clustering algorithms. While reconstructing the electron tracks in the silicon tracker, the effect of bremsstrahlung radiation also needs to be taken in to account. For a majority of charged particles, ionization losses are most important, but for electrons, bremsstrahlung radiation also plays a significant role. About 35% of the electrons lose at least 70% of their initial energy due to bremsstrahlung radiation in the tracker. These radiated photons can further convert into electron-positron pairs, yielding more charged particle tracks. A dedicated Gaussian Sum Fitter (GSF) is used to reconstruct the tracks for electron candidates in the tracker [52]. All the pixel tracks are checked for consistency in position and momentum to an ECAL super-cluster. If these tracks are found to be consistent, then the GSF track reconstruction is performed for such track candidates. This is called ECAL based seeding. The ECAL-driven seeding is quite efficient and provides good momentum measurement for the relatively high momentum electrons which are isolated. A second approach, especially targeted at the low momentum electrons, is used to recover the low momentum, non-isolated electrons. In this approach, CTF tracks with projections matching to an ECAL PF cluster, and with missing tracker hits or poor χ^2 values are considered for the GSF fit.

The final momentum of the electron candidates is estimated using information from tracker as well as ECAL, for the low momentum electrons ($\mathrm{p_T} < 15\ GeV$) the tracker measurements are more precise due to low bremsstrahlung radiation, while for the high momentum electrons the ECAL energy measurements play the leading role. Figure 4.3 shows the effective resolution of electron momentum by using the information of tracker and the ECAL [53].

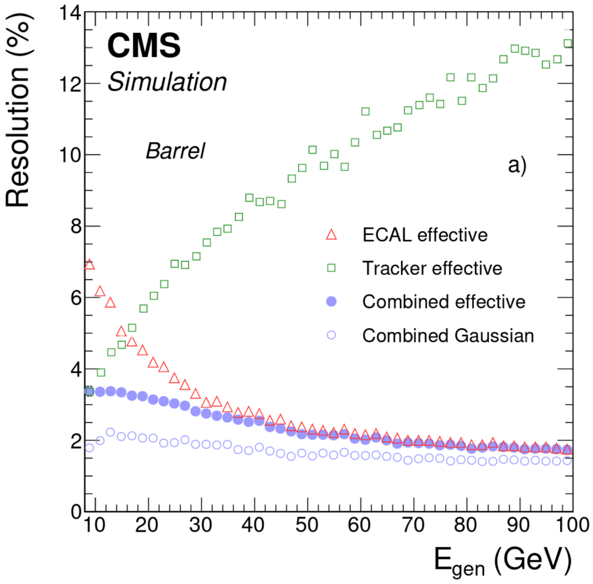


Figure 4.3: Figure shows the effective resolution of electron momentum by using the information of tracker and the ECAL [53].

4.5 Taus

The τ lepton is the heaviest lepton in the standard model. It is heavy enough to decay to leptons as well as hadrons and neutrinos. The leptonic decay of τ leptons to electron and muon are accompanied by the two neutrinos. These are denoted as τ_e and τ_μ respectively. The electrons and muons are reconstructed as described in the previous sections. All other decay of the τ lepton almost always contain the hadrons with a combination of charged and neutral mesons and a τ neutrino (ν_τ). The hadronic

Decay Mode	Resonance	B (%)	Reconstruction
Leptonic decays		35.2	
$\tau^- \to e^- \overline{\nu}_e \nu_\tau$		17.8	e^-
$\tau^- \to \mu^- \overline{\nu}_\mu \nu_\tau$		17.4	μ^-
Hadronic decays		64.8	
$\tau^- \to h^- \nu_\tau$		11.5	1-prong
$\tau^- \to h^- \pi_0 \nu_\tau$	$\rho\{770\}$	25.9	1-prong
$\tau^- \to h^- \pi_0 \pi_0 \nu_\tau$	$a_1\{1260\}$	9.5	1-prong
$\tau^- \to h^- h^+ h^- \nu_\tau$	$a_1\{1260\}$	9.8	3-prong
$\tau^- \to h^- h^+ h^- \pi_0 \nu_\tau$		4.8	Not targeted
Other		3.3	Not targeted

TABLE 4.1: Tau lepton decay modes and their branching fractions are provided. Also listed the corresponding reconstruction modes. The charged hadrons are denoted by the symbol $h^{\pm}$.

decay of the τ lepton is denoted as τ_h. Decay modes with one charged hadron are called 1-prong decays, while the modes with three charged hadrons are called 3-prong decay modes. The detailed description of the τ lepton decay modes and corresponding branching fractions are provided in table 4.1.

The τ_h candidates at CMS are reconstructed using the hadrons-plus-strips(HPS) algorithm [54, 55]. In the HPS algorithm charged hadrons ($h^{\pm}$) and the neutral pions (π_0) used to reconstructed the τ_h object. The charged hadrons are reconstructed using their associated charged particle tracks in the silicon tracker. Most of the time, the π_0 converts to a pair of photons in the tracker region, and the emitted photons can then convert into the electron-positron pair in the ECAL. These electron and photon candidates are clustered in rectangular regions of pseudorapidity and azimuthal angle, $\eta \times \phi$ to obtain the π_0 candidates. These rectangular regions in the calorimeter plane are called “strips”.

At the LHC main challenge in efficiently identifying the τ_h, is to distinguish them from the quark and gluon jet. The technique used to separate the τ_h candidates from the hadronic jets are discussed later in this chapter.

4.6 Jets and Missing Transverse Momentum

The quarks and gluons coming out of proton-proton collision go through fragmentation and hadronization to form color singlet states. These states can further create additional quark-antiquark pairs via gluon emission. This creates a collection of color neutral particles traveling in a narrow cone in the detector. This collection of particles, ideally originating from a single quark/gluon, is referred to as a jet. Charged hadrons carry most of the energy in a typical jet ($\sim$ 60%), photons (coming from the decay of π^0) carry $\sim$ 25% while neutral hadrons carry about $\sim$ 10%. There are many techniques to associate the detector signatures of the particles by forming a jet.

The algorithms used in the formation of jets should be infrared and collinear safe (IRC) [56]. The IRC safe algorithms are immune to the emission of arbitrarily soft (low energy) and collinear particles resulting in measurements that can be compared to fixed-order perturbative QCD calculations. CMS uses a number of IRC safe algorithms to reconstruct jets of various radiuses such as, the SISCone [57], Cambridge/Aachen [58], k_T [59, 60] and anti-k_T [61].

$$\begin{aligned} d_{ij} = min\{1/k_{T,i}^2, 1/k_{T,j}^2\}\frac{(n_i - n_j)^2 + (\phi_i - \phi_j)^2}{R^2} \\ d_{iB} = 1/k_{T,i}^2 \end{aligned} \tag{4.1}$$

Jets clustered using anti-k_T algorithm with radius parameter R=0.4 are used in this thesis. The anti-k_T algorithm computes two distances for clustering the individual particles. These are given in Eq 4.1 as d_{iB} (distance between i^{th} object and beamline) and d_{ij} (distance between i^{th} and j^{th} object). Starting from a given i^{th} object, the combination algorithm looks for a j^{th} object such that $d_{ij} < d_{iB}$, if such an object is found then the algorithm combines these two objects by adding the individual four vectors. If no additional object satisfying $d_{ij} < d_{iB}$ is found, then the i^{th} object is considered as a jet and is removed from the further iterations. These resulting jets are corrected for variation in energy response as a function of $\mathrm{p_T}$ and η, discrepancies in data and simulations comparisons, and the effects of pileup [62].

$$\vec{p}_T^{\,miss} = -\sum_i^{PF} p_{T,i} \tag{4.2}$$

Neutrinos and other possible weakly interacting BSM particles escape the detector without leaving any signature. The presence of such particles is inferred by constructing a quantity called missing transverse momentum ($\mathrm{p_T^{miss}}$). The $\mathrm{p_T^{miss}}$ is also labeled as E_T^{miss}

throughout this thesis. The p_T^{miss} is defined as the negative vector sum of the all the PF candidates in the event as in Eq. 4.2. The definition of the p_T^{miss} for the vectorlike lepton search performed using data collected in 2017 has been modified to mitigate noise effects related to the aging of the CMS ECAL. For 2017 dataset the PF jet candidates with $2.5 < |\eta| < 3.0$ and $p_T < 70$ GeV are excluded from the sum in Eq.4.2. The impact of this change is shown in Fig. 4.4, 4.5 for dilepton event selection enriched in Drell-Yan (DY) + jets events.

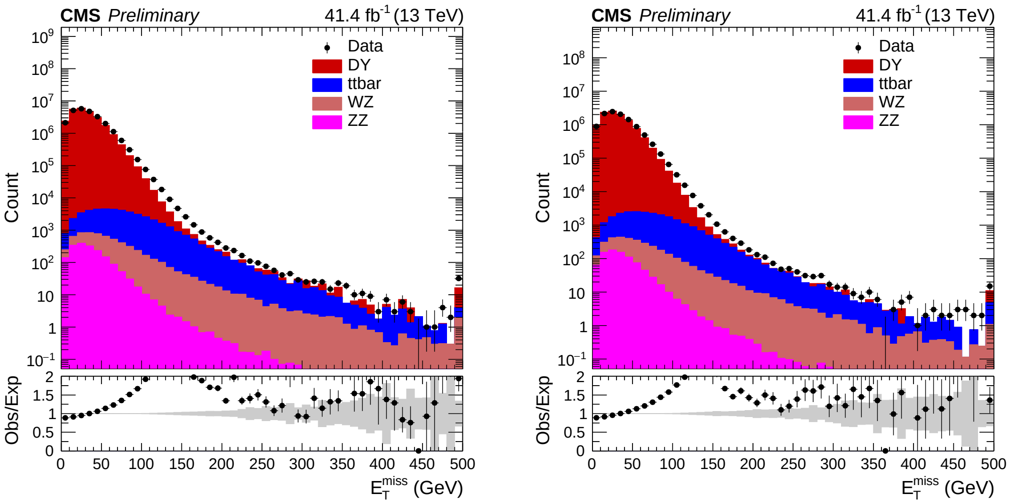


FIGURE 4.4: Distributions of E_T^{miss} in di-muon (left) and di-electron (right) region when dilepton mass is on-Z.

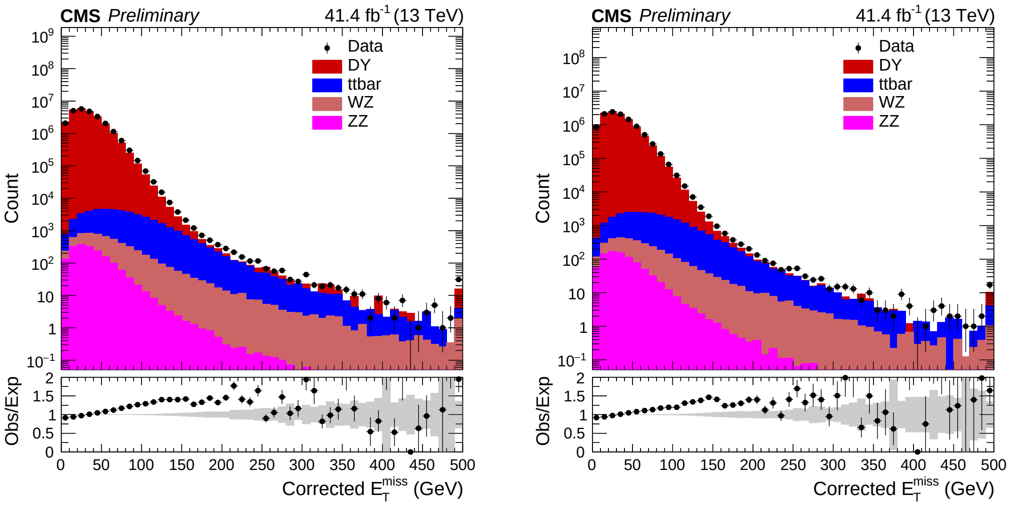


FIGURE 4.5: Distributions of E_T^{miss} in di-muon (left) and di-electron (right) region when dilepton mass is on-Z after correcting the E_T^{miss}.

A much better agreement between data and standard model predictions is observed by using a modified p_T^{miss} definition. The remaining discrepancy in Fig. 4.5 is covered by the corresponding uncertainties on the SM prediction. These uncertainties are discussed in Chapter 5.

4.7 Leptons Discriminations and Selections

The search for new physics relies on the identifications of particles produced in the hard interaction. The leptons originating from the decays of the bosons (W, Z, and Higgs) or new exotic particles produced in hard interaction, are generally isolated, i.e., the additional hadronic activity around them is minimal. This property of these leptons could be utilized to identify such leptons efficiently. In the CMS experiment, we define various isolation quantities in a cone around leptons separately for electron, muon, and taus. The main challenge in defining the isolation variables is to account for the impact of the additional pileup particles present in the isolation cone. For the pileup correction due to charged hadrons, the information from the CMS tracker is used. The charged particles associated with pileup vertices are removed from the isolation sum. For the neutral particles coming from pileup vertices, the tracker information is not available. Average charged to neutral particle ratio called $\Delta\beta$ is used to remove neutral pileup from isolation sum effectively. Other useful parameters in differentiating the real vs. a lepton from within the jets are the transverse (d_{xy}) and longitudinal d_z impact parameters.

Electrons with $\mathrm{p_T} > 20$ GeV and $|\eta| < 2.5$, and muons with $\mathrm{p_T} > 20$ GeV and $|\eta| < 2.4$ are used in the search for the vectorlike leptons. Muons are required to satisfy certain set of idendification requirements summarized in Table 4.2 [63]. The selection corresponds to about 99% efficiency for real muons. Additionally, muon candidates are required to statisfy $\Delta\beta$ corrected relative PF-based isolation in a cone of $\Delta R < 0.4$ around muon candidate to be $< 15\%$. Muon candidates are also required to satisfy $|d_z| < 0.1$ cm and $|d_{xy}| < 0.05$ cm. A loose muon selection is defined where $\Delta\beta$ corrected PF-based isolation requirement is relaxed to $< 100\%$ and the muon ID selections are relaxed.

Table 4.3 summarizes the electron selections [64], corresponding to a efficiency of about 80%. A loose selection is also defined for electrons, where the selections are relaxed. The loose WP corresponds to an efficiency of about 90%.

Muon Variable	Selection
Global Muon	yes
normalized global track χ^2	< 3
Tracker standalone position match	< 12
kick finder quality	< 20
Segment compatibility	> 0.451
d_z (c. m.)	< 0.1
d_{xy} (c. m.)	< 0.05

Table 4.2: Muon ID selections.

Electron Variable	Selection	
	Barrel ($\lvert\eta\rvert \leq$1.479)	Endcap ($\lvert\eta\rvert \geq$1.479)
$\sigma_{i\eta i\eta}$	< 0.00998	< 0.0298
$\lvert\Delta\eta_{seed}\rvert$	< 0.00311	< 0.00609
$\lvert\Delta\phi_{in}\rvert$	< 0.103	< 0.045
H/E	< 0.253	< 0.0878
$\lvert 1/E - 1/P\rvert$	< 0.134	< 0.13
missing inner hits	≤ 1	≤ 1
conversion veto	yes	yes
Relative combined PF isolation with effective area correction	< 0.0695	< 0.0821
d_z (c. m.)	< 0.10	< 0.20
d_{xy} (c. m.)	< 0.05	< 0.10

Table 4.3: The table summarises electron identification selections.

Tau candidates with $p_T > 20$ GeV, $|\eta| < 2.3$ are selected using hadron-plus-strips (HPS) algorithm which reconstructs tau decay to 1 or 3 charged hadrons. The tau candidates are required to pass a multivariate-based isolation discriminator with a working point corresponding to about 40% efficiency. Additionally, it is required that $d_Z < 0.2$ c.m., where d_Z is computed for the leading charged hadron candidate of the tau object. In order to remove misidentified tau candidates originating from electrons or muons, we require all candidates also to pass multivariate-based discriminators designed to distinguish tau candidates from electrons and muons. A looser tau selection for the prediction of fake tau background is also defined; for this selection, the isolation is relaxed corresponding to an efficiency of about 70% [65]. The tau ID selections are summarized in Table 4.4.

Tau Variable	Selection
decayModeFindingOldDMs	yes
byVTightIsolationMVArun2v1DBoldDMwLT	yes
againstElectronTightMVA6	yes
againstMuonTight3	yes
d_z (c. m.)	< 0.2

Table 4.4: The table summarises hadronically decaying tau lepton identification requirements.

Chapter 5

Multileptonic Signatures and Backgrounds

In this chapter, multileptonic signatures at the LHC are discussed. The chapter also describes the techniques used to estimate the standard model backgrounds for the multi-lepton final states. In this chapter, data collected by CMS detector at $\sqrt{s}$=13 TeV using single electron and single muons triggers in the year 2017 is used. The p_T threshold for the electron trigger was 35 GeV, while for muon trigger, it was 27 GeV. The collected dataset corresponds to a luminosity of 41.4 fb^{-1} recorded in the year 2017. A detailed description of the datasets used in the analysis is provided in Appendix A.1.

Multileptons region is divided into several categories as follows, events with three electrons/muons (light leptons) are referred to as 3L, events with four or more light leptons are referred to ass 4L and events with two light leptons, and at least one hadronically decaying tau are labeled as 2L1T. The 2L1T category is further divided based on the relative charge of the light lepton pair. The 2L1T events where both light leptons have the same (opposite) charge are referred to as 2L1TSS (2L1TOS).

5.1 Multileptonic Signatures at LHC

At a hadron colliders such as LHC, leptonic signatures are rare due to low electroweak production cross-section as compared to the strong production. Figure 5.1 shows the production cross-section of various SM processes as a function of the center of mass energy. The leading process with at least one lepton in the final state is the leptonic decay of W boson. The production cross-section for W boson is orders of magnitude smaller than the dijet cross-section at the LHC. The leading SM processes which yield multilepton final state (diboson process WZ, ZZ) have an even smaller cross-section. Having minimal SM background, the multilepton final states are a very clean probe to look for beyond standard model (BSM) physics. The BSM processes resulting in multiple leptons also tend to have a small cross-section; still, the multileptons signatures generally have good sensitivity to a large number of BSM models.

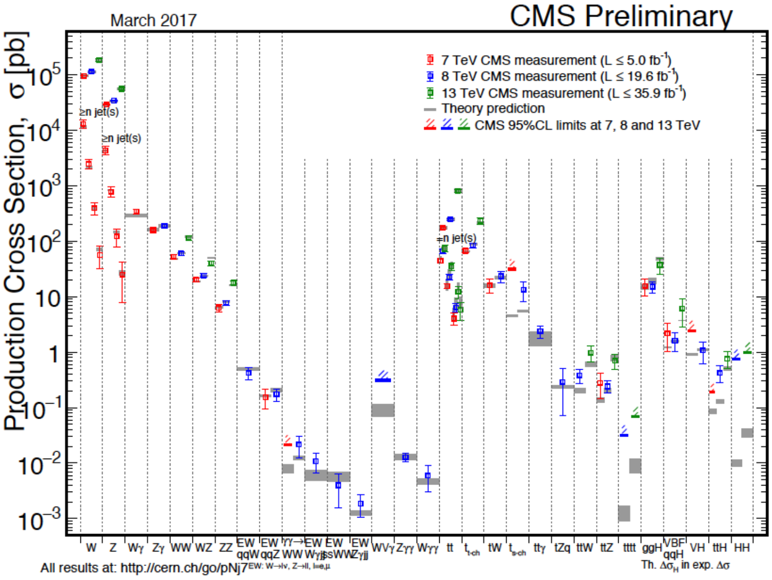


FIGURE 5.1: Production cross section for various standard model processes as a function of center of mass energy.

There are a variety of BSM models that attempt to address the deficiencies of SM. Many of them give rise to multilepton signatures, such as electroweak production of chargino and neutralino in SUSY, pair production of heavy type-III seesaw fermions, pair production of vectorlike leptons. I developed a generic framework to search for models yielding multilepton signatures. For any BSM search, more often than not, the most challenging part is the estimation of SM background. In the next section, I will describe the methods I developed to tackle this problem for a generic multilepton search.

5.2 Backgrounds for Multileptonic Signatures

There are many SM processes that can lead to multiple leptons in the final state, and these constitute the background for our search. For this search, we categorize the leptons based on their origin. The leptons that directly come from the W, Z, Higgs boson decays (or from the direct decay of BSM particles such as vectorlike tau (τ') are called prompt leptons. The leptons coming from the semileptonic hadron decays within jets or some other misidentified detector signatures are labeled as misidentified(fake) leptons. Additionally, a smaller fraction of non-prompt leptons are due to internal or external asymmetric conversions of photons, and such leptons are labeled as conversion leptons.

Based on the lepton compositions, the background for this search could be broadly categorized into two components: irreducible and reducible. The irreducible backgrounds are the processes where all the leptons under consideration are prompt, while the reducible backgrounds are the ones where at least one of the lepton is non-prompt. The non-prompt leptons are also labeled as misidentified leptons.

The leptonic decays of diboson productions (WZ, ZZ) are the primary sources of prompt backgrounds. The leptons from these decays are generally isolated and originate from the primary vertex. These processes are modeled very well using the Monte Carlo(MC) simulations. On the other hand, the non-prompt backgrounds involving leptons from hadron decays and misidentified detector signatures are challenging to model using Monte Carlo methods due to the lack of understanding of low energy QCD and the misidentified detector signatures. The leading processes contributing in this category are DY+jets and $\mathrm{t\bar{t}}$+jets, where two prompt leptons are coming from the decay of bosons, while the third lepton is non-prompt lepton arising from other activity in the event. Due to these limitations of Monte Carlo methods, data-driven techniques are employed for such processes.

Before beginning a search for BSM physics, it is important to study the properties of the physics objects (in our case, leptons) selected in the analysis. In the next section, dilepton regions are used as a cross-check to verify the object selections.

5.3 Dilepton Control Regions

In the signal regions, at least three leptons are required, given this, we can construct dilepton regions to better understand and commission the object selections. In the analysis, the data is collected using single lepton triggers; these regions are also very useful in determining if the triggers have appropriately fired.

The Z boson decays to two leptons with opposite charge in the final state, we create a selection of events enriched in Z + jets (or Drell-Yan (DY) + jets) events by requiring the mass of the opposite sign same flavor (OSSF) dilepton pair to be on-Z and $\mathrm{p_T^{miss}} <$ 50 GeV. The on-Z selection is defined by requiring the invariant mass of the lepton pair ($M_{\ell\ell}$) to be within 15 GeV from the Z mass (91.19 GeV). DY($\rightarrow ee$)+jets and DY($\rightarrow \mu\mu$)+jets processes are studied separately. To study $\mathrm{t\bar{t}}$+jets process, events having an opposite-sign eμ pair and $\mathrm{S_T} > 300$ GeV are selected. $\mathrm{S_T}$ is defined as the scalar sum of $\mathrm{p_T}$ of leptons ($\mathrm{L_T}$), $\mathrm{p_T^{miss}}$ and sum of $\mathrm{p_T}$ of jets(H_T).

Figures 5.2, 5.3 and 5.4 show the lepton $\mathrm{p_T}$, $\mathrm{L_T}$, and $M_{\ell\ell}$ distributions for the DY+jets enriched selections, and the lepton $\mathrm{p_T}$, $\mathrm{L_T}$, and $\mathrm{p_T^{miss}}$ distributions for the $\mathrm{t\bar{t}}$+jets enriched selection. The $\mathrm{L_T}$ is defined as the sum of $\mathrm{p_T}$ of leptons. All plots are normalized to data over the NLO cross-section (DY($\rightarrow$ ee): 0.98, DY($\rightarrow \mu\mu$): 1.05, $\mathrm{t\bar{t}}$: 1.07). The jet multiplicity is normalized to data in order to account for higher order effects that might not be properly estimated by NLO MC. Additional correction factor as a function of Z $\mathrm{p_T}$ (vector sum of lepton $\mathrm{p_T}$) is applied to correct for the mismodelling of Z $\mathrm{p_T}$ in MC [66].

A good agreement between data and the SM predictions is observed in all the lepton parameters. These regions have a lot of events and relatively low statistical uncertainty even in the tails of distributions, suggesting the lepton kinematics is well modeled.

5.4 Misidentified Lepton Backgrounds

The background component involving at least one non-prompt leptons are estimated via a 3-dimensional implementation of a matrix method [67]. The leading SM processes contributing to this component of the background are DY+jets, $\mathrm{t\bar{t}}$+jets, and WW+jets.

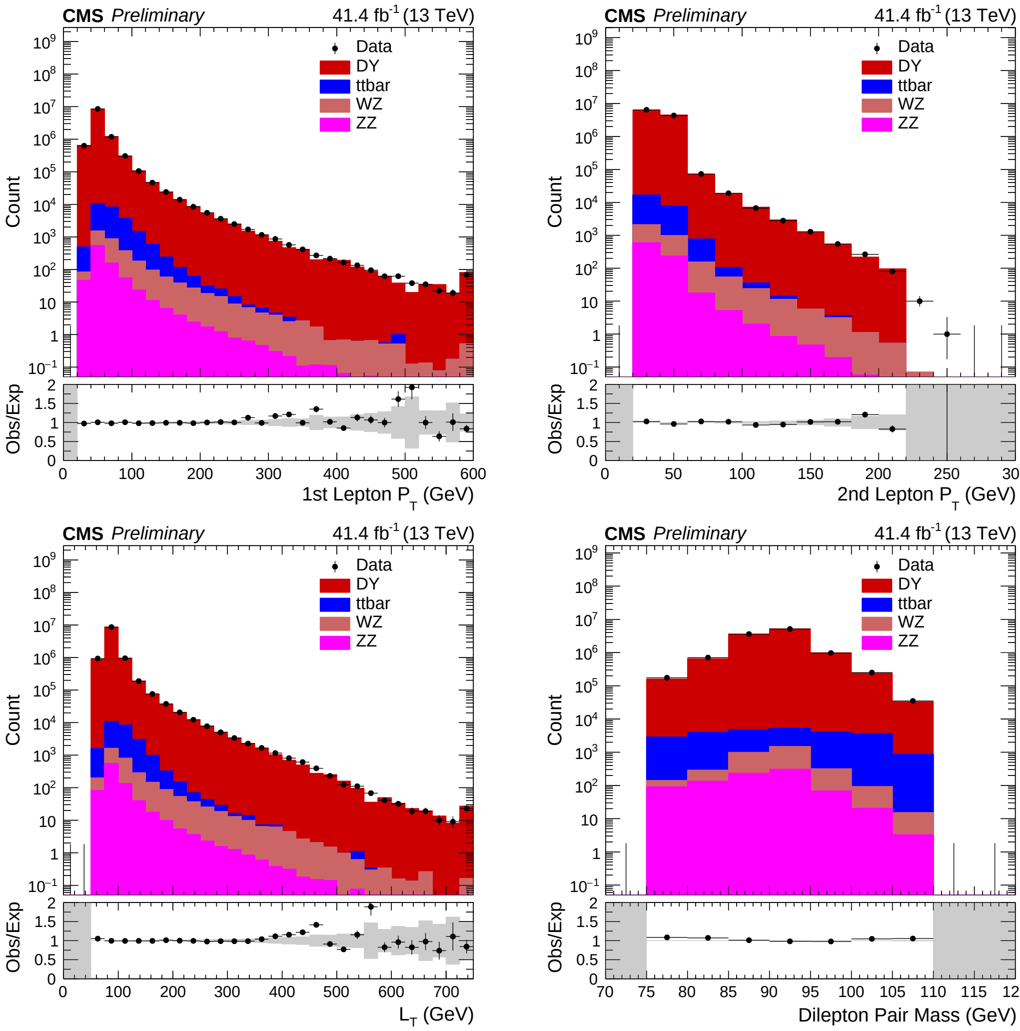


FIGURE 5.2: Leading (upper left) and subleading (upper right) electron $\mathrm{p_T}$, $\mathrm{L_T}$ (lower left) and $M_{\ell\ell}$ (lower right) distributions in the DY→ee enriched dilepton selection.

A unified method is developed to estimate all these processes together. The method also estimates this background for all lepton flavors (viz. $e/\mu/\tau$).

5.5 Matrix Method

Matrix method is a data-driven techniques that relies on the assumption that the probabilities with which prompt and fake leptons pass a tight lepton selection given that they satisfy a loose lepton selection, prompt (p) and fake (f) rates respectively, are universal and can be described as a function of the lepton and event dependent parameters. This

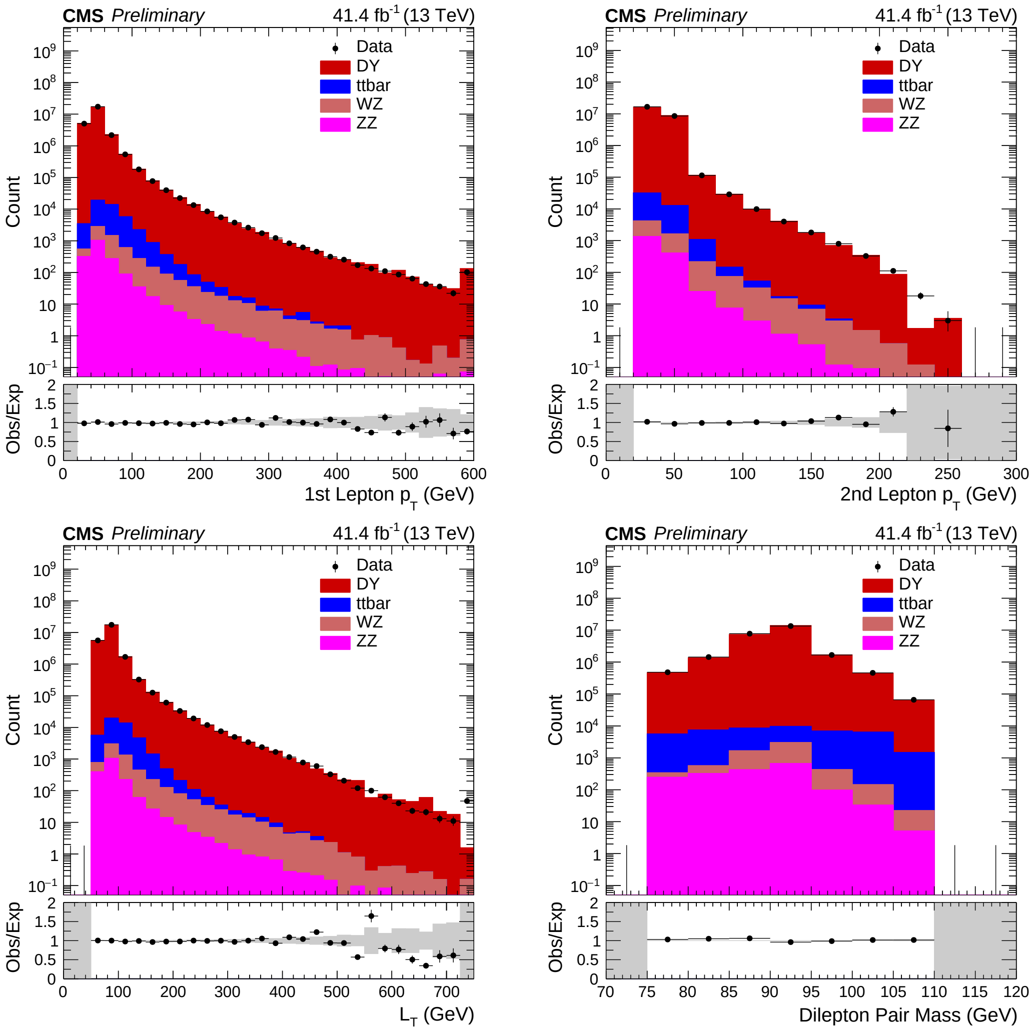


FIGURE 5.3: Leading (upper left) and subleading (upper right) muon p_T, L_T (lower left) and $M_{\ell\ell}$ (lower right) distributions in the DY$\rightarrow \mu\mu$ enriched dilepton selection.

assumption allows the measurement of these rates in signal-depleted regions and then their application to the signal regions.

In the analysis, a 3-dimensional matrix method is used, while for clarity, a 2-dimensional matrix method is explained in this section.

