Adapting the Multilepton Final States to Complex Models

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Shriyansh Ranjan
Registration number: 20201133

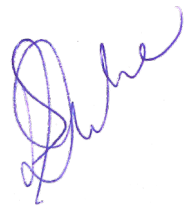
Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2025

Supervisor: Prof. Sourabh Dube

Certificate

This is to certify that this dissertation entitled Adapting the Multilepton Final States to Complex Models towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Shriyansh Ranjan at Indian Institute of Science Education and Research under the supervision of Prof. Sourabh Dube, Professor, Department of Physics , during the academic year 2024-2025.



Prof. Sourabh Dube

Committee:

Prof. Sourabh Dube

Dr. Nilanjana Kumar

This thesis is dedicated to my family.

Declaration

I hereby declare that the matter embodied in the report entitled Adapting the Multilepton Final States to Complex Models are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Sourabh Dube and the same has not been submitted elsewhere for any other degree.

Shriyansh Ranjan

Acknowledgments

I want to express my deepest gratitude to my thesis supervisor, Prof. Sourabh Dube, for mentoring me through every aspect of my college life, whether it is academic or research work, scientific discussions, or life skills. Your guidance and everlasting support throughout this research have been crucial in shaping my scientific career. I am also sincerely grateful to Prof. Nilanjana Kumar for her valuable insights and comments, which significantly contributed to the development of this work.

A special thanks to Mr. Arpit Jain, my high school teacher, for igniting my passion for physics and helping me embark on this amazing journey into physics research.

I want to thank all my senior colleagues, Arnab, Prachurjya, Riya and Yash, for their invaluable guidance and advice throughout my research. I also want to thank my fellow labmates, Shreyas, Kirutheeka and Vaidehi, for making long hours in the lab fun and creating a positive working environment. I couldn't have asked for better colleagues.

To all my friends at IISER - thank you for being with me through thick and thin and keeping me grounded through college. A very special thank you to Sanchi for always being by my side, Jasveer for all the wild adventures, and Sumant for being the best roommate ever.

My utmost appreciation goes to my family, whose unconditional love and support have kept me on my legs through every stage of my life.

Finally, I would like to thank IISER Pune for giving me the opportunity to pursue this research and introducing me to these fantastic people with whom I could work and grow.

Abstract

The standard model of particle physics has several shortcomings, such as the origin of neutrino masses, dark matter candidates and baryon asymmetry. Several searches for beyond standard model (BSM) phenomena have been conducted. One paper probes new phenomena in multilepton final states and searches for new particles arising in type-III seesaw, singlet and doublet models of vector-like leptons and scalar leptoquark models. However, this paper treats each of these models as independent.

Recent theoretical developments propose models with vector-like leptons as well as leptoquarks or models of seesaw fermions with additional scalars. Such models present unique challenges to existing searches due to the higher number of theoretical parameters, as well as complex decay topologies.

In this thesis, we will study such models of mixed particles and how a multilepton search can be designed to have high sensitivity to such models. We aim to build a robust analysis strategy that targets the multiple theoretical parameters in a systematic way.

Contents

Abstract **xi**

1 Introduction **1**

2 The standard model of particle physics **3**

3 The CMS detector **7**

4 Reinterpretation of a CMS multilepton search **11**

4.1 Reinterpretation workflow 13

4.2 Signal Generation 14

4.3 Object Selection 16

4.4 Implementation of Detector Effects 16

4.5 Signal Regions 18

4.6 Calculating Limits 22

4.7 Reinterpreting the VLL μ-like model 23

5 Analysing a Complex Model **27**

5.1 Model Description 27

5.2 Object Selection 28

5.3 Signal Regions 33
5.4 Backgrounds 36
5.5 2L2Fj Final State 36
5.6 3L1Fj Final State 39
5.7 Results 42

6 Conclusion **45**

List of Tables

5.1 Cross-sections for the quintuplet model using different PDFs 29

5.2 Cross-sections of the signal and background processes for the quintuplet model . . 37

5.3 2L2Fj final state cuts for the quintuplet model 39

5.4 3L1Fj final state cuts for the quintuplet model 41

5.5 $S/\sqrt{B}$ ratios and $(A\varepsilon)$ for the two final states of the quintuplet model 43

5.6 Expected limits for the two final states of the quintuplet model 44

List of Figures

2.1 The standard model 4

3.1 A slice of the CMS detector 8

4.1 Limit plot for the VLLD τ' model 12

4.2 Feynman diagrams for the VLLD τ' model 15

4.3 Feynman diagram for the VLLD τ' model in 2L2T final state 15

4.4 Example efficiency maps 18

4.5 $p_T(L_0)$ and S_T before and after efficiency maps 18

4.6 M_{min} and dR_{min} variables for the VLL τ' model after event selections 20

4.7 Fundamental S_T table in 4L final state 21

4.8 Signal yield for the VLLD τ' model 21

4.9 Limits wrt the number of SRs considered for the VLLD τ' model 22

4.10 Reinterpreted limit plot for VLLD τ' 23

4.11 Feynman diagram for the VLLD μ' model 24

4.12 Signal acceptance of different final states for the VLLD μ' model 24

4.13 M_{min} and S_T variables compared for 1 TeV VLL μ' and τ' 25

4.14 Signal yield for the VLLD μ' model 25

4.15 Reinterpreted limit plot for the VLLD μ' model 26

5.1 Feynman diagram for the quintuplet model 28
5.2 Fatjet mass and p_T for soft-drop v/s pruning algorithms 30
5.3 Soft-dropped mass for the fatjets of a signal and several SM processes 31
5.4 τ_{21} and τ_{32} for the fatjets of a signal and several SM processes 32
5.5 Relative mass for the fatjets of several signal and SM processes 32
5.6 $(A\varepsilon)$ compared between the quintuplet and VLL τ-like model 33
5.7 M_{min} variable for two mass points of the quintuplet model 34
5.8 Feynman diagram for the quintuplet model with $4W$ state 35
5.9 M_{min} and S_T for the quintuplet model in the 2L2Fj final state 38
5.10 H_T and JMF in the 2L2Fj final state, after the other cuts 40
5.11 p_T^{miss} and L_T for the quintuplet model in the 3L1Fj final state 41
5.12 $L_T + p_T^{miss}$ and dR_{min} in the 3L1Fj final state, after the other cuts 42

Chapter 1

Introduction

One of the significant open-ended questions in physics is the search for fundamental particles that make up the universe and their interaction. Physicists worldwide strive to explain what the universe is precisely made up of. This motivation has given rise to a branch of physics called particle physics, which aims to study these fundamental particles and interactions that constitute matter and radiation.

The standard model (SM) of particle physics [1] is the most successful and widely accepted theory in explaining the fundamental particles. It can accurately describe the rates of processes involving the fundamental particles across over 12 orders of magnitude [2]. However, it has many limitations, such as failing to explain the source of neutrino masses, dark matter candidates, gravity and matter-antimatter asymmetry. Moreover, several anomalies are found in various particle physics experiments, which hint at physics beyond the standard model (BSM). Multiple high-energy physics experiments have been set up to search for physics beyond the standard model, one of them being CMS. The CMS experiment [3] constitutes a large detector, called the CMS detector, which detects the particles coming out of high-energy proton-proton collisions at its centre. Using this data, scientists carry out various searches for BSM physics.

One such search was conducted by CMS in 2022 [4] with signatures involving multiple leptons. It was a model-independent search, which refers to the search identifying anomalies in a broad and unbiased way, not relying on a specific BSM model. The search resulted in constraints being

placed over the parameter space of many BSM models, such as the vector-like lepton model [5, 6], type-III seesaw model, and leptoquark model. However, the search considers these models as independent of one another.

Recent theoretical developments suggest BSM models with multiple theoretical particles produced either in association or decay. For example, one BSM model [7] proposes seesaw fermions associated with additional scalars. Such models present unique challenges to the existing searches due to a significant increase in theoretical parameters and complex decay chains. Thus, there arises a need to devise an analysis strategy that can tackle the multiple theoretical parameters systematically.

The thesis is divided into multiple chapters described as follows: Chapter 2 discusses the standard model, its limitations, and the motivations for searching for physics beyond the standard model. Chapter 3 describes the CMS detector and the analyses carried out with the data from the detector. Chapter 4 talks about a multilepton search performed by CMS and how its results can be used to analyse other models. I will constrain the VLL μ-like doublet model using the results from the search. Chapter 5 presents a novel complex model, compares it to the conventional models and analyses the model in a systematic way. Finally, chapter 6 summarizes the results, conclusions, and future work.

Chapter 2

The standard model of particle physics

The standard model (SM) of particle physics is a theory that explains the fundamental particles in nature and their interaction [1]. The fundamental particles are classified according to their internal properties, such as charges and spin. The fermions are the particles that make up the matter around us, and the bosons are the particles that carry the fundamental forces, namely, the strong nuclear force (carried by gluon), the electromagnetic force (carried by photon), and the weak nuclear force (carried by W and Z bosons). The addition of the Higgs boson, which gives mass to the other fundamental particles, has completed the standard model. The fermions are further divided into quarks (which carry colour charge) and leptons (which do not carry colour charge).

The standard model is the most effective theory to date in explaining the different building blocks of our universe. It can accurately describe the rates of processes involving the fundamental particles across over 12 orders of magnitude. Figure 2.1 displays a representation of the various particles in the standard model, grouped according to their properties.

Despite the huge success of the standard model, there are many anomalies and limitations that challenge it. Some of the limitations are listed below:

1. **Neutrino masses:** The standard model assumes neutrinos to be massless. However, experiments have shown that neutrinos can exhibit oscillations between their mass and flavour eigenstates [8]. This means that neutrinos must have mass.

2. **Dark matter candidates:** The presence of dark matter is inferred from multiple astrophysical observations, such as the rotation curves in a galaxy and excess gravitational lensing around the galaxy. It is said to occupy 85% of all the mass in the universe. However, no particle in the standard model can act as a dark matter candidate [9].

3. **Gravity:** Gravity is one of the four fundamental forces of nature. However, there is no boson in the standard model that carries this force. One of the reasons for this is the absence of a renormalisable quantum description of gravity [10].

4. **Matter-antimatter asymmetry:** It is proven that our universe is matter-dominant. The standard model predicts this asymmetry to some extent via CP violation; however, it is not enough to match the observed asymmetry [11].

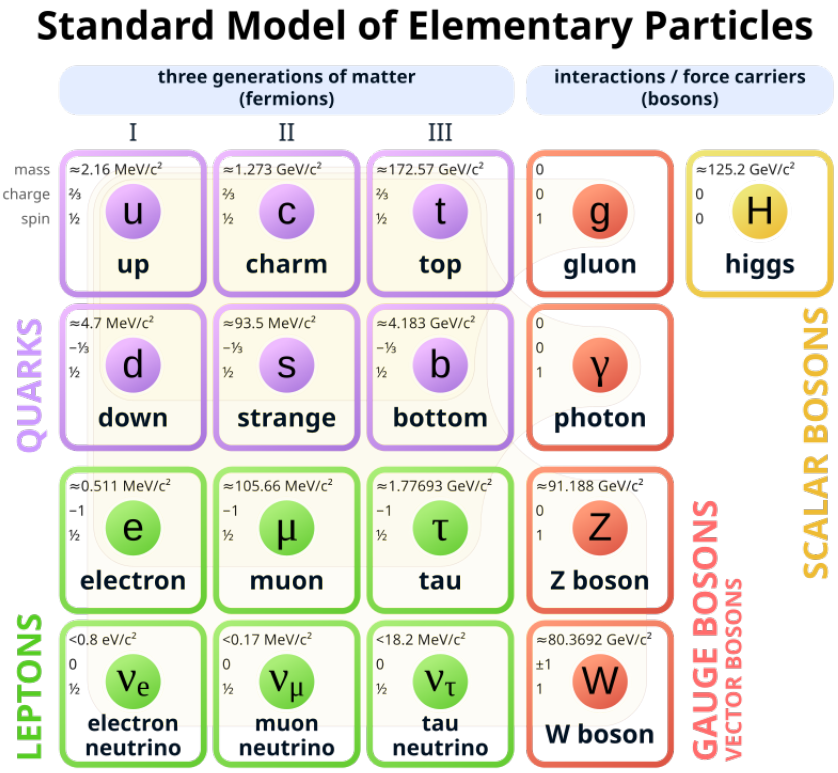


Figure 2.1: The standard model of particle physics. The figure shows the 12 fundamental fermions and 5 fundamental bosons. (Image Courtesy: Wikipedia [12])

To address the limitations of the standard model, many BSM models have been theorised, which are extensions of the standard model. For example, the vector-like lepton (VLL) model [5, 6] predicts additional non-chiral fermions, meaning their left and right-handed components transform in the same sence under gauge symmetry. VLLs can provide a dark matter candidate and explain the anomalous magnetic moment of the muon. Moreover, the supersymmetric (SUSY) theory [13] introduces theoretical bosonic partners of all the fermions and vice versa. Supersymmetry can explain some of the shortcomings of the standard model, such as providing dark matter candidates while also serving as an aesthetic motivation for the standard model being symmetric. The theories mentioned above are just a tiny subset of all the existing theories, with each theory predicting new particles to be found.

