SEARCH FOR NEW PHYSICS AT THE LARGE HADRON COLLIDER IN A TOPOLOGY WITH A SINGLE TAU LEPTON, MISSING TRANSVERSE MOMENTUM, AND A BOOSTED JET.

By

Jethro Taylor Gaglione

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

August 11th, 2023

Nashville, Tennessee

Approved:

Paul Sheldon, Ph.D.

Alfredo Gurrola, Ph.D.

Will Johns, Ph.D.

Thomas Kephart, Ph.D.

Norman Tolk, Ph.D.

ACKNOWLEDGMENTS

I want to dedicate this dissertation to my mom, Wendy, who taught me to be relentless in my pursuits and to approach the world with a positive curiosity. My dad, Renato, for his unconditional support through this process. My mom and dad gave up a lot of their personal aspirations so that my siblings and I could have a quality education, and I'm grateful for that. I want to thank my wife, Jess, for being my rock and partner in this journey through thick and thin. She especially held down our household these last few months of extra hard work while also carrying our child. I also want to dedicate this dissertation to our future daughter. I hope you get to read this one day and that we get to have a good conversation about it.

I want to thank my advisor and his wife, Dr. Paul Sheldon and Sue Sheldon, for treating me like family during my stay at Vanderbilt. I learned a lot of physics from Paul during the first crucial months of learning particle physics, and also picked up some useful knowledge of good wine and whiskey. I equally want to thank my co-advisor, Dr. Alfredo Gurrola, who guided me through most of this work. He is the definition of an expert in his field and an incredible, patient teacher. I want to thank Dr. David Ernst for believing in me as a physicist. He played a crucial role in my acceptance to the Vanderbilt program. Finally, I want to thank my classmates and colleagues in the Vanderbilt High Energy Physics Group for their support and friendship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Introduction

Though particle physics is traditionally thought of as a means for understanding only the smallest constituents of our universe, its implications can also shed light on the mechanisms by which large-scale features in our universe operate, many of which are poorly understood. Among these big-picture observations that physicist currently seek to explain are those concerning dark matter (DM) - its origin, properties, and how its interactions with "visible" matter resulted in the universe we experience today. Current measurements via astronomical observations estimate that 84% of the matter in the universe is DM [26], and current efforts in particle physics (including that which concerns this dissertation) aim to provide some insight into how and why that is.

The Standard Model (SM) of particle physics has been the primary theoretical framework that physicists rely on to make predictions about elementary particle interaction. It has been incredibly accurate since its inception in the 1970s, and continues to be one of the most successful and useful tools in the field. Despite its success, there are phenomena in the universe that we observe to be inconsistent with SM predictions. For one, the SM doesn't provide a viable DM candidate, let alone an explanation of its properties. The SM also lacks a complete explanation of gravity, as well as more nuanced phenomena such as neutrino oscillations, and rare b-meson decays [27, 28]. Aside from its shortcomings in this respect, the SM raises various questions regarding why some of its inner workings and parameters are as they are. The Hierarchy Problem is an example of such a concern.

It is clear that there is a need for a theoretical extension to the SM, and there are many such theories that aim to fit the role. These are collectively known as "beyond standard model" (BSM) theories. One of the most popular and elegant BSM contenders is Supersymmetry (SUSY). There are many proposed versions of SUSY, but they all generally provide an explanation to the Hierarchy Problem and simultaneously provide viable DM particle candidates. Alternatively, other various BSM theories predict the existence of a particle called a "leptoquark" (LQ), which provides a means for two types of SM elementary particles, leptons and quarks, to interact with each other directly. Interactions via LQs could also provide explanations for DM, as well as rare b-beson decays. Though there are other well-motivated BSM theories, the work described in this dissertation is a search for experimental evidence in support of SUSY and/or the existence of a generic LQ that interacts with DM.

The experimental vessel by which we conduct this search is the Large Hadron Collider (LHC) in Geneva Switzerland. At the time of writing, it is the largest particle accelerator in the world, and its collision energies

make it uniquely suited to search for particles that may have been created in similar conditions to those in the early universe. The run period considered in this dissertation (Run II) constitutes an integrated luminosity of $\mathscr{L}_{int} = 137.6\ fb^{-1}$ over three years (2016, 2017, and 2018), where the proton-proton collisions are conducted at a center-of-mass energy $\sqrt{s} = 13\ TeV$.

The Compact Muon Solenoid (CMS) is an "all-purpose" detector located on the north side of the LHC accelerator loop in France. It measures at 25 x 15 x 15 meters in volume and is at one of the four collision points in the LHC. Its main features are a 4 *Tesla* solenoid magnet, pixel and microstrip silicon trackers, a crystal electromagnetic calorimeter, a hadron calorimeter, and muon chambers. During Run II, CMS saw collisions at rate of about 40 MHz that were then sorted through a complex system of triggers, subjected to various levels of processing, and stored in various formats to form the data that we use in this study. Though primarily designed to cleanly detect muons, CMS stores a wealth of collision information that has proved useful in important work across many categories in particle physics, including the discovery of the Higgs boson.

It seems appropriate to begin the description of the work contained in this dissertation by discussing some of its theoretical foundations and motivations, starting with The Standard Model, and followed by the specifics of the BSM theories we aim to investigate. We will then describe relevant details of the CMS detector and our experimental approach in general before presenting the results and conclusions.