All the dilepton events are divided into four categories. The dilepton events in which both the leptons satisfy the tight lepton criteria are referred to as TT events, events where only one lepton satisfy the tight criteria while the other lepton only satisfies a loose criterion and fails the tight criteria are referred to as LT and TL. The letters are ordered based on the p_T of leptons; for example, if the lepton leading in p_T satisfies the tight criteria and the trailing lepton only satisfies loose criteria; then the events

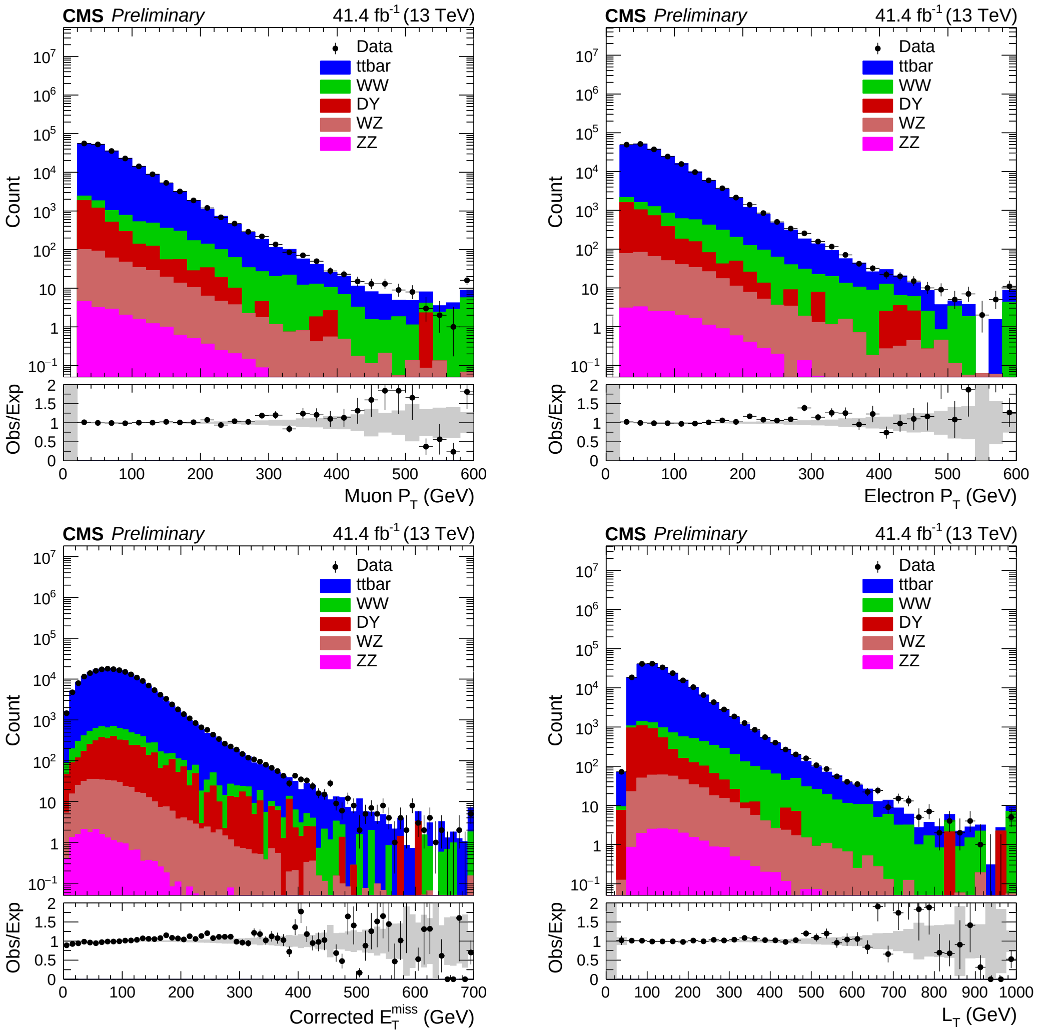


FIGURE 5.4: Muon p_T (upper left), electron p_T (upper right), p_T^{miss} (lower left) and L_T (lower right) distributions in the $t\bar{t} \rightarrow e\mu$ enriched dilepton selection.

are labeled as TL. The events where both the leptons fail the tight selections while still passing the loose selection are referred to as LL events. The events where both the leptons are tight, TT, are used to define the signal regions while the TL, LT, and LL events are used to estimate the misidentified background using the matrix method. Another categorization of these events can also be done based on the origin of these leptons. Events where both leptons are prompt and are labeled as PP, both fake as FF and one fake, one prompt as PF and FP. It should be noted that the total number of events is the same after loose selections; this implies:

$$N_{FF} + N_{PF} + N_{FP} + N_{PP} = N_{TT} + N_{TL} + N_{LT} + N_{LL} \tag{5.1}$$

We can define $\hat{f}_i = 1 - f_i$ and $\hat{p}_i = 1 - p_i$, using these definations the two categorizations can also be related by a transformation matrix given in Eq. 5.2.

$$\begin{pmatrix} N_{LL} \\ N_{LT} \\ N_{TL} \\ N_{TT} \end{pmatrix} = \begin{pmatrix} \hat{f}_1.\hat{f}_2 & \hat{f}_1.\hat{p}_2 & \hat{p}_1.\hat{f}_2 & \hat{p}_1.\hat{p}_2 \\ \hat{f}_1.f_2 & \hat{f}_1.p_2 & \hat{p}_1.f_2 & \hat{p}_1.p_2 \\ f_1.\hat{f}_2 & f_1.\hat{p}_2 & p_1.\hat{f}_2 & p_1.\hat{p}_2 \\ f_1.f_2 & f_1.p_2 & p_1.f_2 & p_1.p_2 \end{pmatrix} \begin{pmatrix} N_{FF} \\ N_{FP} \\ N_{PF} \\ N_{PP} \end{pmatrix} \tag{5.2}$$

This matrix relates the actual origin of the leptons to the observations using the detector. The matrix could be inverted to get the number of origin based quantities as a function of numbers and parameters that can be measured experimentally. While inverting the matrix, it is assumed that $p_i \neq f_i$. The inverted matrix is given in Eq. 5.3.

$$\begin{pmatrix} N_{FF} \\ N_{FP} \\ N_{PF} \\ N_{PP} \end{pmatrix} = \frac{1}{(p_1 - f_1)(p_2 - f_2)} \begin{pmatrix} p_1.p_2 & -p_1.\hat{p}_2 & -\hat{p}_1.p_2 & \hat{p}_1.\hat{p}_2 \\ -p_1.f_2 & p_1.\hat{f}_2 & \hat{p}_1.f_2 & -\hat{p}_1.\hat{f}_2 \\ -f_1.p_2 & f_1.\hat{p}_2 & \hat{f}_1.p_2 & -\hat{f}_1.\hat{p}_2 \\ f_1.f_2 & -f_1.\hat{f}_2 & -\hat{f}_1.f_2 & \hat{f}_1.\hat{f}_2 \end{pmatrix} \begin{pmatrix} N_{LL} \\ N_{LT} \\ N_{TL} \\ N_{TT} \end{pmatrix} \tag{5.3}$$

The number of events with at least one fake in the TT category can be estimated using Eq. 5.3 and the fact that the total fake contribution is coming from events with one fake lepton and two fake leptons. The total fake background can be expressed as:

$$N_{TT}^{Fake} = f_1 f_2 N_{FF} + f_1 p_2 N_{FP} + p_1 f_2 N_{PF} \tag{5.4}$$

We can now use Eq. 5.3 and Eq. 5.4 to predict the fake contribution, given that the fake and prompt rates are already measured. This method can be extended to include the third lepton, as well. In this analysis, its 3-dimensional form is used. The 3-dimensional matrix method can predict background contributions due to events with up to 3 simultaneous misidentified leptons. In signal regions with four or more leptons, the 3-dimensional matrix method is applied on the 2^{nd}, 3^{rd}, and 4^{th} leading $\mathrm{p_T}$ leptons, and the lepton with the highest $\mathrm{p_T}$ in the event is considered a prompt lepton.

5.6 Measurement of the Prompt and Misidentification Rates

The measurement of the prompt and misidentification or fake rates is performed in the signal depleted control regions and in MC simulated samples. The prompt rates are defined as the probability that a given lepton satisfies the tight lepton selection, given that it also fulfilled the loose lepton selections. Depending on the origin of the lepton, this probability is referred to as "fake" or "prompt" rates.

In the detector, one of the main differences between the leptons arising from bosonic decays as opposed to leptons from jets and other hadronic activity is the lepton isolation, i.e., the amount of energy around the lepton candidate. Moreover, even for fake leptons, the amount of energy deposited in the immediate neighborhood of the lepton might change depending on the origin of the fake lepton. It is also worth noting that the properties of lepton changes depending on how many particles are depositing given energy and how close they are to the lepton. The other important parameter is the displacement of the lepton from the vertex. The dominant processes contributing to the misidentified background are DY + jets and $\mathrm{t\bar{t}}$ +jets. In DY + jets, the fake lepton mostly comes from additional jets of light flavor or a gluon jets, on the other hand, fakes generally originate from the b quarks jets in the $\mathrm{t\bar{t}}$ events. The light jets and the b-jets have very different fragmentations, and moreover, the b-jets are coming from the decays of top quarks and tend to have large $\mathrm{p_T}$ compared to light jets in the DY + jets process. This leads to a unique challenge of two very different processes contributing to misidentified lepton background. The rate at which loose leptons coming from the b-jets in $\mathrm{t\bar{t}}$ pass the light lepton selection are quite different as compared to the DY + jets leptons. Additionally, DY + jets events have a very different topology compared to the $\mathrm{t\bar{t}}$ events. In the DY process, the jet giving rise to the fake lepton essentially recoils against the Z boson system(the prompt lepton pair). This means the property of the fake lepton in the DY processes directly correlates to the momentum of the Z boson. If the Z has high momentum, the recoiling jet would be more collimated due to the momentum conservation, this will hugely impact the fake rate for the fake lepton coming from the recoiled jet. This lepton is more likely to fail the tight isolation requirement used in the tight selection, resulting in a much lower fake rate. These observations suggest that for parameterization of the fake rates, we need something more than just the lepton specific properties such as $\mathrm{p_T}$ and η. To mitigate the issues discussed above, we utilize the $\mathrm{p_T}$ of the AK4 jet matched to the given fake lepton (also referred to as mother jet) for the fake rate parameterization in addition to lepton $\mathrm{p_T}$ and η. This not only helps in reducing the differences between rates measured in $\mathrm{t\bar{t}}$ and DY but also helps in capturing the nuisances of the event topology.

Prompt rates for leptons are measured using a tag-and-probe method in MC and data. For MC measurements, a set of events enriched in prompt leptons from $\mathrm{Z} \rightarrow \ell\ell$ or $\mathrm{t\bar{t}}$ decays is created by requiring the reconstructed leptons to be kinematically matched to a generator level prompt lepton ($\Delta R < 0.1$) to be labeled as prompt objects. In data, we require to have an on-Z, OSSF pair, and $\mathrm{p_T^{miss}} < 50$ GeV. The leading $\mathrm{p_T}$ lepton is chosen as the tag and is also required to satisfy the tight lepton selection, and the subleading $\mathrm{p_T}$ one is chosen as the probe.

The rates are measured separately for each lepton flavor. The prompt rates are parametrized in bins of the lepton $\mathrm{p_T}$, lepton $|\eta|$, and the number of tracks in an event. Lepton fake rates are measured inclusively for trilepton events in MC. In data, we require to have an on-Z, OSSF pair satisfying the tight lepton selection, $\mathrm{p_T^{miss}} < 50$ GeV, and a single additional probe lepton satisfying the loose lepton selection. A fake probe lepton enriched selection of events is created in simulated samples by imposing an inverted $\Delta\mathrm{R}$ requirement among the probe lepton and the prompt generator level leptons in the event ($\Delta\mathrm{R} > 0.1$). Additionally, fake probe leptons matching to a generator level photon are also vetoed.

In all rate measurements conducted in data, contributions due to leptons of the undesired origin (prompt leptons for fake rate measurements and vice versa) are estimated and subtracted using MC methods, whereas a simple binomial ratio of tight over loose leptons is taken in MC. These quantities are calculated as given in Eq. 5.5.

$$\begin{aligned} p^{\mathrm{MC}} &= \frac{\mathrm{N}^{\mathrm{MC}}_{\mathrm{tight\ prompts}}}{\mathrm{N}^{\mathrm{MC}}_{\mathrm{loose\ prompts}}}, \qquad & p^{\mathrm{data}} &= \frac{\mathrm{N}^{\mathrm{data}}_{\mathrm{tight\ prompts}}}{\mathrm{N}^{\mathrm{data}}_{\mathrm{loose\ prompts}}}, \\ f^{\mathrm{MC}} &= \frac{\mathrm{N}^{\mathrm{MC}}_{\mathrm{tight\ fakes}}}{\mathrm{N}^{\mathrm{MC}}_{\mathrm{loose\ fakes}}}, \qquad & f^{\mathrm{data}} &= \frac{\mathrm{N}^{\mathrm{data}}_{\mathrm{tight}} - \mathrm{N}^{\mathrm{MC}}_{\mathrm{tight\ non-fakes}}}{\mathrm{N}^{\mathrm{data}}_{\mathrm{loose}} - \mathrm{N}^{\mathrm{MC}}_{\mathrm{loose\ non-fakes}}}. \end{aligned} \tag{5.5}$$

Since the fake rate measurement in data is statistically limited and contaminated with the prompt background, we do not use these fake rates directly, but to correct the fake rates measured in MC. This is done only for fake rate bins in lepton $\mathrm{p_T}$ & lepton $|\eta|$ with significant non-prompt contribution. We calculate the ratio of $k^{\mathrm{f}}_{\mathrm{data}} = f^{\mathrm{data}}/f^{\mathrm{DY\ MC}}$ and apply this correction factor back to f^{MC} in the corresponding bins of the $\mathrm{p_T}$ of the mother jet of the probe lepton. Prompt rates are corrected in each lepton $\mathrm{p_T}$, lepton $|\eta|$ and number of tracks bin by $k^{\mathrm{p}}_{\mathrm{data}} = p^{\mathrm{data}}/p^{\mathrm{DY\ MC}}$

Taking into account the variation of the rates in simulated and data enriched DY and only simulated $\mathrm{t\bar{t}}$ events, a single prompt, and the fake rate is defined for each bin as given in Eq. 5.6.

$$\begin{aligned} p &= k^{\mathrm{p}}_{\mathrm{data}} \cdot \frac{p^{\mathrm{DY\ MC}} + p^{\mathrm{t\bar{t}\ MC}}}{2}, \\ f &= k^{\mathrm{f}}_{\mathrm{data}} \cdot \frac{f^{\mathrm{DY\ MC}} + f^{\mathrm{t\bar{t}\ MC}}}{2}. \end{aligned} \tag{5.6}$$

The electron and muon prompt rates are measured to be $\gtrsim 80\%$, whereas the fake rates vary in the range of 5-25% for both lepton flavors.

The electrons prompt rates measured in the data are shown in Fig. 5.5.

Electron prompt rates in the simulated samples for the DY + jets and $\mathrm{t\bar{t}}$ are shown in Fig. 5.6.

Electron fake rates measured as a function of $\mathrm{p_T}$ of AK4 jet matched to the electron in bins of electron $\mathrm{p_T}$ for barrel and endcap regions are provided in Fig 5.7.

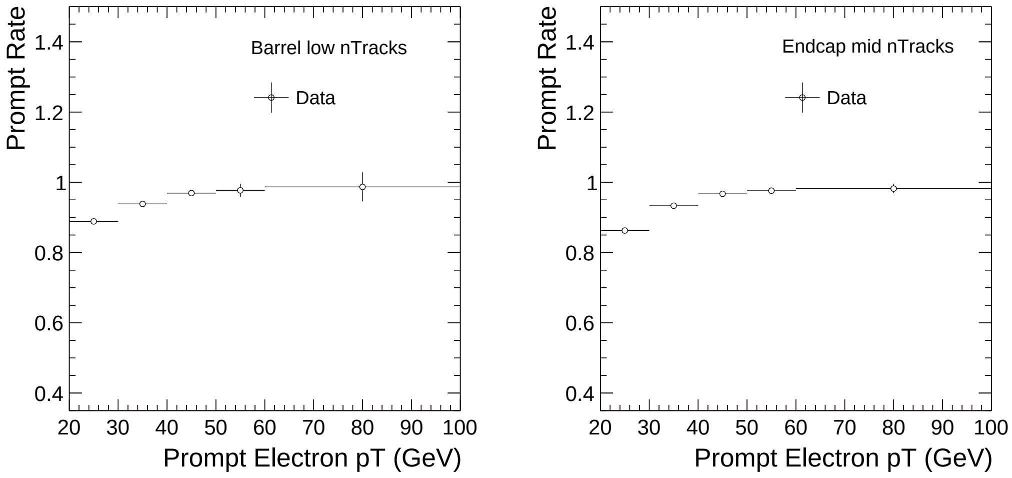


FIGURE 5.5: Electron data prompt rates in barrel for $N_{Track}s < 20$ (left), and endcap $20 < N_{Tracks} < 50$ (right).

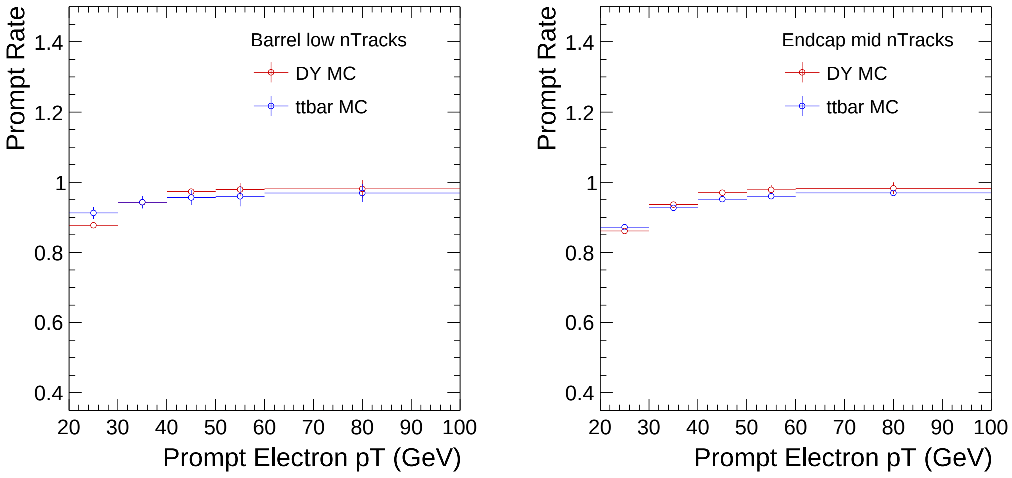


FIGURE 5.6: Electron MC prompt rates in barrel for $N_{Tracks} < 20$ (left), and in endcap for $20 < N_{Tracks} < 50$ (right) .

5.7 Misidentified Tau Backgrounds

In the regions with two light lepton and a tau, the most dominant background comes from processes with two prompt light leptons and a fake tau, and fully leptonic decays of Z and W boson in DY + jets and $t\bar{t}$ + jets are the most significant processes resulting in this final state.

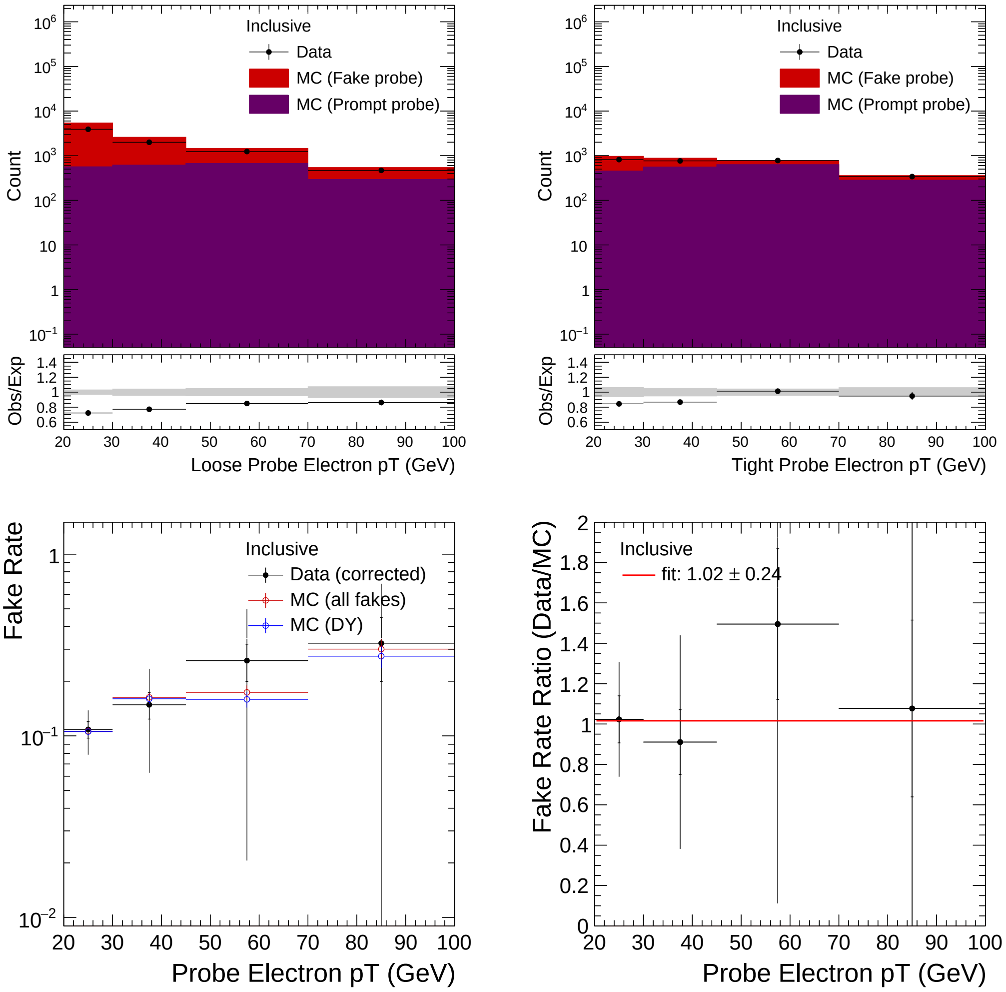


FIGURE 5.7: Electron data fake rates: loose probe lepton selection (upper left), tight probe lepton selection (upper right), fake rates (bottom left) and data to DY MC fake rate ratios (bottom right).

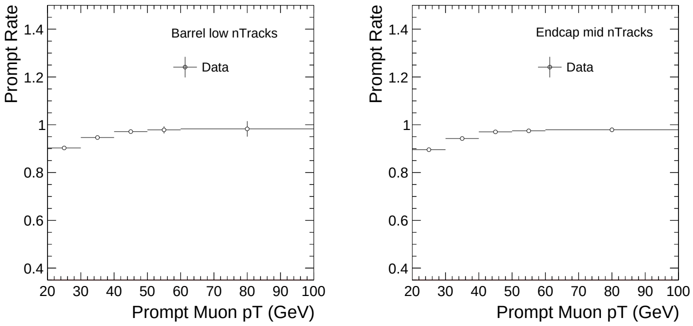


FIGURE 5.8: Muon data prompt rates in barrel for $N_{Tracks} < 20$ (left), and in endcap for $20 < N_{Tracks} < 50$ (right).

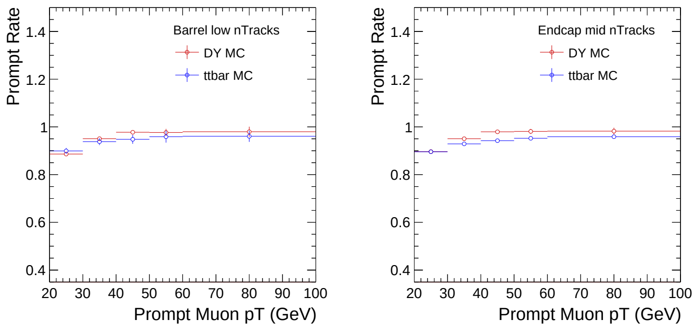


FIGURE 5.9: Muon MC prompt rates in barrel for $N_{Tracks} < 20$ (left), and in endcap for $20 < N_{Tracks} < 50$ (right).

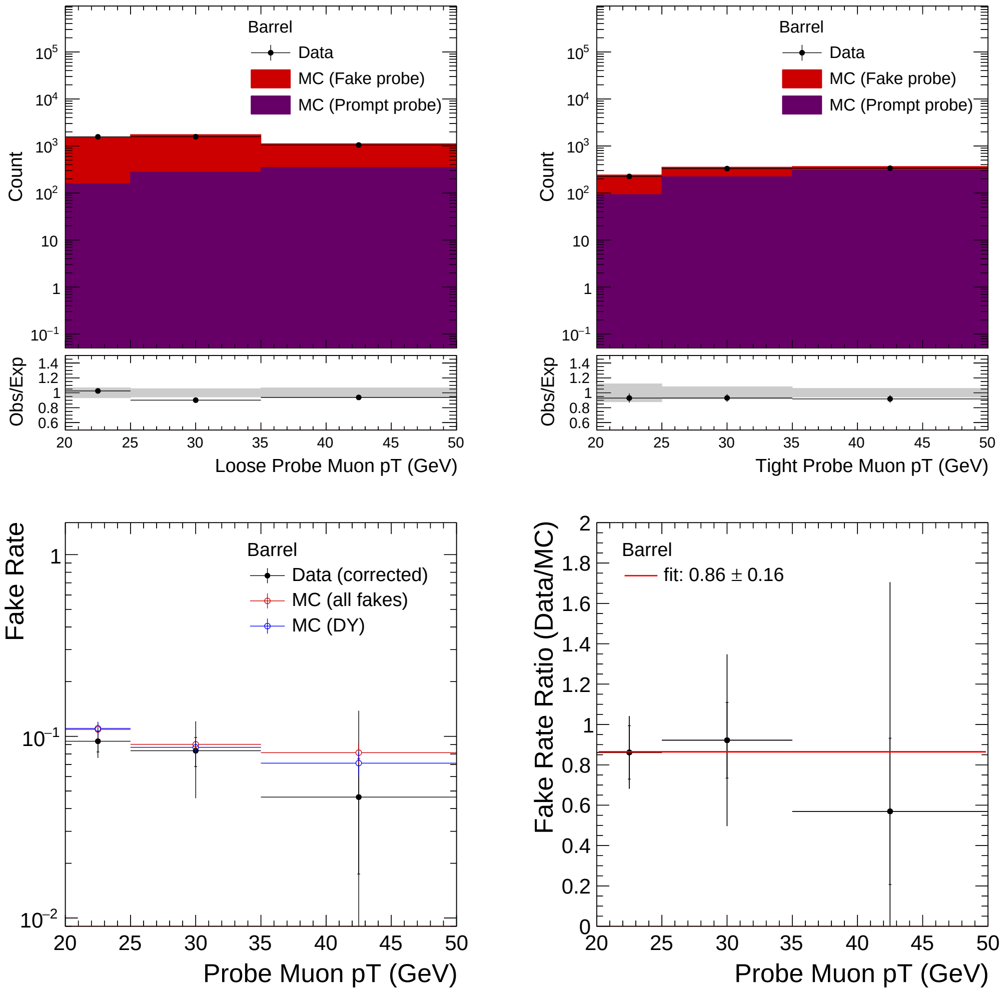


FIGURE 5.10: Muon data fake rates in barrel: loose probe lepton selection (upper left), tight probe lepton selection (upper right), fake rates (bottom left) and data to DY MC fake rate ratios (bottom right).

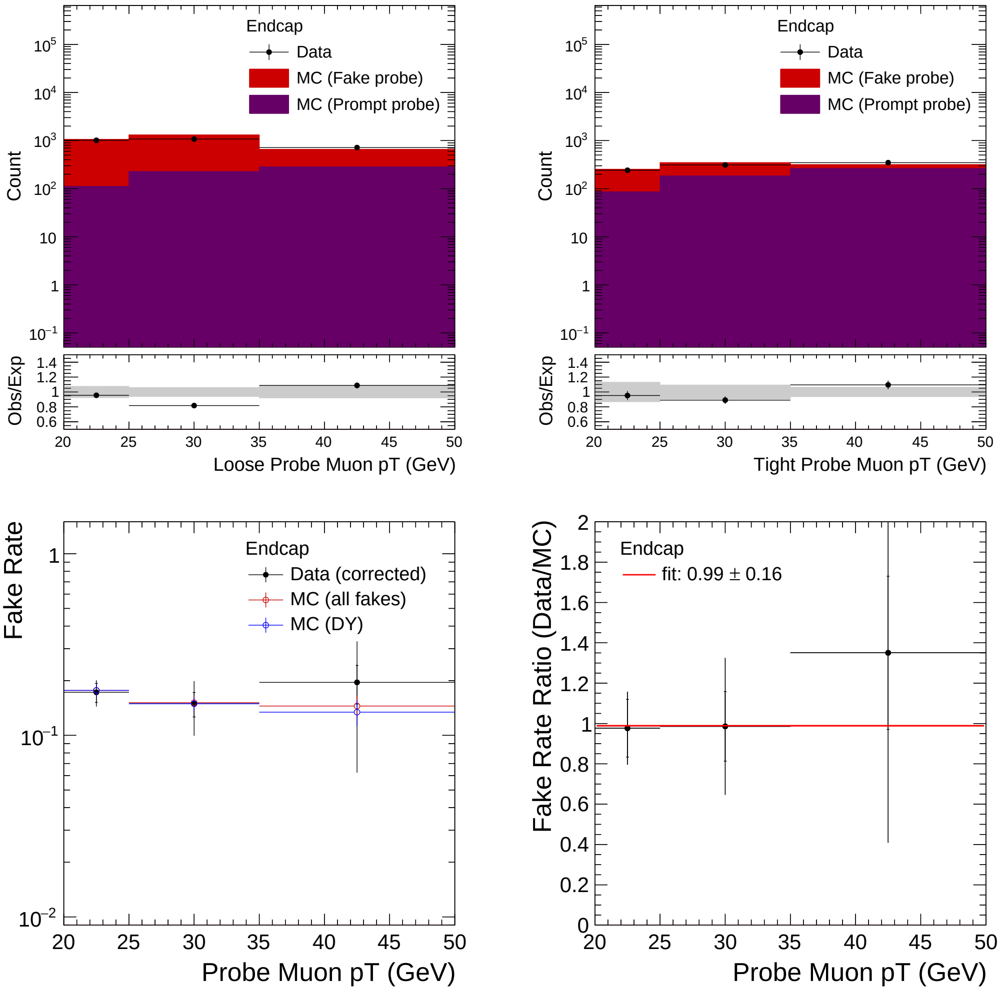


FIGURE 5.11: Muon data fake rates in endcap: loose probe lepton selection (upper left), tight probe lepton selection (upper right), fake rates (bottom left) and data to DY MC fake rate ratios (bottom right).

5.8 Tau Fake and Prompt Rates

Tau fake rates are measured in events with two opposite sign, same flavor leptons satisfying tight lepton selection with an additional loose tau lepton. Light lepton pair invariant mass is required to be on-Z. We also impose a $p_T^{miss} < 50$ GeV requirement to create a signal depleted, DY + jets enriched selection of events to both measure and validate tau fake rates. A fake tau enriched selection of events is created in simulated samples by imposing an inverted ΔR requirement among the probe tau and the prompt generator level tau in the event ($\Delta R > 0.2$).

We measure prompt rates for taus in simulated DY + jets samples. We require that a loose tau candidate matches to a gen tau within a $\Delta R < 02.$ and measure rate at which it also satisfies the tight working point.

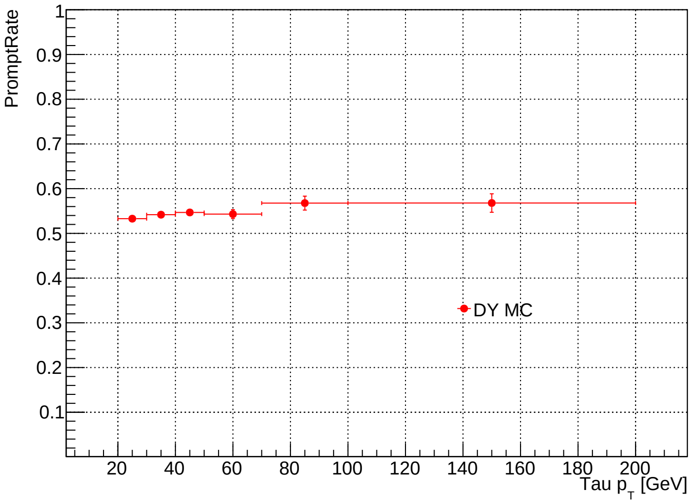


FIGURE 5.12: Tau prompt rates as a function of tau p_T as measured in simulated DY + jets samples

Fake rates are measured separately for taus reconstructed with one charged hadron (1 prong) and with three charged hadrons (3 prong) in the barrel ($|\eta| < 1.46$) and endcap ($|\eta| > 1.56$) region of the detector. Then we use tau p_T and p_T of the mother jet (AK4 PF jet matched to tau candidate) for fake rate parametrization. We measure tau fake rate as a function of mother jet p_T in regions with tau $p_T < 30, 30-50$ and $50-100$ GeV. For tau $p_T > 100$ GeV, fake rates are measured inclusively irrespective of the decay mode and η position. The mother jet p_T parameterization is used both to minimize the spread of fake rates in DY vs. $t\bar{t}$ events, and also to capture a particular recoil effect which is most visible in DY+1jet processes where the jet is misidentified as a fake tau. Since tau isolation is defined by an MVA that is based on absolute (rather than relative) isolation quantities, we observe that tau candidates with a large mismatch between tau p_T and the mother jet pt are more likely to fail tighter isolations requirements. This effect is always present but most visible in DY+1jet events where the recoil of the Z boson is almost directly correlated with the mother jet pt of the fake tau object, and it affects the tau fake rate irrespective of the tau p_T. Our current parametrization accounts for such topologies, and the L_T distribution is accurately modeled by the data-driven method. A few examples of the tau fake rates as measured in the data and DY+jets and $t\bar{t}$

simulation samples are provided in Fig. 5.13 and 5.14, all the remaining tau fake rates are given in Appendix A.2.

An additional correction factor to tau fake rates is derived to account for the hadronic activity dependence of rates. We measure fake rates as a function of the number of high purity tracks in the events and divide that with the average tau fake rate. This gives us a correction to tau fake rates as a function of the number of tracks as shown in Figure 5.15. We fit a second-degree polynomial to extract this correction factor. The resultant fit parameters for the function are $p_0 = 1.367 \pm 0.038$, $p_1 = -0.01247 \pm 0.0017$, $p_2 = 0.000066 \pm 0.000017$ where the functional form of the fit for the correction factor is $p_0 + p_1(N_{Trk}) + p_2(N_{Trk})^2$.

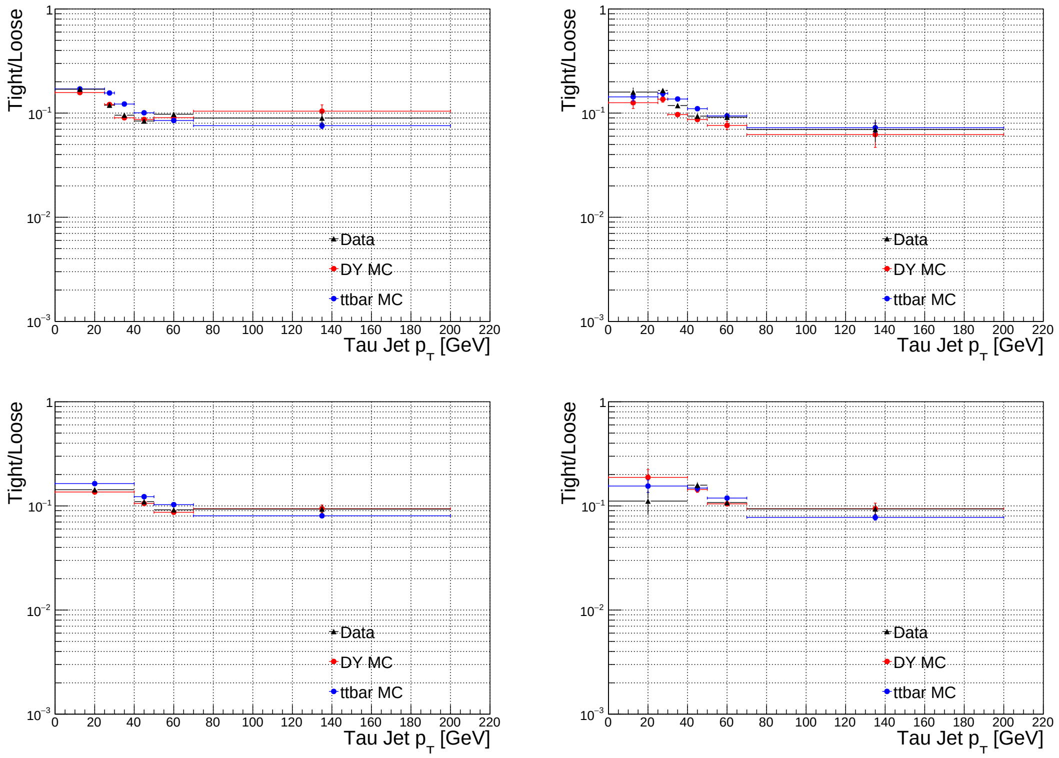


FIGURE 5.13: Upper left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with $p_T < 30$ GeV in barrel. Upper right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with $p_T < 30$ GeV in endcap. Lower left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $30 - 50$ GeV in barrel. Lower right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $30 - 50$ GeV in endcap.