Chapter 3

The CMS detector

The CMS experiment [3] is a high-energy experiment situated at one of the collision points of the Large Hadron Collider (LHC) [14] at CERN in Geneva, Switzerland. The Large Hadron Collider is the most powerful particle accelerator in the world. It comprises a 27 km long ring placed 100 m underground, which accelerates and collides two proton beams at the centre of mass energy of 13.6 TeV. These collisions produce multiple particles due to their immense energies, which are then detected by the CMS detector.

The CMS detector is a 15 m wide and 21 m long cylindrical detector. It consists of a superconducting solenoid of an internal diameter of 6 m, capable of producing a 3.8 T magnetic field. The purpose of the solenoid is to bend the charged particles produced from the collision, thus providing information about their momenta. Moreover, it consists of several subdetectors, each serving different purposes. The innermost detector is a silicon pixel and strip tracker, which produces tracks of charged particles passing through it and measures their trajectories. The next layer holds a lead tungstate ($PbWO_4$) crystal electromagnetic calorimeter (ECAL), which stores the total energy of any electron or photon hitting it, measuring their positions and energies. After that comes the hadron calorimeter (HCAL), which comprises layers of brass plates and scintillation tiles. Similar to the ECAL, the HCAL stores the total energy of charged and neutral hadrons and measures their positions. The outermost layer is made up of drift tubes, cathode strip chambers, resistive-plate chambers and gas electron multiplier detectors. These detectors work together to identify and track muons. They also measure the muons' momenta. Figure 3.1 shows a transverse section of the CMS

detector, with the multiple subdetectors and how different particles interact with each subdetector.

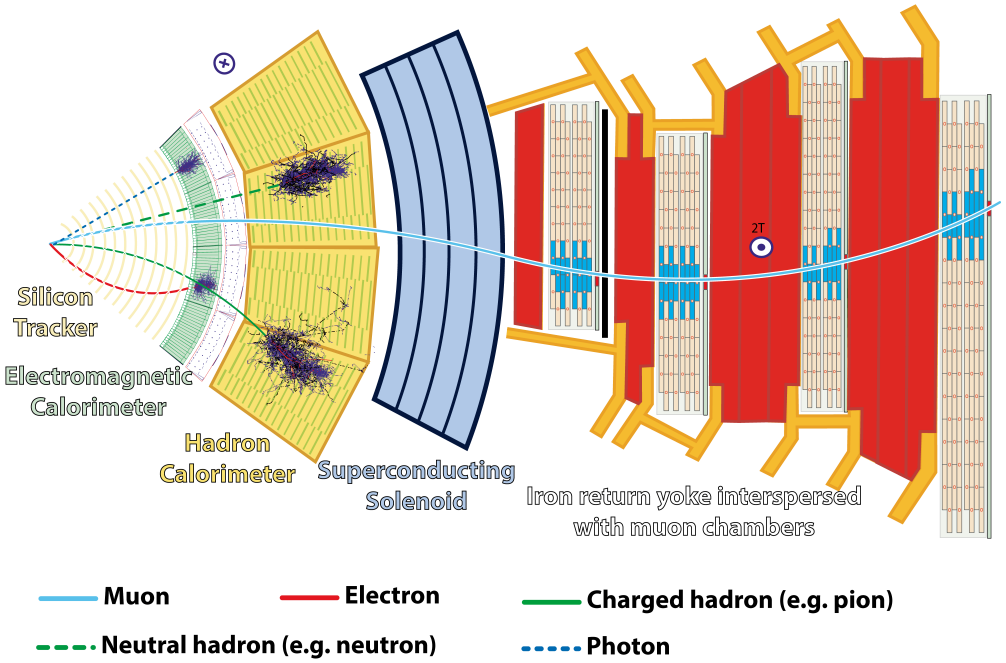


Figure 3.1: A slice of the CMS detector, showing its subdetectors and their interaction with various particles [15].

After every proton-proton collision, these detector signatures are collected, and a particle flow algorithm is applied to the signatures to reconstruct the objects. Additionally, jet clustering algorithms are run over the objects to form jets, which are clusters of particles generally formed by the decays of quarks or gluons. Moreover, jet clustering algorithms with larger dimension parameter form fatjets, which are bigger clusters than jets. This data is stored for every event over a long run to make a dataset. The datasets are then analysed, often taking the events with a specific composition of the objects in the final state.

A counting experiment is one method to conduct searches for BSM phenomena. In a counting experiment, one or more signal regions (SRs) are defined such that these have high signal acceptance and low SM backgrounds. In the SRs, the expected number of events is predicted for the null hypothesis (only SM background) and the alternative hypothesis (SM background + signal). This predicted number of events is then compared to the observation in data to determine which

hypothesis is supported. If the null hypothesis is true (i.e. no signal is observed), then constraints on the signal parameter space are obtained. These constraints are typically upper limits on the cross-section of the BSM model.

Chapter 4

Reinterpretation of a CMS multilepton search

The CMS collaboration performed a search [4] for physics beyond the standard model in 2022, with the signatures involving multiple leptons in an event. The analysis utilized the CMS proton-proton collision data at the centre of mass energy of 13 TeV, gathered from 2016 to 2018, which corresponds to an integrated luminosity of 138 fb^{-1}. The search was conducted in seven different final states with respect to the number of light leptons (e and μ) and hadronically decaying τ leptons (τ_h) in the event. These final states are: 4L, 3L1T, 2L2T, 1L3T, 3L, 2L1T and 1L2T, where L denotes the light leptons and T denotes τ_h.

The search is model-independent, meaning it is a generic analysis designed to look for any deviation from the standard model, not motivated by a specific model. Thus, it has probed BSM physics in a broad phase space, considering around 1200 signal regions (SRs).

In every signal region, the observed data aligns with the predicted background and no excess has been observed. Constraints have been set on the parameter space of several BSM models, such as the vector-like lepton (VLL) τ-like singlet and doublet model, the type-III seesaw model and the leptoquark model. For example, the upper limits on the production cross-section of the VLL τ-like doublet process are shown in Fig. 4.1 [16]. The vector-like τ lepton model shows an exclusion for $m_{\tau'} < 1045$ GeV at 95% CL.

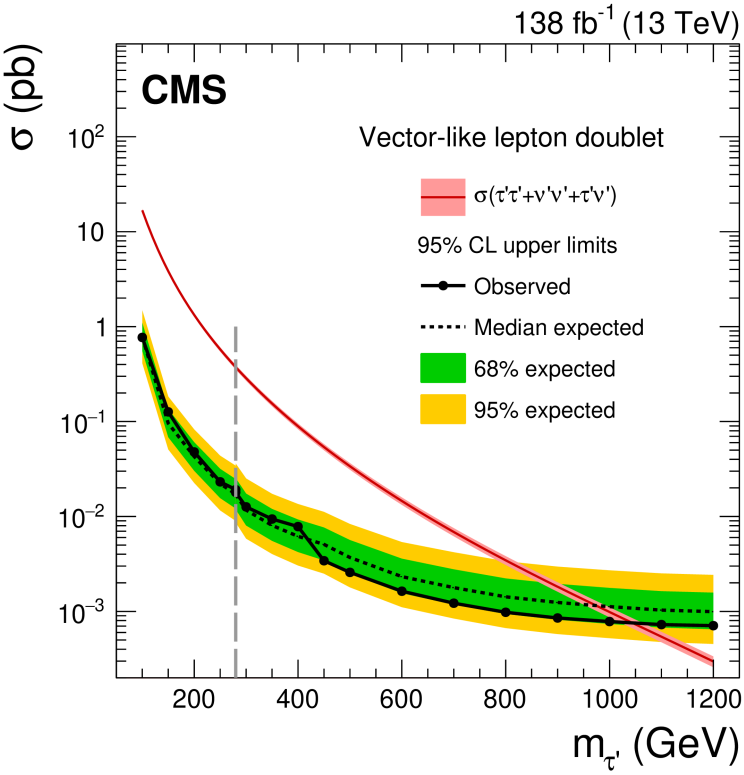


Figure 4.1: Observed and expected upper limits on the production cross-section of the VLL τ' particles at 95% CL [4]. From the figure, we can infer an observed and expected exclusion for doublet VLL τ' upto mass of 1045 GeVand 975 GeVrespectively.

In addition to the constraints, the counting experiment results are provided for the 1200 signal regions corresponding to the various final states and distributions in several event variables. These include the observation, the background and the uncertainty in the background for each signal region. The provided results can be used to analyse other BSM models. This is called reinterpretation of the available results.

4.1 Reinterpretation workflow

Reinterpretation is the process of repurposing the results from a previous search to analyse a specific new model. Reinterpreting a result to constrain a new model is much more efficient than designing a new analysis for the model, which would take a lot of human effort. Moreover, reinterpreting a result elevates the scientific value of the result as we can extract more information from the existing data.

The published results from CMS include a prescription for reinterpreting their results [16]. Following is a description of the methodology of reinterpretation:

1. **Simulation of signal events for the new model:** The generated simulation samples are used to find the signal acceptance and efficiency with respect to the search. To have an adequate number of events that meet the selection criteria/cuts in the search, we need to generate enough events to start with. Moreover, several samples are generated for different masses of the hypothetical particles in the model.

2. **Estimation of reconstructed yield from generator-level yield:** The particles obtained from the simulations are generator-level particles, which do not consider detector effects and efficiencies. These detector effects are mimicked by efficiency maps (as shown in Fig. 4.4), which predict the reconstructed-level p_T (transverse momentum) of a generator-level particle [16]. This enables us to compare the signal yield (from simulation) to the data/background (from search).

3. **Implementation of the object and event-level selections:** Several cuts and selections are made on the reconstructed particles precisely as described in the paper. These cuts form the various signal and control regions used in the search. Some examples of the variables that are generally cut on are p_T (transverse momentum), η (pseudorapidity), invariant masses, scalar p_T sums of several particles, etc.

4. **Calculation of signal yield for each SR:** The signal yield (N_{sig}) is calculated for further analysis via the formula:

$$N_{sig} = \mathscr{L} \cdot (\sigma B)_{th} \cdot (A\varepsilon) \tag{4.1}$$

where L is the luminosity of the collisions, $(\sigma B)_{th}$ is the theoretical cross-section of the process in the model, and $(A\varepsilon)$ is the signal acceptance for the SR. L is obtained from the paper, $(\sigma B)_{th}$ is specific to the model, and $(A\varepsilon)$ is the number of events in the SR divided by the total number of events generated in the simulations.

5. **Limit calculation for the SR:** For each SR, the N_{obs} (number of observed SR events), N_{bkg} (number of estimated background events) and δN_{bkg} (uncertainty in the number of background events) is obtained from the search. These numbers are then used to calculate the limit, i.e. the maximum amount of signal in the data that remains undetected. Furthermore, the signals from different SRs can be statistically added to improve the overall limit. The limit can also be calculated in terms of the upper limit in cross-section of the process instead of the number of events.

6. **Constraining the model:** This limit is calculated for multiple mass points and then compared with the theoretical cross-section of the process. The mass points where the theoretical cross-section is higher than the upper limit are excluded. Therefore, the model is constrained.

This recipe can be used to constrain new models, such as the VLL doublet μ-like model. However, before applying the methodology to a new model, I must validate my selections and check if my workflow is trustworthy. To tackle this problem, I will first constrain the VLL τ-like model and compare my calculated limits to the limits in the paper.

4.2 Signal Generation

The signal samples for the VLL τ-like model were generated via Monte Carlo simulation softwares, namely MADGRAPH5_AMC@NLO [17], PYTHIA 8.3 [18, 19] and DELPHES [20]. MADGRAPH5_AMC@NLO software generates events and calculates cross-sections for a given process. MADGRAPH already contains information on all the standard model particles and their interactions. However, to generate a BSM process, a UFO file has to be given to MADGRAPH, which includes the particles from the specific BSM model, their properties and their interactions with other particles. The output of MADGRAPH is fed into PYTHIA for the hadronisation of the quarks and gluons. DELPHES simulates the interaction of the detector with the particles generated by the above softwares. The output of the DELPHES software is a sample that contains the different

objects and their properties for many events in the ROOT format.