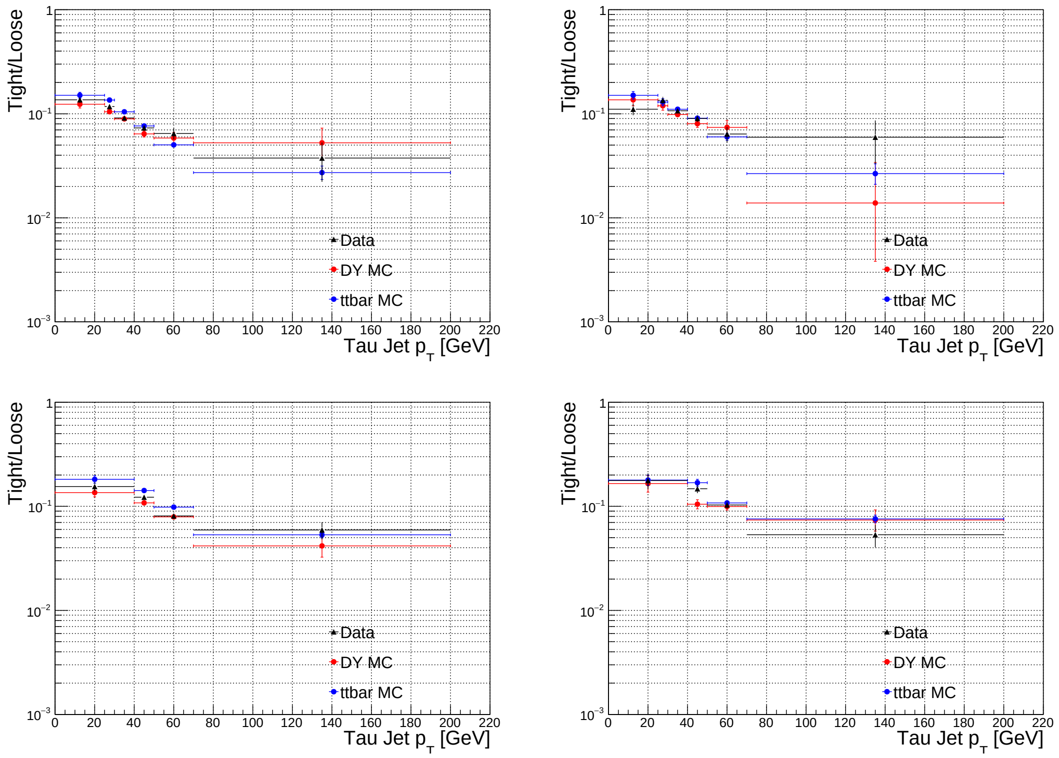


FIGURE 5.14: Upper left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 3 prong taus with $p_T < 30$ GeV in barrel. Upper right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 3 prong taus with $p_T < 30$ GeV in endcap. Lower left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 3 prong taus with p_T $30 - 50$ GeV in barrel. Lower right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 3 prong taus with p_T $30 - 50$ GeV in endcap.

5.9 Uncertainties on the Misidentified Lepton Background

The rates measured in the simulated samples are tested for self-consistency and to make sure the crucial kinematics distributions are well modeled. The simulated DY+jets and $t\bar{t}$+jets samples are used where at least one fake lepton is required as part of the event selection. The simulated samples are treated as data, and the sidebands in the simulated samples are used to estimate the fake contribution. This estimation is then plotted against the actual event, which has at least one lepton that does not match to a genuine lepton at the generator level. Figure 5.16 includes L_T distributions showing the observed yield of the fake events in the simulated samples and the estimation of the fake lepton yield using the matrix method. The L_T distribution agrees quite well in for DY+jets and $t\bar{t}$ samples for both electron and muon fake events.

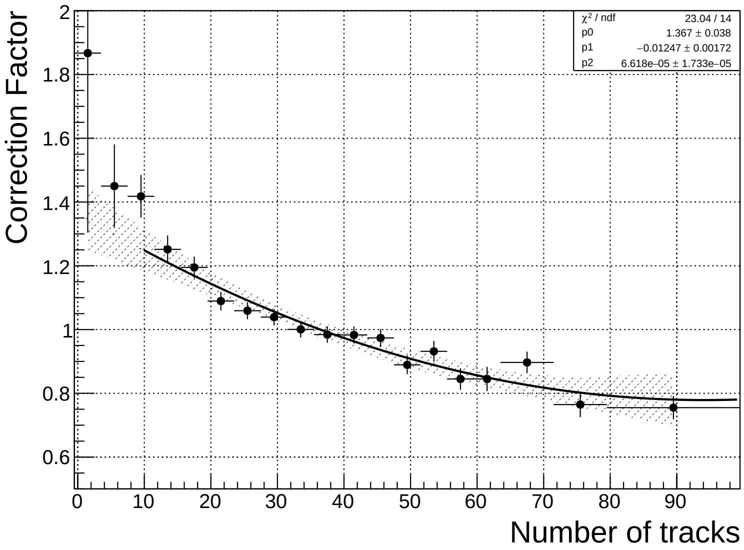


FIGURE 5.15: Fake rate correction factor as a function of the number of high purity tracks in the event. This correction factor is used to account for the hadronic activity dependence of tau fake rate. We fit a second-degree polynomial to extract the correction factor.

A similar test is also performed in the data in a region where the data fake rates are measured. This region is mostly dominated by DY+jets like events. The data rates are used to predict the misidentified background yield, and the estimate is compared to the observed events in the data. Figure 5.17 shows an excellent agreement between the matrix method prediction and the observations.

Various other distributions such as leptons $\mathrm{p_T}$, η, $\mathrm{p_T^{miss}}$, $\mathrm{H_T}$ are also inspected to verify the method. We observe that all these distributions are also well modeled.

The measured fake rates for light leptons have about $5-25\%$ uncertainty due to the limited statistics in the measurement regions and the contamination of the fake rates measurement regions by prompt sources (WZ, ZZ). Additional uncertainty on the rates arises due to the process dependent (DY vs. $\mathrm{t\bar{t}}$) variations ($10-15\%$) and flavor $\mathrm{p_T}$ ordering of the fake lepton [68]. Considering these uncertainties on the light lepton fake rates, we vary the fake rates by $20-25\%$ depending on the region to estimate the impact of the fake rates uncertainties on the misidentified background yield. Figure 5.18 shows the change in data-driven background yield in bins of $\mathrm{L_T}$ as the rates are varied up and down relative to the central rates. We observe the background yield changes about 30% by varying rates, and this is used as systematic uncertainty on this background component.

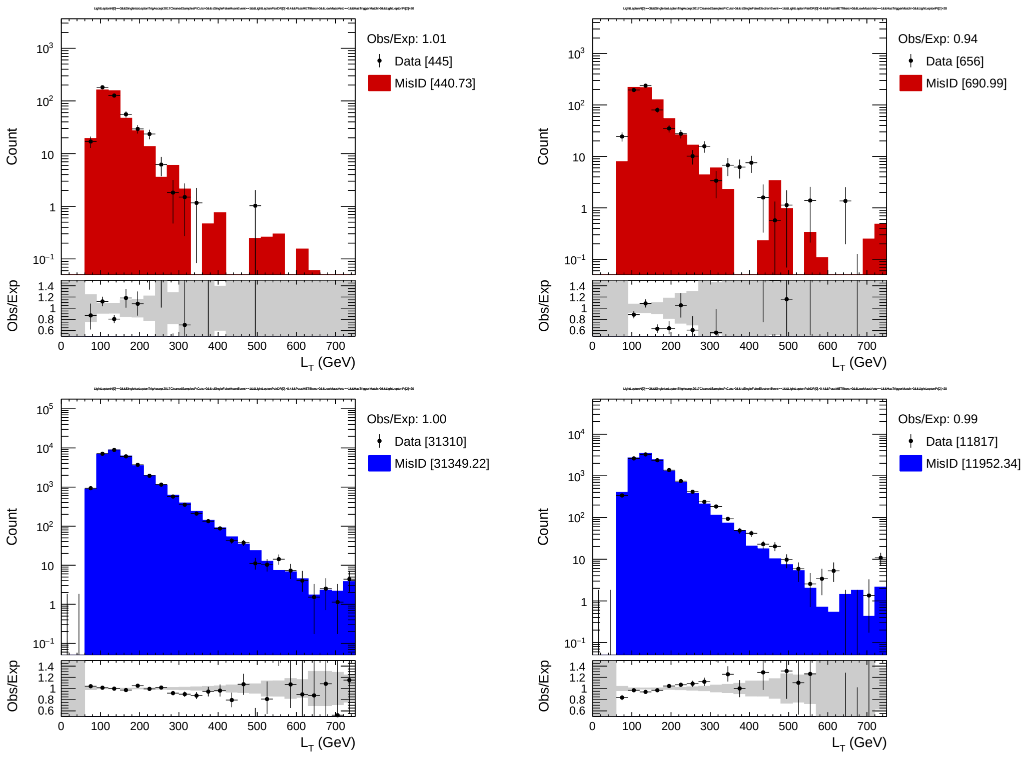


FIGURE 5.16: MC closure tests in simulated DY+jets and $t\bar{t}$+jets where at least one fake lepton is required as a part of event selection and DY or $t\bar{t}$ MC based prompt and fake rates are used accordingly. L_T distributions in simulated DY+jets events with a fake muon (upper left) and fake electron (upper right), and in simulated $t\bar{t}$+jets events with a fake muon (lower left) and fake electron (lower right) selection, where "Data" entries represent the observed MC events in all plots. A good agreement is observed in all distributions between the expected and observed number of events.

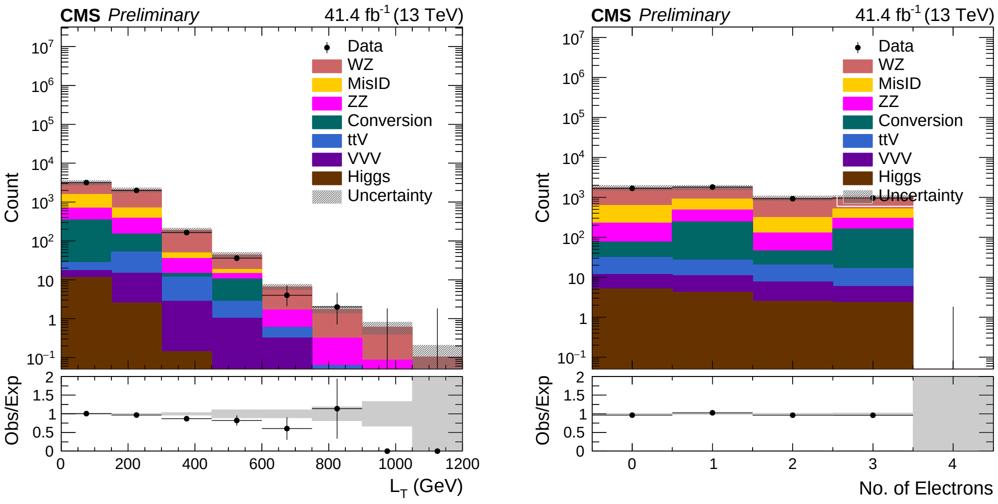


FIGURE 5.17: Self-closure tests in DY enriched data for the L_T and electron multiplicity distributions.

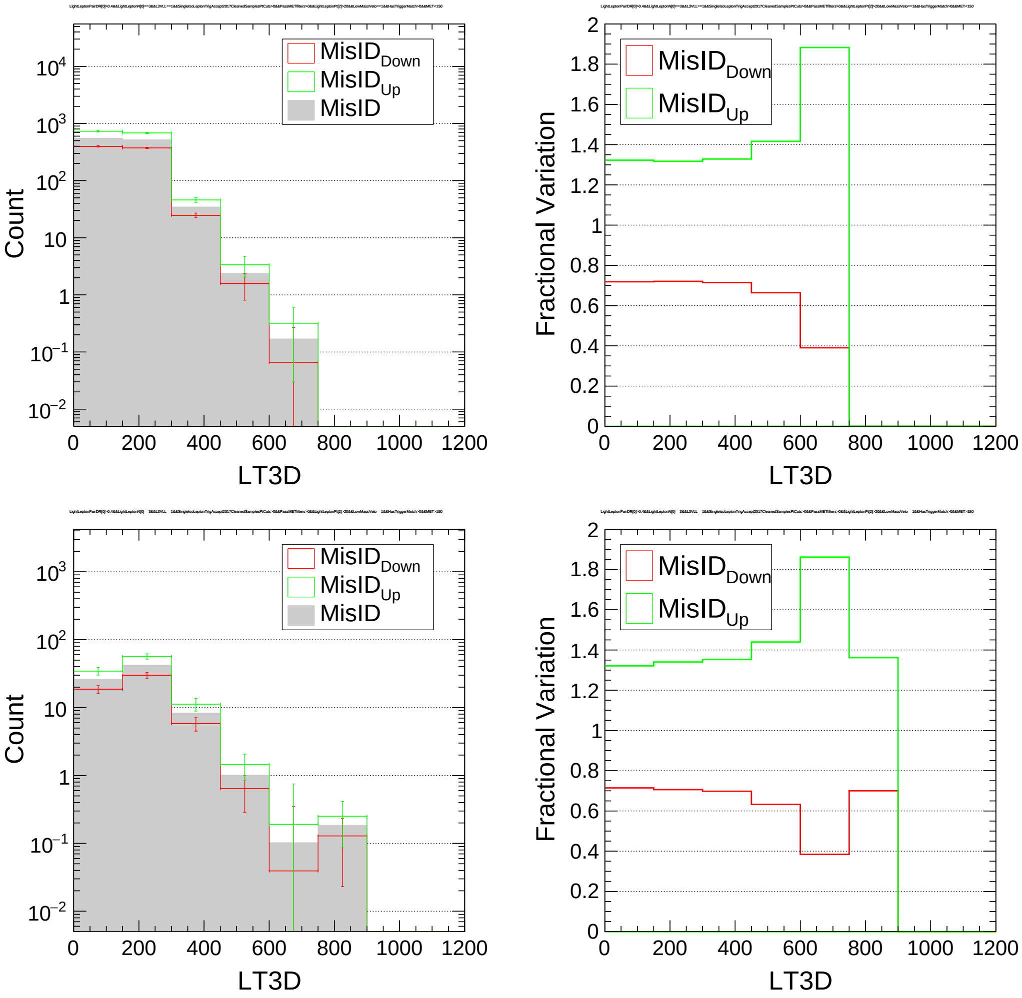


FIGURE 5.18: Impact of variations of prompt and fake rates in the trilepton signal regions: low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

The tau fake rates are measured in the region with two electrons and muons consistent with a Z boson and $p_T^{miss} < 50$ *GeV* with additional hadronic tau being present. This region has relatively higher statistics due to the higher rate of hadronic jets faking the hadronic tau lepton. The tau fake rates have typical uncertainty of $5 - 10\%$ due to the statistics and prompt contamination. The process dependent (DY vs. $t\bar{t}$) variations are of the order of $5 - 15\%$, while the correction factor based on the number of the tracks in the event also has an uncertainty of about 10%. Considering these variations of the fake rates, we vary the tau fake rates around central value by 20% for tau $p_T < 100$ *GeV* and by 40% for $p_T > 100$ *GeV* (the high p_T taus have larger statistical uncertainties) and observe the impact on the misidentified background yield. Figure 5.19, 5.20 shows the variation in the background yield in bins of L_T for events with a pair of opposite and same sign e/μ + τ_h respectively.

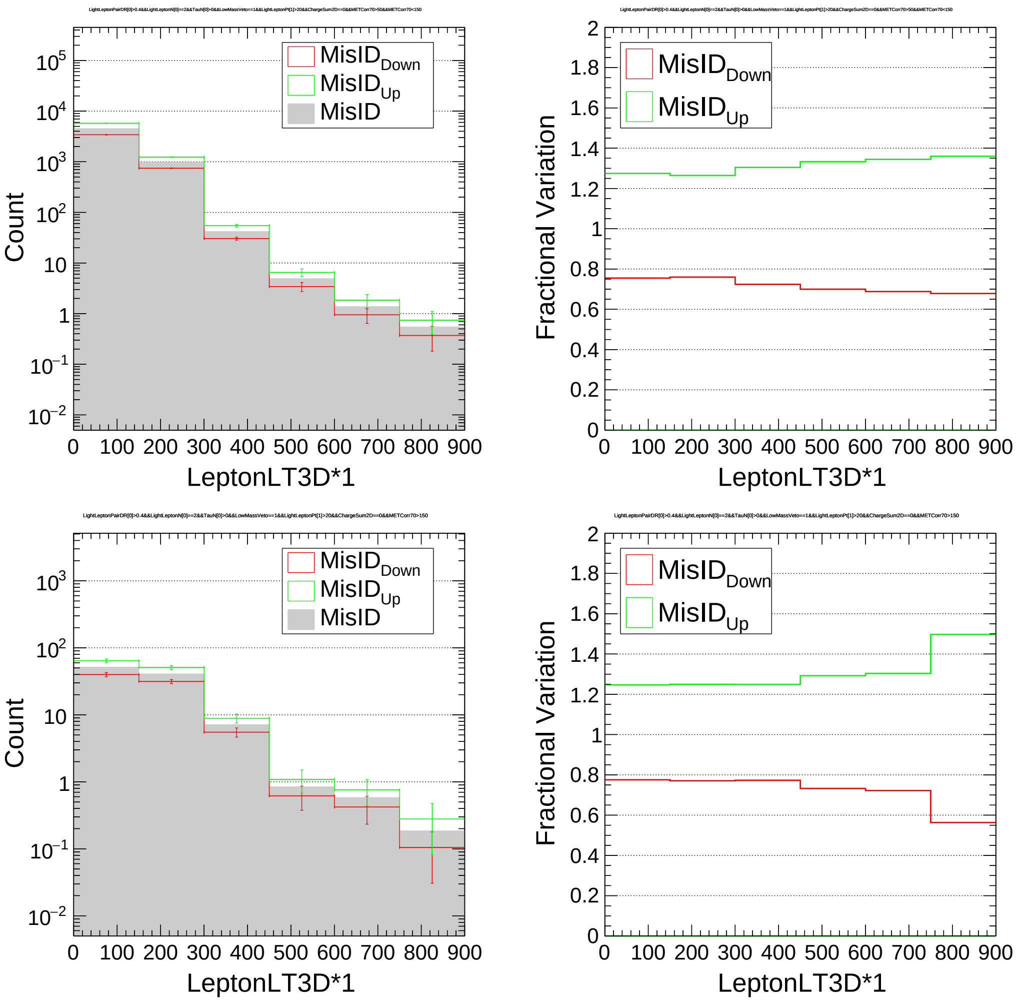


FIGURE 5.19: Impact of varying tau rates rates on $\mathrm{L_T}$ distribution in two leptons of oposite charge + hadronic tau channels in low $\mathrm{p_T^{miss}}$ (upper plots) and high $\mathrm{p_T^{miss}}$ (bottom plots).

For the 2L1TOS channel, the fake background component is hugely dominated by the tau lepton fakes while for the 2L1TSS channel the light lepton fakes also become important. We also see the impact of varying the light lepton fake rates in the tau channels. For the 2L1TOS channel seen in Fig. 5.21 the change in background yield is 3%, while for the 2L1TSS channel the impact is around 15-20%. Due to this reason, two separate nuisances are considered for misidentified background arising from the light leptons and taus.

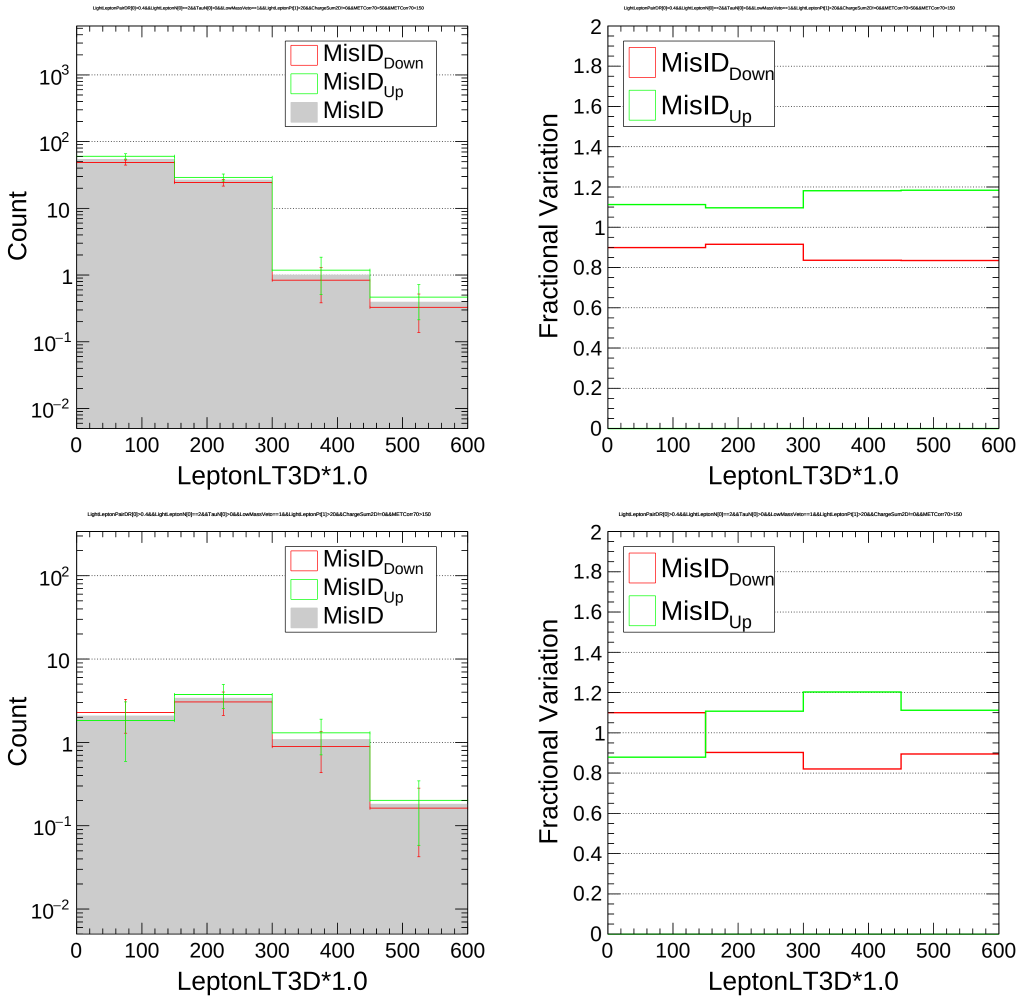


FIGURE 5.20: Impact of varying tau rates on L_T distribution in two leptons of the same charge + hadronic tau channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

5.10 WZ Background

The WZ $\rightarrow 3l\nu$ process constitutes one of the major backgrounds in this analysis in the three lepton channel. The background from the WZ process is estimated using aMC@NLO generated NLO MC sample after normalizing in the dedicated control region in data. The normalization is performed to not rely on the theory cross-section of the process but using data to get the proper cross-section making the analysis more robust and less sensitive to the theoretical calculations. A WZ $\rightarrow 3l\nu$ ($\geq 80\%$) enriched selection of events is created by requiring exactly 3 leptons with an on-Z OSSF pair and 50 GeV $\leq p_T^{miss} \leq$ 100 GeV. The WZ MC normalization scale factor over the NLO cross-section is then calculated as the ratio of non-WZ subtracted data events

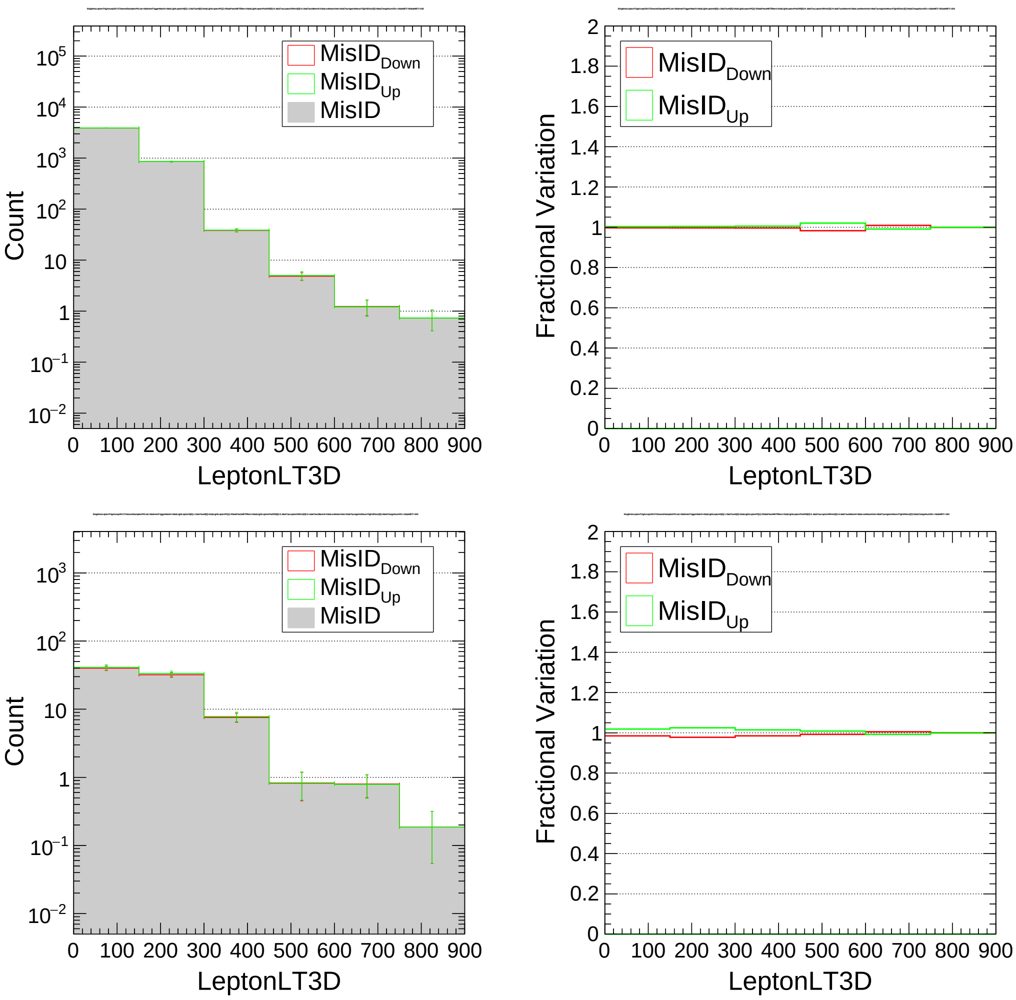


FIGURE 5.21: Impact of varying light lepton rates rates on L_T distribution in two leptons of oposite charge + hadronic tau channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

over WZ MC events, yielding a value of 1.24 ± 0.08 (statistical and systematic), and this normalization correction factor is applied to all WZ MC events. This leads to a relative normalization uncertainty of 7%. While normalizing to data ensures the total yield from this MC samples will be correct, it is also important to make sure that the key distributions are well modeled. We examine many such distributions; some example distributions corresponding to lowest lepton p_T, L_T, H_T, and electron multiplicity are shown in Figure 5.23.

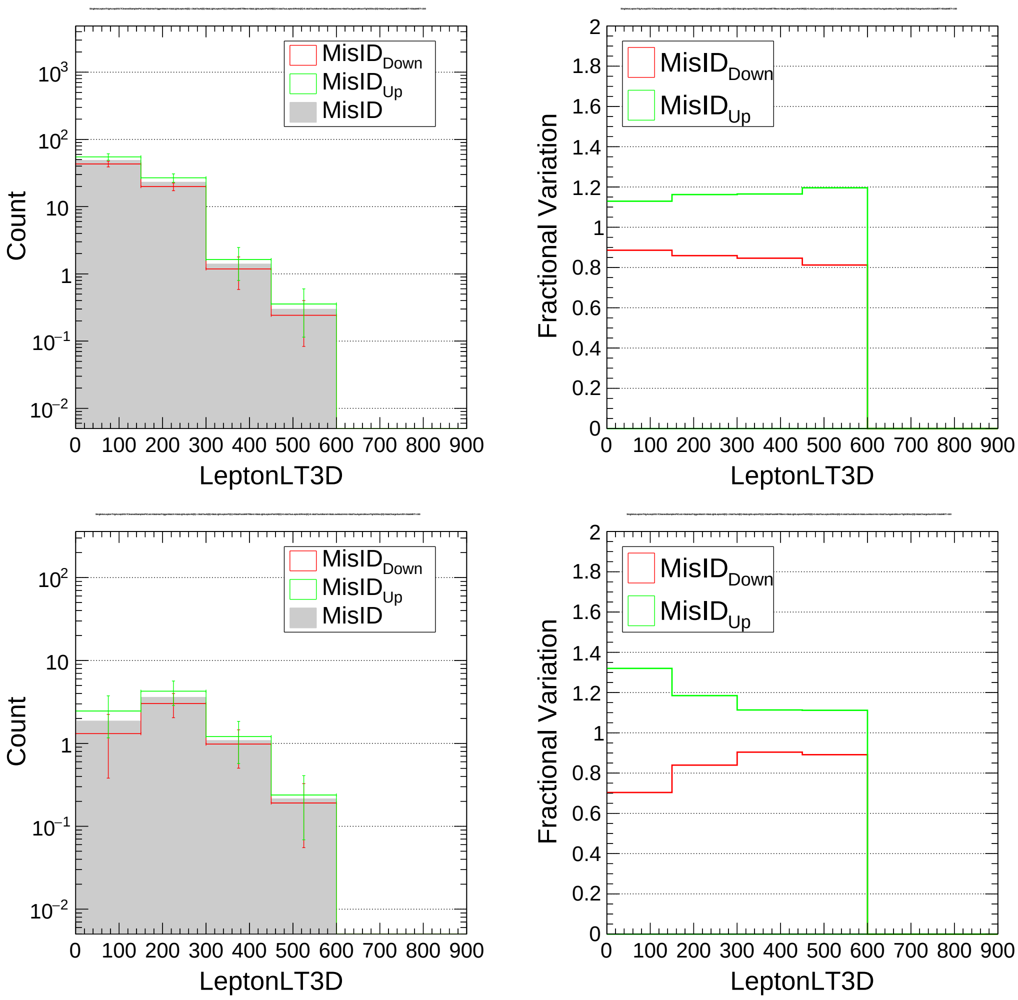


FIGURE 5.22: Impact of varying light lepton rates rates on $\mathrm{L_T}$ distribution in two leptons of same charge + hadronic tau channels in low $\mathrm{p_T^{miss}}$ (upper plots) and high $\mathrm{p_T^{miss}}$ (bottom plots).

5.11 ZZ Background

The ZZ $\rightarrow 4l$ background dominates ≥ 4 lepton signal regions. ZZ contributions are estimated using a Powheg generated NLO MC sample that is normalized to data in a dedicated ZZ-enriched selection of events. This selection is defined by requiring exactly four leptons that form two distinct on-Z OSSF pairs and $\mathrm{p_T^{miss}} < 50$ GeV, and yields to a set of events 99% pure in ZZ $\rightarrow 4l$. The ZZ MC normalization scale factor over the NLO cross-section is calculated as the ratio of non-ZZ subtracted data events over ZZ MC events, yielding a value of 0.91 ± 0.05 (statistical and systematic), which leads to a 6% relative uncertainty. For the ZZ enriched selection $M_{4L}, \mathrm{L_T}, \mathrm{H_T}$ and electron multiplicity distributions are provided in Fig 5.24.

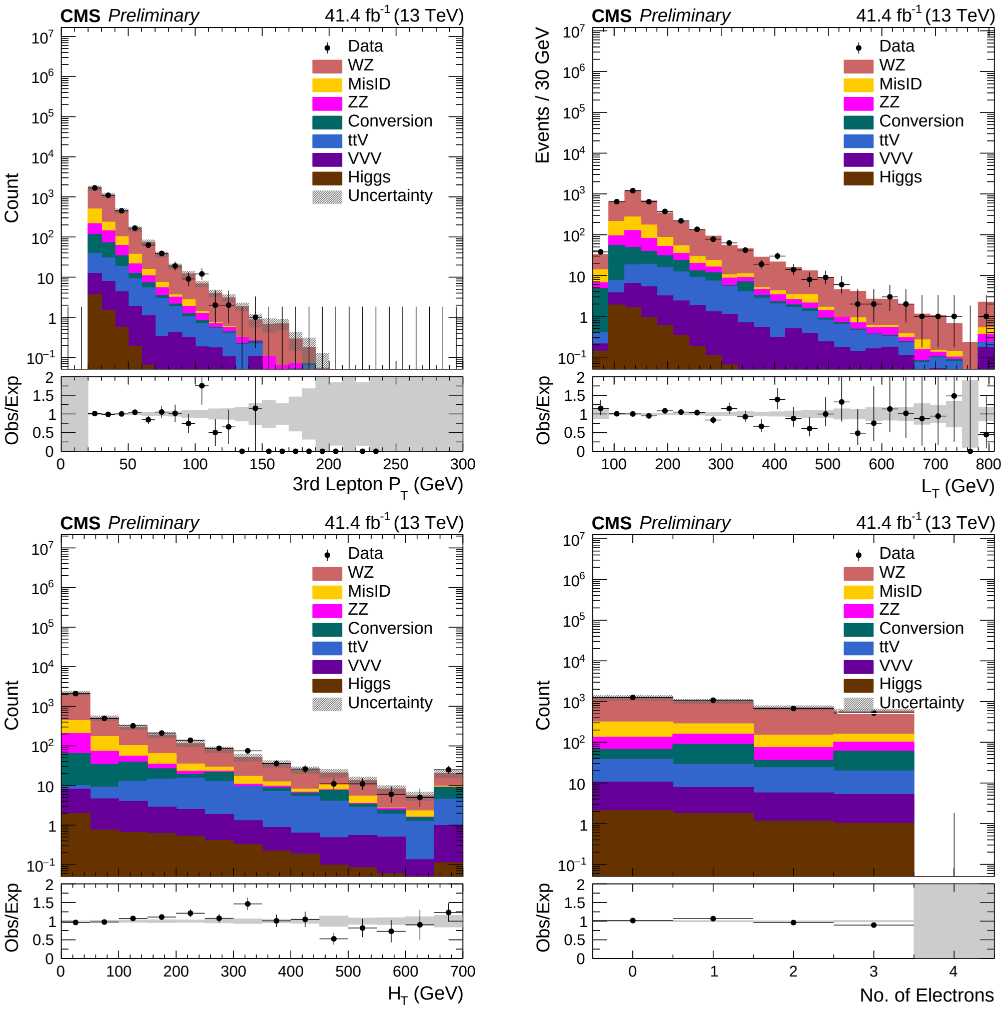


FIGURE 5.23: Distributions of 3rd lepton p_T , L_T, H_T and electron multiplicity in the WZ enriched region.

5.12 Conversion backgrounds

In process with two prompt leptons, an internal or external photon conversion might result in 2 additional leptons, which might also pass the promptness criteria. If the two conversion leptons were reasonably symmetric in $\mathrm{p_T}$, the standard conversion filters might reject those leptons. In cases where the conversion is asymmetric, one of the leptons might not result in a reconstructed lepton or not pass our quality criteria. In such cases, the final state consists of 3 leptons, and this results in a small background component to the three lepton signal regions. We estimate this background using MC samples. A selection constrained by the invariant mass of the two leading leptons ($M_{\ell\ell} \leq$ 76 GeV, below-Z), the invariant trilepton mass on-Z and $\mathrm{p_T^{miss}} \leq 50$ GeV results

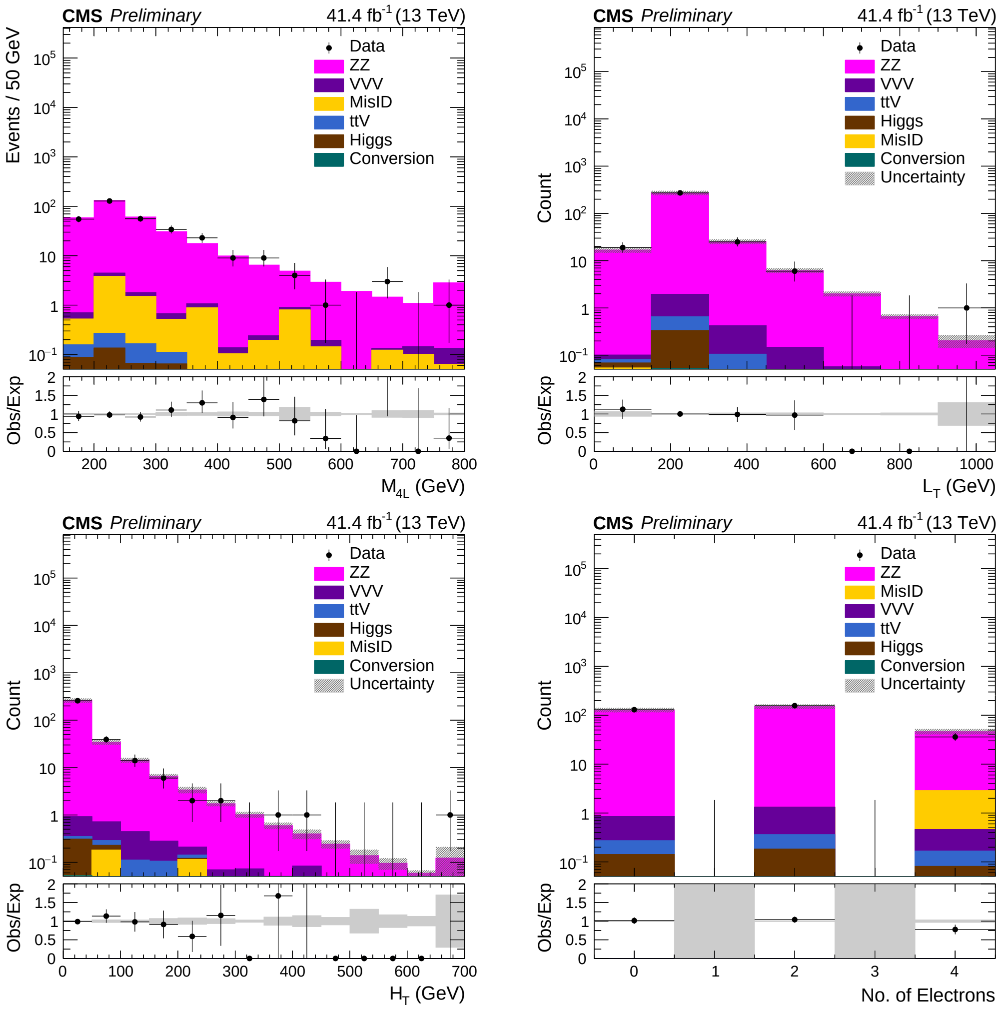


FIGURE 5.24: Distributions of M_{4L}, L_T, H_T, and electron multiplicity in the ZZ enriched region.

in a conversion dominant control region. The major contribution to this background comes from the Drell-Yan ($Z\gamma^*$) process and taken from a Drell-Yan+jets aMC@NLO generated NLO MC sample. Other processes like $t\bar{t}\gamma$ and WW can also contribute to this background and for their estimation we use $t\bar{t}$ and WW MC samples given in Table A.2. The overall MC normalization scale factor over the NLO cross-section is calculated as the ratio of non-conversion events subtracted data events over conversion events, yielding a value of 0.96 ± 0.08 (statistical and systematic), which leads to an 8% relative uncertainty. This normalization will be applied to all conversions processes in all signal regions. For this region, M_{3l}, leading lepton $\mathrm{p_T}$, jet and electron multiplicity distributions are provided in Fig 5.25. The DY MC is the most important MC sample for this control region, the DY samples have an effective luminosity of about 20 fb^{-1}, resulting in the non-smooth tails of distributions.

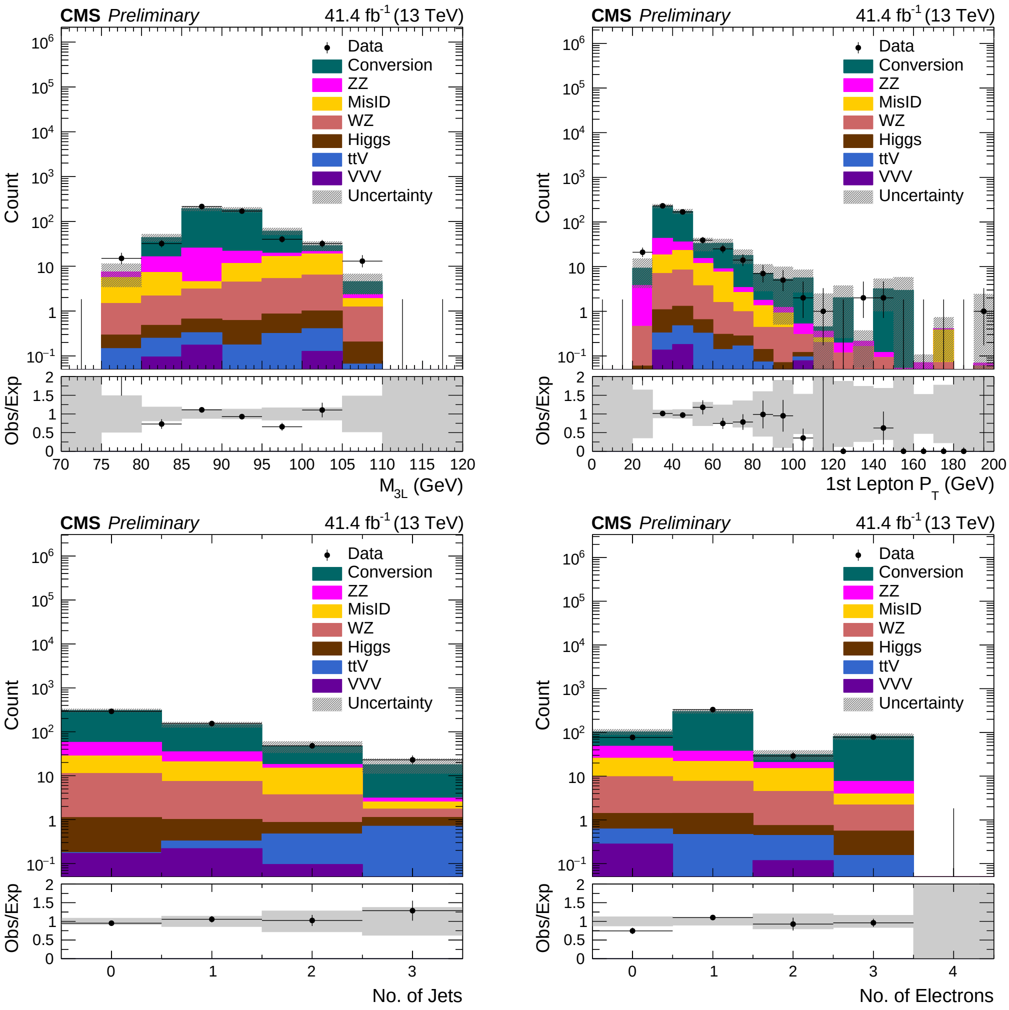


FIGURE 5.25: Distributions of M_{3l}, leading lepton p_T , jet and electron multiplicity in the Conversion enriched region.

5.13 Rare and Higgs Backgrounds

A relatively small contribution to the backgrounds comes from the rare standard model processes. The leading processes in this category are VVV, $t\bar{t}V$, and Higgs (e.g., VH, $t\bar{t}H$), where V refers to W, Z, H boson. The $t\bar{t}Z$ process contributes significantly to the 4L high p_T^{miss} channel. These processes are estimated using high luminosity Powheg, or aMC@NLO generated NLO MC samples. These samples are normalized by the appropriate theoretical cross-sections calculated at NLO accuracy.

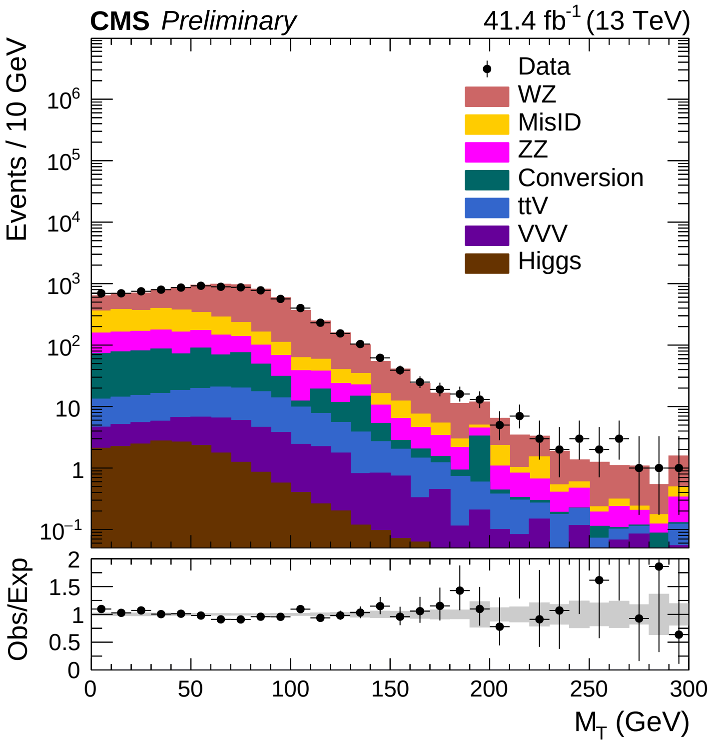


FIGURE 5.26: The M_T distribution in a selection of events with three light leptons and one OSSF pair with mass on-Z, $p_T^{miss} < 100$ GeV. The total SM background is shown as a stack of all contributing processes, and the gray band in the lower panel represents the statistical uncertainty on the expected background.

5.14 Validation Regions

After all the background methods are ready, we test the background prediction in the signal depleted regions, referred to as validation regions. These regions are not used to examine the signal and are removed from the signal regions. Figure 5.26 demonstrates the agreement between the expected background and observed data yields as a function of transverse mass M_T in signal-depleted 3 e/μ selections of events, where $M_T = (2p_T^{miss} {p_T}^{\ell}[1 - \cos\measuredangle(p_T^{miss}, {p_T}^{\ell})])^{1/2}$ and is calculated using the lepton that is not a part of the on-Z OSSF pair.

As described earlier, we measure fake rates in 2L1TOS, $p_T^{miss} < 50$ GeV, on-Z region. Figure 5.27 shows the invariant mass distribution of light leptons in the events 2L1TOS, $p_T^{miss} < 50$ GeV region. We do a closure test in data in the measurement region, as given in Figure 5.28. As an independent test of the background methods, we test our predictions in 2L1TOS, $p_T^{miss} < 50$ GeV regions with on-Z requirement reversed (off-Z) as given in Figure 5.29-5.30. We also test our predictions in 2L1TSS, $p_T^{miss} < 50$ GeV region as given in Figure 5.31. Both of these regions are completely orthogonal to the fake rate measurement region. We see good agreement between data and background predictions in all these regions. Since we use 2L1T OS/SS, $p_T^{miss} < 50$ GeV regions to measure and validate our fake rates, we exclude these from the signal regions.

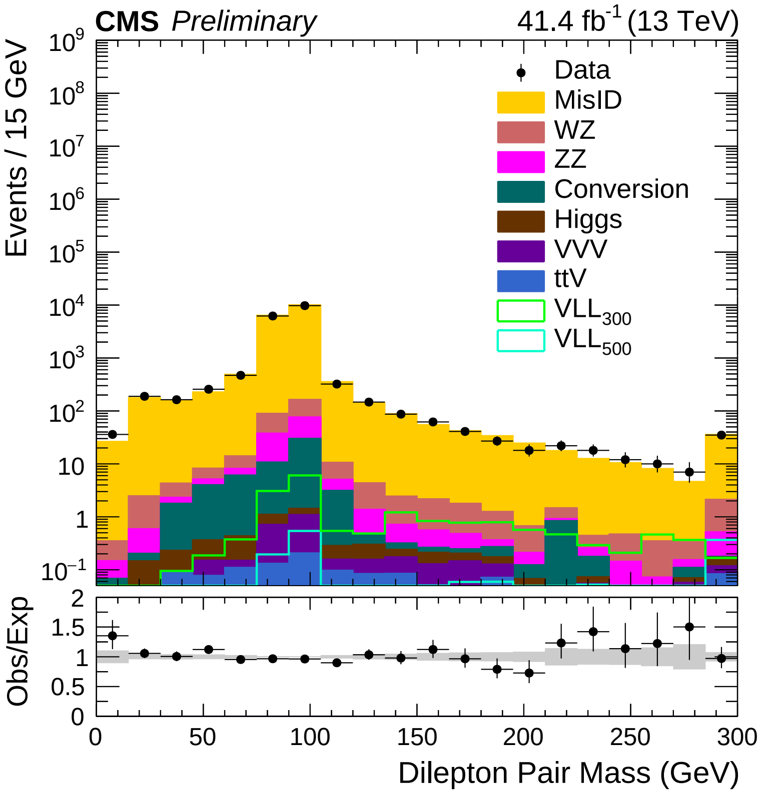


FIGURE 5.27: Dilepton Mass distribution in 2L1TOS, $p_T^{miss} < 50$ GeV region.

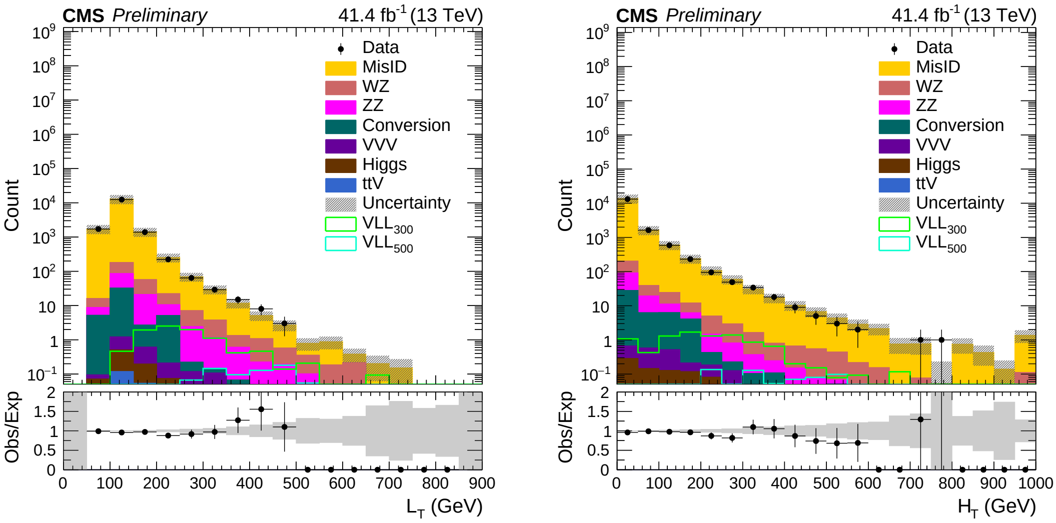


FIGURE 5.28: L_T (left) and H_T (right) distributions in 2L1TOS, $p_T^{miss} < 50$ GeV on-Z region. Grey band represents the statistical uncertainties.

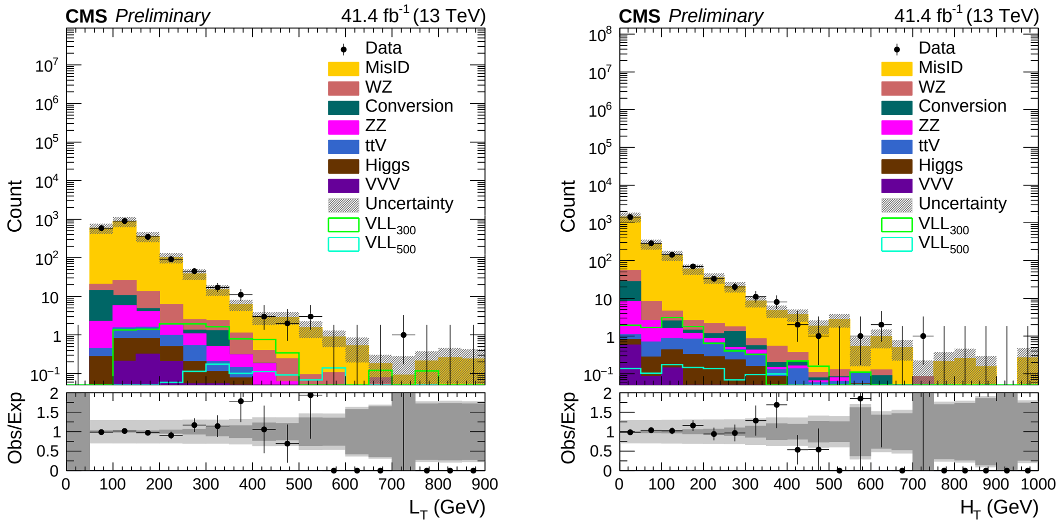


FIGURE 5.29: L_T (left) and H_T (right) distributions in 2L1TOS, $p_T^{miss} < 50$ GeV off-Z region.

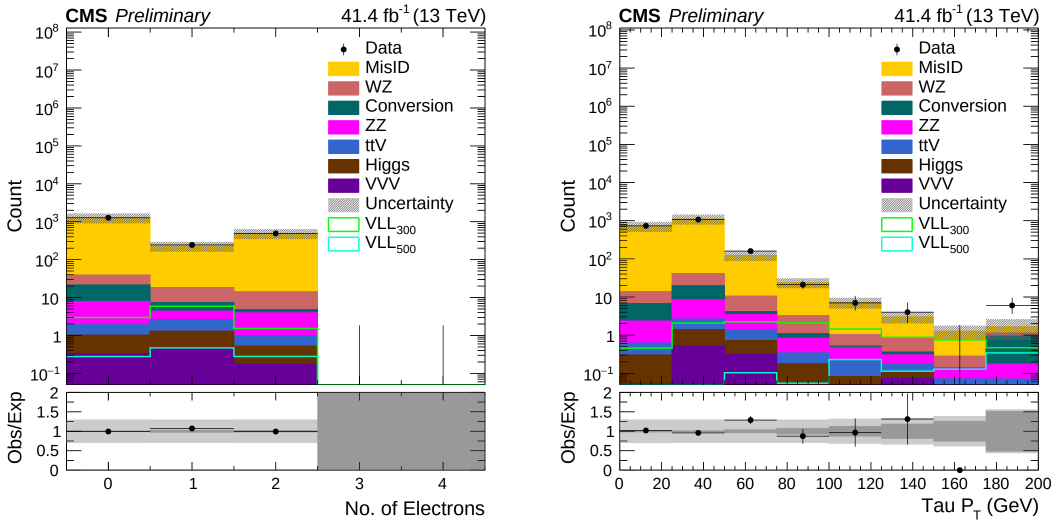


FIGURE 5.30: Number of electron distributions in 2L1TOS, $p_T^{miss} < 50$ GeV off-Z region.

By examining various validation regions, it is clear that the background methods we developed are performing very well and can be used for the search for BSM physics. However, before we test the predictions in the signal region, we need to estimate uncertainties on each of the background component. The systematic uncertainties for this search are summarized in the next section.

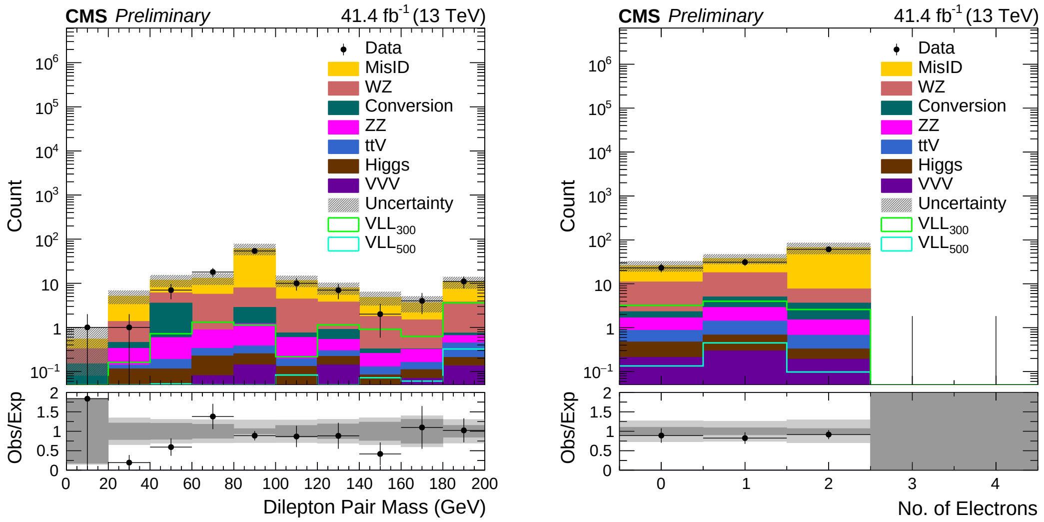


FIGURE 5.31: Invariant mass of dilepton pair (left) and number of electrons distributions in 2L1TSS, $p_T^{miss} < 50$ GeV region.

5.15 Pileup Uncertainties

The number interaction in a given bunch crossing is called pileup, and it is directly proportional to the total min-bias cross-section. We vary the min-bais cross-section up and down by the recommended uncertainty of 5%, from the central value and observe the effect on MC yield in various control/signal regions. Depending on the amount of pileup, the reconstruction of the physics objects gets affected. In most cases, the reconstruction and identification probability goes down with the increases. For composite objects such as p_T^{miss}, it is especially important, an incorrect pileup in the MC simulation may lead to imprecise p_T^{miss} calculation. To estimate the impact of the pile reweighting on the MC samples, we look at key distributions in various signal regions. Figure 5.32-5.33 show the impact of these variations in various signal regions.

We also check how the reconstructed number of vertices distribution looks if we vary the min-bias cross-section up and down. Figure 5.34 shows the number of reconstructed vertices in WZ enriched selections of events, assuming the nominal min-bias cross-section.

Figure 5.35 shows the number of reconstructed vertices in WZ enriched selections of events when the min-bias cross-section varied up and down by 5%. We observe that within the uncertainties arising from these variations, the data and the SM prediction agree well.

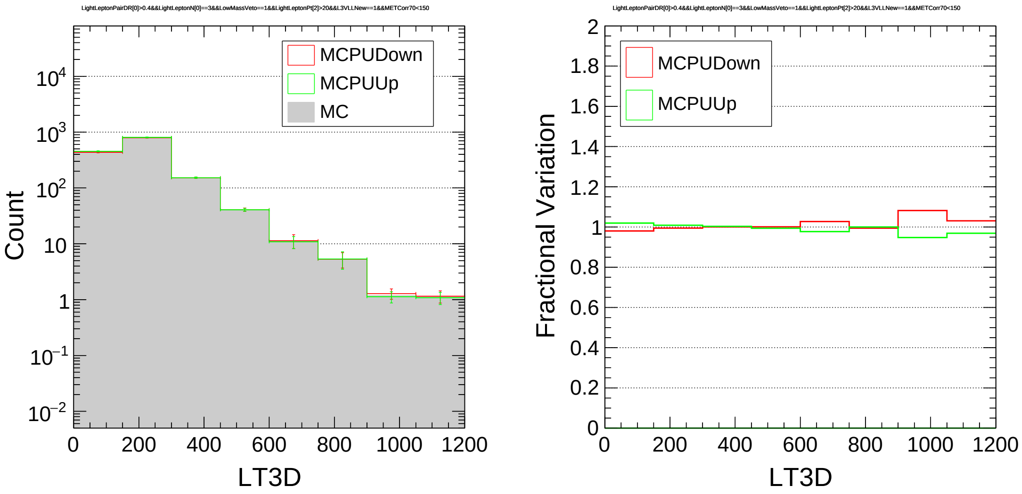


FIGURE 5.32: Impact of varying min-bias cross-section in 3L, low p_T^{miss} signal region.

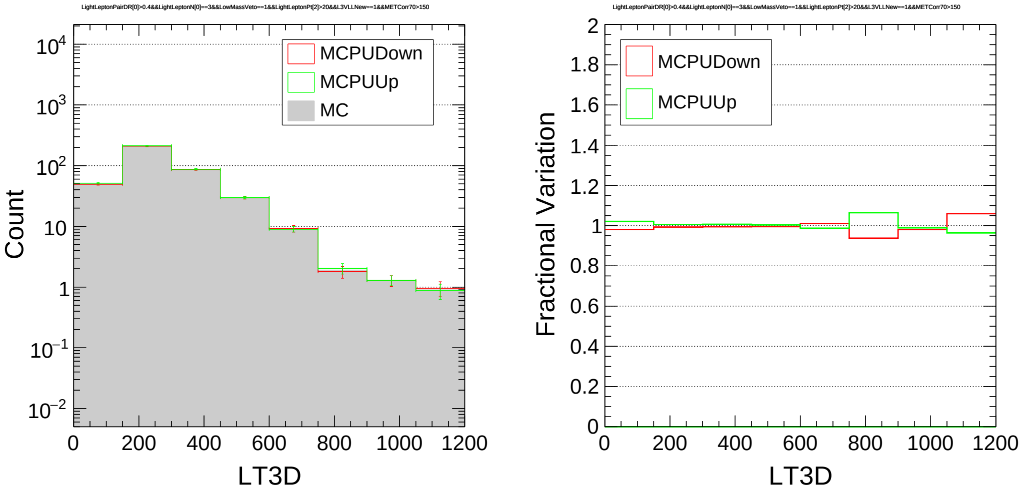


FIGURE 5.33: Impact of varying min-bias cross-section in 3L, high p_T^{miss} signal region.

5.16 Object Energy Scale and Resolution Uncertainties

We also evaluate the impact of varying the energy scale and resolutions of various objects used in the analysis. These parameters for each object are varied within their respected uncertainty, and the change in the yield of the background is observed in each signal region. Figure 5.36 shows some of the example variations for a subset of the signal regions, remaining variations are provided in the Appendix A. The impact of these variations is found to be 10% across signal regions.

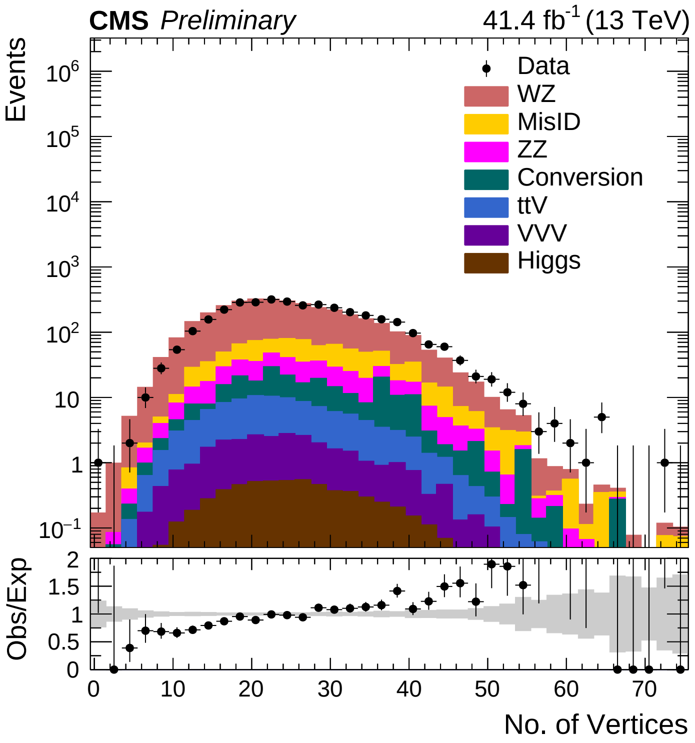


FIGURE 5.34: Number of vertices distribution in WZ enriched control region.

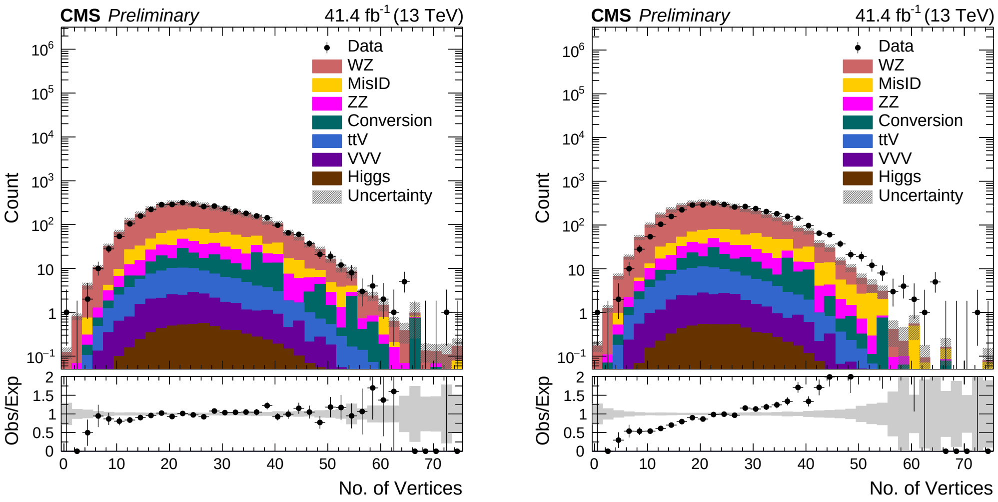


FIGURE 5.35: Number of vertices distribution in WZ enriched control region when the minbias cross section is increased by 5% (left), and when the minbias cross section is decreased by 5% (right).

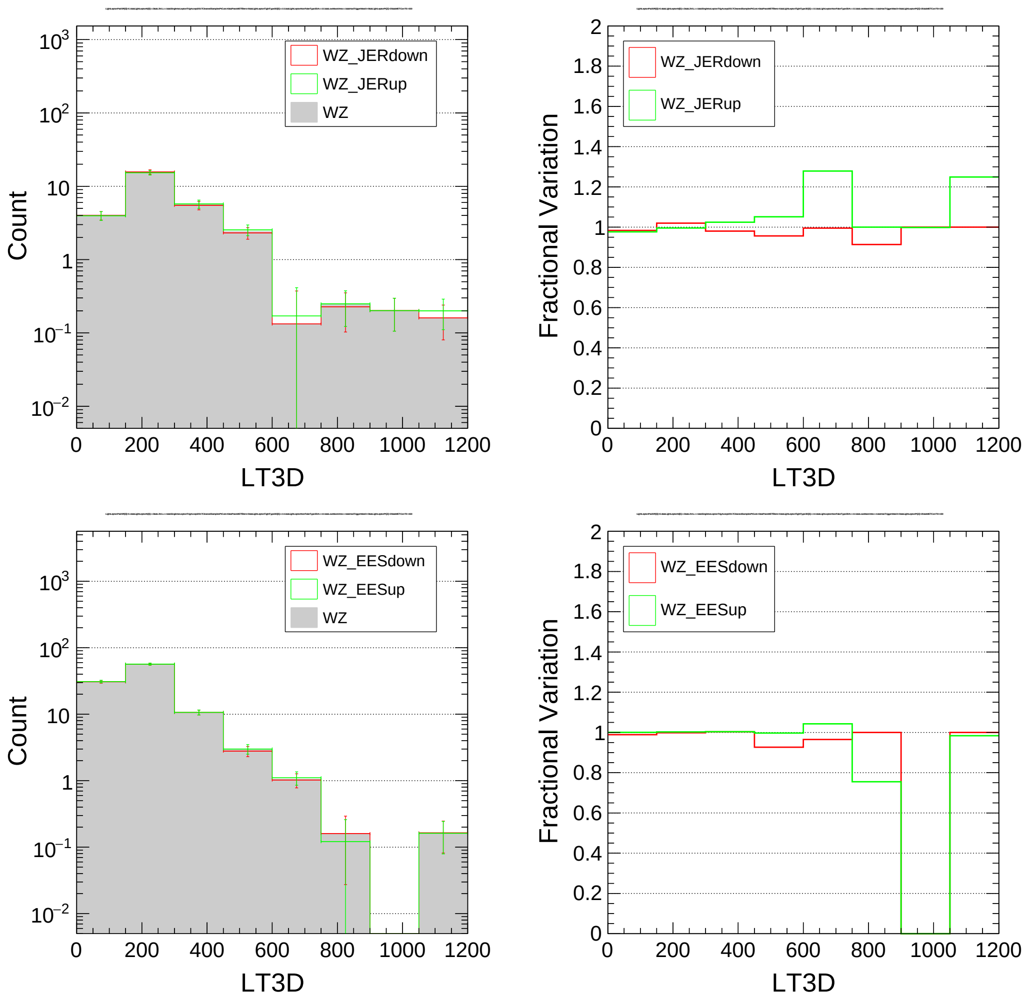


FIGURE 5.36: Impact of varying jet energy resolution (JER) on $\mathrm{L_T}$ distribution in three lepton channel with $\mathrm{p_T^{miss}} > 150$ *GeV* channel (upper plots) and the impact of varying electron energy scale (EES) in three lepton channel with $\mathrm{p_T^{miss}} < 150$ *GeV* channel (bottom plots).

5.17 Summary of the Systematics Uncertainties

The misidentified background contribution estimated via the matrix method has associated systematic uncertainties that arise due to the uncertainties of the respective prompt, and fake rates used in this method, as well as process-dependent variations to the rates. We vary the rates within the individual uncertainties and observe the change in background yield in all signal regions. These variations result in 20-40% systematic uncertainty on the misidentified lepton background. Corresponding plots for the 3L channel are given in Figure 5.18. For the tau channels, the plots are given in Fig. 5.19-5.20.

The WZ and ZZ backgrounds are normalized in dedicated control regions as described in section 5.10 and 5.11 respectively. We measure a relative uncertainty of 7(6)% on WZ(ZZ) background. Since background contributions due to all-prompt lepton events are estimated via simulated samples, many different sources of systematic uncertainty are considered to account for differences between MC and data events.

The uncertainties on the muon and electron identification, isolation efficiency scale factors are 2%, the tau identification and isolation uncertainties are 3%, which are applied per-lepton in each event. The trigger efficiency scale factors have an overall <2-3% impact per event. The jet, unclustered energy, and lepton energy (or momentum) scale uncertainties as well as jet and lepton resolution uncertainties are applied at the per-object level, where the corresponding object p_Ts are varied up and down within the recommended uncertainty range. The impact of these variations on the kinematic quantities is then observed, corresponding plots are given in Appendix A.3.

For the subdominant, rare background processes such as $t\bar{t}V$, triboson, or associated Higgs production, a 50% systematic uncertainty is applied on the theoretical normalization cross-sections to cover both PDF as well as renormalization and factorization scale uncertainties. Additionally, the recommended luminosity uncertainty of 2.3% [69] is applied to the MC based rare background estimates as well as the signal yields since these are not normalized to data but the theoretical cross-sections.

A summary of the uncertainty sources considered in this analysis and the corresponding typical variations in the affected processes is provided in Table 5.1.

Source of uncertainty	Typical impact on background yield (%)	Processes
MisID lepton backgrounds (matrix method)	20-40	Misidentified
Rare MC backgrounds ($t\bar{t}$V, VVV, Higgs)	50	Rare
Conversion MC backgrounds normalization	8	DY/$t\bar{t}$
WZ MC normalization	7	WZ
ZZ MC normalization	6	ZZ
Single lepton trigger	<3	WZ/ZZ/Rare/Signal
Lepton ID & isolation	$6-8$	WZ/ZZ/Rare/Signal
Electron energy scale and resolution	2-5	WZ/ZZ/Rare/Signal
Muon momentum scale and resolution	2-10	WZ/ZZ/Rare/Signal
Tau Energy Scale	<5	WZ/ZZ/Rare/Signal
Jet energy scale	5-10	WZ/ZZ/Rare/Signal
Jet energy resolution	5-10	WZ/ZZ/Rare/Signal
Unclustered energy scale	2-10	WZ/ZZ/Rare/Signal
Integrated luminosity	2.3	Rare/Signal
Pileup	<4	WZ/ZZ/Rare/Signal

TABLE 5.1: Systematic uncertainty sources and tentative typical variations observed in the affected background and signal yields.