The UFO file for the VLL model was provided by the authors. Using the UFO file and the simulation softwares, I generated samples of 100000 events for five mass hypotheses of the VLL τ-like particle: 200 GeV, 400 GeV, 600 GeV, 800 GeV, and 1 TeV. Figure 4.2 shows the Feynman diagrams for the production and decay of the VLL τ-like particle. The τ' and ν'_τ can be pair-produced or associated-produced and always decay to a standard model τ and a boson. The bosons can further decay leptonically or hadronically.

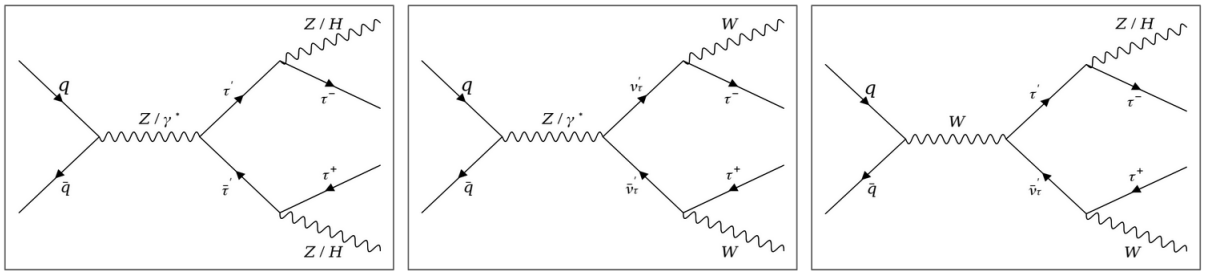


Figure 4.2: The Feynman diagrams for the doublet VLL τ-like model. The VLLs can be pair-produced or associated-produced and always decay to a standard model τ and a boson.

Figure 4.3 shows an example Feynman diagram of the VLL τ-like model where the Z further decays to electrons and W decays to quarks, giving jets. The two taus decay hadronically (τ_h), giving us a 2L2T final state.

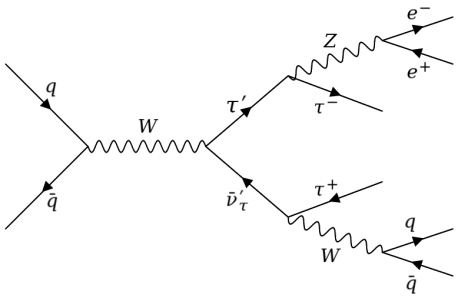


Figure 4.3: An example Feynman diagram for the doublet VLL τ-like model. The Z further decays to electrons and W decays to quarks, giving jets. The two taus decay hadronically (τ_h), giving us a 2L2T final state.

4.3 Object Selection

Several selections/cuts have to be made for the objects in an event so that "good" objects are selected to be fed into the efficiency maps. I followed the reinterpretation recipe to implement the selections. They are as follows:

- $p_T > 5$ GeV, $|\eta| < 2.4$ for e's and μ's
- $p_T > 15$ GeV, $|\eta| < 2.3$ for τ_h's
- Only the leptons originating from a signal particle, a boson or light leptons originating from τ leptons are considered.
- $p_T > 30$ GeV, $|\eta| < 2.4$ for jets
- The jets need to have an angular separation of 0.4 or more from the selected leptons in the event.

4.4 Implementation of Detector Effects

The leptons passing the object selections are generator-level. They do not take the detector effects into account. For example, a particle hitting a calorimeter might not deposit its whole energy into the calorimeter. A part of the energy might leak out of the detector, in which case the detector would report an energy lower than the actual value of the particle. The background and observation already consider these effects, so we must also mimic the detector effects on the signal.

We use efficiency maps to implement the detector effects onto the leptons. These maps are generated from the ZZ process simulations. The efficiency map presents the probability distribution of the reconstructed p_T corresponding to a generator-level p_T input. The efficiency map accounts for the reconstruction and identification efficiencies of the lepton. There are a lot of efficiency maps for different properties of the leptons and the event.

- There are separate maps for e's, μ's and 1-prong and 3-prong τ_h's. Moreover, there are separate maps for the light leptons arising from tau and boson decays.

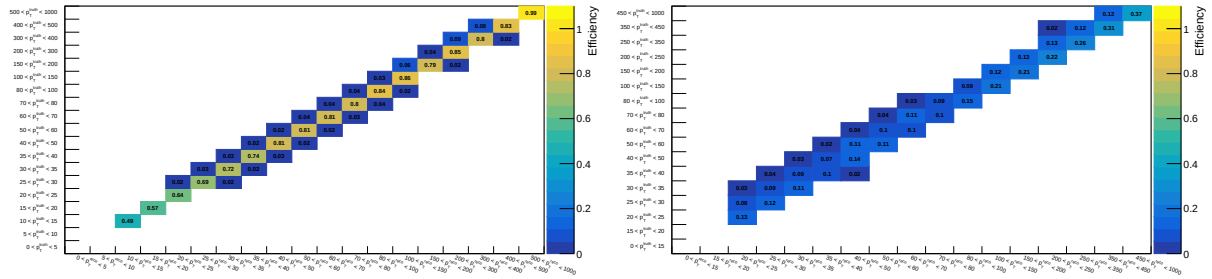
- There are different maps for the regions in the detector where the lepton hits: barrel ($|\eta| \leqslant 1.1$), transition ($1.1 < |\eta| \leqslant 1.6$) and endcap ($|\eta| > 1.6$) regions.

- Additionally, there are separate maps for different event topologies, such as the number of generator-level jets (N_{jet}) and the minimum angular distance between any pair of selected light leptons(dR_{min}).

Implementing efficiency maps to obtain the reconstructed yield is a non-trivial and involved process. There are over 50 efficiency maps according to the lepton and event properties mentioned above. The user has to first select the correct map using these properties. Then, a random number between 0 and 1 is compared to the probabilities in the reconstructed p_T bins corresponding to the particular generator-level p_T bin. According to the probabilities, a reconstructed p_T bin is assigned to the lepton, and its p_T is set to the mean of the bin's upper and lower bounds. Additionally, there is a finite possibility that no reconstructed p_T bin is allotted to the lepton, in which case the lepton is said to fail reconstruction/identification and is not considered for further analysis.

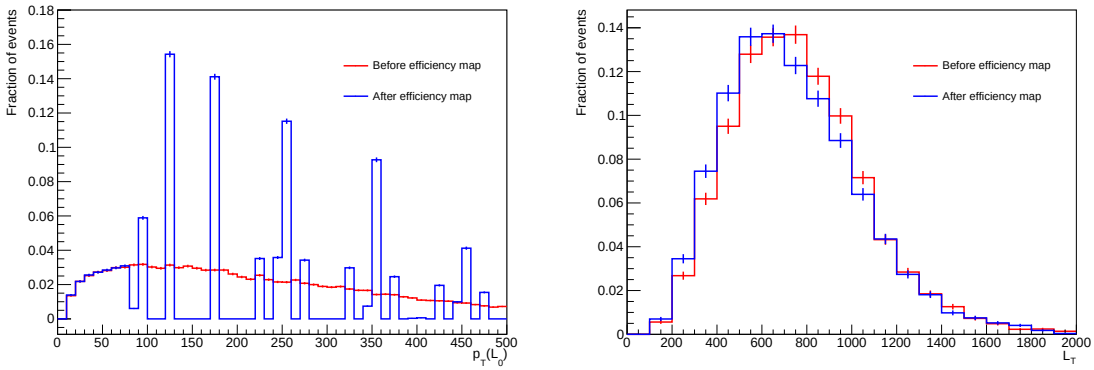
Figure 4.4 shows two efficiency maps [16]: one for a μ coming from τ decay, hitting the detector in the barrel region with $dR_{min} > 0.4$, and the other for a 1-prong τ_h from boson decay, hitting the endcap region. The x and y axes depict the reconstructed and generated lepton p_T bins. The maps show that a τ_h has lower reconstruction efficiencies than a μ in terms of their respective properties.

Figure 4.5 shows the p_T of the leading light lepton in the event before and after the efficiency maps have been applied. The curves are normalised such that their integrals are 1. We can see that the efficiency maps provide discrete p_T values (the mean of the lower and higher bounds of the bin) instead of a continuous distribution. The figure also shows the L_T variable before and after the efficiency maps have been applied. L_T of the event slightly decreases after the efficiency maps because the probability of the lepton being reconstructed with lower p_T is marginally higher than that with higher p_T.



(a) μ efficiency map (b) τ_h efficiency map

Figure 4.4: Efficiency maps for a μ in the barrel region of the CMS detector, coming from a tau decay with $dR_{min} > 0.4$ and a τ_h from boson decay, hitting the endcap region [16]. The x and y axes are the reconstructed-level (reco) and generator-level (gen) p_T. The maps show that a τ_h has lower reconstruction efficiencies than a μ in terms of their respective properties.



(a) $p_T(L_0)$ (b) L_T

Figure 4.5: The figure shows the p_T of the leading light lepton and L_T variables for the VLL τ' model of mass 1 TeV. The p_T variable shows discrete values after the efficiency maps have been applied, corresponding to the bin width. The L_T variable slightly decreases after applying the efficiency map.

4.5 Signal Regions

After the detector effects are implemented to obtain the reconstructed leptons, several event-level selections are made using various kinematic variables to select the desired events. Moreover, signal regions are also formed via further selections of the kinematic variables. These kinematic variables are listed below:

- Missing Transverse Momentum (p_T^{miss}): By the virtue of the conservation of momentum, the vector sum of the transverse momenta of all the particles coming out of a proton-proton collision should be 0. However, neutrinos and other undetectable particles (possibly BSM particles) might not be detected by CMS, which results in in non-zero vector sum of the momenta. p_T^{miss} is the residual momentum needed to bring the sum back to 0. It is defined as:

$$p_T^{miss} = \left| \sum_{i=observed} \overrightarrow{p_T}^i \right|$$

- L_T: L_T is defined as the scalar p_T sum of all the charged leptons that constitute the final state.

$$L_T = \sum_{i=leptons} \left| \overrightarrow{p_T}^i \right|$$

- H_T: H_T is defined as the scalar p_T sum of all the jets.

$$H_T = \sum_{i=jets} \left| \overrightarrow{p_T}^i \right|$$

- S_T: S_T is the sum of L_T, H_T and p_T^{miss}.

$$S_T = L_T + H_T + p_T^{miss}$$

- OSSFn: OSSFn is counted as the number of distinct oppositely-charged same-flavour lepton pairs in an event.

- M_{min}: M_{min} is defined as the minimum invariant mass of all the dilepton pairs, irrespective of their charge or flavour.

- M_{OSSF}: M_{OSSF} is defined as the mass of the OSSF electron or muon pair closest to the Z-boson mass at 91 GeV.

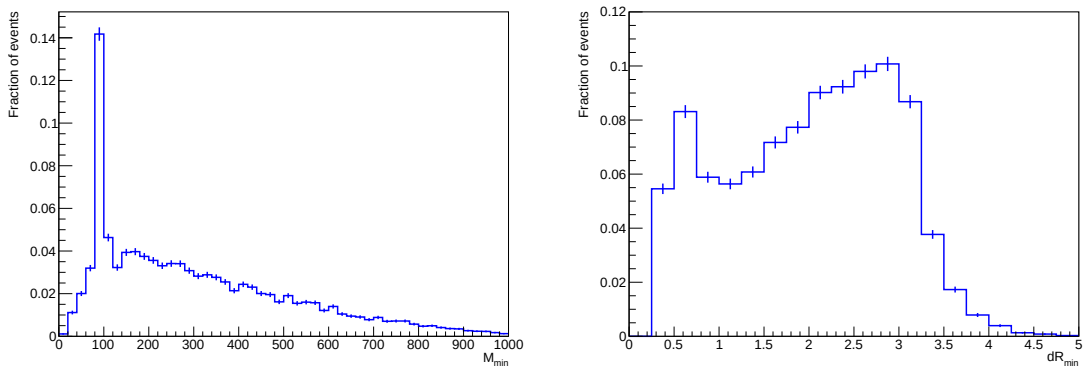
- dR_{min}: dR_{min} is defined as the minimum angular distance between all the dilepton pairs in the event.

- $dR_{min}^{\tau_h}$: $dR_{min}^{\tau_h}$ is defined as the minimum angular distance between all the dilepton pairs, where at least one of the leptons is a τ_h.

- N_{jet}: N_{jet} is defined as the number of jets in the event.

- N_b: N_b is defined as the number of b-tagged jets in the event.

For event selection, several selections are made on kinematic variables. The events with $M_{min} < 12$ GeV, $dR_{min} < 0.2$ or $dR_{min}^{\tau_h} < 0.5$ are vetoed. In each event, the presence of at least one muon with $p_T > 26$ GeV or one electron with $p_T > 35$ GeV is required. Figure 4.6 shows a few of the kinematic variables, namely M_{min} and dR_{min} plotted after the event selections. We can observe that the M_{min} peaks at the Z boson mass, implying that many of these lepton pairs are results of Z boson decays.



(a) M_{min} (b) dR_{min}

Figure 4.6: The figure shows the M_{min} and dR_{min} variables for the VLL τ' model of mass 1 TeV after event selections are made. The M_{min} variable shows a peak at the Z boson mass, implying that many of these lepton pairs are results of Z boson decays.

After the event selections are made, the events can be categorised into different signal regions using selections mentioned in the paper. These signal regions are parametrised in the different final states using distributions of two variables: $L_T + p_T^{miss}$ and S_T. Figure 4.7 [16] shows a subset of signal regions, specifically the 4L final state using the S_T distribution, with all backgrounds, uncertainties, and observations. The N_{sig}, N_{obs}, N_{bkg} and δN_{bkg} are obtained for each SR. Figure 4.8 shows the signal yield for the VLL doublet τ-like model with $m_{\tau'} = 1$ TeV in the same SRs displayed in Fig. 4.7.

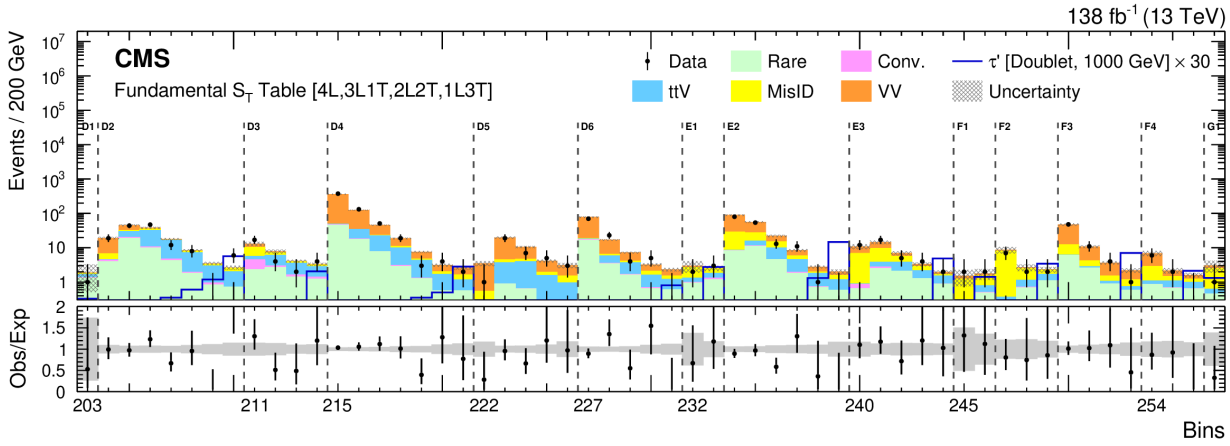


Figure 4.7: Signal regions used in [16]. The signal regions are a part of the fundamental S_T distribution in the 4L final state. The coloured stack represents the background, whereas the black markers represent the observation. In the lower panel, the ratio of the data to the background for each signal region is presented, with the gray band representing the uncertainty in the background estimate. For illustration purposes, a signal of the VLL doublet τ-like model for the mass of the τ' being 1 TeV is overlaid, represented by the blue line.

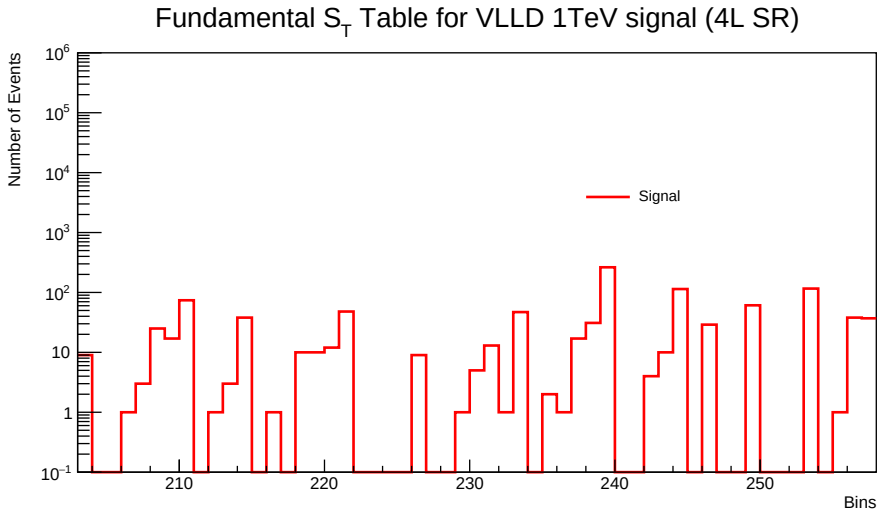


Figure 4.8: Signal yield for the VLL doublet τ' of 1 TeV mass in the 4L final state, using S_T distribution.

4.6 Calculating Limits

The N_{obs}, N_{bkg} and δN_{bkg} obtained from the search results [16] and the N_{sig} obtained using eq 4.1 for each SR can be used to calculate the upper limits in the VLL doublet τ' particle production cross-section. The limits are calculated by the CMS COMBINE software [21] using the Asymptotic Limit method.

The SRs are arranged in the decreasing order of their signal-to-background ratio ($S/\sqrt{B}$). Figure 4.9 shows the limit on the cross-section for a 1 TeV τ' particle with respect to the number of SRs with highest $S/\sqrt{B}$ considered to calculate the limit. Calculating the limits for all the SRs combined gives the best limits. However, it is more computationally expensive. Moreover, the limit plateaus out at a high number, whereas the time to calculate the limits increases exponentially. Therefore, the best strategy is to find the sweet spot with a good enough limit and a low enough computational expense. However, the time spent to calculate the limits is of the order of a few minutes, so we are not constrained by the number of SRs considered.

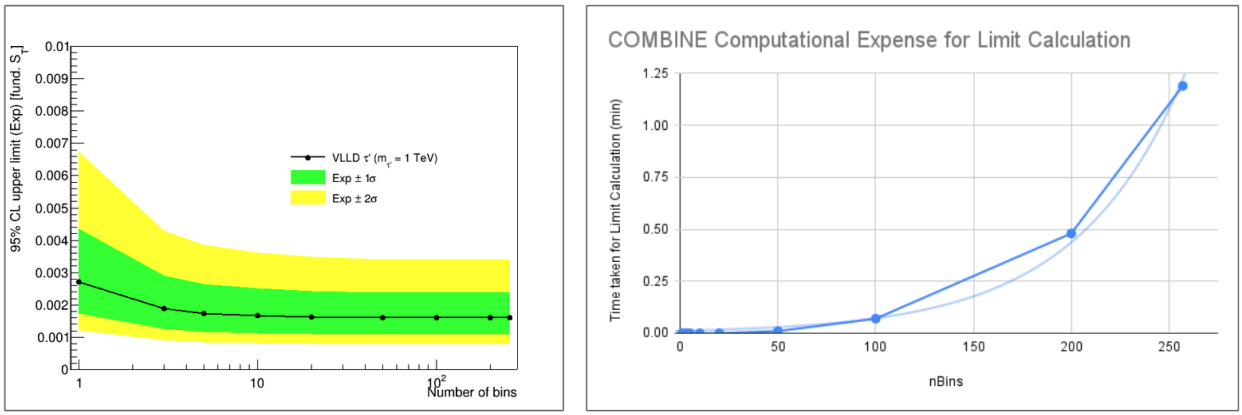


Figure 4.9: (a) (σB) limit for a 1 TeV τ' particle with respect to the number of SRs considered. (b) Time taken to calculate the limits for each number of SRs considered. As more bins are considered, the limit improves till a minimum, but the time taken to calculate the limits increases exponentially.

Figure 4.10 shows the expected cross-section limits from the reinterpretation using S_T and $L_T + p_T^{miss}$ distributions overlaid on top of the expected limits from the paper. The figure also shows the production cross-section for the VLL doublet model [5, 6]. We can observe an excellent agreement between the reinterpreted limits produced here and those from the paper. There is a slight

deviation at lower masses; this is attributed to the fact that I consider only statistical uncertainties and not systematic ones. This results in the reinterpreted limits being slightly stronger than the published ones. This agreement validates my reinterpretation workflow, ensuring that all effects are appropriately implemented.

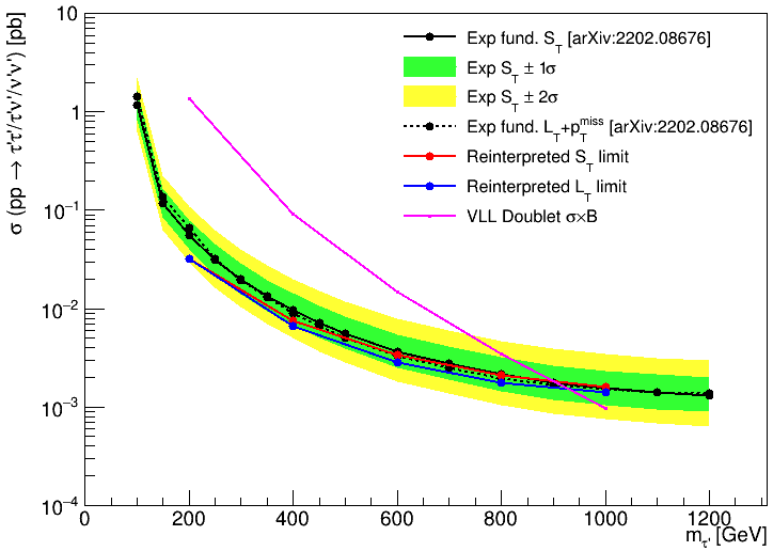


Figure 4.10: The figure shows the expected upper limits on the cross-section of the VLL doublet τ' particle calculated using the S_T and $L_T + p_T^{miss}$ tables compared to the corresponding limits from [4]. Additionally, the production cross-section for the VLL doublet model is also shown. The figure shows an exclusion for τ' of mass < 900 GeV consistent with the reinterpreted limits and those from the paper.

4.7 Reinterpreting the VLL μ-like model

With the framework of reinterpretation in place, I can now proceed to reinterpret the results for the VLL doublet μ-like model. Figure 4.11 shows the Feynman diagram for the production and decay of the VLL doublet μ' particle. It is produced and decays similarly to τ' (as described in Sec. 4.2), the difference being that it decays into a muon and a boson instead of a tau and a boson. However, this signal gives signatures significantly different from the former due to the final state having a majority of light leptons instead of taus. Figure 4.12 shows the signal acceptance $(A\varepsilon)$ of different final states for various mass hypotheses of the μ' particle. It shows that the 3L final state has the

most $(A\varepsilon)$ among the different final states.

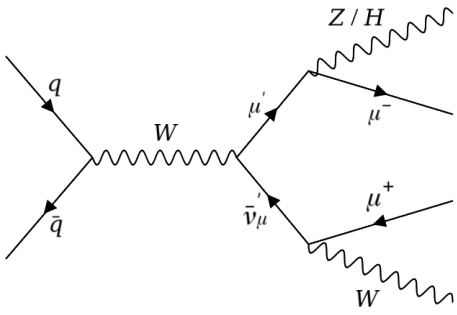


Figure 4.11: An example Feynman diagram for the doublet VLL μ-like model. The VLLs can be pair-produced or associated-produced and always decay to a standard model μ and a boson.

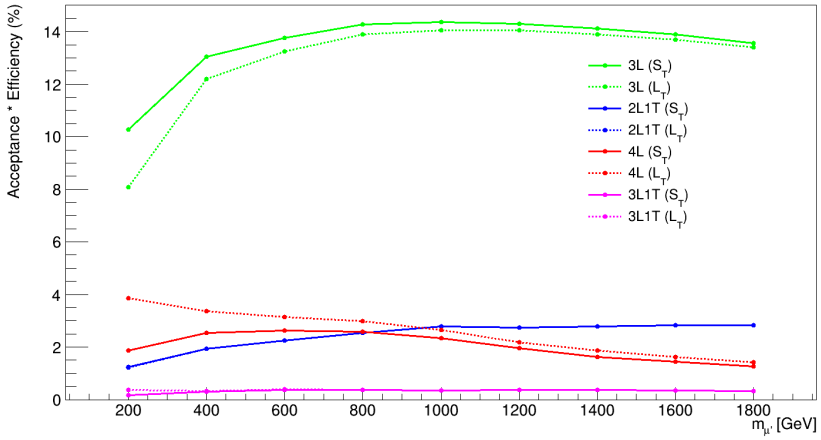
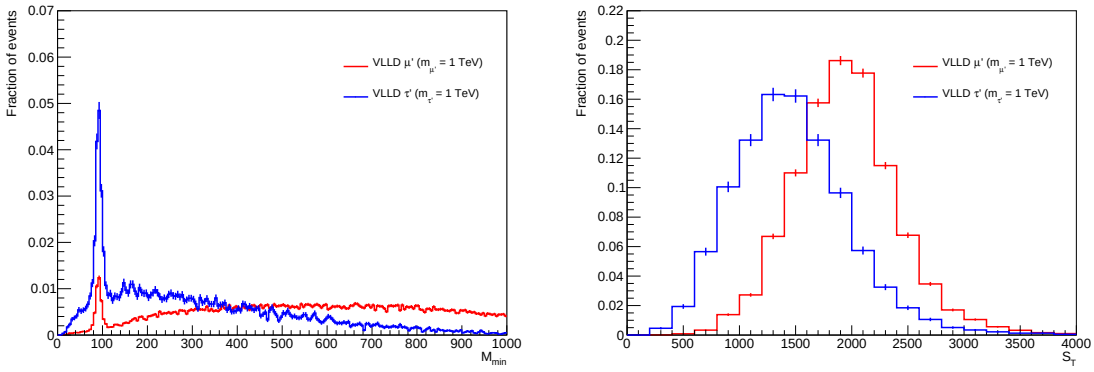


Figure 4.12: The signal acceptance of different final states for various mass hypotheses of the doublet VLL μ-like model. The 3L final state shows clear majority for every mass hypothesis.