Chapter 6

Search for Vectorlike Leptons

This chapter describes the search for the vectorlike leptons performed in multilepton final states using 41.4 fb^{-1} of proton-proton collision data collected by the CMS detector in 2017. The techniques used to estimate the standard model background are discussed in Chapter 5. Details of the model under consideration, the search strategy, and the results are presented in this chapter.

6.1 Vectorlike Leptons

A search for new phenomena beyond the standard model using events with multiple leptons, including hadronically decaying τ leptons, is performed. We are looking for vectorlike leptons (VLL) from a model [70] that is expected to produce signatures with energetic multilepton final states at the CERN LHC. VLLs are non-chiral, color-singlet, fermionic extensions of the SM leptons, whose left and right-handed components transform similarly under the SM gauge symmetries, and arise in a wide variety of models ranging from supersymmetry to extra dimensions [71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81]. VLLs are often classified by the SM generation that they primarily couple to and have lepton numbers similar to that of SM leptons.

The VLL models can be broadly classified into two categories as SU(2) singlet or doublet models. From an experimental perspective, the difference between these models is that the singlet VLL model allows only charged VLLs, while the doublet VLL model allows charged and neutral VLL. For example, the singlet model adds only one charged vectorlike electron (or muon or tau) and its antiparticle to the first generation. On the other hand, the doublet model also adds one neutral vectorlike neutrino along with its antiparticle. Therefore, the singlet model has two new particles, whereas the doublet model has four.

We are looking for an SU(2) doublet VLL extension to the SM with couplings to the third generation leptons [82]. The model introduces a vectorlike tau ($\tau^{'-}$), its antiparticle ($\tau^{'+}$) and a neutrino $\nu_\tau^{'}$ along with its antiparticle $\nu_{tau}^{'}$. At tree level, the $\tau^{'}$ and $\nu_\tau^{'}$ are mass degenerate whereas higher order radiative corrections predict $< 0.3\%$ relative mass splitting among these two states for VLL masses above 100 GeV. Therefore the vectorlike $\tau^{'}$ and $\nu_\tau^{'}$ are assumed to be degenerate in this analysis. At the LHC, these can be pair-produced in $\tau^{'\pm}\nu_\tau^{'}$, $\tau^{'+}\tau^{'-}$ and $\nu_\tau^{'}\nu_\tau^{'}$ channels, with subsequent decays of $\tau^{'}$ to $Z\tau$ or $H\tau$ and $\nu_\tau^{'}$ to $W\tau$ where the mass of the VLL is the only free parameter both in the production cross-section and the branching ratio calculations. We pursue a search for VLL in final states with multiple leptons (e, μ, τ) using proton-proton collision data collected by the CMS detector at the LHC in 2017.

The L3 collaboration placed a lower bound on additional heavy leptons around 100.2 GeV [83]. The ATLAS collaboration performed a search for heavy lepton resonances decaying into a Z boson and a lepton in a multilepton final state at $\sqrt{s} = 8$ TeV with 20.8 fb^{-1} of luminosity [84]. This search constrained the singlet VLL model and excluded VLL in the mass range of 114-176 GeV. However, there are no such constraints on the doublet VLL model. Given the existing constraints, we consider VLLs with masses more than 100 GeV.

6.2 Search Strategy

Events are primarily categorized as those with two light leptons and at least one hadronically decaying tau lepton, three light leptons, and four or more light leptons. In the 2L1T channel, we have a further division based on whether the two light leptons are of opposite-sign (OS) or same-sign (SS).

Events in all categories are selected using a single electron (muon) trigger with a p_T threshold of 35 (27) GeV. Also, mention the heavy signal here. The single lepton triggers are used to have a redundancy in firing the trigger for all categories resulting in overall higher trigger efficiencies and small correction factors for simulated samples. The leading light lepton is required to satisfy a p_T threshold of 28 GeV if it is a muon, and 38 GeV if it is an electron, for the corresponding single lepton triggers to be efficient. All the other trailing leptons are required to satisfy the p_T threshold of 20 GeV. In 2L1T and 4L categories, in the presence of more than one tau or more than four light lepton candidates, only the leading tau and the leading four light leptons are selected and used in the calculation of L_T.

The event classifications are done using the loose light lepton and tau definitions. Given the relatively high momenta of bosons and leptons originating from the decays of massive VLL particles and given the presence of neutrinos in the final state, we use L_T and p_T^{miss} to discriminate signal from SM contributions.

In order to achieve optimum sensitivity for the signal model, in each of the 2L1T OS/SS, 3L, and 4L categories the events are divided into low and high p_T^{miss} regions, where the L_T distribution in 150 GeV (50 GeV for 4L) bins are then utilized as the final discriminant. All control regions are vetoed in the signal region selections. Additionally, all events with a light-lepton pair mass < 12 GeV are vetoed regardless of the flavor and sign of the pair to suppress low mass quarkonia resonances. These search regions are detailed in Table 6.1, where OSSF refers to an opposite-sign same-flavor pair.

6.3 Results

We use L_T as a variable to examine the presence of the signal by comparing the standard model prediction to the observations in high and low p_T^{miss} regions for 3L, 4L, and 2Ll1T channels. Figure 6.2-6.3 show the L_T distribution in 3L and 4L signal regions. No statistically significant deviation from the standard model prediction is seen in these distributions.

Using these results the pair production of the VLLs is constrained using the CL_S modified frequentist technique at 95% confidence level [85, 86, 87, 88]. Multi-bin counting experiment approach is utilized for the statistical analysis. The uncertainties on the signal and background processes are characterized using nuisance parameters. The systematic

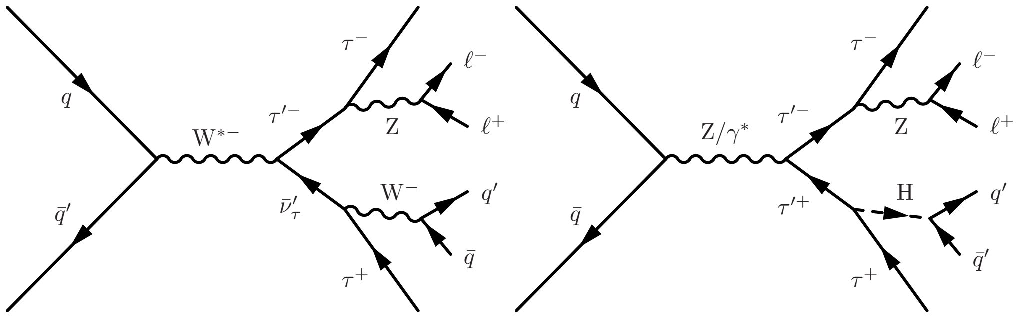


FIGURE 6.1: Associated production of (τ', ν'_τ) pairs via an off-shell W (left) and Z (right) boson, as well as possible subsequent decay chains that result in multileptonic final states.

N_{leptons}	p_T^{miss} (GeV)	Control Region Veto
$\geq 4\ e/\mu$	< 50 > 50	2 OSSF pairs on-Z and $p_T^{\text{miss}} < 50$ GeV
$3\ e/\mu$	< 150 > 150	OSSF pair on-Z and $p_T^{\text{miss}} < 100$ GeV, or OSSF pair below-Z and $p_T^{\text{miss}} < 50$ GeV, or OSSF pair below-Z and M_{3l} on-Z
$2\ e/\mu$ OS + $\geq 1\ \tau$	< 150 > 150	$p_T^{\text{miss}} < 50$ GeV
$2\ e/\mu$ SS + $\geq 1\ \tau$	< 150 > 150	$p_T^{\text{miss}} < 50$ GeV

TABLE 6.1: Signal regions as defined in this analysis. The on-Z selection is defined as an OSSF lepton pair with mass between 76 and 106 GeV.

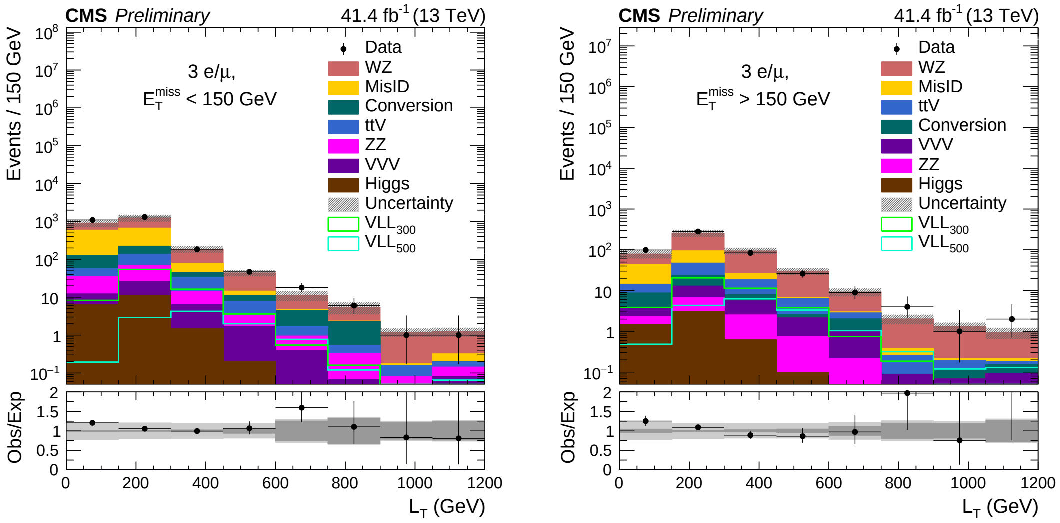


FIGURE 6.2: The L_T distribution for events with three light leptons with $p_T^{\text{miss}} < 150$ GeV (left) and with $p_T^{\text{miss}} > 150$ GeV (right). The total SM background is shown as a stack of all contributing processes. The predictions for signal models with $m_{\tau'/\nu'} = 300$ GeV and $m_{\tau'/\nu'} = 500$ GeV (sum of all production and decay modes) are also shown as solid lines. The hatched gray band in the upper panel and the dark and light gray bands in the lower panel represent the total, statistical, and systematic uncertainties on the expected background, respectively.

Signal Region	Discriminating variable							
	L_T (0 – 150 GeV)		L_T (150 – 300 GeV)		L_T (300 – 450 GeV)		L_T (450 – 600 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
$\geq 4\ e/\mu$, $p_T^{miss} < 50$	8	8.23 ± 0.73	46	45.22 ± 4.44	9	11.03 ± 1.19	3	2.85 ± 0.30
$\geq 4\ e/\mu$, $p_T^{miss} > 50$	6	5.5 ± 1.28	98	70.92 ± 16.24	24	14.85 ± 3.78	5	4.16 ± 1.11
$3\ e/\mu$, $p_T^{miss} < 150$	1098	897.72 ± 145.65	1314	1218.53 ± 153.76	185	179.64 ± 19.20	47	42.41 ± 4.86
$3\ e/\mu$, $p_T^{miss} > 150$	99	76.02 ± 10.65	281	247.27 ± 26.5	84	90.79 ± 10.14	26	28.67 ± 3.57
$2\ e/\mu$ OS + $\geq 1\ \tau$, $p_T^{miss} < 150$	4115	4486.89 ± 1308.15	1001	1062.74 ± 284.41	72	54.30 ± 12.42	7	7.14 ± 1.75
$2\ e/\mu$ OS + $\geq 1\ \tau$, $p_T^{miss} > 150$	50	53.49 ± 14.91	65	56.89 ± 11.63	11	14.82 ± 2.57	1	2.56 ± 0.55
$2\ e/\mu$ SS + $\geq 1\ \tau$, $p_T^{miss} < 150$	73	73.41 ± 16.76	47	40.3 ± 8.54	7	3.38 ± 0.71	0	0.72 ± 0.25
$2\ e/\mu$ SS + $\geq 1\ \tau$, $p_T^{miss} > 150$	8	3.38 ± 1.18	7	6.59 ± 1.56	1	1.85 ± 0.60	0	0.41 ± 0.17
	L_T (600 – 750 GeV)		L_T (750 – 900 GeV)		L_T (900 – 1050 GeV)		L_T (> 1050 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
$\geq 4\ e/\mu$, $p_T^{miss} < 50$	2	0.84 ± 0.08	0	0.6 ± 0.16	0	0.24 ± 0.03		
$\geq 4\ e/\mu$, $p_T^{miss} > 50$	3	1.36 ± 0.35	0	0.46 ± 0.11	0	0.25 ± 0.06		
$3\ e/\mu$, $p_T^{miss} < 150$	18	10.88 ± 2.33	6	5.26 ± 1.36	1	1.18 ± 0.28	1	1.03 ± 0.28
$3\ e/\mu$, $p_T^{miss} > 150$	9	8.35 ± 1.25	4	1.97 ± 0.44	1	1.23 ± 0.27	2	0.82 ± 0.25
$2\ e/\mu$ OS + $\geq 1\ \tau$, $p_T^{miss} < 150$	1	1.90 ± 0.60	0	0.82 ± 0.31				
$2\ e/\mu$ OS + $\geq 1\ \tau$, $p_T^{miss} > 150$	0	1.22 ± 0.34	0	0.41 ± 0.18				

TABLE 6.2: Signal regions table showing the yields for observed and expected events.

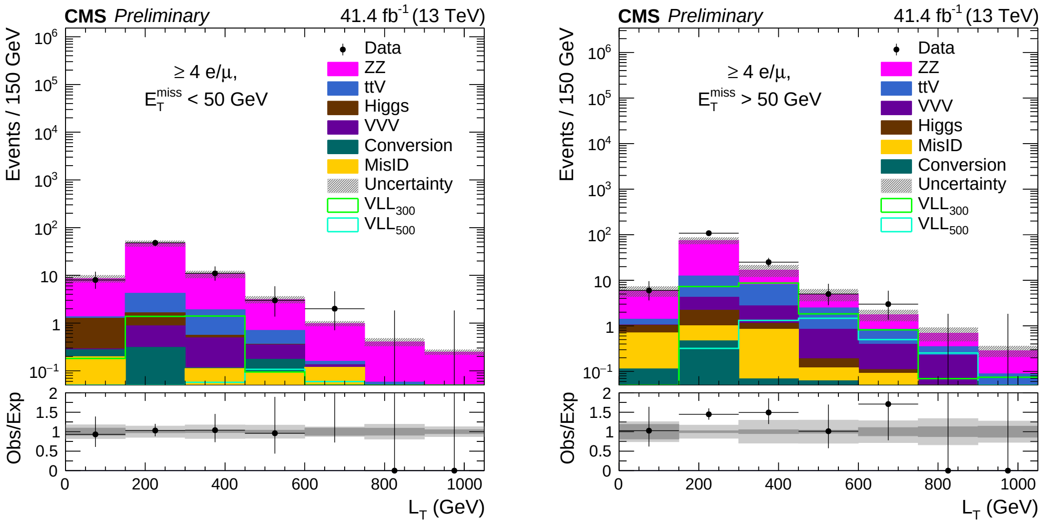


FIGURE 6.3: The L_T distribution for events with four or more light leptons with $p_T^{miss} < 50$ GeV (left) and with $p_T^{miss} > 50$ GeV (right). The total SM background is shown as a stack of all contributing processes. The predictions for signal models with $m_{\tau'/\nu'} = 300$ GeV and $m_{\tau'/\nu'} = 500$ GeV (sum of all production and decay modes) are also shown as solid lines. The hatched gray band in the upper panel and the dark and light gray bands in the lower panel represent the total, statistical, and systematic uncertainties on the expected background, respectively.

nuisance parameters are modeled by log-normal distributions, while gamma distributions are used for uncertainties, which are statistical. All the systematic uncertainties on the background and signal are considered fully correlated, while the statistical uncertainties are considered to be entirely uncorrelated across all channels. The observed and expected limits calculated in the 3L, 4L channels are shown in Figure 6.4. The 3L and 4L channels individually do not exclude the vectorlike leptons.

Figure 6.5-6.6 shows the L_T distribution in various 2L1T regions. We do not observe any sign of new physics even in these regions, as there is no statistically significant deviation from the standard model estimates. These results are also used to constrain the vectorlike leptons. The observed and expected limits in 2l1T channels are shown in Figure 6.7. The tau channels are very sensitive to the VLL signal under consideration. Using these channels, the VLLs with a mass between 160-660 GeV are excluded at a 95% confidence level.

Now all the individual channels are statistically combined to constrain the VLL model.

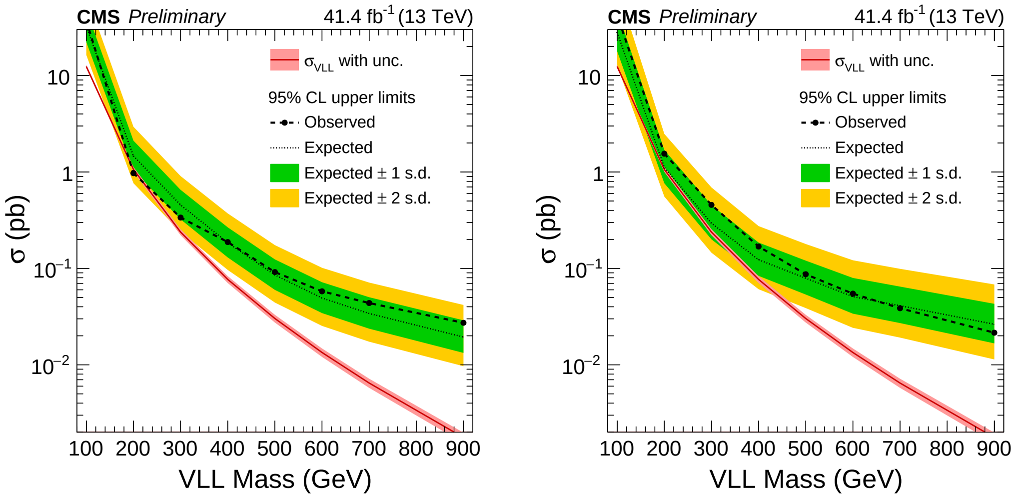


FIGURE 6.4: The 95% confidence level upper limits on the cross section for production of VLL pairs ($\tau'^{+}\tau'^{-}, \tau'^{\pm}\nu'$ and $\nu'\nu'$). Left in 3L channel, right in 4L channels. Also shown is the theoretical prediction for the section of the VLL pair production.

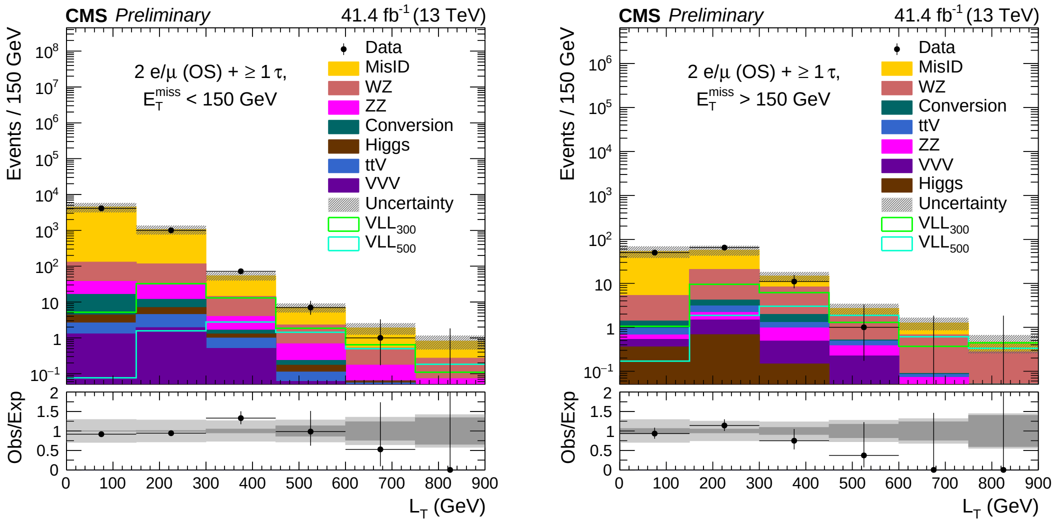


FIGURE 6.5: The L_T distribution for events with two opposite charge light leptons and a hadronically decaying τ lepton with $50 < p_T^{miss} < 150$ GeV (left) and with $p_T^{miss} > 150$ GeV (right). The total SM background is shown as a stack of all contributing processes. The predictions for signal models with $m_{\tau'/\nu'} = 300$ GeV and $m_{\tau'/\nu'} = 500$ GeV (sum of all production and decay modes) are also shown as solid lines. The hatched gray band in the upper panel and the dark and light gray bands in the lower panel represent the total, statistical, and systematic uncertainties on the expected background, respectively.

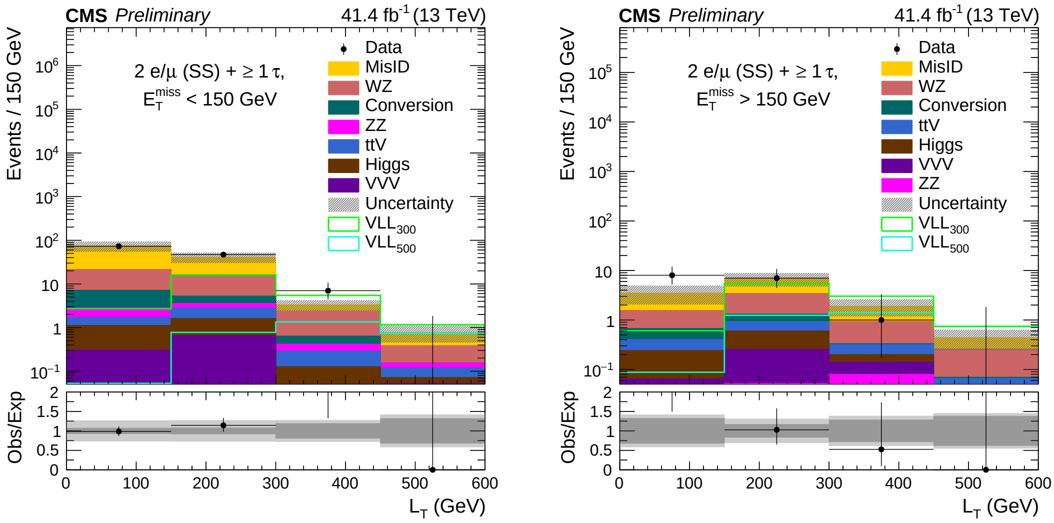


FIGURE 6.6: The L_T distribution for events with two same charge light leptons and a hadronically decaying τ lepton with $50 < p_T^{miss} < 150$ GeV (left) and with $p_T^{miss} > 150$ GeV (right). The total SM background is shown as a stack of all contributing processes. The predictions for signal models with $m_{\tau'/\nu'} = 300$ GeV and $m_{\tau'/\nu'} = 500$ GeV (sum of all production and decay modes) are also shown as solid lines. The hatched gray band in the upper panel and the dark and light gray bands in the lower panel represent the total, statistical, and systematic uncertainties on the expected background, respectively.

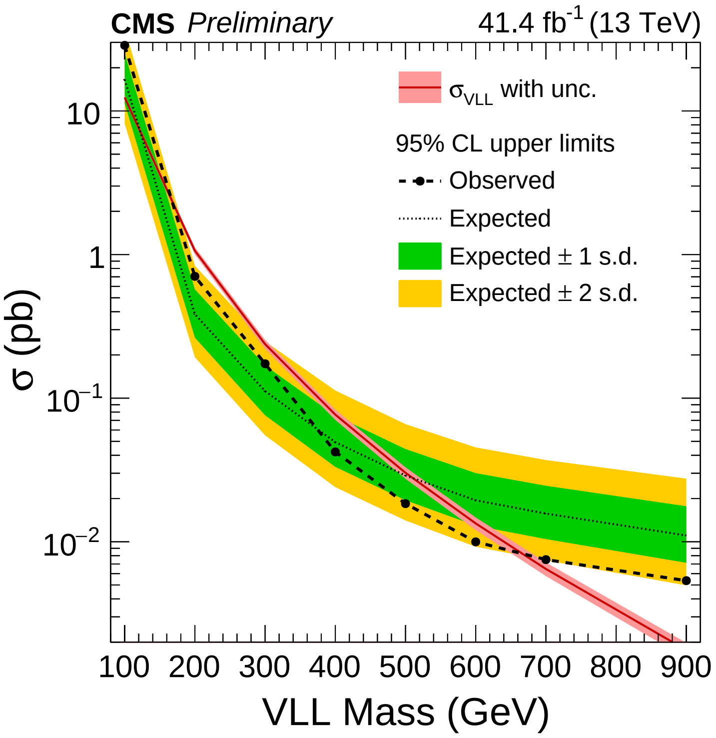


FIGURE 6.7: The 95% confidence level upper limits on the cross section for production of VLL pairs ($\tau'^{+}\tau'^{-}, \tau'^{\pm}\nu'$ and $\nu'\nu'$) in 2L1T channel. Also shown is the theoretical prediction for the section of the VLL pair production.

Figure 6.8 shows the expected significance of all the channels combined, and 95% confidence level limits on the VLL production are shown in Fig. 6.9. From the figure, we observe that observed (expected) limits on the VLL pair production are 130-690 GeV(110-560).

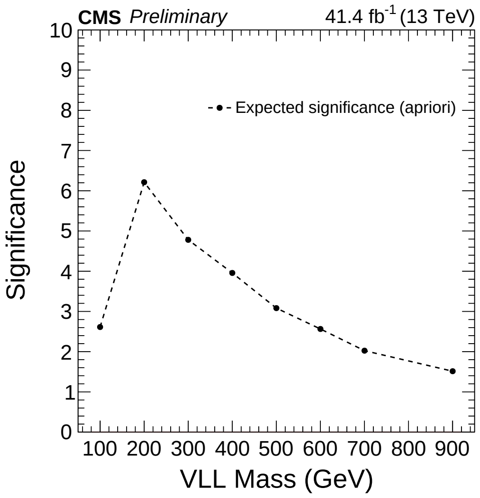


FIGURE 6.8: The expected signal significance as a function of VLL mass after statistically combing all the 3L, 4L and 2L1T channels in the analysis.

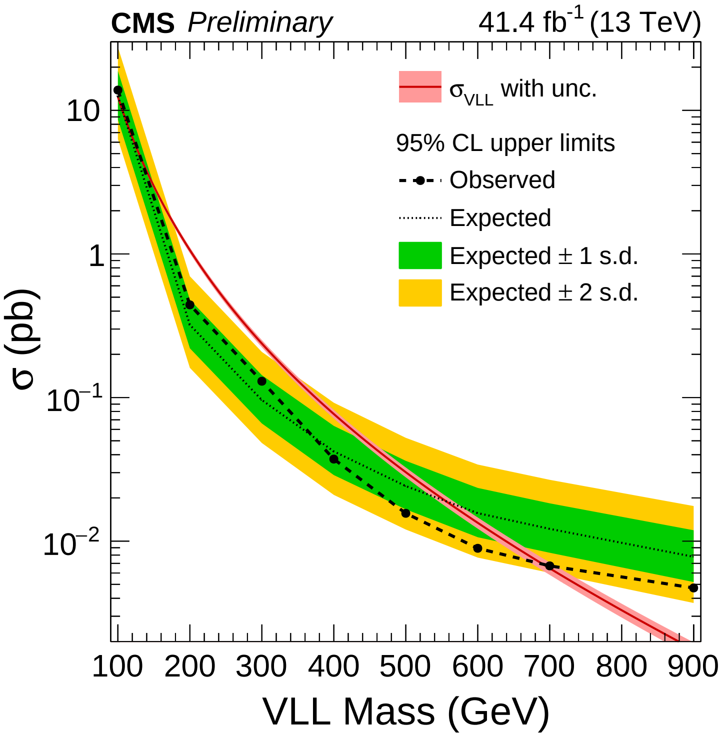


FIGURE 6.9: The 95% confidence level upper limits on the cross section for production of VLL pairs ($\tau'^{+}\tau'^{-}$, $\tau'^{\pm}\nu'$ and $\nu'\nu'$) after statisticallt combing the 3L, 4L and 2L1T channels. The theoretical prediction for the cross section of the VLL pair production is also shown. VLLs in mass range 130-690 GeV are excluded in this analysis.

Chapter 7

Search for the Evidence of Type-III Seesaw Mechanism

This chapter describes a search for the evidence of the type-III seesaw mechanism conducted in the channels with at least three electrons or muons using 35.9fb^{-1} of $\sqrt{s}$=13 TeV data collected by CMS experiment at LHC in 2016. These results are published in reference [68]. The techniques used for the estimation of the standard model background are very similar to the ones described in Chapter 5. Here I will discuss the strategy and results for this search.

7.1 Type-III Seesaw Model

In the SM neutrinos are massless; however, the observation of the neutrino oscillations by various experiments have already shown that the neutrinos have a non-zero mass[89]. This observation means that the SM is not complete, and there must be some physics beyond SM that can explain the experimental results. Type-III seesaw mechanism is one such BSM model that accounts for the small non-zero neutrino mass by the addition of new massive Σ fermions. In the seesaw model under consideration [90] neutrinos are taken to be Majorana particles, and mediation of the heavy fermions gives rise to their mass. The heavy fermions are charged Dirac and neutral Majorana lepton triplet of a SU(2) symmetry group. The heavy fermions couple to both leptons and the Higgs.

At LHC, these heavy seesaw fermions can be produced via both charged and neutral electroweak currents. Production of these fermions is depicted in Fig. 7.1.

The pairs of $\Sigma^0\Sigma^\pm$ and $\Sigma^+\Sigma^-$ decay via electroweak and Higgs bosons into lighter SM particles, naturally leading to a final state with many leptons. There are total of 27 production and decay models summarized in Table 7.1. A search for type-III seesaw fermions is performed in channels with at least three light leptons. One example of decay mode of seesaw fermions with three leptons in the final state is shown in Fig. 7.2.

The CTEQ6.6 and MSTW2008nlo90cl parton distribution functions (PDFs) [91, 92] with an accuracy of NLO + NLL, are used to determine the production cross-section of the heavy fermion pairs. In the current implementation of this model, the heavy fermions mix equally with all three SM lepton flavors. The mixing to each lepton is taken to be $V_\mathrm{e} = V_\mu = V_\tau = 10^{-6}$ to ensure the decays of the seesaw fermions are prompt.

The signal samples are produced centrally by CMS MC team using the FeynRules Model file of Ref. [93]. The events are generated using MadGraph 5_aMC@NLO v5.2.2 using the NNPDF30_lo_as_0130_nf_4 pdf [94]. The decays of boson in the events and hadronization of the colored particles are done using Pythia 8.

7.2 Search Strategy

This search only utilizes the events with only electrons and muons, the 3L and 4L categories defined in Chapter 5. All the combinations of electrons and muons are considered. For example, the 3L region has $\mu\mu\mu$, μ, and eee events; similarly, all combinations in 4L region are considered as well. The data and MC samples and triggers used in this search are given in Appendix A.1. The leading and subleading $\mathrm{p_T}$ leptons are required to satisfy $\mathrm{p_T}$ thresholds of 25 GeV and 15 GeV. All other leptons in the event are required to fulfill a $\mathrm{p_T}$ threshold of 10 GeV. The events further categorized by lepton flavor, charge, and other kinematic quantities, as detailed in Table 7.2.

Each event is primarily classified by the maximum number of distinct opposite-sign same-flavor (OSSF) dilepton pairs that can be formed; i.e., both $\mu^+\mu^-\mu^-$ and $\mu^+\mu^-e^-$ are OSSF1, $\mu^+\mu^+e^-$ is OSSF0, and $\mu^+\mu^-e^+e^-$ is OSSF2. The invariant mass of the OSSF dilepton pair closest to the mass of Z boson is used and referred to as MOSSF. If

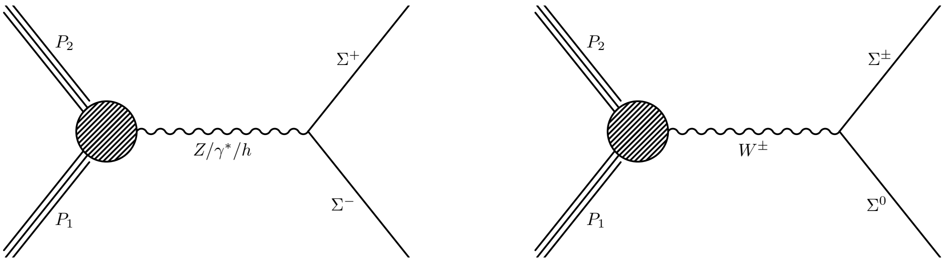


FIGURE 7.1: Feynman diagrams corresponding to the production of the type-III seesaw fermions in electroweak interactions.

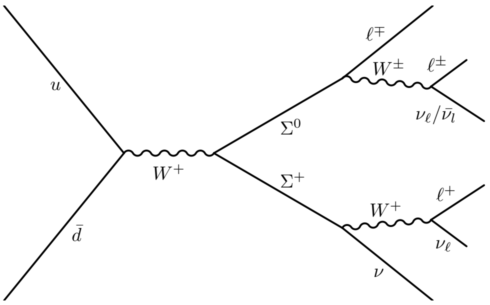


FIGURE 7.2: Feynman diagram example of the fermion production and decay in the Type-III Seesaw model.

Production Channel	Decay Mode	BR(%)
$\Sigma^+\Sigma^0$, *or* $\Sigma^-\Sigma^0$, *or* $\Sigma^+\Sigma^-$	$\rightarrow H\ell^+H\nu$	5.7
	$\rightarrow H\ell^+W^\pm\ell^\mp$	12.1
	$\rightarrow H\ell^+Z\nu$	6.0
	$\rightarrow W^+\nu H\nu$	12.1
	$\rightarrow W^+\nu W^\pm\ell^\mp$	25.8
	$\rightarrow W^+\nu Z\nu$	12.9
	$\rightarrow Z\ell^+H\nu$	6.0
	$\rightarrow Z\ell^+W^\pm\ell^\mp$	12.9
	$\rightarrow Z\ell^+Z\nu$	6.4

TABLE 7.1: The signal production modes for the seesaw samples and the branching fraction of seesaw particles to various final states are shown. The charged and the neutral seesaw particles are considered to be mass degenerate, and their mass is taken to be ($\Sigma \gtrsim 700$ GeV) for these calculations.

$\mathrm{N}_{\mathrm{leptons}}$	OSSF Pair	Discriminating variable	Control Region Veto
3	on-Z	M_T	$\mathrm{p}_\mathrm{T}^\mathrm{miss} > 100$ GeV
	above-Z	$L_\mathrm{T} + \mathrm{p}_\mathrm{T}^\mathrm{miss}$	—
	below-Z	$L_\mathrm{T} + \mathrm{p}_\mathrm{T}^\mathrm{miss}$	$\mathrm{p}_\mathrm{T}^\mathrm{miss} < 50$ and $M(3\ell)$ is on-Z
	none	$L_\mathrm{T} + \mathrm{p}_\mathrm{T}^\mathrm{miss}$	—
≥ 4	1 pair	$L_\mathrm{T} + \mathrm{p}_\mathrm{T}^\mathrm{miss}$	—
	2 pairs	$L_\mathrm{T} + \mathrm{p}_\mathrm{T}^\mathrm{miss}$	$\mathrm{p}_\mathrm{T}^\mathrm{miss} > 50$ GeV if on-Z

TABLE 7.2: Signal regions as used in the search for the heavy seesaw fermions are shown.

MOSSF in an event is found to be within Z window (91 ± 10) GeV, the event is labeled as on-Z. If the MOSSF is below the Z mass window, the event is called below-Z. If MOSSF is above the Z window, the event is referred to as above-Z. The control regions which were utilized in the background estimation are vetoed from the signal regions as given in Table 7.2. In this analysis, a veto to suppress contributions due to low-mass dilepton resonances is also applied. If the mass of any opposite sign lepton pair found to be less than 12 GeV, the event is not considered.