Figure 4.13 shows a few kinematic variables for the VLL μ' model compared to the VLL τ' model, both with masses of 1 TeV. Each distribution is scaled so that the area under its curve is 1. The M_{min} variable has a more prominent peak at the Z mass for the τ' model, whereas that for the μ' model is spread towards higher values. The μ' model also displays higher S_T values than the τ' model. This difference also manifests in the signal yield in the SRs. Figure 4.14 shows the signal

yield for the VLL doublet μ-like model with $m_{\mu'} = 1$ TeV in the same SRs displayed in Fig 4.8. The signal yield varies across the different SRs for the two models.



(a) M_{min} (b) S_T

Figure 4.13: The figure shows the M_{min} and S_T variables for the VLL μ' model compared to the VLL τ' model. The M_{min} variable has a more prominent peak at the Z mass for the τ' model, whereas that for the μ' model is spread towards higher values. The S_T variable has higher values for the μ' model than the τ' model.

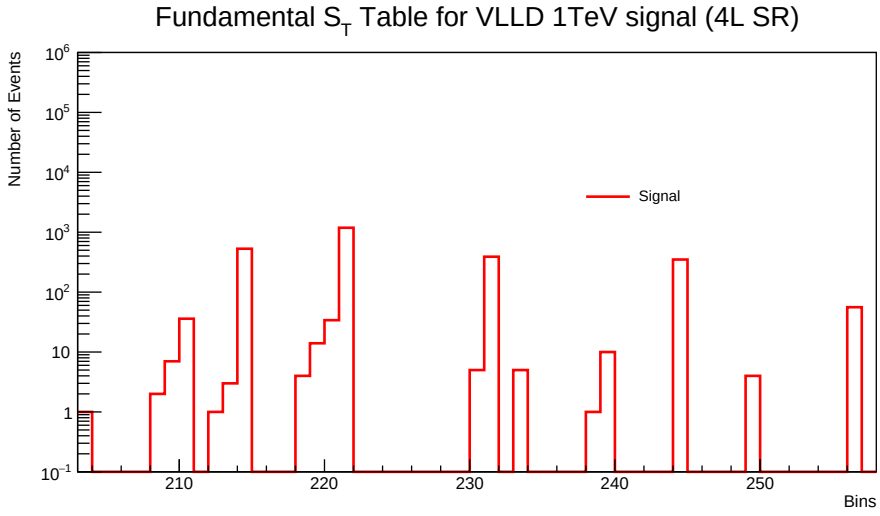


Figure 4.14: Signal yield for the VLL doublet μ' of 1 TeV mass in the 4L final state, using S_T distribution. We can observe a difference in the signal yield in the SRs between the VLL doublet μ' model and the VLL doublet τ' model (displayed in Fig 4.8).

The theoretical production cross-section for the μ' particle can be obtained at leading order (LO)

for any mass hypothesis from the MADGRAPH simulations. The internal documentation of the paper [4] provides the next-to-leading order (NLO) cross-section values for ≤ 1 TeV mass hypotheses. To obtain the NLO cross-section values at higher mass hypotheses, we first calculate the ratio of NLO to LO cross-section for the lower mass hypotheses. This correction is then applied to the LO cross-section to estimate the NLO cross-section for higher mass hypotheses of the μ' particle.

Figure 4.15 shows the expected limits from the reinterpretation of the search results on the VLL doublet μ-like model using S_T and $L_T + p_T^{miss}$ distributions, with the theoretical NLO cross-sections. From the figure, we can conclude that the VLL doublet μ' model is excluded for mass < 1200 GeV. This constraint is novel and is comparable to the constraints published by the ATLAS collaboration [22] (1270 GeV).

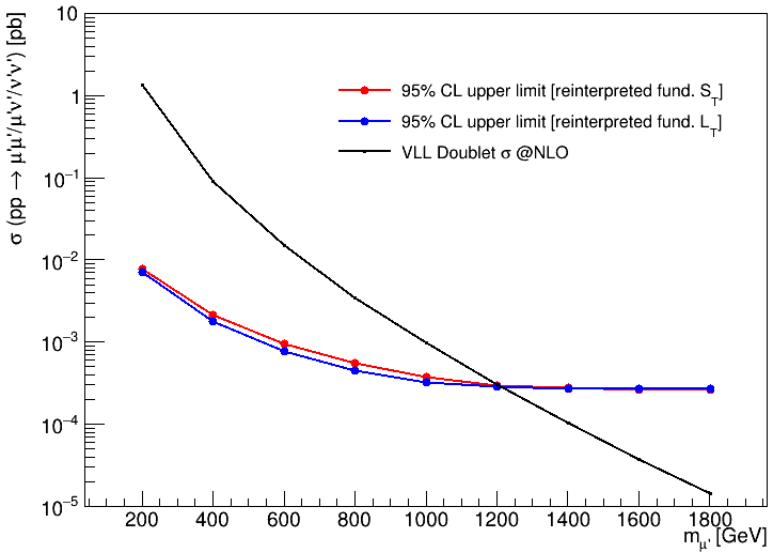


Figure 4.15: The figure shows the expected upper limit on the cross-section of the VLL doublet μ' particle calculated using the reinterpretated S_T and $L_T + p_T^{miss}$ tables. Additionally, the production cross-section for the VLL doublet model is also shown. The figure shows an exclusion for μ' of mass < 1200 GeV

Chapter 5

Analysing a Complex Model

The quintuplet model [7] describes a BSM model with quintuplet fermions and quartet scalars, where the fermions decay into the scalars. It is a complex model with multiple theoretical parameters due to the two groups. Thus, the model is non-trivial to analyse and needs a robust analysis strategy to target the various parameters systematically.

5.1 Model Description

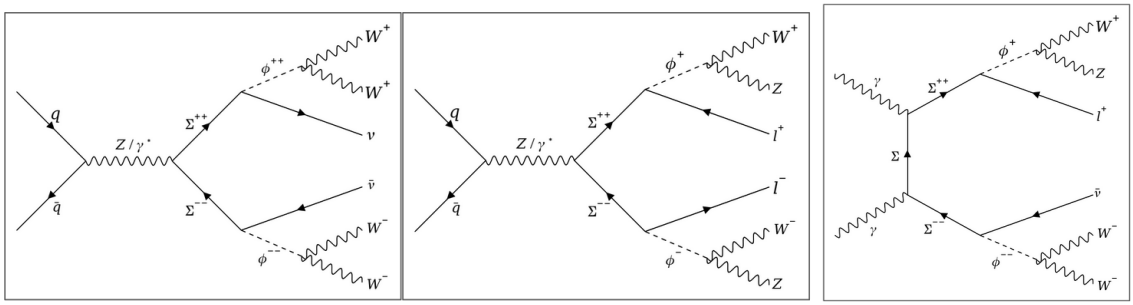
The model is derived from a cascade seesaw-like scenario and comprises two multiplets:

- Quintuplet fermions: $\Sigma_R = [\Sigma_1^{++}, \Sigma_1^{+}, \Sigma^{0}, \Sigma_2^{-}, \Sigma_2^{--}]_R^T$
- Quartet scalars: $\Phi_4 = (\Phi^{++}, \Phi_2^{+}, \Phi^{0}, \Phi_1^{-})^T$

According to the model, $(\Sigma_1^{\pm\pm}, \Sigma_2^{\pm\pm})$ and $(\Sigma_1^{\pm}, \Sigma_2^{\pm})$ are combined to form $\Sigma^{\pm\pm}$ and $\Sigma^{\pm}$ respectively. Moreover, only $\Sigma^{\pm\pm}$ is considered for further analysis due to the production cross-section of $\Sigma^{\pm}$ ($p\ p \rightarrow \Sigma^{\pm}\ \Sigma^{\pm}$) being way smaller than that of $\Sigma^{\pm\pm}$ ($p\ p \rightarrow \Sigma^{\pm\pm}\ \Sigma^{\pm\pm}$) for a hadron collider.

$\Sigma^{\pm\pm}$ can either decay to a doubly-charged $\Phi^{\pm\pm}$ ($\Sigma^{\pm\pm} \rightarrow \Phi^{\pm\pm}\ \nu(\overline{\nu})$) or a singly-charged $\Phi^{\pm}$ ($\Sigma^{\pm\pm} \rightarrow \Phi_2^{\pm}\ l^{\pm}$). $\Phi^{\pm\pm}$ decays as ($\Phi^{\pm\pm} \rightarrow W^{\pm}\ W^{\pm}$), whereas $\Phi_2^{\pm}$ (taken as $\Phi^{\pm}$ from now) decays as ($\Phi^{\pm} \rightarrow W^{\pm}\ Z$). The photon parton distribution function (PDF) for the proton is also added

to the initial state of the process to boost the production cross-section. Figure 5.1 shows the Feynman diagrams of the quintuplet model, showing the various production and decay modes.



(a) Quark Initial State (b) Photon Initial State

Figure 5.1: The figure shows the Feynman diagrams of the quintuplet model, showing the various the production and decay modes, as described in section 5.1.

The inclusion of photon PDF raises the question of the choice of PDF to use when generating simulation samples. To tackle this problem, multiple samples are generated with only-quark initial state (IS), only-photon initial state and inclusive initial state using three different PDFs for two mass hypotheses of the $\Sigma^{\pm\pm}$ particle. For this, the LHAPDF [23] software is incorporated with MADGRAPH at the point of process generation. Table 5.1 shows the production cross-sections and the relative increase in cross-sections by adding photon PDF for the three PDFs and two mass hypotheses.

From the table, we can infer that there is a significant increase in the cross-section when the photon PDF is added at high mass but a faint increase at low mass for all the PDFs. Moreover, the 'NNPDF23_nlo_as_0119_qed' generates the highest cross-sections consistently, and will be used in further analysis.

5.2 Object Selection

The object selections followed here are similar to those for the VLL doublet model. However, due to the high mass of $\Sigma^{\pm\pm}$ resulting in a boosted scenario, fatjets (described in Chap. 3) are also

Mass of $\Sigma^{\pm\pm}$ (GeV)	PDF used	Production cross-section (pb)			Percentage increase in cross-section (%)
		Only quark IS	Only photon IS	Inclusive IS	
400	NNPDF23_lo_as_0130_qed	0.1616	0.0085	0.1697	5.012
	NNPDF23_lo_as_0119_qed	0.1595	0.0079	0.1676	5.078
	NNPDF23_nlo_as_0119_qed	0.1851	0.0094	0.1944	5.024
300	NNPDF23_lo_as_0130_qed	0.001649	0.0003004	0.001951	18.314
	NNPDF23_lo_as_0119_qed	0.001726	0.0002802	0.002005	16.165
	NNPDF23_nlo_as_0119_qed	0.001767	0.0003269	0.002089	18.223

Table 5.1: Production cross-sections of two mass hypotheses of the $\Sigma^{\pm\pm}$ particle for various initial states, considering three different PDFs. The uncertainties in the cross-sections are typically $< 1\%$. We observe a significant increase in the cross-section when the photon PDF is added at high mass but a faint increase at low mass for all the PDFs. Moreover, the 'NNPDF23_nlo_as_0119_qed' generates the highest cross-sections.

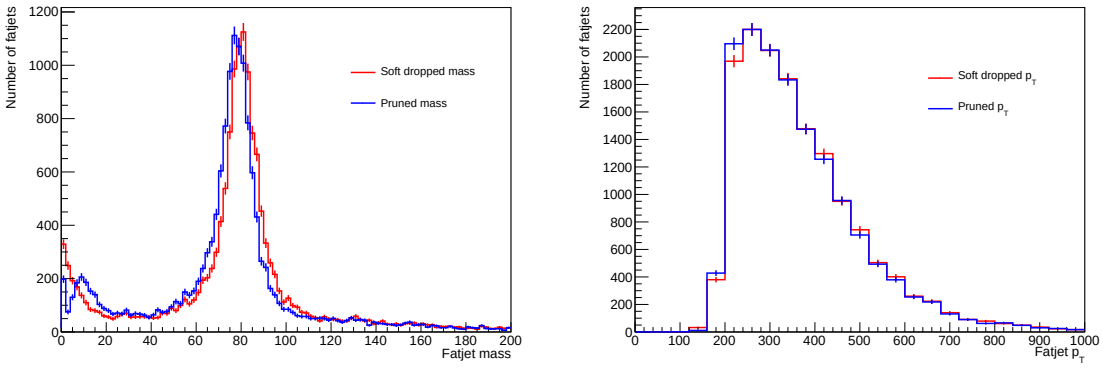
considered and selected. They are as follows:

- $p_T > 10$ GeV, $|\eta| < 2.4$ for e's and μ's
- $p_T > 20$ GeV, $|\eta| < 2.3$ for τ_h's
- Only the leptons originating from a signal particle, a boson or light leptons originating from τ leptons are considered.
- The leading lepton has to satisfy $p_T > 30$ GeV.
- $p_T > 30$ GeV, $|\eta| < 2.4$ for jets
- The jets need to have an angular separation of 0.4 or more from the selected leptons in the event.
- $p_T > 200$ GeV, $|\eta| < 2.4$ for fatjets
- The fatjets need to have an angular separation of 0.4 or more from the selected leptons in the event.

Apart from the object selections, the analysis uses several fatjet substructure variables to select the desired fatjets in the signal processes. The variables are:

- Groomed jet variables
- Jet n-subjettiness
- ρ variable

Due to the fatjets being significantly larger than jets, radiation from an underlying event or pileup can result in an overestimation of the fatjet invariant mass and loss in the mass resolution. **Jet grooming** algorithms are generally applied to the fatjets to account for these effects by removing unwanted soft radiation. There are two methods of jet grooming: soft-drop algorithm [24] and jet pruning [25]. To compare the two methods, the resulting fatjet masses and p_T's after grooming are overlaid for the signal sample (obtained from W boson decay), shown in Fig. 5.2a and 5.2b respectively.



(a) Fatjet mass

(b) Fatjet p_T

Figure 5.2: The figures show the fatjet mass and p_T for the fatjets of a signal sample groomed with two different algorithms: soft-drop algorithm and jet pruning. The soft-drop mass peaks closer to the W boson mass (80 GeV) and the soft-drop p_T is slightly higher as compared to the pruned counterparts.

The above figures show that the two algorithms perform similarly. One can say that the soft-drop algorithm works slightly better as the jet mass is closer to the W boson mass (80 GeV) for same compared to jet pruning algorithm. Moreover, the jet p_T is slightly higher for soft-dropped fatjets. Thus, soft-dropped fatjets are used for further analysis. Figure 5.3 shows the behaviour of soft-dropped fatjet mass for several SM processes and signal process. The soft-dropped mass peaks

around the W boson mass for the signal, expected due to the fatjet coming from a W.

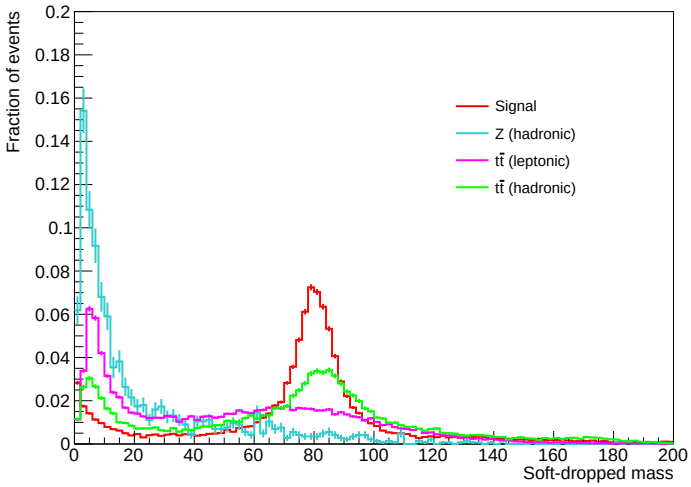
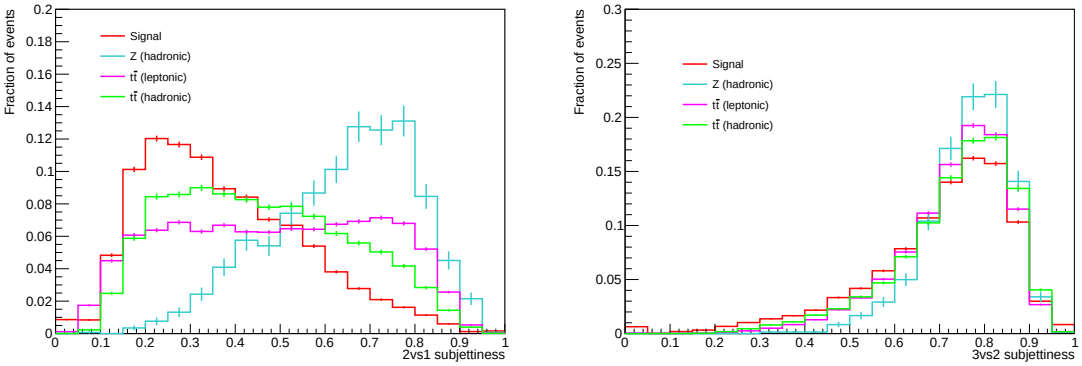


Figure 5.3: The figure shows the soft-dropped fatjet mass for the fatjets of a signal sample and several SM processes. The soft-dropped mass peaks around the W boson mass for the signal.

The **n-subjettiness** varibles (τ_n) [26] do not provide with much information on their own. However, the ratios of subsequent n-subjettiness ($\tau_{p(p-1)} = \tau_p/\tau_{p-1}$) presents a relative probablity of the jet being p-pronged (containing p subjets) or (p-1)-pronged (containing (p-1) subjets), according to the value being closer to 0 or 1 respectively. There are two varibales used in this regime: τ_{21} and τ_{32}. Figures 5.4a and 5.4b show the behaviour of τ_{21} and τ_{32} variables for several SM processes and signal process. The τ_{21} and τ_{32} peak at low and high values respectively for the signal, expected due to the fatjet coming from a W, which has 2 prongs.

The **relative mass** (ρ) [27] is calculated and filtered for each fatjet to avoid regions where non-perturbative effects result in inconsistencies in generator predictions. It is constructed as: $\rho = 2\ln(m_{SD}/p_T)$. Figure 5.5 shows the behaviour of the ρ variable for several SM processes and signal process. The ρ variable peaks at high values for the signal, however, the different mass points of the signal behave differently in the ρ distributions.



(a) τ_{21} (b) τ_{32}

Figure 5.4: The figures show the τ_{21} and τ_{32} variables for the fatjets of a signal sample and several SM processes. The soft-dropped mass peaks around the W boson mass for the signal, expected due to the fatjet coming from a W. The τ_{21} and τ_{32} peak at low and high values respectively for the signal.

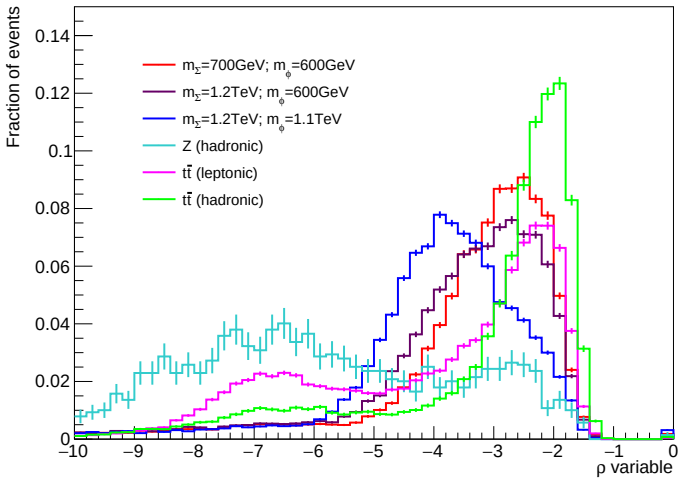


Figure 5.5: The figure shows the relative mass for the fatjets of several signal sample and SM processes. The ρ variable peaks at high values for the signal, however, the different mass points of the signal behave differently in the ρ distributions.

5.3 Signal Regions

The quintuplet model, being a complex model, behaves differently than the VLL model. Using the CMS multilepton search to analyse the quintuplet model has several disadvantages. Figure 5.6 shows the signal acceptance $(A\varepsilon)$ for a quintuplet model with a mass point ($m_{\Sigma^{\pm\pm}} = 400$ GeV, $m_{\phi^{\pm\pm}} = 300$ GeV) compared to a VLL τ-like model with a mass of 400 GeV. The $(A\varepsilon)$ for the quintuplet model is different for the final states compared to the VLL model. The quintuplet model has lower $(A\varepsilon)$ than the VLL model as the multilepton search was not designed for it. In addition, due to multiple mass parameters in hand, the quintuplet model is sensitive to changes in any parameter. For example, Fig. 5.7 shows the distribution of M_{min} for two mass points: ($m_{\Sigma^{\pm\pm}} = 1000$ GeV, $m_{\phi^{\pm\pm}} = 300$ GeV) and ($m_{\Sigma^{\pm\pm}} = 1000$ GeV, $m_{\phi^{\pm\pm}} = 900$ GeV). We can see that changing a single parameter results in a drastic change in the sample properties. Furthermore, the multilepton search does not include jets or fatjets in its final state. Therefore, there is a need for a new analysis strategy.

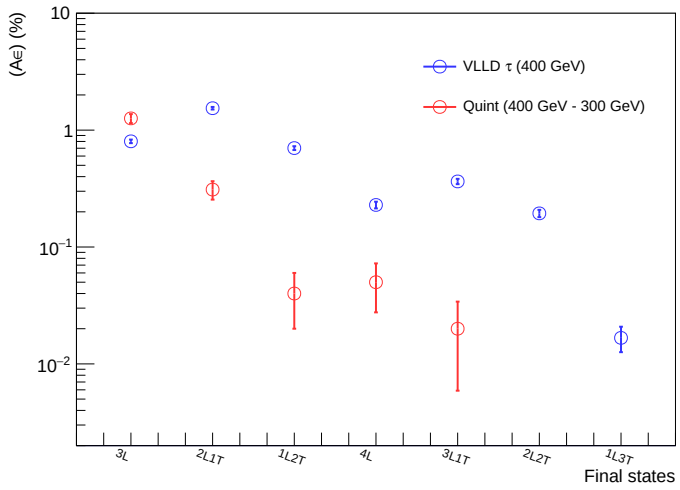


Figure 5.6: The figure shows the signal acceptance $(A\varepsilon)$ for a quintuplet model with a mass point ($m_{\Sigma^{\pm\pm}} = 400$ GeV, $m_{\phi^{\pm\pm}} = 300$ GeV) compared to a VLL τ-like model with a mass of 400 GeV. The $(A\varepsilon)$ for the quintuplet model in 2L2T and 1L3T final states are 0, hence not shown in the figure. The $(A\varepsilon)$ for the quintuplet model is different (and generally lower) for the final states compared to the VLL model.

The desired events are selected using several kinematic variables and signal regions are formed via further selections on the kinematic variables. These kinematic variables are listed below:

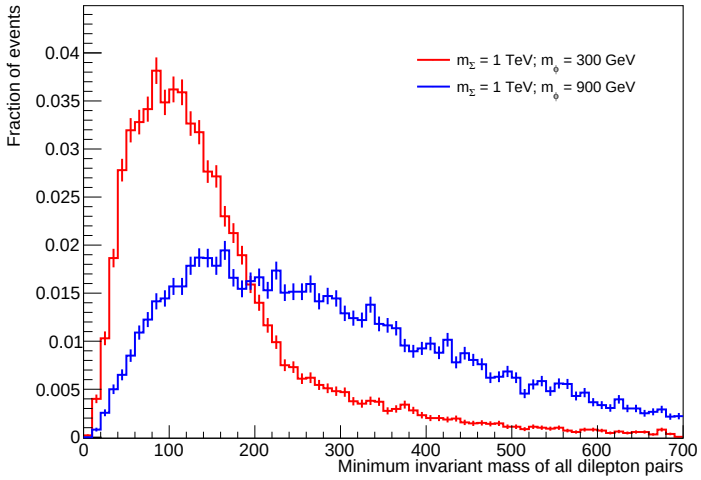


Figure 5.7: The figure shows the M_{min} variable for two mass points of the quintuplet model: ($m_{\Sigma^{\pm\pm}} = 1000$ GeV, $m_{\phi^{\pm\pm}} = 300$ GeV) and ($m_{\Sigma^{\pm\pm}} = 1000$ GeV, $m_{\phi^{\pm\pm}} = 900$ GeV). We can see that changing just the $m_{\phi^{\pm\pm}}$ parameter results in a drastic change in the M_{min} distribution.