For distinguishing the signal from the SM background contributions, kinematic quantities such as L_{T} play a major role due to the relatively high momenta of lepton in the event. This is especially true for decay modes such as $\Sigma^{\pm} \rightarrow \ell^{\pm} Z \rightarrow \ell^{\pm}\ \ell'^{\pm}\ell'^{\mp}$ where all of the daughter particles of the heavy fermion are visible in the detector. However, for certain other decay modes where there are multiple neutrinos in the final states such as $\Sigma^0 \rightarrow H\nu \rightarrow WW\nu$, $\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ become very sensitive. To remain sensitive for all decay modes, $L_{\mathrm{T}} + \mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ is used as the final discriminant. However, M_{T} is chosen as the discriminating variable for the events having a topology similar to the leptonic WZ decays. In each region the $L_{\mathrm{T}} + \mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ (M_{T}) distribution is divided on to 8 bins of 150 GeV in the range of 0-1200 GeV. The M_{T} distribution is divided into eight bins of 100 GeV in the range of 0-800 GeV. The last bin in all distributions contains the overflow events. This division yields a total of 48 orthogonal signal regions as summarized in Table 7.2

7.3 Backgrounds and Uncertainties

Chapter 5 describes the techniques used to estimate multileptonic backgrounds using the data collected in 2017. To a large extent, the background estimation techniques remain very similar to those described in Chapter 5. In this section, I will summarize the backgrounds for type-III seesaw search while focusing on the important differences to the Chapter 5.

The primary source of the prompt background for three leptons final state is the WZ process, while for the four-lepton channel, the dominant contribution comes from ZZ production. Simulated samples estimate the prompt backgrounds. Before using these samples in the signal regions, the cross-section for simulated samples is normalized using a dedicated control sample in data for both WZ and ZZ. A control sample in data with

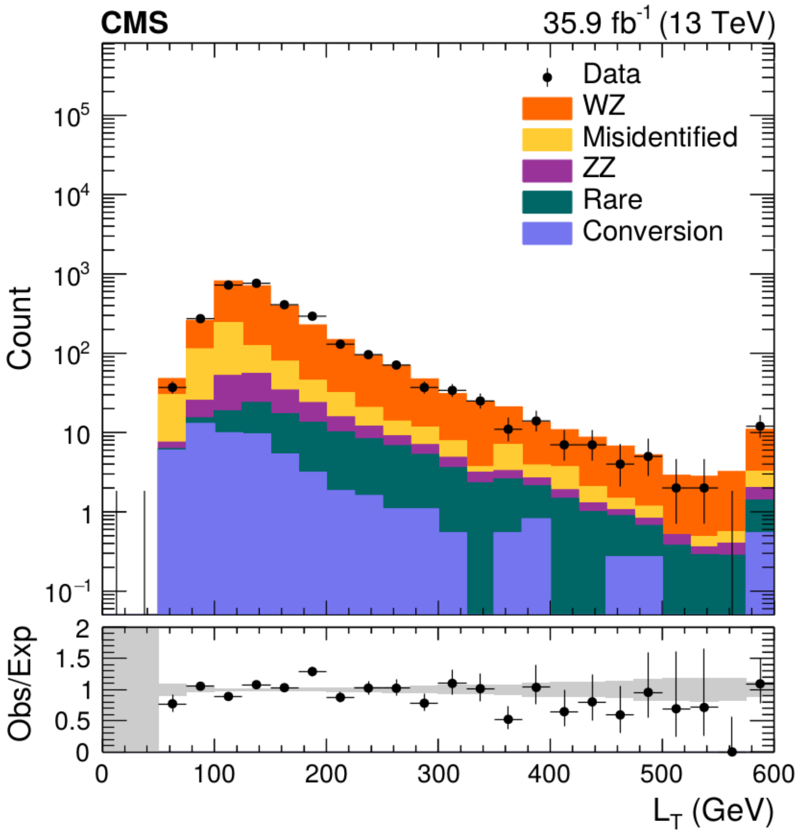


FIGURE 7.3: L_T distribution in WZ enriched selection of events is shown.

3 e/μ events with $50p_T^{miss} < 100$ GeV and $M_T > 30$ is created to normalize the WZ simulation, the normalization scale factor is found to be 1.15±0.08. Figure 7.3 shows L_T of three leptons in this region after applying the appropriate normalization to the WZ simulation sample. To normalize the ZZ simulated sample, a selection of events having $4e/\mu$ with $p_T^{miss} < 50$ and 2 OSSF onZ pairs is created. The normalization factor for the ZZ sample is found to be 1.20±0.06. Figure 7.4 shows the invariant mass of the four-lepton in this region after applying the appropriate normalization factor to the ZZ sample. Other smaller backgrounds such as Higgs, $t\bar{t}V$ are also estimated using simulated samples. For the misidentified lepton backgrounds, the method is very similar to that described in Chapter 5. The significant difference lies in the parameterization of the fake rate for electrons and muons. Here we use the number of particles in a jet matched to the fake lepton for the fake rate parametrization to minimize the spread in rates for DY+jets and $t\bar{t}$ samples. For the VLL search, we used the p_T of the jet matched to the fake lepton as a parameter in fake rate measurement. Other implementation of the matrix method remains precisely the same.

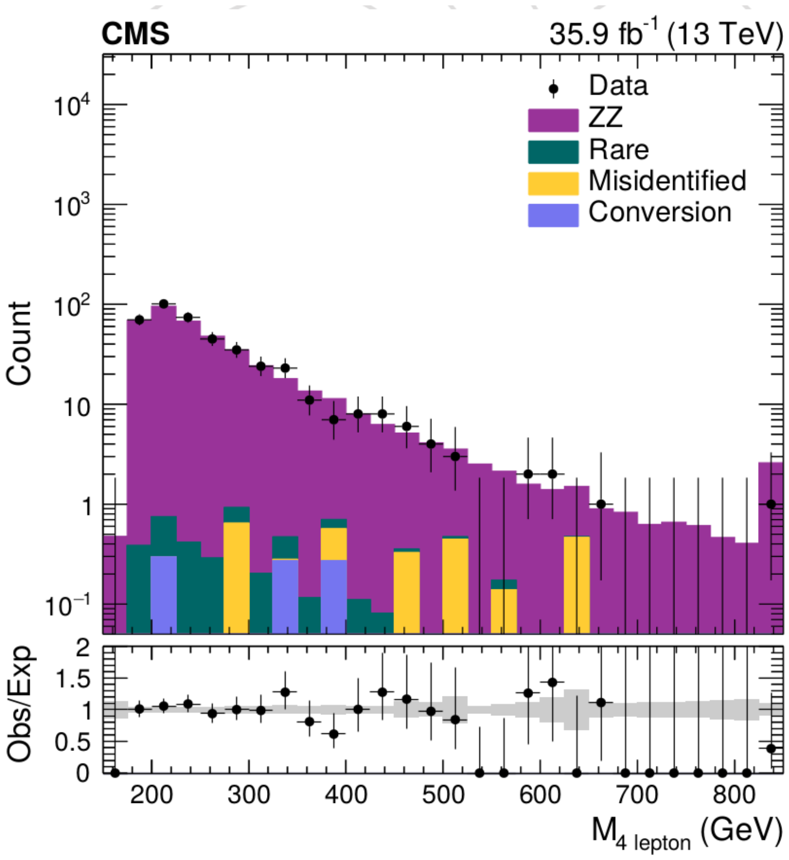


FIGURE 7.4: The invariant mass of the 4 leptons in a ZZ enriched 4L region is shown.

For a smaller photon conversion background, a photon-proxy method [68] is used, which relies on the ratio of reconstructed photons to leptons (measured in conversion enriched region) to predict the contribution from photon conversions to the background in the signal regions. The total contribution from photon conversions is less than 5% of the total background.

7.4 Results

The 3 lepton, below-Z signal region is presented in Fig. 7.5, which shows the $L_{\mathrm{T}} + \mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ distribution in the 3L below Z signal region. We observe a good agreement between the data and standard model prediction. Additionally, L_{T}, $\mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$, primary vertex multiplicity, and jet multiplicity distributions for this selection are shown in Fig. 7.6, all these distributions also show a good agreement. The seesaw signal with mass 380 GeV and 700 GeV are overlayed on top of the stacked SM backgrounds to demonstrate the expected signal yield. Fig. 7.7-7.9 depicts the $L_{\mathrm{T}} + \mathrm{p}_{\mathrm{T}}^{\mathrm{miss}}$ (or M_{T}) distributions in 3L (on-Z, above-Z, and OSSF0) and 4L (OSSF1, OSSF2) regions.

We observe no statistically significant excess in the various signal regions. The CLs method [85, 86, 87, 88] is used to constrain the production of the heavy seesaw fermions at 95% confidence level, where a counting experiment is performed in each signal bin for the statistical analysis. The uncertainties on the mean values of the expected signal and background yields are treated as nuisance parameters modeled by log-normal and gamma distributions for systematic and statistical uncertainties, respectively. Statistical uncertainties on the signal and background yields are assumed to be fully uncorrelated, whereas all other systematic uncertainties are considered to be fully correlated among the 48 signal channels. The expected signal efficiencies and the expected significances as a function of the Σ mass by combining all 48 bins are shown in Fig. 7.10.

The upper limit on the production cross-section of the Σ pairs are shown in Fig. 7.11, assuming the Σ particles couple equally to all SM generation of leptons. The Σ particles with the mass below 840 GeV are excluded at 95% CL, using this data. The uncertainty on the signal yield due to the choice of scales and PDF are also shown in Fig. 7.11. The uncertainties are calculated to be around 5-15%. Figure 7.12 to 7.13 shows the exclusion limits when the couplings of the seesaw fermions to the SM lepton generations are varied on b_e-b_τ axis. It is observed that the limits become weaker and weaker as the coupling to third generation (τ lepton) is increased.

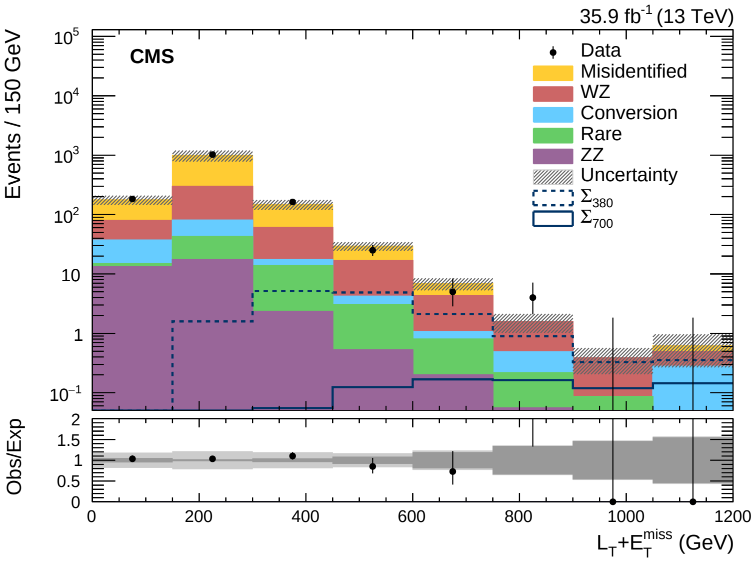


FIGURE 7.5: $L_T + p_T^{miss}$ distribution for the 3 lepton, below-Z signal region as defined in Table 7.2. The expected yields of the Seesaw signal model with $M_\Sigma = 380$ GeV and $M_\Sigma = 700$ GeV are overlaid. The last bin includes overflow events, and the uncertainty bands in the ratio plots include statistical (dark) and systematic (light) uncertainties.

Signal Region	Discriminating variable							
	$L_T + p_T^{miss}$ (0 − 150 GeV)		$L_T + p_T^{miss}$ (150 − 300 GeV)		$L_T + p_T^{miss}$ (300 − 450 GeV)		$L_T + p_T^{miss}$ (450 − 600 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
L3BelowZ	183	177 ± 32	1022	990 ± 210	163	148 ± 28	25	29.4 ± 4.9
L3AboveZ	313	260 ± 56	1038	930 ± 160	246	235 ± 34	59	60.8 ± 8.1
L3OSSF0	228	232 ± 56	654	710 ± 180	107	93 ± 22	16	18.1 ± 4.4
L3OSSF1	2	6.2 ± 1.8	18	28.2 ± 5.5	14	10.5 ± 3.5	6	4.3 ± 1.6
L3OSSF2	21	28.2 ± 6.1	57	54 ± 11	25	21.1 ± 4.6	6	5.9 ± 1.5
	$L_T + p_T^{miss}$ (600 − 750 GeV)		$L_T + p_T^{miss}$ (750 − 900 GeV)		$L_T + p_T^{miss}$ (900 − 1050 GeV)		$L_T + p_T^{miss}$ (> 1050 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
L3BelowZ	5	6.9 ± 1.6	4	1.57 ± 0.50	0	0.38 ± 0.17	0	0.61 ± 0.34
L3AboveZ	16	20.9 ± 2.7	6	6.88 ± 1.17	3	2.78 ± 0.65	3	3.58 ± 0.73
L3OSSF0	7	4.3 ± 1.2	3	1.97 ± 0.65	0	0.34 ± 0.18	0	0.71 ± 0.36
L3OSSF1	1	1.39 ± 0.55	1	0.41 ± 0.24	0	0.78 ± 0.65	0	0.08 ± 0.05
L3OSSF2	2	2.09 ± 0.56	0	0.99 ± 0.27	0	0.48 ± 0.18	1	0.32 ± 0.10
	M_T (0 − 100 GeV)		M_T (100 − 200 GeV)		M_T (200 − 300 GeV)		M_T (300 − 400 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
L3OnZ	816	840 ± 150	137	115 ± 20	22	19.1 ± 3.4	5	7.2 ± 1.5
	M_T (400 − 500 GeV)		M_T (500 − 600 GeV)		M_T (600 − 700 GeV)		M_T (> 700 GeV)	
	Obs.	Exp	Obs.	Exp	Obs.	Exp	Obs.	Exp
L3OnZ	2	1.67 ± 0.51	1	1.24 ± 0.44	1	0.77 ± 0.28	0	0.63 ± 0.27

TABLE 7.3: The table shows the expected and observed signal region yields.

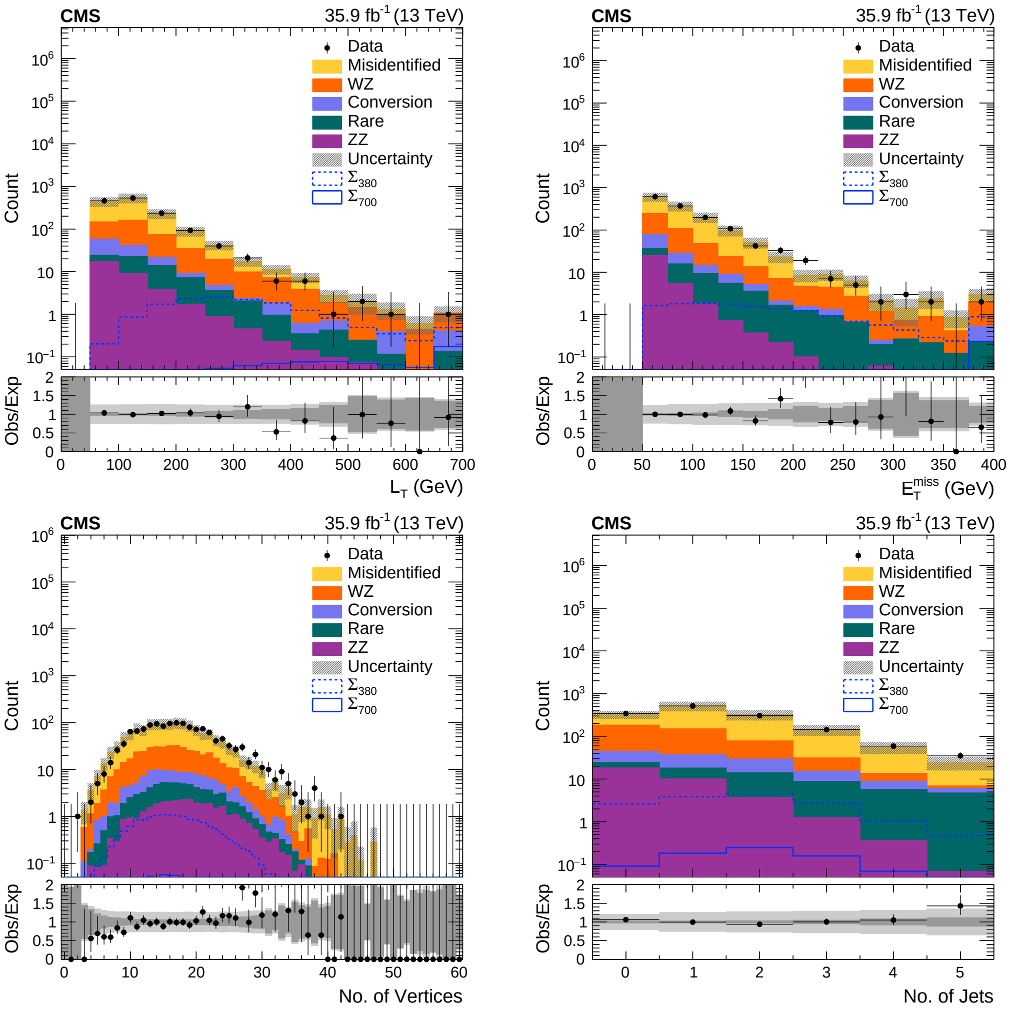


FIGURE 7.6: $\mathrm{L_T}$ (upper left), $\mathrm{p_T^{miss}}$ (upper right), primary vertex multiplicity (lower left), and jet multiplicity (lower right) distributions for the 3 lepton, below-Z signal region as defined in Table 7.2. The expected yields of the Seesaw signal model with $\mathrm{M}_\Sigma = 380$ GeV and $\mathrm{M}_\Sigma = 700$ GeV are overlaid. The last bins include overflow events, and the uncertainty bands in the ratio plots include statistical (dark) and systematic (light) uncertainties.

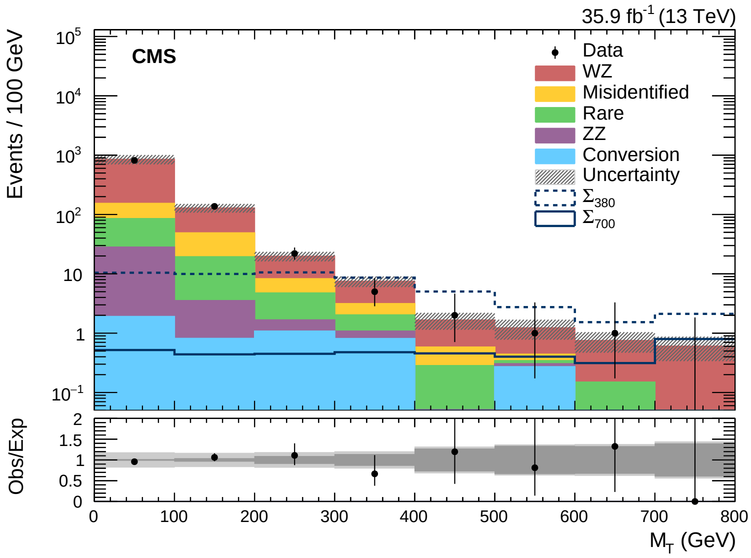


FIGURE 7.7: M_T distribution in the 3 lepton on-Z signal region as defined in Table 7.2. The expected yields of the Seesaw signal model with $M_\Sigma = 380$ GeV and $M_\Sigma = 700$ GeV are overlaid. The last bin includes overflow events, and the uncertainty bands in the ratio plot include statistical (dark) and systematic (light) uncertainties.

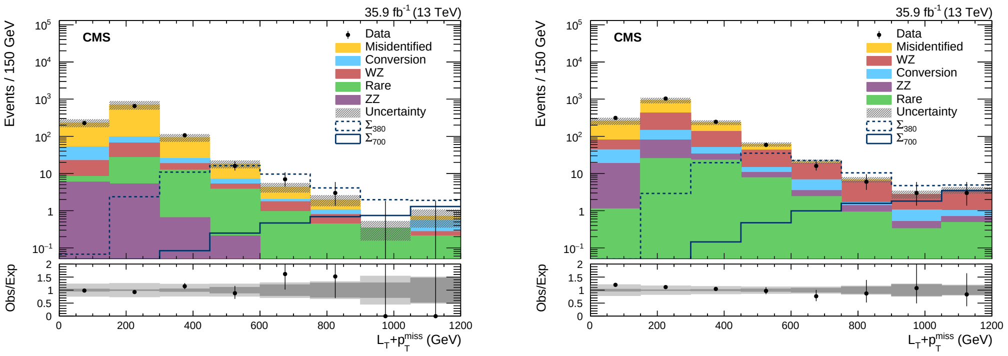


FIGURE 7.8: $L_T + p_T^{miss}$ distributions in the 3 lepton OSSF0 (left) and above-Z (right) signal regions as defined in Table 7.2. The expected yields of the Seesaw signal model with $M_\Sigma = 380$ GeV and $M_\Sigma = 700$ GeV are overlaid. The last bins include overflow events, and the uncertainty bands in the ratio plots include statistical (dark) and systematic (light) uncertainties.

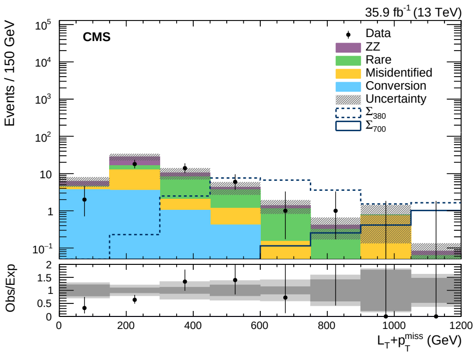


FIGURE 7.9: $L_T + p_T^{miss}$ distributions in the 4 lepton OSSF1 (left) and OSSF2 (right) signal regions as defined in Table 7.2. The expected yields of the Seesaw signal model with $M_\Sigma = 380$ GeV and $M_\Sigma = 700$ GeV are overlaid. The last bins include overflow events, and the uncertainty bands in the ratio plots include statistical (dark) and systematic (light) uncertainties.

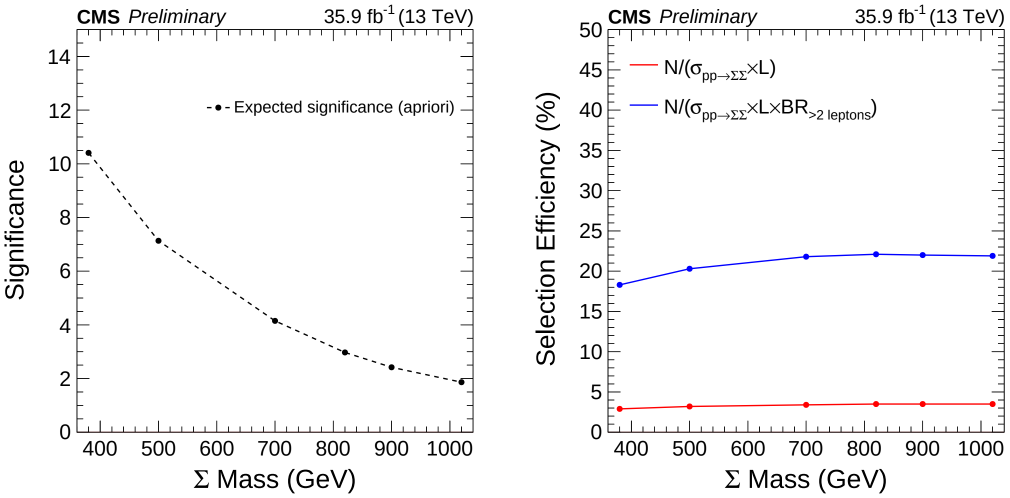


FIGURE 7.10: Expected (apriori) signal significance calculated using the profileLikelihood mode of the Higgs Combination Tool is shown as function of the Σ mass (left). The expected signal efficiency as a function of the Σ mass is shown (right) assuming the democratic mixing scenario.

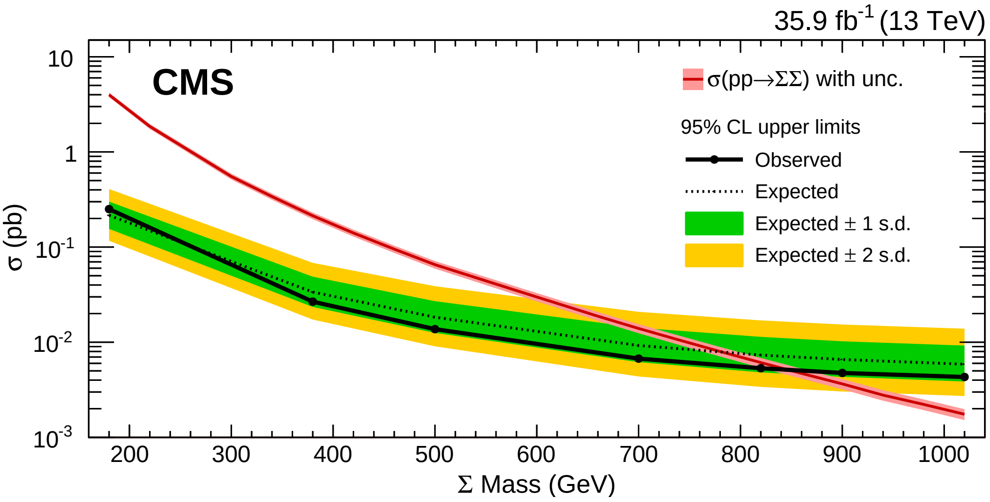
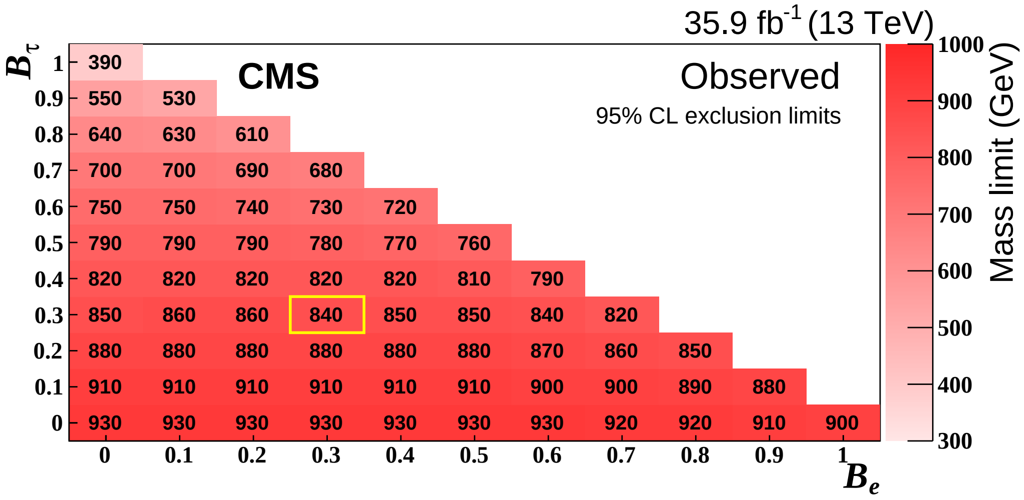


FIGURE 7.11: Observed exclusion limit at 95% CL on the production cross-section of Σ pairs, as calculated with the asymptotic CLs mode of the Higgs Combination Tool. All systematic uncertainties listed in Table 5.1 are included.



B_τ \ B_e	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	390										
0.9	550	530									
0.8	640	630	610								
0.7	700	700	690	680							
0.6	750	750	740	730	720						
0.5	790	790	790	780	770	760					
0.4	820	820	820	820	820	810	790				
0.3	850	860	860	840	850	850	840	820			
0.2	880	880	880	880	880	880	870	860	850		
0.1	910	910	910	910	910	910	900	900	890	880	
0	930	930	930	930	930	930	930	920	920	910	900

FIGURE 7.12: Two dimensional electron-tau branching ratio dependent observed exclusion limits at 95% CL on the production cross-section of Σ pairs, as calculated with the asymptotic CLs mode of the Higgs Combination Tool.

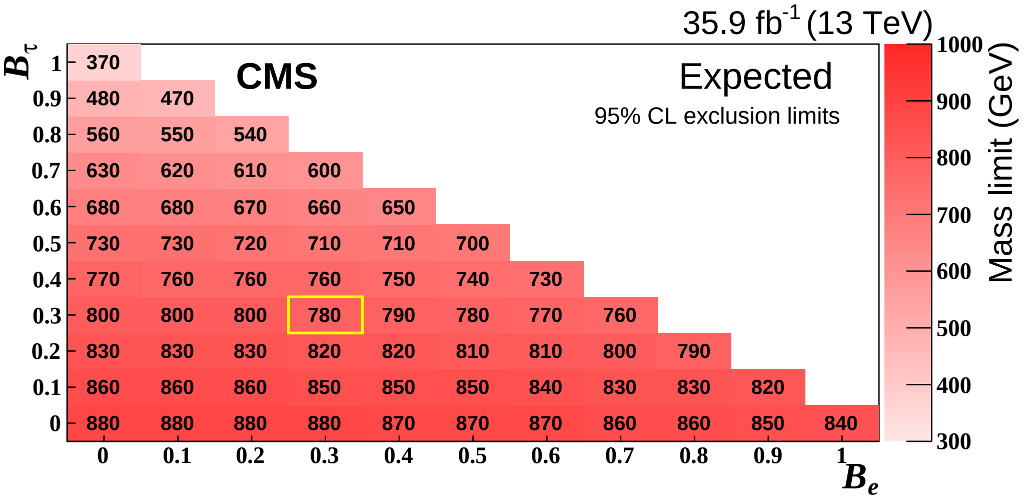


FIGURE 7.13: Two dimensional electron-tau branching ratio dependent Expected exclusion limits at 95% CL on the production cross-section of Σ pairs, as calculated with the asymptotic CLs mode of the Higgs Combination Tool.

Chapter 8

Summary

Searches for beyond standard model physics are carried out in multilepton signatures at LHC using data collected by CMS detector. The search regions are categorized based on the number of electrons, muons and hadronically decaying taus, resulting in regions with two light leptons and at least one hadronic tau, three light leptons and at least four light leptons. The tau regions are further divided into two channels based on the charge of the light lepton pair. Sum of the lepton transverse momentum and the missing transverse momentum are used as discriminating variables to distinguish signal from the background.

The primary background for the search comes from the diboson standard model processes, WZ, and ZZ for three and four-light lepton channels, respectively. These backgrounds are estimated using simulated samples. For the tau channels, the important background comes from the DY+jets and $\mathrm{t\bar{t}}$ processes, where one lepton is hadronic activity in detector misidentified as a lepton. The misidentified backgrounds are estimated using a data-driven matrix method.

Unfortunately, no signs of new physics have been observed, and data is consistent with the standard model prediction. The results are used to constrain models of beyond standard model physics. The vectorlike leptons coupling to the third generation (tau) of standard model leptons is excluded in the mass range of 130-690 GeV. Proton-proton collision dataset corresponding to 41.4 fb^{-1} collected in 2017 at the center of mass energy of 13 TeV is used for the search [95]. Another search for the evidence of the type-III seesaw mechanism is performed in the light lepton channels using a 35.9 fb^{-1} of data collected in 2016 at 13 TeV center of mass energy. The heavy seesaw fermions with the mass below 840 GeV are excluded assuming the seesaw fermions couple equally to the standard model leptons [68].

In the year 2018, LHC experiments have collected close to 60 fb^{-1} of additional data. Moreover, LHC is expected to deliver about 300 fb^{-1} of data in its phase -1 operation. The additional dataset will make it possible to extend the VLL search below 130 GeV. For heavy vectorlike leptons and seesaw fermions, most of the sensitivity comes from the regions with small statistics resulting in large uncertainties. Additional data will help in minimizing those uncertainties and improve the sensitivity of the search. Backgrounds such as $\mathrm{t\bar{t}W}$, $\mathrm{t\bar{t}Z}$, and backgrounds, including Higgs boson, have a large uncertainty of 50%. With more data, these backgrounds could be better estimated using the dedicated control samples. For signals which couple to the third generation of the lepton, the misidentified tau background is most dominant. In the current analysis, it has systematic uncertainty between 30-40% due to the large uncertainty on the fake rates. With more data, fake rates for the leptons can be measured much more precisely, which in turn would mean lower uncertainties on this background component. Precise estimation of backgrounds would significantly enhance our ability to look for BSM physics, and to measure standard model processes more precisely.

LHC is delivering more and more data each year with an increase in the instantaneous luminosity of the proton collisions. I believe we have fascinating times ahead of us in search of the unknown.

Appendix A

Appendix

A.1 Datasets, Triggers, and Simulated Samples

In this section, the details of the datasets, trigger and simulated samples used in this thesis are discussed. For the search for vectorlike lepton, we used the 41.4 fb^{-1} of proton-proton collision data collected by CMS. The list of datasets is provided in Table A.1. The data is collected by single muon and single electron triggers having a $\mathrm{p_T}$ threshold of 27 and 35 GeV respectively. The collected data is stored in SingleMuon and SingleElectron primary datasets. Every single muon and single electron primary dataset is split into five periods based on the time of the data collection. For each of these periods, the run number and the integrated luminosity collected is also provided. The run number provides a unique identity to data collected in a specific period by the CMS detector. Before selecting the events for analysis, some event filters are applied to filter out spurious events due to the electronic noise, not ideal beam conditions and other known hardware issues. For reconstructing the events, CMS has developed a software package called CMS Software (CMSSW) [96]. CMSSW is used to reconstruct every event used in the analysis.

The prompt backgrounds and the signal used in this thesis are modeled using the MC samples. The background samples are produced using CMSSW centrally by the collaboration, while the signal samples for the vectorlike leptons are produced privately. The privately produced samples are validated with respect to a few signal samples produced

Dataset Name	Run Range	L [pb^{-1}]
/SingleMuon/Run2017B-17Nov2017-v1/MINIAOD	297020-299329	4802
/SingleMuon/Run2017C-17Nov2017-v1/MINIAOD	299337-302029	9630
/SingleMuon/Run2017D-17Nov2017-v1/MINIAOD	302030-303434	4235
/SingleMuon/Run2017E-17Nov2017-v1/MINIAOD	303435-304826	9268
/SingleMuon/Run2017F-17Nov2017-v1/MINIAOD	304911-306462	13433
/SingleElectron/Run2017B-17Nov2017-v1/MINIAOD	297020-299329	4802
/SingleElectron/Run2017C-17Nov2017-v1/MINIAOD	299337-302029	9630
/SingleElectron/Run2017D-17Nov2017-v1/MINIAOD	302030-303434	4235
/SingleElectron/Run2017E-17Nov2017-v1/MINIAOD	303435-304826	9268
/SingleElectron/Run2017F-17Nov2017-v1/MINIAOD	304911-306462	13433

Table A.1: Single lepton data samples and corresponding luminosities in 2017.

centrally. Table A.2 lists the simulation samples used in the analysis as well as the theoretical cross section for each sample.

While searching for the seesaw fermions, the data collected in 2016 was used. This data was collected using di-lepton triggers given in Table A.4. Each trigger path stores the data into MuonEG, DoubleMuon, and DoubleEG primary datasets respectively. Each of the primary datasets are divided into eight periods based on the time of the data collection. The list of all the datasets used in this search is given in Table A.3. The table also contains the corresponding run number and the integrated luminosity for each dataset. A list of simulated samples used in this search is given in Table A.5. The table also lists the corresponding production cross-section of each sample at LHC.

A.2 Lepton Prompt and Fake Rates

In this section, the prompt and fake rates used in the matrix method for the prediction of the misidentified background are presented. The measurement of the prompt rates for electrons and muons is performed as a function of lepton $\mathrm{p_T}$, η and the number of tracks (N_{Tracks}) in the event under consideration. The fake rates for the light leptons are parameterized as a function if lepton $\mathrm{p_T}$, η and the $\mathrm{p_T}$ of jet nearest to lepton($\Delta R < 0.4$). Some of the measurements are already discussed in the Chapter 5, remaining measurements are provided here.

Figure A.1 shows the prompt rates for electrons as measured in data while Fig. A.2 depicts the measurements performed in MC samples. The electron fake rates are as measured in MC samples are provided in Fig. A.3 and A.4. The muon prompt rate

Sample Name	Cross-section [pb]
/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	4.43
/ZZTo4L_13TeV_powheg_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	1.256*1.1
/WW_TuneCP5_13TeV-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	115
/WW_DoubleScattering_13TeV-pythia8_TuneCP5/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	1.64
/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	87.31
/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	87.31
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	5765.4
/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10_ext1-v1/MINIAODSIM	5765.4
/ST_tW_top_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	19.56
/ST_tW_antitop_5f_NoFullyHadronicDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	19.56
/TTZToLL_M-1to10_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.0493
/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.2043
/WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.0001471
/WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X$_m c2017\_realistic\_v10 - v1/MINIAODSIM$	0.0002339
/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.000652
/ttH_HToZZ_4LFilter_M125_13TeV_powheg2_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.000337
/GluGluToContinToZZTo2e2mu_13TeV_MCFM701_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.0032*1.7
/GluGluToContinToZZTo2e2tau_13TeV_MCFM701_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.0032*1.7
/GluGluToContinToZZTo2mu2tau_13TeV_MCFM701_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.0032*1.7
/GluGluToContinToZZTo4e_13TeV_MCFM701_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.0016*1.7
/GluGluToContinToZZTo4mu_13TeV_MCFM701_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.0016*1.7
/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.2529
/TTTT_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-PU2017_94X_mc2017_realistic_v11-v1/MINIAODSIM	0.0091
/TTTJ_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.00048
/TTTW_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v3/MINIAODSIM	0.00086
/TTWW_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10_ext1-v2/MINIAODSIM	0.0078
/TTWZ_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.00294
/TTZZ_TuneCP5_13TeV-madgraph-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v2/MINIAODSIM	0.00156
/ttHToNonbb_M125_TuneCP5_13TeV-powheg-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.2151
/bbH_HToZZTo4L_M125_13TeV_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.00032
/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11_ext1-v1/MINIAODSIM	0.01212
/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUGenV7011_pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11_ext1-v1/MINIAODSIM	0.001034
/WWW_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11-v1/MINIAODSIM	0.2086
/WWZ_4F_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11-v1/MINIAODSIM	0.1651
/WZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11-v1/MINIAODSIM	0.05565
/ZZZ_TuneCP5_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v11-v1/MINIAODSIM	0.01398
/tZq_ll_4f_ckm_NLO_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17MiniAOD-94X_mc2017_realistic_v10-v1/MINIAODSIM	0.09418

TABLE A.2: Background MC samples.

Dataset Name	Run Range	L [pb^{-1}]
/MuonEG/Run2016B-03Feb2017_ver2-v2/MINIAOD	272007 - 275376	5788
/MuonEG/Run2016C-03Feb2017-v1/MINIAOD	275657 - 276283	2573
/MuonEG/Run2016D-03Feb2017-v1/MINIAOD	276315 - 276811	4248
/MuonEG/Run2016E-03Feb2017-v1/MINIAOD	276831 - 277420	4009
/MuonEG/Run2016F-03Feb2017-v1/MINIAOD	277772 - 278808	3102
/MuonEG/Run2016G-03Feb2017-v1/MINIAOD	278820 - 280385	7540
/MuonEG/Run2016H-03Feb2017_ver2-v1/MINIAOD	280919 - 284035	8391
/MuonEG/Run2016H-03Feb2017_ver3-v1/MINIAOD	284036 - 284044	215
	Total luminosity:	35867
/DoubleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD	272007 - 275376	5788
/DoubleMuon/Run2016C-03Feb2017-v1/MINIAOD	275657 - 276283	2573
/DoubleMuon/Run2016D-03Feb2017-v1/MINIAOD	276315 - 276811	4248
/DoubleMuon/Run2016E-03Feb2017-v1/MINIAOD	276831 - 277420	4009
/DoubleMuon/Run2016F-03Feb2017-v1/MINIAOD	277772 - 278808	3102
/DoubleMuon/Run2016G-03Feb2017-v1/MINIAOD	278820 - 280385	7540
/DoubleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD	280919 - 284035	8391
/DoubleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD	284036 - 284044	215
	Total luminosity:	35867
/DoubleEG/Run2016B-Run2016B-03Feb2017_ver2-v2/MINIAOD	272007 - 275376	5788
/DoubleEG/Run2016C-03Feb2017-v1/MINIAOD	275657 - 276283	2573
/DoubleEG/Run2016D-03Feb2017-v1/MINIAOD	276315 - 276811	4248
/DoubleEG/Run2016E-03Feb2017-v1/MINIAOD	276831 - 277420	4009
/DoubleEG/Run2016F-03Feb2017-v1/MINIAOD	277772 - 278808	3102
/DoubleEG/Run2016G-03Feb2017-v1/MINIAOD	278820 - 280385	7540
/DoubleEG/Run2016H-03Feb2017_ver2-v1/MINIAOD	280919 - 284035	8391
/DoubleEG/Run2016H-03Feb2017_ver3-v1/MINIAOD	284036 - 284044	215
	Total luminosity:	35867

Table A.3: Dilepton data samples and corresponding luminosities in 2016.

Dataset	HLT Path Name	Run Range
DoubleEG	HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ	272007-284044
DoubleMuon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL	272007-280385
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL	272007-280385
	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ	280919-284044
	HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ	280919-284044
MuonEG	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL	272007-280385
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL	272007-280385
	HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL_DZ	280919-284044
	HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL_DZ	280919-284044

Table A.4: List of dilepton triggers used in the type-III seesaw analysis. All triggers are unprescaled and a logical or is to be understood over those valid in the same dataset and run range.

Sample Name	Cross-section [pb]
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†_ext1-v2/‡	5765.4
/DY1JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†-v1/‡	1016×1.54527
/DY2JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†-v1/‡	331.4×1.54527
/DY3JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†-v1/‡	96.36×1.54527
/DY4JetsToLL_M-50_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†-v1/‡	51.40×1.54527
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†-v1/‡	87.31
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/†_ext1-v1/‡	87.31
/WWTo2L2Nu_13TeV-powheg/†-v1/‡	12.178
/WZTo3LNu_TuneCUETP8M1_13TeV-powheg-pythia8/†-v1/‡	4.42965
/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/†-v1/‡ (cross-check)	4.712
/ZZTo4L_13TeV_powheg_pythia8/†-v1/‡	1.256
/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/†-v1/‡	0.2086
/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/†-v1/‡	0.1651
/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/†-v1/‡	0.05565
/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/†-v1/‡	0.01398
/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/†_ext1-v1/‡	0.2529
/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/†_ext1-v3/‡	0.2043
/TTTT_TuneCUETP8M2T4_13TeV-amcatnlo-pythia8/†-v1/‡	0.009103
/ttH_HToZZ_4LFilter_M125_13TeV_powheg2_JHUgenV6_pythia8/†-v1/‡	0.000337
/GluGluHToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/†-v1/‡	0.01212
/VBF_HToZZTo4L_M125_13TeV_powheg2_JHUgenV6_pythia8/†-v1/‡	0.001034
/WminusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV6_pythia8/†-v1/‡	0.0001471
/WplusH_HToZZTo4L_M125_13TeV_powheg2-minlo-HWJ_JHUgenV6_pythia8/†-v1/‡	0.0002339
/ZH_HToZZ_4LFilter_M125_13TeV_powheg2-minlo-HZJ_JHUgenV6_pythia8/†-v1/‡	0.000652

† RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6
‡ MINIAODSIM

Table A.5: Background MC samples used in type-III seesaw analysis.

measurements are shown in Fig. A.5 and A.6 in data and MC samples respectibly. The fake rates for muon measured in the MC samples are shown in Fig. A.7-A.8. The tau leptons that decay hadronically the fake rate is measured as separately for 1-prong and 3-prong decays. For each tau category (1-prong/3-prong) the fake rates are measured in bins of tau $\mathrm{p_T}$, η and the $\mathrm{p_T}$ of a jet matched to the tau candidate within $\Delta R < 0.4$. An additional correction factor as a function of the number of tracks in the event is also measured as discussed in Chapter 5. Figure A.9-A.11 show the fake rates for taus measured in data as well as DY+jets and $\mathrm{t\bar{t}}$ simulated samples.

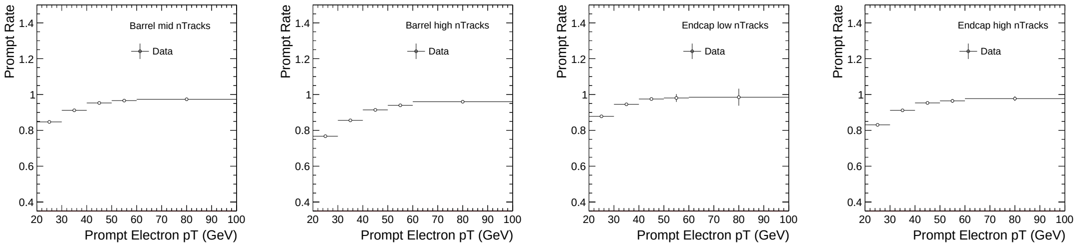


FIGURE A.1: Electron prompt rates as measured in data.

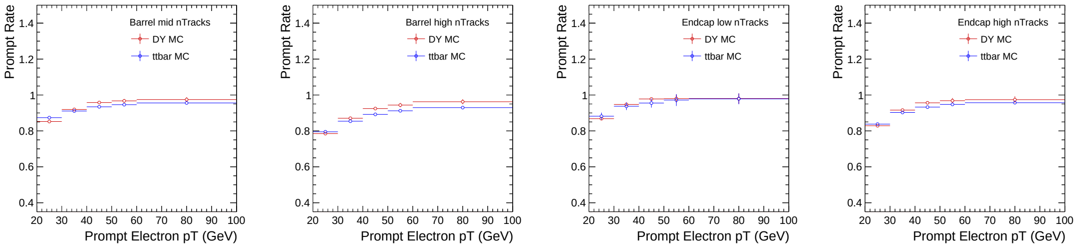


FIGURE A.2: Electron prompt rates as measured in simulated samples.

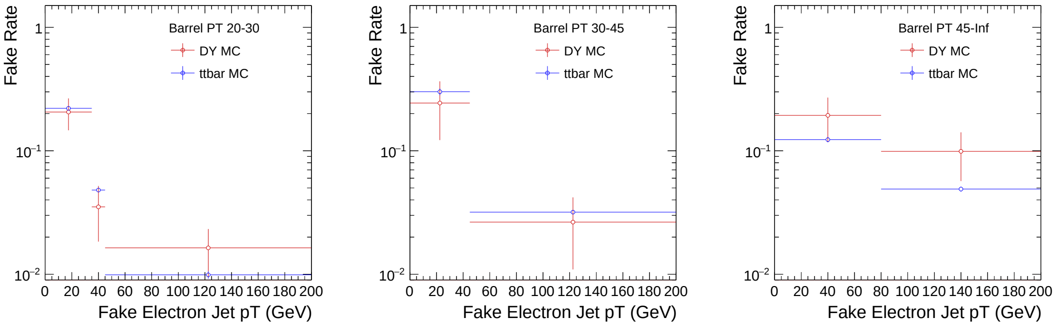


FIGURE A.3: Electron fake rates measured in MC in barrel for $\mathrm{p_T} < 30$ GeV (left), $30 < \mathrm{p_T} < 45$ GeV (middle) and $\mathrm{p_T} > 45$ GeV (right).

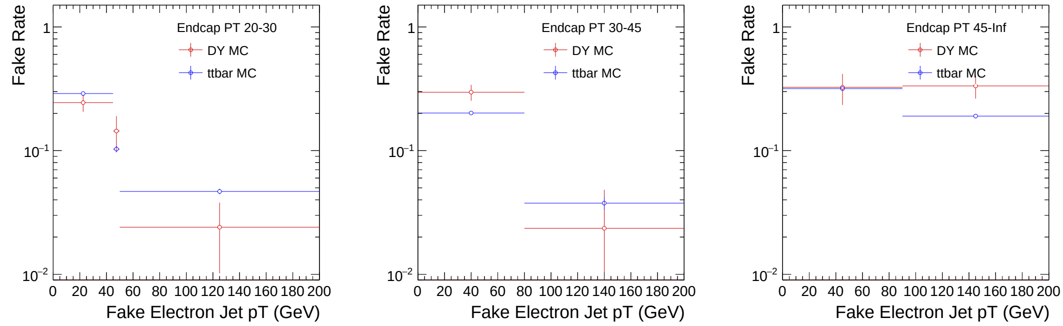


Figure A.4: Electron MC fake rates in endcap for $p_T < 30$ GeV (left), $30 < p_T < 45$ GeV (middle) and $p_T > 45$ GeV (right).

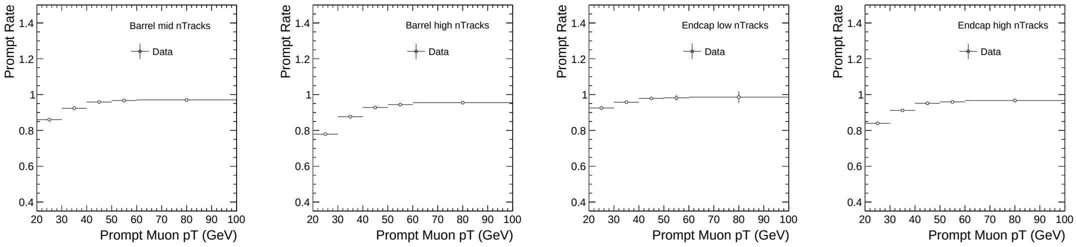


Figure A.5: Muon prompt rates as measured in the single muon dataset.

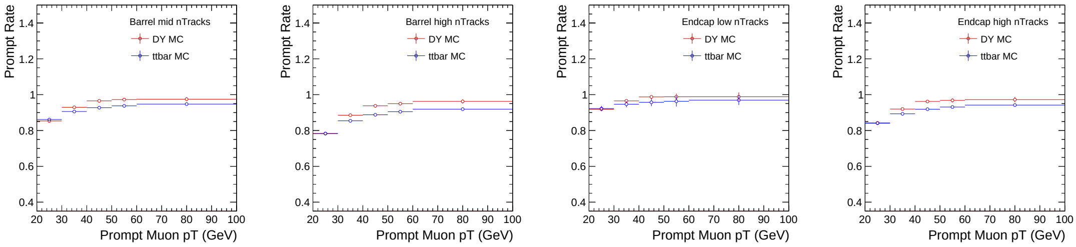


Figure A.6: Muon prompt rates measured in the simulated samples.

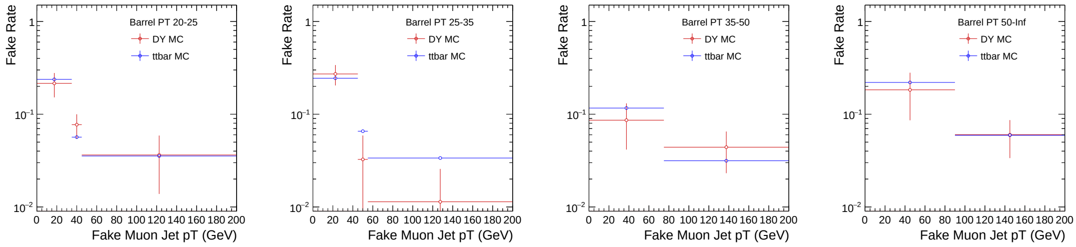


Figure A.7: Muon MC fake rates in barrel for $p_T < 25$ GeV, $25 > p_T > 35$ GeV, $35 < p_T < 50$ GeV and $p_T > 50$ GeV from left to right respectively.

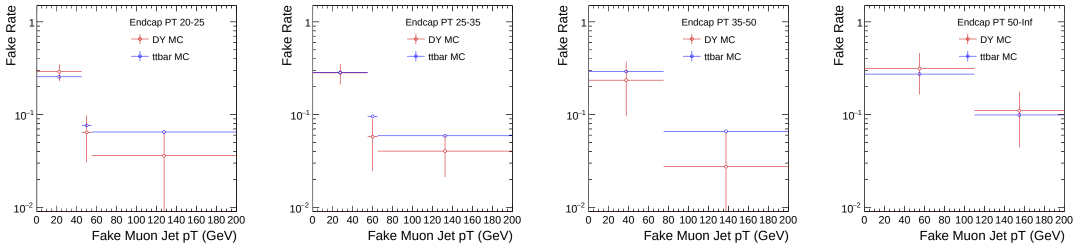


FIGURE A.8: Muon MC fake rates in endcap for $p_T < 25$ GeV, $25 > p_T > 35$ GeV, $35 < p_T < 50$ GeV, and $p_T > 50$ from left to right respectively.

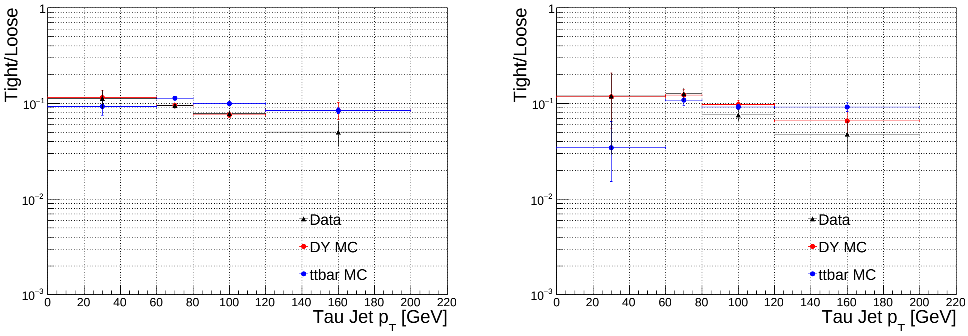


FIGURE A.9: Left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $50 - 100$ GeV in barrel. Right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $50 - 100$ GeV in endcap.

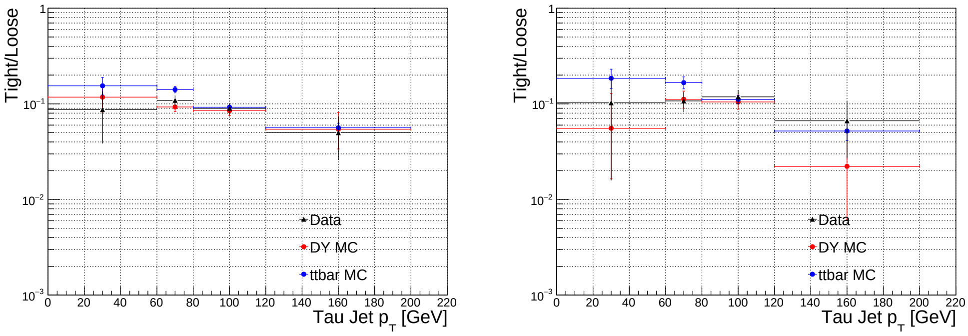


FIGURE A.10: Left: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $50 - 100$ GeV in barrel. Right: Tau fake rates in DY enriched data, and in simulated DY+jets and $t\bar{t}$ + jets events as a function of the mother jet p_T, for 1 prong taus with p_T $50 - 100$ GeV in endcap.

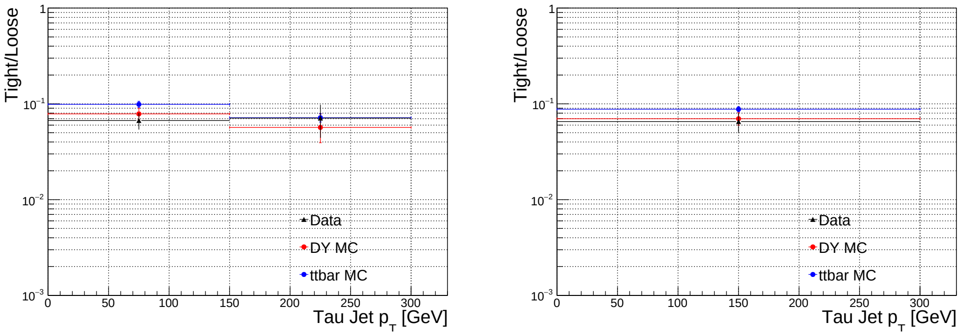


FIGURE A.11: Left: Tau fake rates in DY enriched data, and in simulated DY+jets and $\mathrm{t\bar{t}}$ + jets events as a function of the mother jet $\mathrm{p_T}$, taus with $\mathrm{p_T}$ $100 - 125$ GeV. Right: Tau fake rates in DY enriched data, and in simulated DY+jets and $\mathrm{t\bar{t}}$ + jets events as a function of the mother jet $\mathrm{p_T}$, taus with $\mathrm{p_T} > 125$ GeV.

A.3 Uncertainties on the MC based backgrounds

Some of the major backgrounds for this search are estimated using the simulated samples. For example, the WZ process is the most dominant background in the 3L channels, while ZZ is the primary background for the 4L channels. To make sure there aren't any biases in the simulation, we evaluate the impacts of various uncertainties on the MC samples. The main sources considered here are the energy and scale variations of the various objects used in the analysis, viz. leptons, *jet*. The quantities considered are jet energy resolution (JER), jet energy scale (JES), uncluttered energy scale (UES), muon energy and resolution scale (MERS), electron energy scale (EES) and tau energy scale (TES). The impacts of these uncertainties are accessed by varying each quantity within one standard deviation of their central values and observing the change in the background yield for particular MC samples in all the signal regions. Figure A.12-A.17 depicts the variation in the background yields while we vary the quantities listed above.

A.4 Trigger Efficiencies

The trigger efficiencies are measured using the Tag and Probe method on Z resonance by selecting the dilepton events by a single lepton trigger. The tag and probe objects are required to pass tight selection requirement and match to an HLT level object. The tight selection requirements for each lepton favor are described in Chapter 4. The probe object is also required to be opposite in charge compared to tag and is at least $\Delta R = 0.4$

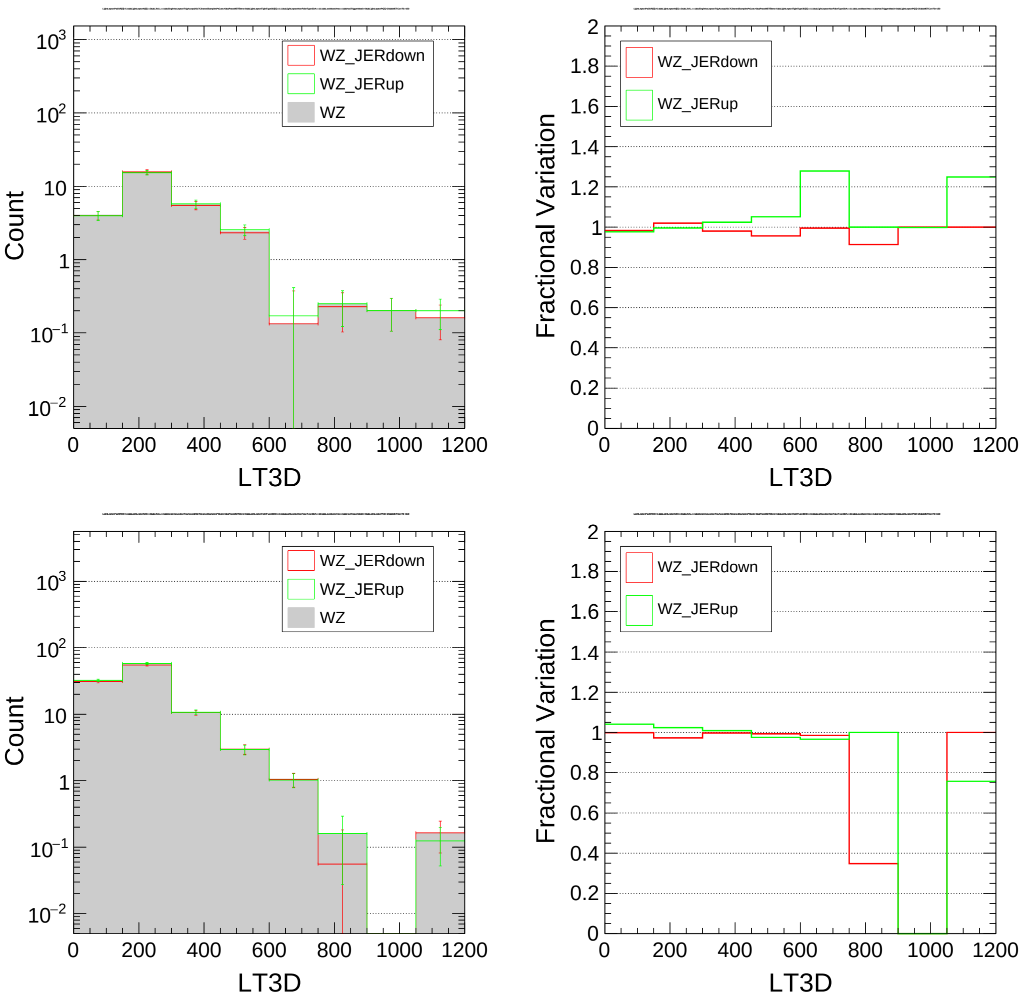


FIGURE A.12: Impact of varying JER on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

away from the tag object. The probability at which this probe object also matches to HLT level object (within $\Delta R < 0.2$) defines the trigger efficiency for the that HLT path. The trigger efficiencies are measured in bins of lepton p_T and η. Figure A.18 and A.19 show the trigger efficiencies for single muon “IsoMu27” and single electron “Ele35WPTight” respectively.

To evaluate the impact of the uncertainties on the trigger efficiencies measured in the data simulated samples we vary the trigger efficiency up and down by 2% for both data and MC in the opposite direction and observe the change in MC yield. Figure A.20-A.21 show the impact of these variations in various signal regions. We observe the impact of the uncertainty on the trigger efficiency is minimal, especially in the regions with 3 or more light leptons due to the redundancy in the objects that can fire the trigger. The

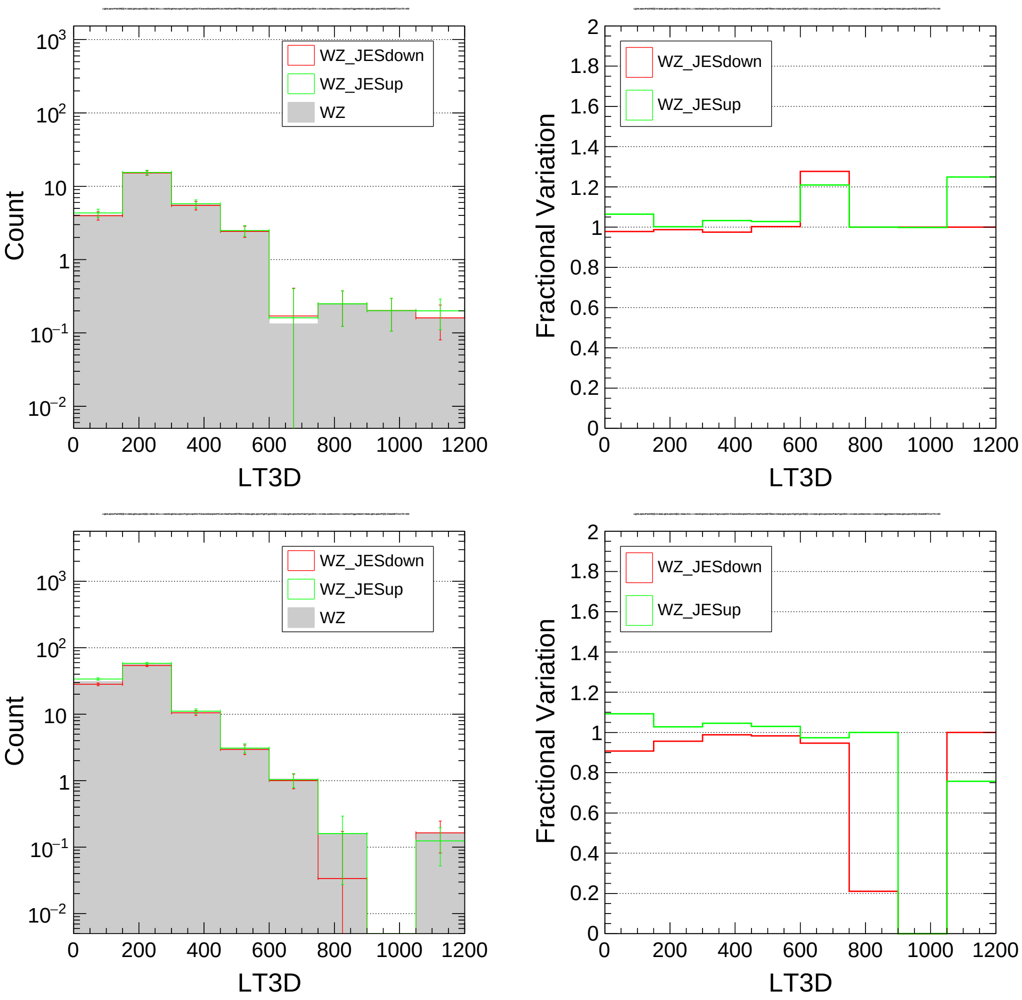


FIGURE A.13: Impact of varying JES on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

impact is most visible (< 4%) in 2L1T channels as seen Figure A.20-A.21. The trigger efficiency for the trigger paths used in the 2016 seesaw analysis as given in Table A.4 are shown in Figures A.22-A.25

A.5 Impact of L1 pre-firing

We evaluated the impact of Level-1 (L1) pre-firing using a combination of single muon and single electron 2017 datasets (Run B, D, and F). We use a selection of events with at least two e/μ, and a recipe provided by the L1 group [1]. We observe a minimal

---

[1] `https://indico.cern.ch/event/732212/contributions/3026917/attachments/1663881/2666673/Level1_DPG_UpdatePreFire_PPD_7June2018.pdf`

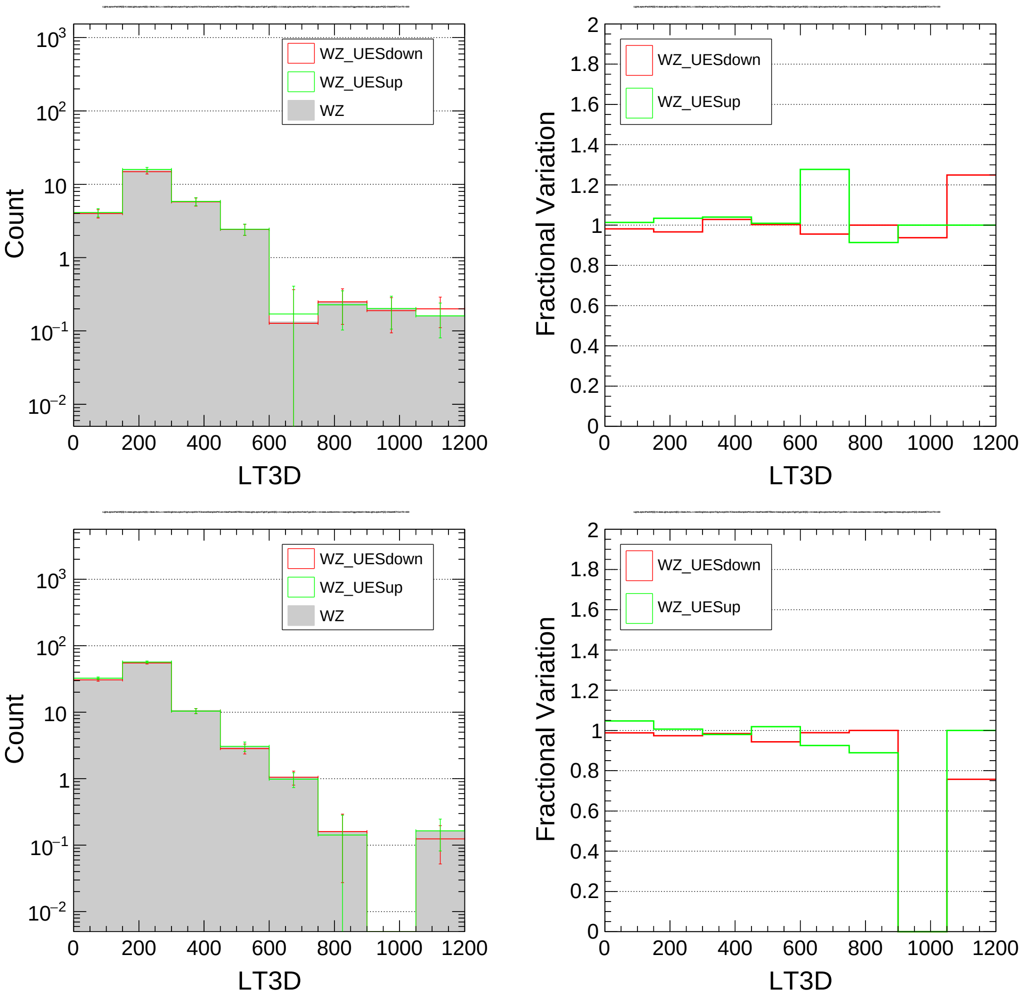


FIGURE A.14: Impact of varying UES on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

impact (about 1.2%) of this issue in Run B+D+F. The impact on entire 2017 dataset is expected to be even smaller. In trilepton selection, we have very few events, but our observations are statistically consistent with the dilepton observations.

We also checked the impact of this issue on the signal MC using an in-efficiency curve provided by the L1 group as a function of jet p_T in 2.4< $|\eta|$ <3.4. Figure A.26 shows the inefficiency curve as a function of jet p_T in 2.4< $|\eta|$ <3.4 (left) as provided by the L1 group and p_T of leading AK4 PF jet (right) in 2.4< $|\eta|$ <3.4 for VLL signal with mass 600 GeV. We estimate the impact of this issue on VLL signal is <1%.