- p_T^{miss}: p_T^{miss} is defined as the magnitude of the vector sum of p_T of all the observed particles in an event.

$$p_T^{miss} = \left| \sum_{i=observed} \overrightarrow{p_T}^i \right|$$

- L_T: L_T is defined as the scalar p_T sum of all the charged leptons that constitute the final state.

$$L_T = \sum_{i=leptons} \left| \overrightarrow{p_T}^i \right|$$

- H_T: H_T is defined as the scalar p_T sum of all the jets.

$$H_T = \sum_{i=jets} \left| \overrightarrow{p_T}^i \right|$$

- S_T: S_T is the sum of L_T, H_T and p_T^{miss}.

$$S_T = L_T + H_T + p_T^{miss}$$

- M_{min}: M_{min} is defined as the minimum invariant mass of all the dilepton pairs, irrespective of their charge or flavour.

- dR_{min}: dR_{min} is defined as the minimum angular distance between all the dilepton pairs in the event.

- Jet momentum fraction (JMF): JMF is defined as the ratio of the scalar p_T sum of the selected fatjets to the scalar p_T sum of all the jets.

$$JMF = \frac{\sum_{i=fatjets} \left|\overrightarrow{p_T}^i\right|}{\sum_{i=jets} \left|\overrightarrow{p_T}^i\right|}$$

Figure 5.8 shows the $4W$ state obtained from the signal process, which is the most common state due to the branching fraction of $\Sigma^{\pm\pm} \to \Phi^{\pm\pm}$ being 98.4%. The W's obtained are boosted due to high masses of $\Sigma^{\pm\pm}$ and $\phi^{\pm\pm}$, thus giving boosted fatjets in the case of hadronic decay. Due to boosted scenario in the samples, two final states have been chosen for fatjet analysis: 2L2Fj and 3L1Fj with respect to the number of hadronically and leptonically decaying W's. Here, L denotes the leptons and Fj denotes the fatjets. Moreover, three mass points of the signal sample are considered for analysis, named by the mean of the masses of $\Sigma^{\pm\pm}$ and $\phi^{\pm\pm}$: **low-mass**($m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV), **medium-mass**($m_\Sigma = 1200$ GeV, $m_\phi = 600$ GeV) and **high-mass**($m_\Sigma = 1200$ GeV, $m_\phi = 1100$ GeV).

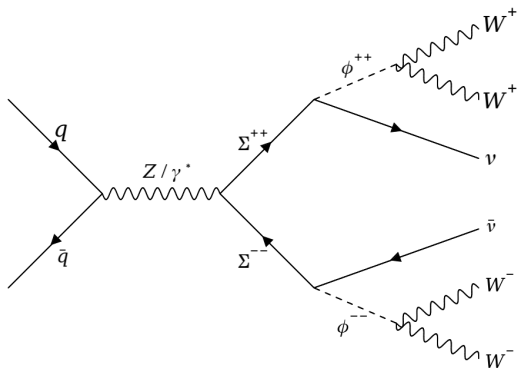


Figure 5.8: The figure shows the Feynman diagram for the quintuplet model with the most common, $4W$ state. The W's can further decay leptonically or hadronically according to the standard model.

5.4 Backgrounds

The standard model processes which can give rise to the 2L2Fj or 3L1Fj final states constitute their backgrounds. For each final state, three background processes with the highest cross-sections are estimated. Other unaccounted background processes are assumed to contribute insignificantly to the total background. Following is a list of the background processes considered for each final state:

- 2L2Fj final state:
  - $t\bar{t}$ (leptonic final state)
  - $t\bar{t}Z$ (leptonically decaying Z)
  - ZVV (leptonically decaying Z, hadronically decaying V)
- 3L1Fj final state:
  - $WZ + jets$ (leptonically decaying W and Z)
  - $t\bar{t}Z$ (leptonic decays)
  - $t\bar{t}W$ (leptonic decays)

Here, V refers to vector bosons ($W^{\pm}$ or Z). These background processes are generated via the simulation softwares: MADGRAPH5_AMC@NLO, PYTHIA 8.3 and DELPHES. Table 5.2 shows the cross-sections, number of events generated and luminosities for each signal and background samples.

5.5 2L2Fj Final State

For the 2L2Fj final state, three backgrounds have been estimated. These are:

- $t\bar{t}$ (leptonic final state)
- $t\bar{t}Z$ (leptonically decaying Z)

Final State	Process	Mass Point	Cross-section (pb)	Number of events generated	Luminosity (fb^{-1})
2L2Fj	Quintuplet	$m_\Sigma = 1200$ GeV, $m_\phi = 1100$ GeV	1.455×10^{-4}	5×10^4	10^5
		$m_\Sigma = 1200$ GeV, $m_\phi = 600$ GeV	1.455×10^{-4}	5×10^4	10^5
		$m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV	3.627×10^{-3}	5×10^4	10^4
	$t\bar{t}$	-	4.384	10^6	228
	$t\bar{t}Z$	-	5.13×10^{-2}	5×10^5	10^4
	ZVV	-	1.017×10^{-3}	2×10^6	10^6
3L1Fj	Quintuplet	$m_\Sigma = 1200$ GeV, $m_\phi = 1100$ GeV	4.405×10^{-5}	5×10^4	10^6
		$m_\Sigma = 1200$ GeV, $m_\phi = 600$ GeV	4.405×10^{-5}	5×10^4	10^6
		$m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV	1.098×10^{-3}	5×10^4	10^4
	$WZ + jets$	-	5.538×10^{-2}	5×10^5	10^4
	$t\bar{t}Z$	-	6.745×10^{-3}	5×10^5	10^5
	$t\bar{t}W$	-	2.239×10^{-3}	5×10^5	10^5

Table 5.2: The cross-sections, number of events generated and luminosities of the signal and background processes for the quintuplet model.

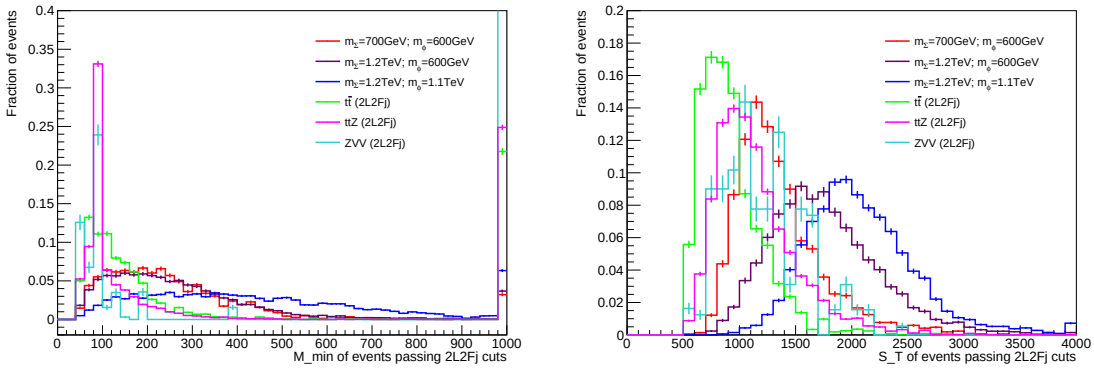
- ZVV (leptonically decaying Z, hadronically decaying V)

Preliminary event selections are made to retain a large signal acceptance. The selections are as follows:

- $p_T^{miss} > 50$ GeV
- $m_{min} > 50$ GeV
- $H_T > 200$ GeV

- $S_T > 500$ GeV
- $L_T + p_T^{miss} > 100$ GeV

After the initial cuts, several kinematic variables are plotted and overlaid for signal and background samples. For example, Fig. 5.9a and 5.9b show the M_{min} and S_T variables respectively for signal and background processes, normalised such that the integral of the curves is 1. The shapes of the distributions are analysed and a new set of cuts is proposed.



(a) M_{min} (b) S_T

Figure 5.9: The figures show the M_{min} and S_T for the signal and background processes in the 2L2Fj final state post preliminary cuts, analysed for the quintuplet model. The curves are normalised, having an integral of 1. Inferring from the shapes of the distributions, a new set of cuts is proposed.

After analysing the plots, a new set of cuts is proposed to enhance the signal-to-background ratio for the final state. Table 5.3 tabulates the cuts for different mass points.

To check the relevance of these cuts, the corresponding variables are plotted after implementing all the other cuts except the specific cut. In this way, the redundancy of the cut is tested. For example, Fig. 5.10a shows the H_T variable for the signal and background processes, after all the other medium-mass cuts are applied. Similarly, Fig. 5.10b shows the JMF variable for the signal and background processes, after all the other low-mass cuts are applied. Note that the backgrounds are stacked on top of each other and the signal is overlaid. Furthermore, the signal and background curves shown are cross-section-weighted, and are measures of their respective yields.

Variable	Low-mass cuts	Med-mass cuts	High-mass cuts
H_T	> 500 GeV	> 500 GeV	> 700 GeV
L_T	> 100 GeV	> 100 GeV	> 200 GeV
$L_T + p_T^{miss}$	> 300 GeV	> 500 GeV	> 500 GeV
S_T	> 900 GeV	> 1100 GeV	> 1400 GeV
M_{min}	> 100 GeV		
dR_{min}	> 1		
p_T^{miss}	> 100 GeV		
JMF	$0.7 - 1.1$		

Table 5.3: Cuts implemented on the signal and background processes in the 2L2Fj final state, analysed for the quintuplet model.

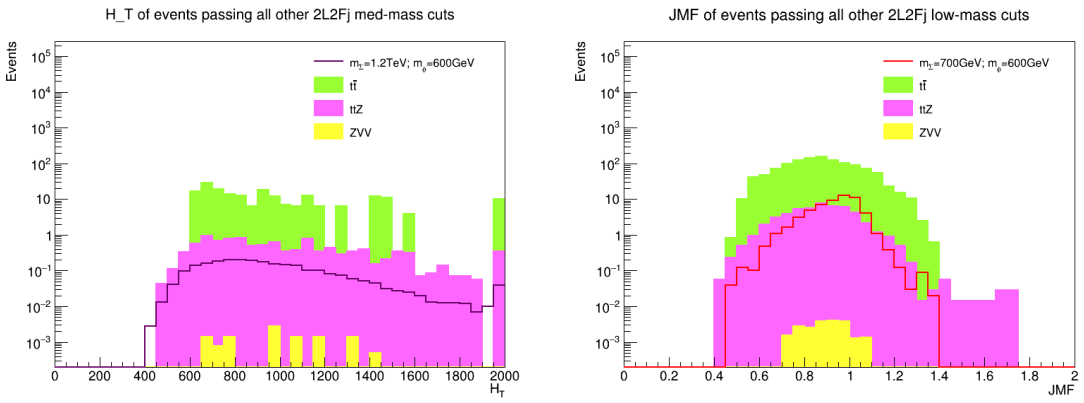
After analysing the plots, we conclude that most of the cuts are relevant. All the cuts are retained to exclude any unaccounted background. The fatjet substructure variables are also cut on to obtain the desired fatjets. These cuts are applied only to the two leading fatjets. The cuts are:

- 60 GeV $< m_{SD} < 100$ GeV
- $\tau_{21} < 0.8$
- $\tau_{32} > 0.3$
- $\rho > -5.5$

5.6 3L1Fj Final State

For the 3L1Fj final state, three backgrounds have been estimated. These are:

- $WZ + jets$ (leptonically decaying W and Z)
- $t\bar{t}Z$ (leptonic decays)
- $t\bar{t}W$ (leptonic decays)



(a) H_T (b) JMF

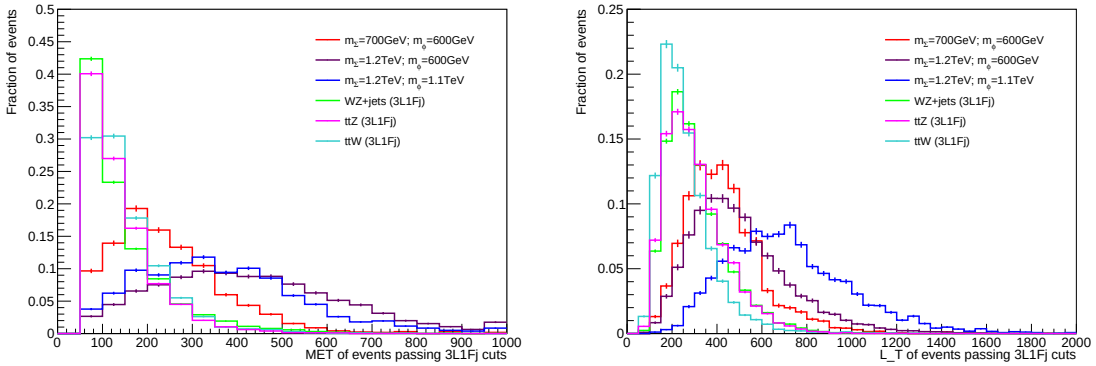
Figure 5.10: The figures show the H_T and JMF for the signal and background processes in the 2L2Fj final state, analysed for the quintuplet model, after all the other respective mass cuts are applied. The curves are cross-section-weighted, being measures of their respective yields.