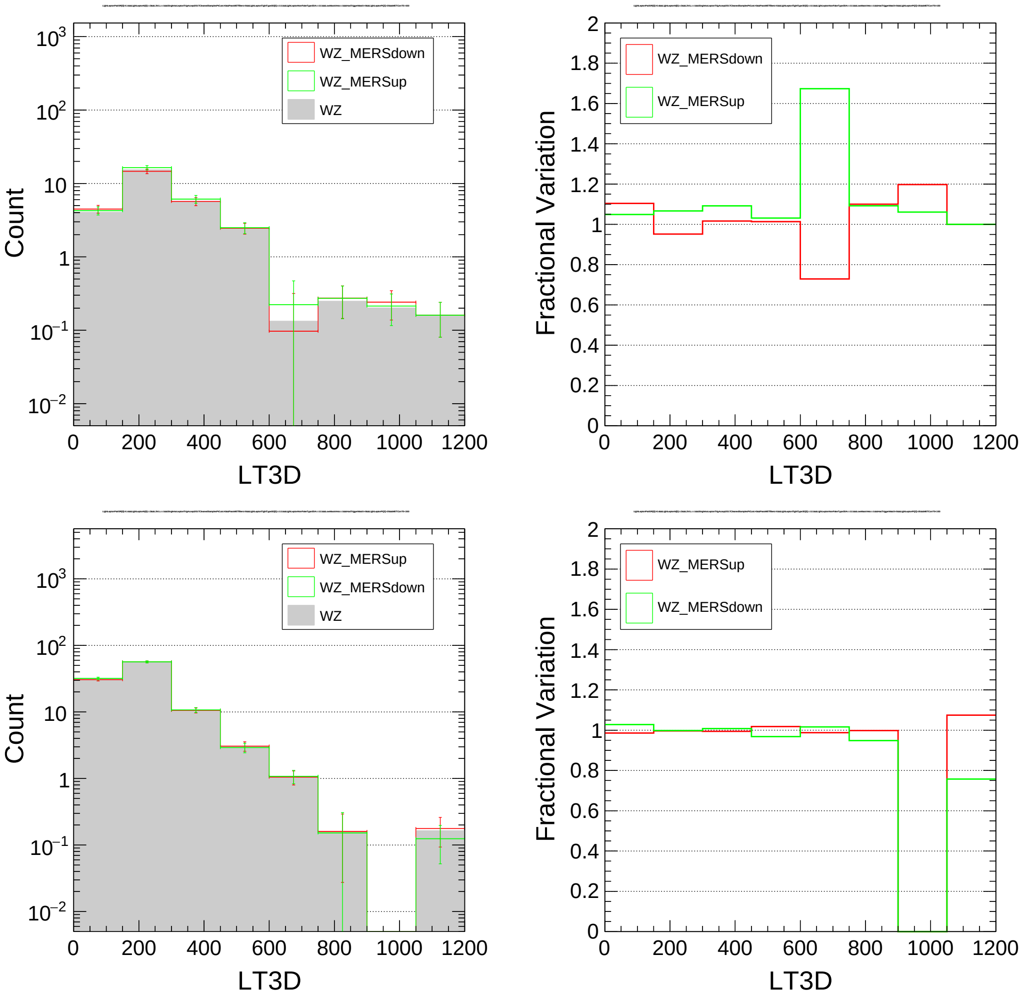


FIGURE A.15: Impact of varying MERS on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

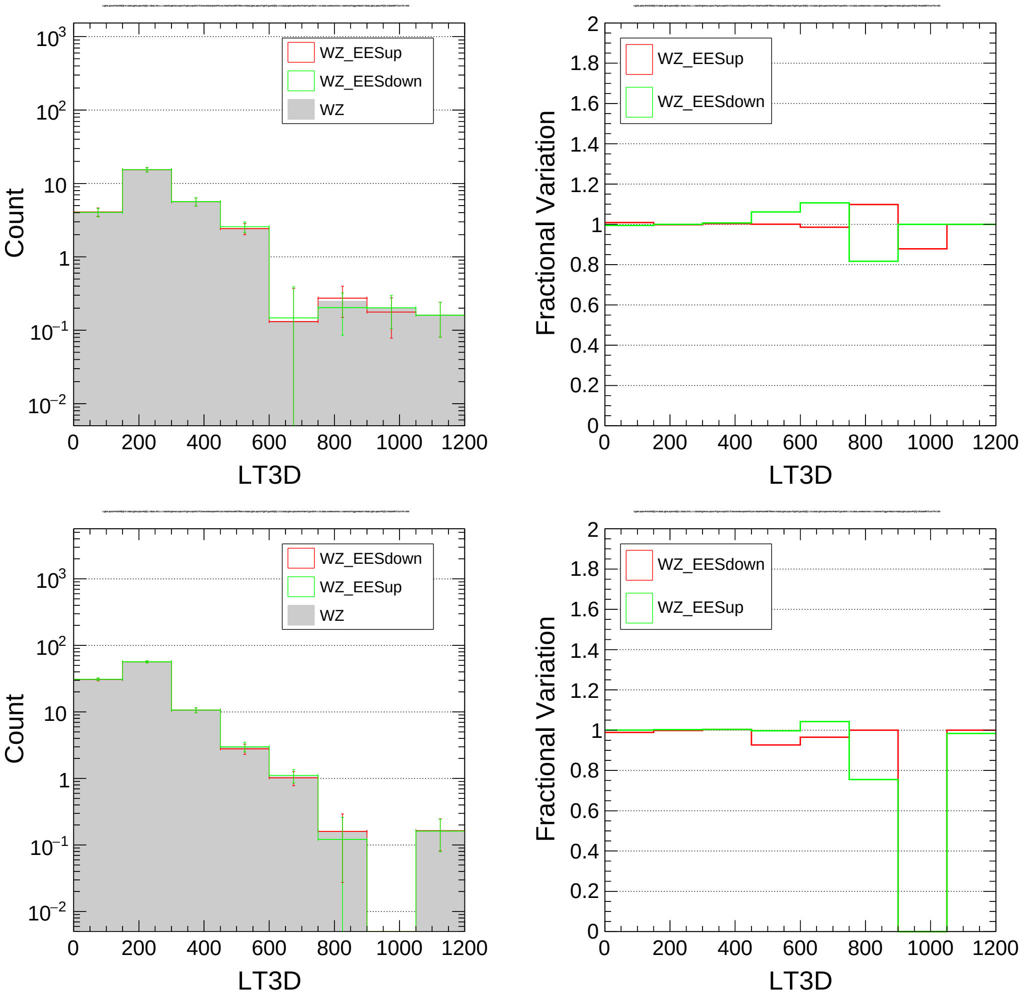


FIGURE A.16: Impact of varying EES on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

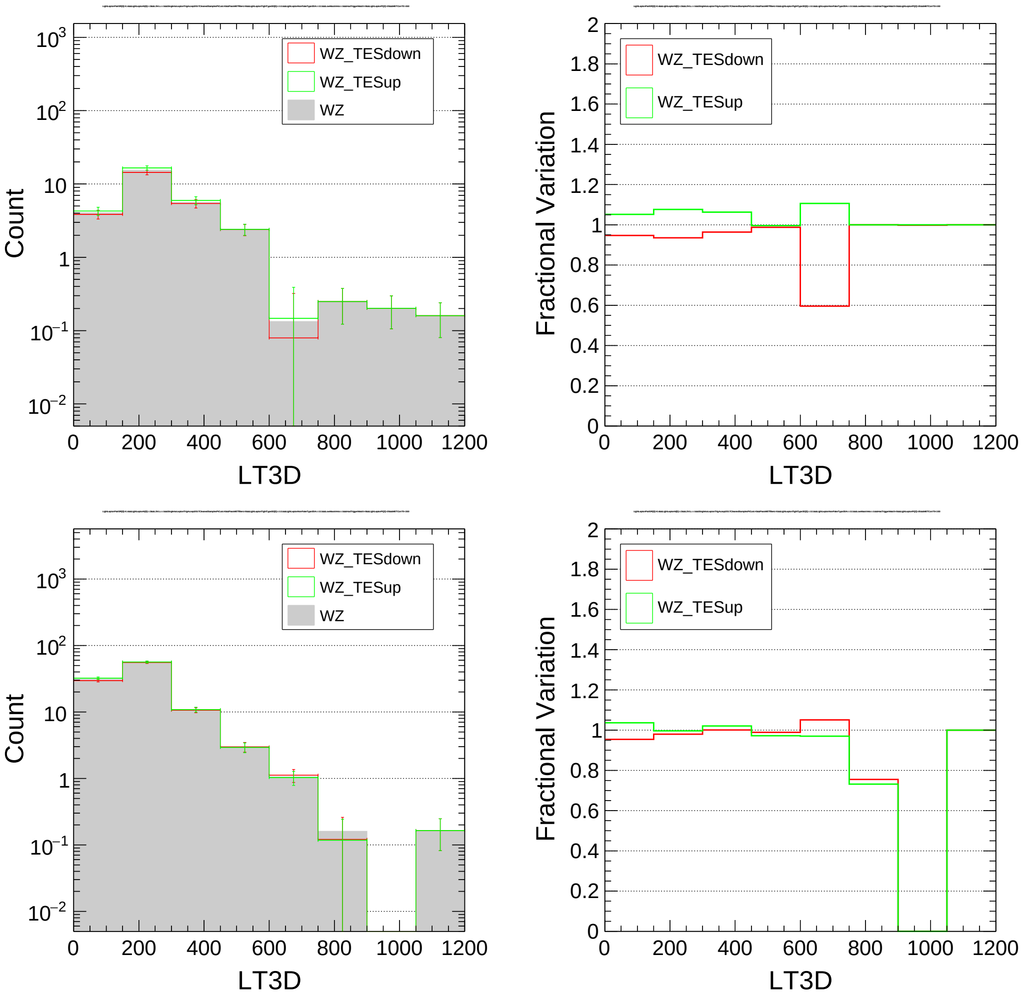


FIGURE A.17: Impact of varying TES on LT distribution in 3L channels in low p_T^{miss} (upper plots) and high p_T^{miss} (bottom plots).

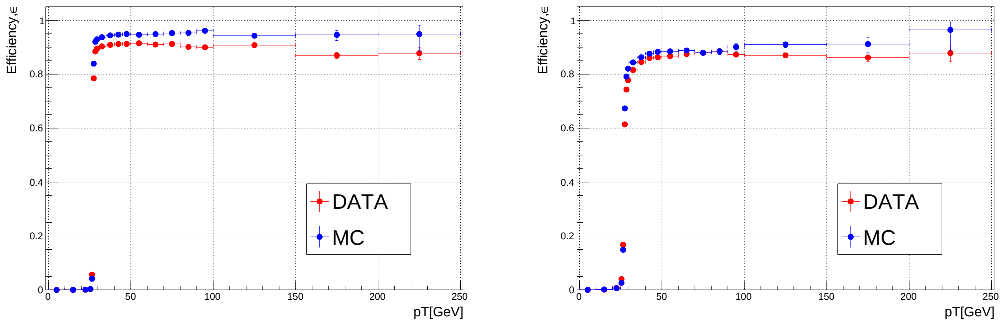


FIGURE A.18: Trigger efficiency in Data and MC as a function of p_T of tight muon objects in two broad η bins : barrel which is for $|\eta| \leq 1.2$ (left) and endcap which is for $|\eta| > 1.2$ (right). These are obtained in dimuon events collected by HLT_IsoMu27 trigger path.

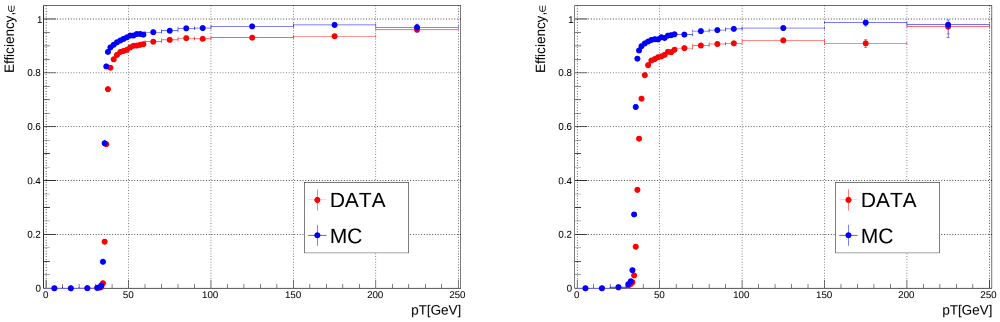


FIGURE A.19: Trigger efficiency in Data and MC as a function p_T of tight electron objects in two broad η bins : barrel which is for $|\eta| \leq 1.479$ (left) and endcap which is for $|\eta| > 1.479$ (right). These are obtained in dielectron events collected by HLT_Ele35_WPTight_Gsf trigger path.

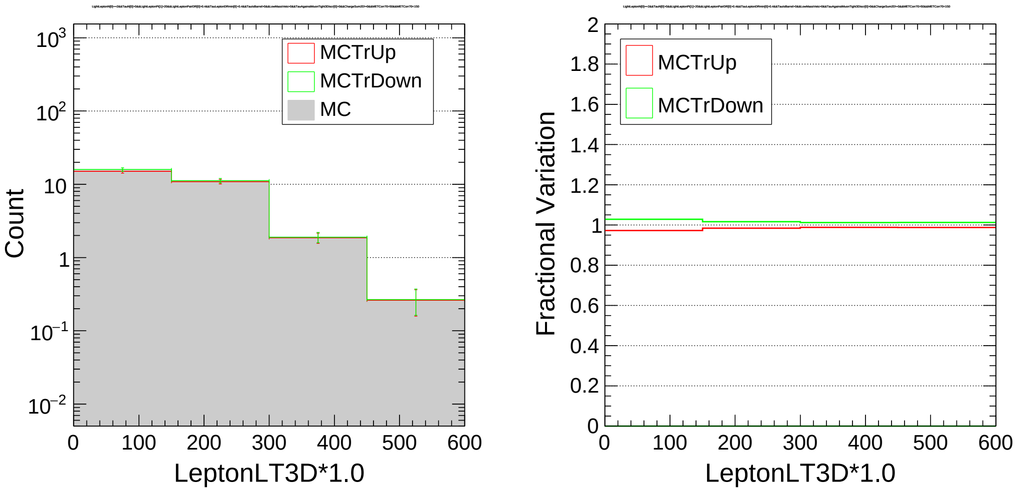


FIGURE A.20: Impact of varying trigger efficiencies in 2L1T SS, low p_T^{miss} signal region.

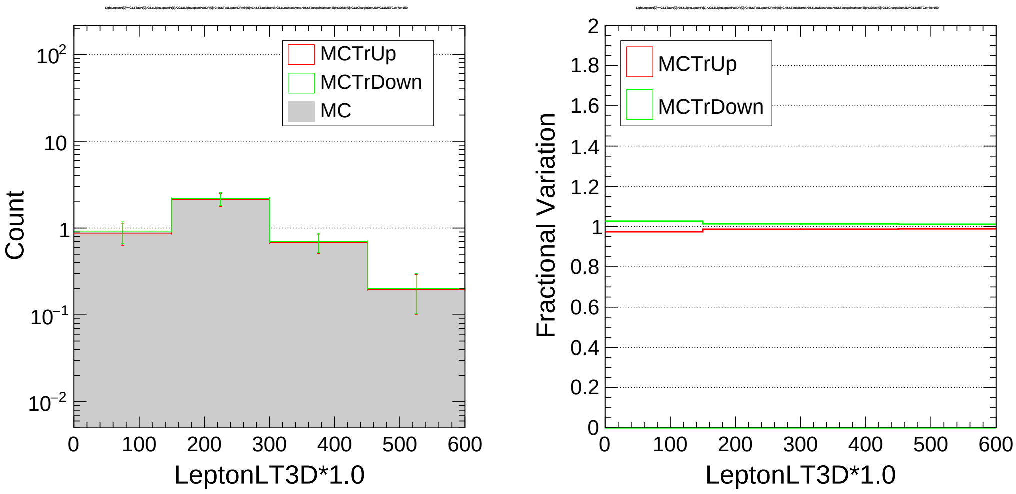


FIGURE A.21: Impact of varying trigger efficiencies in 2L1T SS, high p_T^{miss} signal region.

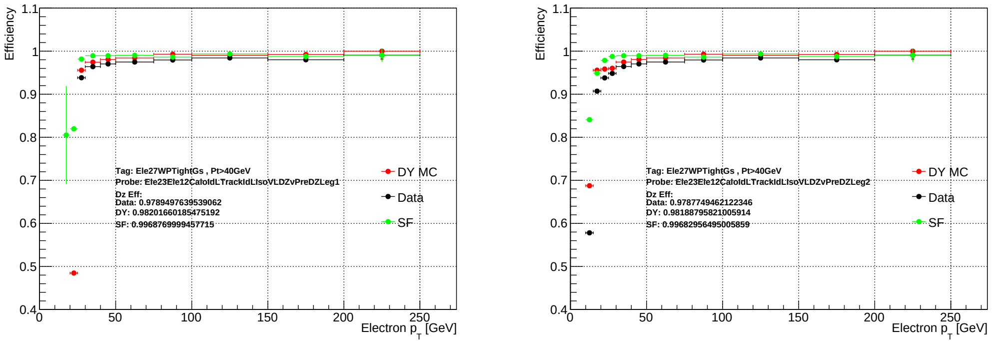


FIGURE A.22: Leading (left) and subleading (right) electron leg efficiencies for the HLT_Ele23_Ele12_CaloIdL_TrackIdL_IsoVL_DZ trigger path as measured in events with a dielectron pair in the full 2016 dataset and in MC, collected by a single electron trigger. The DZ filter efficiency is found to be 0.98 both in data and MC.

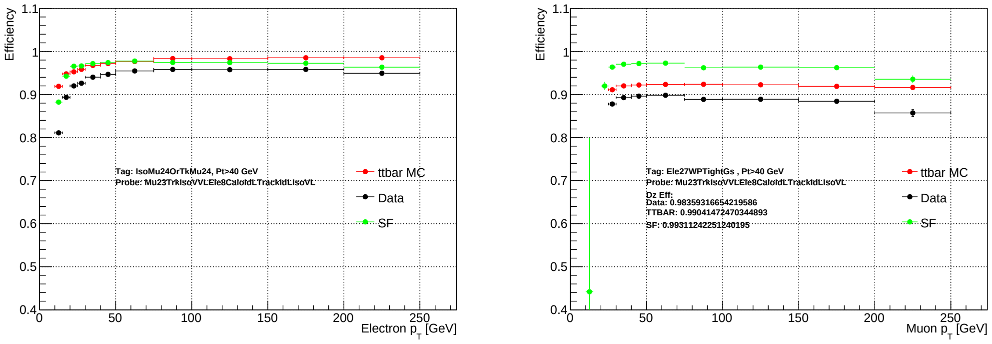


FIGURE A.23: Electron (left) and muon (right) leg efficiencies for the HLT_Mu23_TrkIsoVVL_Ele8_CaloIdL_TrackIdL_IsoVL trigger path as measured in events with a muon+electron pair in the full 2016 dataset and in MC, collected by a single muon or a single electron trigger, respectively. The DZ filter efficiency is found to be 0.98 in data and 0.99 in MC.

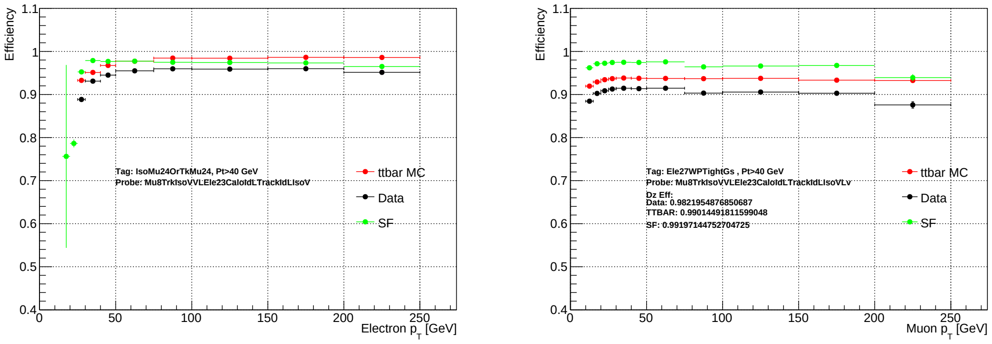


FIGURE A.24: Electron (left) and muon (right) leg efficiencies for the HLT_Mu8_TrkIsoVVL_Ele23_CaloIdL_TrackIdL_IsoVL trigger path as measured in events with a muon+electron pair in the full 2016 dataset and in MC, collected by a single muon or a single electron trigger, respectively. The DZ filter efficiency is found to be 0.98 in data and 0.99 in MC.

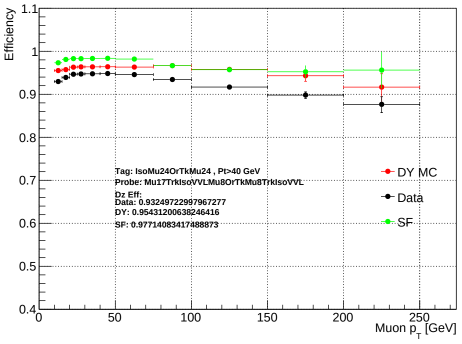


FIGURE A.25: Muon leg trigger efficiency for the HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ trigger path as measured in events with a dimuon pair in the full 2016 dataset and MC, collected by a single muon trigger. The DZ filter efficiency is found to be 0.93 in data and 0.95 in MC.

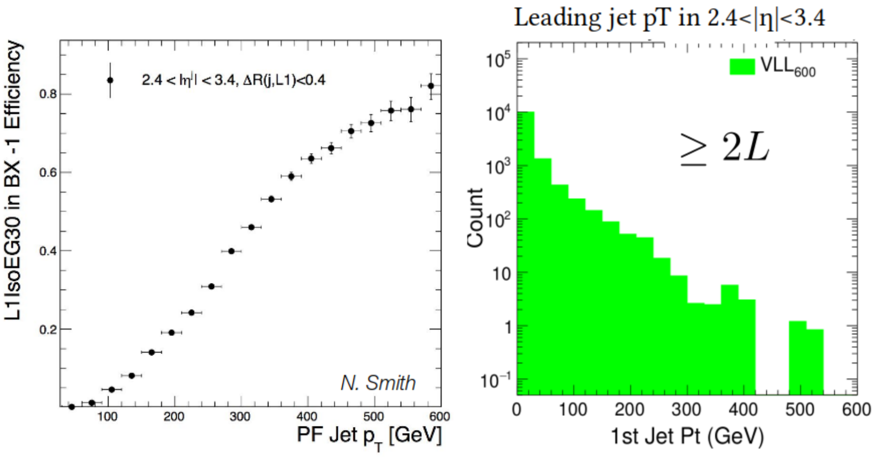


FIGURE A.26: The inefficiency curve as a function of jet p_T in 2.4< $|\eta|$ <3.4 (left) as provided by the L1 group and p_T of leading AK4 PF jet (right) in 2.4< $|\eta|$ <3.4 for VLL signal with mass 600 GeV. We estimate the impact of this issue on VLL signal is <1%.

Bibliography

[1] Fehling Dave. The Standard Model of Particle Physics: A Lunchbox's Guide. The Johns Hopkins University. 2008.

[2] Sheldon L. Glashow. Partial-symmetries of weak interactions. *Nuclear Physics*, 22(4):579 – 588, 1961.

[3] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

[4] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc.*, C680519:367–377, 1968.

[5] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973.

[6] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973.

[7] H. Fritzsch, M. Gell-Mann, and H. Leutwyler. Advantages of the color octet gluon picture. *Physics Letters B*, 47(4):365 – 368, 1973.

[8] F. Halzen and A. D. Martin. *Quarks and Leptons: An Introductory Course in Modern Particle Physics.* John Wiley and Sons, New York, NY, USA, 1984.

[9] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory.* Westview Press, 1995.

[10] A. Bettini. *Introduction to Elementary Particle Physics.* Cambridge University Press, 2008.

[11] UA1 collaboration. Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s}$=540 GeV. *Physics Letters B*, 122(1):103 – 116, 1983.

[12] UA2 collaboration. Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the cern pp collider. *Physics Letters B*, 122(5):476 – 485, 1983.

[13] G. Arnison et al. Experimental Observation of Lepton Pairs of Invariant Mass Around 95-GeV/c^2 at the CERN SPS Collider. *Phys. Lett.*, B126:398–410, 1983. [7.55(1983)].

[14] P. Bagnaia et al. Evidence for Z0 —¿ e+ e- at the CERN anti-p p Collider. *Phys. Lett.*, B129:130–140, 1983. [,7.69(1983)].

[15] Particle Data Group. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.

[16] J. Goldstone. Field theories with superconductor solutions. *Il Nuovo Cimento (1955-1965)*, 19(1):154–164, Jan 1961.

[17] Yoichiro Nambu. Quasi-particles and gauge invariance in the theory of superconductivity. *Phys. Rev.*, 117:648–663, Feb 1960.

[18] Jeffrey Goldstone, Abdus Salam, and Steven Weinberg. Broken symmetries. *Phys. Rev.*, 127:965–970, Aug 1962.

[19] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.

[20] Abdelhak Djouadi. The anatomy of electroweak symmetry breaking: Tome i: The higgs boson in the standard model. *Physics Reports*, 457(1):1 – 216, 2008.

[21] B. Odom, D. Hanneke, B. D'Urso, and G. Gabrielse. New measurement of the electron magnetic moment using a one-electron quantum cyclotron. *Phys. Rev. Lett.*, 97:030801, Jul 2006.

[22] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[23] SNO Collaboration. Direct evidence for neutrino flavor transformation from neutral-current interactions in the sudbury neutrino observatory. *Phys. Rev. Lett.*, 89:011301, Jun 2002.

[24] SNO Collaboration. Measurement of day and night neutrino energy spectra at sno and constraints on neutrino mixing parameters. *Phys. Rev. Lett.*, 89:011302, Jun 2002.

[25] M. N. Rebelo. On quasidegeneracy of Majorana neutrinos and the observed pattern of Leptonic mixing. In *17th Lomonosov Conference on Elementary Particle Physics Moscow, Russia, August 20-26, 2015*, 2016.

[26] T. S. van Albada, J. N. Bahcall, K. Begeman, and R. Sancisi. Distribution of dark matter in the spiral galaxy NGC 3198. , 295:305–313, August 1985.

[27] Katherine Garrett and Gintaras Duda. Dark Matter: A Primer. *Adv. Astron.*, 2011:968283, 2011.

[28] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.

[29] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.

[30] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:S08001, 2008.

[31] CMS Collaboration. *Technical proposal.* LHC Tech. Proposal. CERN, Geneva, 1994. Cover title : CMS, the Compact Muon Solenoid : technical proposal.

[32] CMS Collaboration. CMS technical design report, volume II: Physics performance. *J. Phys. G*, 34:995–1579, 2007.

[33] Serguei Chatrchyan et al. Description and performance of track and primary-vertex reconstruction with the CMS tracker. *JINST*, 9(10):P10009, 2014.

[34] A Dominguez, D Abbaneo, K Arndt, N Bacchetta, A Ball, E Bartz, W Bertl, G M Bilei, G Bolla, H W K Cheung, M Chertok, S Costa, N Demaria, Daniel Dominguez Vazquez, K Ecklund, W Erdmann, K Gill, G Hall, K Harder, F Hartmann, R Horisberger, W Johns, H C Kaestli, K Klein, D Kotlinski, S Kwan, M Pesaresi,

H Postema, T Rohe, C SchÃfer, A Starodumov, S Streuli, A Tricomi, P Tropea, J Troska, F Vasey, and W Zeuner. CMS Technical Design Report for the Pixel Detector Upgrade. Technical Report CERN-LHCC-2012-016. CMS-TDR-11, Sep 2012. Additional contacts: Jeffrey Spalding, Fermilab, Jeffrey.Spalding@cern.ch Didier Contardo, Universite Claude Bernard-Lyon I, didier.claude.contardo@cern.ch.

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

[36] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997.

[37] CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. 2000.

[38] L. Cadamuro. The cms level-1 trigger system for lhc run ii. *Journal of Instrumentation*, 12(03):C03021, 2017.

[39] Sergio Cittolin, Attila Rácz, and Paris Sphicas. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002.

[40] The CMS Collaboration. The cms experiment at the cern lhc. *Journal of Instrumentation*, 3(08):S08004, 2008.

[41] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014.

[42] Johan Alwall, Michel Herquet, Fabio Maltoni, Olivier Mattelaer, and Tim Stelzer. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.

[43] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159, 2015.

[44] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250, 2003.

[45] Sezen Sekmen. Recent Developments in CMS Fast Simulation. *PoS*, ICHEP2016:181, 2016.

[46] CMS Collaboration. CMS Physics. 2006.

[47] T. Miao, H. Wenzel, F. Yumiceva, and N. Leioatts. Beam position determination using tracks. 2007.

[48] CMS Collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical Design Report CMS. CERN, Geneva, 2006. There is an error on cover due to a technical problem for some items.

[49] R. Fruhwirth, W. Waltenberger, and P. Vanlaer. Adaptive vertex fitting. *J. Phys.*, G34:N343, 2007.

[50] A. M. Sirunyan et al. Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13(06):P06015, 2018.

[51] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{}$s= 8 TeV. *Journal of Instrumentation*, 10(06):P06005–P06005, June 2015.

[52] W Adam, R Frühwirth, A Strandlie, and T Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005.

[53] Vardan Khachatryan et al. Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at s = 8 TeV. *JINST*, 10(06):P06005, 2015.

[54] CMS Collaboration. Performance of -lepton reconstruction and identification in cms. *Journal of Instrumentation*, 7(01):P01001, 2012.

[55] Vardan Khachatryan et al. Reconstruction and identification of lepton decays to hadrons and at CMS. *JINST*, 11(01):P01019, 2016.

[56] Gavin P. Salam. Towards Jetography. *Eur. Phys. J.*, C67:637–686, 2010.

[57] Gavin P. Salam and Gregory Soyez. A Practical Seedless Infrared-Safe Cone jet algorithm. *JHEP*, 05:086, 2007.

[58] Yuri L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber. Better jet clustering algorithms. *JHEP*, 08:001, 1997.

[59] S. Catani, Yu.L. Dokshitzer, M.H. Seymour, and B.R. Webber. Longitudinally-invariant k-clustering algorithms for hadron-hadron collisions. *Nuclear Physics B*, 406(1):187 – 224, 1993.

[60] Stephen D. Ellis and Davison E. Soper. Successive combination jet algorithm for hadron collisions. *Phys. Rev.*, D48:3160–3166, 1993.

[61] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-k_t jet clustering algorithm. *JHEP*, 04:063, 2008.

[62] The CMS collaboration. Determination of jet energy calibration and transverse momentum resolution in cms. *Journal of Instrumentation*, 6(11):P11002, 2011.

[63] CMS Collaboration. Baseline muon selections for Run-II. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2?rev=26`. Accessed: January 28, 2019.

[64] CMS Collaboration. Cut Based Electron ID for Run 2. `https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2?rev=39`. Accessed: January 28, 2019.

[65] CMS Collaboration. TauID for 13 TeV run: recommendation from the Tau POG. `https://twiki.cern.ch/twiki/bin/viewauth/CMS/TauIDRecommendation13TeV`. Accessed: January 28, 2019.

[66] CMS Collaboration. Measurements of inclusive and differential Z boson production cross sections in pp collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-SMP-15-011, CERN, Geneva, 2016.

[67] CMS Collaboration. Measurement of the $t\bar{t}$ production cross section in the dilepton channel in pp collisions at $\sqrt{s} = 7$ TeV. *JHEP*, 11:067, 2012.

[68] CMS Collaboration. Search for evidence of the type-iii seesaw mechanism in multilepton final states in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 119:221802, Dec 2017.

[69] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-LUM-17-004, CERN, Geneva, 2018.

[70] Adam Falkowski, David M. Straub, and Avelino Vicente. Vector-like leptons: Higgs decays and collider phenomenology. *JHEP*, 05:092, 2014.

[71] Takeo Moroi and Yasuhiro Okada. Radiative corrections to Higgs masses in the supersymmetric model with an extra family and antifamily. *Mod. Phys. Lett.*, A7:187–200, 1992.

[72] K. S. Babu, Ilia Gogoladze, Mansoor Ur Rehman, and Qaisar Shafi. Higgs Boson Mass, Sparticle Spectrum and Little Hierarchy Problem in Extended MSSM. *Phys. Rev.*, D78:055017, 2008.

[73] Stephen P. Martin. Extra vector-like matter and the lightest Higgs scalar boson mass in low-energy supersymmetry. *Phys. Rev.*, D81:035004, 2010.

[74] Motoi Endo, Koichi Hamaguchi, Sho Iwamoto, and Norimi Yokozaki. Higgs Mass and Muon Anomalous Magnetic Moment in Supersymmetric Models with Vector-Like Matters. *Phys. Rev.*, D84:075017, 2011.

[75] Marc Sher. Charged leptons with nanosecond lifetimes. *Phys. Rev.*, D52:3136–3138, 1995.

[76] Scott D. Thomas and James D. Wells. Phenomenology of Massive Vectorlike Doublet Leptons. *Phys. Rev. Lett.*, 81:34–37, 1998.

[77] Paul H. Frampton, P. Q. Hung, and Marc Sher. Quarks and leptons beyond the third generation. *Phys. Rept.*, 330:263, 2000.

[78] Radovan Dermisek. Insensitive Unification of Gauge Couplings. *Phys. Lett.*, B713:469–472, 2012.

[79] James Halverson, Nicholas Orlofsky, and Aaron Pierce. Vectorlike Leptons as the Tip of the Dark Matter Iceberg. *Phys. Rev.*, D90(1):015002, 2014.

[80] Radovan Dermisek and Aditi Raval. Explanation of the Muon g-2 Anomaly with Vectorlike Leptons and its Implications for Higgs Decays. *Phys. Rev.*, D88:013017, 2013.

[81] Radovan Dermisek. Unification of gauge couplings in the standard model with extra vectorlike families. *Phys. Rev.*, D87(5):055008, 2013.

[82] Nilanjana Kumar and Stephen P. Martin. Vectorlike leptons at the Large Hadron Collider. *Phys. Rev.*, D92(11):115018, 2015.

[83] P. Achard et al. Search for heavy neutral and charged leptons in e^+e^- annihilation at LEP. *Phys. Lett.*, B517:75–85, 2001.

[84] Georges Aad et al. Search for heavy lepton resonances decaying to a Z boson and a lepton in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 09:108, 2015.

[85] Thomas Junk. Confidence Level Computation for Combining Searches with Small Statistics. *Nucl. Instrum. Meth.*, A434:435, 1999.

[86] Alexander L. Read. Modified frequentist analysis of search results (The CL(s) method). In *Workshop on confidence limits, CERN, Geneva, Switzerland, 17-18 Jan 2000: Proceedings*, page 81, 2000.

[87] A. L. Read. Presentation of search results: The CL(s) technique. *J. Phys. G*, 28:2693, 2002.

[88] ATLAS and CMS Collaborations. Procedure for the lhc higgs boson search combination in summer 2011. Technical Report CMS NOTE-2011/005, ATL-PHYS-PUB-2011-11, 2011.

[89] F. Capozzi, G.L. Fogli, E. Lisi, A. Marrone, D. Montanino, and A. Palazzo. Status of three-neutrino oscillation parameters, circa 2013. *Phys. Rev. D*, 89:093018, 2014.

[90] R. Foot, H. Lew, X.G. He, and G.C. Joshi. See-saw neutrino masses induced by a triplet of leptons. *Z. Phys. C*, 44, 1989.

[91] Benjamin Fuks, Michael Klasen, David R. Lamprea, and Marcel Rothering. Gaugino production in proton-proton collisions at a center-of-mass energy of 8 TeV. *JHEP*, 1210:081, 2012.

[92] Benjamin Fuks, Michael Klasen, David R. Lamprea, and Marcel Rothering. Precision predictions for electroweak superpartner production at hadron colliders with Resummino. *Eur. Phys. J. C*, 73:2480, 2013.

[93] C. Biggio and F. Bonnet. Implementation of the type iii seesaw model in feynrules/madgraph and prospects for discovery with early lhc data. *Eur. Phys.J. C*, 72:1899, 2012.

[94] Richard D. Ball et al. Parton distributions for the lhc run ii. *JHEP*, 04:040, 2015.

[95] CMS Collaboration. Search for vector-like leptons in multilepton final states in pp collisions at $\sqrt{s} = 13$ TeV. Technical Report CMS-PAS-EXO-18-005, CERN, Geneva, 2018.

[96] CMS Collaboration. CMS Software. `https://cms-sw.github.io`. Accessed: January 28, 2019.