Preliminary event selections are made to retain a large signal acceptance. The selections are as follows:

- $p_T^{miss} > 50$ GeV
- $m_{min} > 50$ GeV
- $H_T > 200$ GeV
- $S_T > 400$ GeV
- $L_T + p_T^{miss} > 200$ GeV

After the initial cuts, several kinematic variables are plotted and overlaid for signal and background samples. For example, Fig. 5.11a and 5.11b show the p_T^{miss} and L_T variables respectively for signal and background processes, normalised such that the integral of the curves is 1. The shapes of the distributions are analysed and a new set of cuts is proposed.

After analysing the plots, a new set of cuts is proposed to enhance the signal-to-background ratio for the final state. Table 5.4 tabulates the cuts for different mass points.



(a) p_T^{miss} (b) L_T

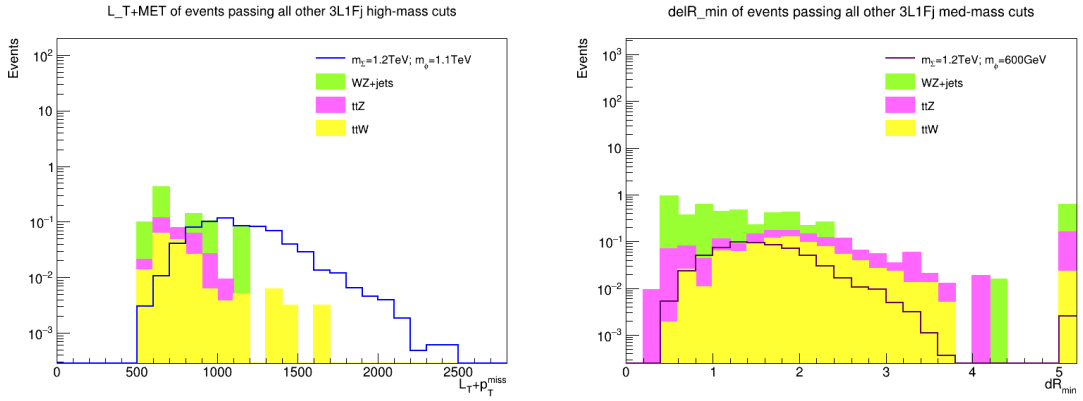
Figure 5.11: The figures show the p_T^{miss} and L_T for the signal and background processes in the 3L1Fj final state post preliminary cuts, analysed for the quintuplet model. The curves are normalised, having an integral of 1. Inferring from the shapes of the distributions, a new set of cuts is proposed.

Variable	Low-mass cuts	Med-mass cuts	High-mass cuts
H_T	> 200 GeV	> 200 GeV	> 400 GeV
L_T	> 200 GeV	> 200 GeV	> 400 GeV
$L_T + p_T^{miss}$	> 400 GeV	> 500 GeV	> 700 GeV
S_T	> 700 GeV	> 800 GeV	> 1100 GeV
M_{min}	> 100 GeV		
dR_{min}	> 1		
p_T^{miss}	> 100 GeV		
JMF	$0.7 - 1.1$		

Table 5.4: Cuts implemented on the signal and background processes in the 3L1Fj final state, analysed for the quintuplet model.

To check the relevance of these cuts, the corresponding variables are plotted after implementing all the other cuts except the specific cut. In this way, the redundancy of the cut is tested. For example, Fig. 5.12a shows the $L_T + p_T^{miss}$ variable for the signal and background processes, after all the other high-mass cuts are applied. Similarly, Fig. 5.12b shows the dR_{min} variable for the signal and background processes, after all the other low-mass cuts are applied. Note that the backgrounds are

stacked on top of each other and the signal is overlaid. Furthermore, the signal and background curves shown are cross-section-weighted, and are measures of their respective yields.



(a) $L_T + p_T^{miss}$ (b) dR_{min}

Figure 5.12: The figures show the $L_T + p_T^{miss}$ and dR_{min} for the signal and background processes in the 3L1Fj final state, analysed for the quintuplet model, after all the other respective mass cuts are applied. The curves are cross-section-weighted, being measures of their respective yields.

After analysing the plots, we conclude that most of the cuts are relevant. All the cuts are retained to exclude any unaccounted background. The fatjet substructure variables are also cut on to obtain the desired fatjets. These cuts are applied only to the two leading fatjets. The cuts are:

- 60 GeV $< m_{SD} <$ 100 GeV
- $\tau_{21} < 0.8$
- $\tau_{32} > 0.3$
- $\rho > -5.5$

5.7 Results

After obtaining the signal and background yields (N_{sig} and N_{bkg}) for both the final states and all three mass points, the signal-to-background ($S/\sqrt{B}$) ratios are calculated to check the sensitivity of the signal process to the backgrounds. Moreover, the signal acceptance is calculated for both

the final states to check the applicability of the signal on the final states. Table 5.5 records the signal-to-background ratios and signal acceptances for different signal mass points in the two final states.

Signal Mass Point	Final State	$S/\sqrt{B}$	Acceptance (%)
$m_\Sigma = 1200$ GeV, $m_\phi = 1100$ GeV	2L2Fj	2.534	12.6
	3L1Fj	3.907	8.3
$m_\Sigma = 1200$ GeV, $m_\phi = 600$ GeV	2L2Fj	1.119	6.44
	3L1Fj	0.496	6.41
$m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV	2L2Fj	5.736	5.75
	3L1Fj	8.962	5.94

Table 5.5: The signal-to-background ratios and signal acceptances for different signal mass points in the two final states for the quintuplet model. In the $S/\sqrt{B}$ ratio, the 2L2Fj final state performs better in the medium mass point, and the 3L1Fj final state performs better in the low and high mass points. However, in the signal acceptance, the 2L2Fj final state performs better in the high mass point, and the 3L1Fj final state performs better in the low mass point.

From the table, we can conclude that in the $S/\sqrt{B}$ ratio, the 2L2Fj final state performs better in the medium mass point, and the 3L1Fj final state performs better in the low and high mass points. Moreover, in the signal acceptance, the 2L2Fj final state performs better in the high mass point, and the 3L1Fj final state performs better in the low mass point.

Table 5.6 shows the expected upper limits on the cross-sections of the different final states and mass points of the quintuplet model, with a 10%, 50% and 100% background uncertainty taken, along with the theoretical cross-sections for the respective final states. We conclude that according to the above selections and backgrounds considered, the mass point ($m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV) is excluded.

Signal Mass Point	Final State	Uncertainty (%)	Expected Limit ($\times 10^{-4}$ pb)	Theoretical cross-section ($\times 10^{-4}$ pb)
$m_\Sigma = 1200$ GeV, $m_\phi = 1100$ GeV	2L2Fj	10	1.9609	1.455
		50	1.9922	
		100	2.0391	
	3L1Fj	10	1.8125	0.4405
		50	1.8281	
		100	1.8672	
$m_\Sigma = 1200$ GeV, $m_\phi = 600$ GeV	2L2Fj	10	4.1719	1.455
		50	4.3906	
		100	4.7344	
	3L1Fj	10	3.4219	0.4405
		50	3.4844	
		100	3.6094	
$m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV	2L2Fj	10	15.125	36.27
		50	23.6875	
		100	29.875	
	3L1Fj	10	4.1875	10.98
		50	4.4219	
		100	4.8125	

Table 5.6: The figure shows the expected upper limits on the cross-sections of the different final states and mass points of the quintuplet model, with a 10%, 50% and 100% background uncertainty taken, besides the theoretical cross-sections for the respective final states. We conclude that according to the above selections and backgrounds considered, the mass point ($m_\Sigma = 700$ GeV, $m_\phi = 600$ GeV) is excluded.

Chapter 6

Conclusion

In this thesis, I have explored the applicability of complex models in multilepton final states. Initially, we looked at a model-independent multilepton search performed by CMS and investigated how to repurpose the search results to constrain other models using the reinterpretation procedure. I constrained a novel VLL μ-like model using the reinterpretation recipe to give an exclusion for masses below 1200 GeV and compared them to the ATLAS results.

Following that, I demonstrated how a complex model such as the quintuplet model behaves compared to the conventional VLL models. Due to this, there arises a need to execute a targeted analysis to analyse the complex model systematically. I considered two final states containing fatjets along with leptons to aid the analysis using fatjet substructure variables. Finally, I probed the applicability of the quintuplet model to the two final states and calculated the expected upper limits on the cross-sections of the process. Further projections of the limits for more collision data or higher collision energies can act as guidelines for the possible discovery or exclusion at future colliders.

Bibliography

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

[2] CMS Collaboration. Stairway to discovery: A report on the CMS programme of cross section measurements from millibarns to femtobarns. *Physics Reports*, 2024.

[3] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[4] CMS Collaboration. Inclusive nonresonant multilepton probes of new phenomena at $\sqrt{s}$=13 TeV. *Phys. Rev. D*, 105(11):112007, 2022.

[5] Nilanjana Kumar and Stephen P. Martin. Vectorlike leptons at the Large Hadron Collider. *Phys. Rev. D*, 92:115018, Dec 2015.

[6] Prudhvi N. Bhattiprolu and Stephen P. Martin. Prospects for vectorlike leptons at future proton-proton colliders. *Phys. Rev. D*, 100:015033, Jul 2019.

[7] Nilanjana Kumar and Vandana Sahdev. Alternative signatures of the quintuplet fermions at the LHC and future linear colliders. *Phys. Rev. D*, 105:115016, Jun 2022.

[8] 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, May 2014.

[9] J. Silk et al. *Particle Dark Matter: Observations, Models and Searches*. Cambridge Univ. Press, Cambridge, 2010.

[10] Assaf Shomer. A pedagogical explanation for the non-renormalizability of gravity, 2007.

[11] James M. Cline. Baryogenesis. In *Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime*, 9 2006.

[12] Wikimedia Commons. Standard model of elementary particles, 2019.

[13] A. Salam and J. Strathdee. Super-symmetry and non-Abelian gauges. *Physics Letters B*, 51(4):353–355, 1974.

[14] Lyndon Evans and Philip Bryant. LHC Machine. *Journal of Instrumentation*, 3(08):S08001, aug 2008.

[15] David Barney. CMS Slice. 2015.

[16] CMS Collaboration. Inclusive nonresonant multilepton probes of new phenomena at $\sqrt{s}$ = 13 TeV. HEPData (collection), 2024. https://doi.org/10.17182/hepdata.110691.

[17] J. Alwall et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *Journal of High Energy Physics*, 2014(7):79, Jul 2014.

[18] Christian Bierlich et al. A comprehensive guide to the physics and usage of PYTHIA 8.3. *SciPost Phys. Codebases*, page 8, 2022.

[19] Christian Bierlich et al. Codebase release 8.3 for PYTHIA. *SciPost Phys. Codebases*, pages 8–r8.3, 2022.

[20] J. de Favereau et al. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2):57, Feb 2014.

[21] CMS Collaboration. The CMS statistical analysis and combination tool: COMBINE. *Computing and Software for Big Science*, 8(1):19, Nov 2024.

[22] ATLAS Collaboration. Search for vector-like leptons coupling to first- and second-generation Standard Model leptons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, 2024.

[23] Andy Buckley et al. LHAPDF6: parton density access in the LHC precision era. *The European Physical Journal C*, 75(3):132, Mar 2015.

[24] Andrew J. Larkoski, Simone Marzani, Gregory Soyez, and Jesse Thaler. Soft drop. *Journal of High Energy Physics*, 2014(5):146, May 2014.

[25] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. Recombination algorithms and jet substructure: Pruning as a tool for heavy particle searches. *Phys. Rev. D*, 81:094023, May 2010.

[26] Davide Napoletano and Gregory Soyez. Computing N-subjettiness for boosted jets. *Journal of High Energy Physics*, 2018(12):31, Dec 2018.

[27] G. Aad et al. Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV. *Phys. Rev. D*, 101:052007, Mar 2020.