Taking the opportunity to provide insight into the physics work necessary before a detector-level search such as this, I will also provide a brief description of phenomenology studies I have been involved in and worked on in parallel to what constitutes my main thesis work.

CHAPTER 2

Theory and Motivation

2.1 The Standard Model

2.1.1 Overview

From a physicist's standpoint, the current understanding of how the universe works at the smallest known scales is rooted in the concept of particles. Sometimes what we refer to as particles are complex entities with smaller constituents, as is the case of the proton, but there is a collection of particles that we consider to be fundamental and indivisible. These are known as the elementary particles, and the Standard Model of particle physics (SM) is the most successful theory we have in describing the properties of the elementary particles and how they interact. In the SM, these particles are split into two categories: $fermions$, which are the primary constituents of matter, and *bosons*, which are responsible for mediating particle interactions. Tables 2.1 and 2.2 include a list of the SM fermions and bosons, respectively, and their general properties.

The interactions amongst particles take the form of what are known as forces, and the fundamental forces of nature are often said to be the electromagnetic, weak, strong, and gravitational forces. The electromagnetic and weak forces are understood to be manifestations of the same "electroweak" force, and only appear as separate in the limit of energies accessible in every day experience. The electromagnetic force is mediated by the *photon* (γ), the weak force by the W and Z bosons, and the strong force by the *gluon* (g). Particles are subject to interact via the electromagnetic and strong forces if they carry the corresponding *charge*. For the strong force this is known as color charge, and for the electromagnetic force the electric charge. The SM does not provide an explanation for gravity. The leading theory on gravitation is Einstein's general relativity, and an ongoing problem in modern physics is the reconciliation of these two school of thought.

The fermions contain two sub-types, quarks and leptons, which are split into three generations, each of which have different particle "flavors" with similar charge and spin but different masses. The quarks are characterized by the fact that they have fractional electric charge and have color charge, thus are subject to the electromagnetic and strong force. The first generation is composed of the *up* (u) and *down* (d) quarks, the second generation of the *charm* (c) and *strange* (s), and the third of the *top* (t) and *bottom* (b). The leptons have have an electric charge of 1 or 0, and have no color charge, thus are not affected by strong force interactions.

In the case of the leptons we have the electron (e), muon (μ), and tau (τ). Each gets respectively heavier, but all have an electric charge of $+1$, and spin of $1/2$. For each of these charged leptons there is also a

corresponding neutrino, which carries no electric charge. Both leptons and quarks also have corresponding anti-particles that carry opposite electric charge but same mass and spin.

Leptons

Generation	Particle	Name	Mass	Charge
1st	e	electron	0.51 MeV	-1
	ν_e	electron neutrino	< 1.1 eV	0
2nd	μ	muon	105.66 MeV	-1
	ν_μ	muon neutrino	< 0.19 MeV	0
3rd	τ	tau	1.78 GeV	-1
	ν_τ	tau neutrino	< 18.4 MeV	0

Quarks

Generation	Particle	Name	Mass	Charge
1st	u	up	2.16 MeV	$\frac{2}{3}$
	d	down	4.67 MeV	$-\frac{1}{3}$
2nd	c	charm	1.27 GeV	$\frac{2}{3}$
	s	strange	93.0 MeV	$-\frac{1}{3}$
3rd	t	top	172.6 GeV	$\frac{2}{3}$
	b	bottom	4.18 GeV	$-\frac{1}{3}$

Table 2.1: Standard Model Fermions

Bosons

Particle	Name	Mass	Charge	Spin
γ	photon	$< 1 \times 10^{-18}$ eV	0	1
g	gluon	< 1.3 MeV	0	1
$W^\pm$	W	80.39 GeV	± 1	1
Z^0	Z	91.19 GeV	0	1
H	higgs	125.18 GeV	0	0

Table 2.2: Standard Model Bosons

2.1.2 Symmetries and the Standard Model [1]

The SM is formulated on a framework known as Quantum Field Theory (QFT), where particles arise as excitations in fields that exist in space-time. An important mathematical quantity at the center of QFT is one that relates the configuration of a set of fields to a system's measurable energy. This is known as the Lagrangian, and it exhibits some important symmetries, i.e. it remains unchanged upon particular changes in the underlying fields. This carries a great deal of significance because, as proven by Emmy Noether in 1918, symmetries of a Lagrangian carry a corresponding physical conservation law (Noether's Theorem). Once a Lagrangian is in place, we can deduce the mathematics that govern the dynamics of a system, which gives us predictive power and the ability to compare with experiment. One can think of the SM as being contained in its Lagrangian along with the fields that compose it.

In understanding the construction of the SM Lagrangian, we use as a guide the quantities we see as conserved in nature. Roughly speaking, those include electric and color charge, as well as the invariance of the laws of physics with translations in space and time (there are other more subtle conserved quantities that will be discussed shortly). These symmetries are often expressed in the language of a branch of mathematics know as group theory. Space-time translations, rotations, and boosts form part of the $SO(3,1)$ (Poincaré) group, and internal symmetries related to charge and isospin conservation are in the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge group. Each of the fields out of which our SM particles arise have a different representation in each of the individual groups mentioned. The $SU(2)_L \times U(1)_Y$ groups are the symmetries related to the combined electromagnetic and weak (electroweak) force, and $SU(3)_C$ symmetries are those responsible for the strong force. Because the leptons are unaffected by the strong force, for example, we expect that the representation of the fields describing leptons in the $SU(3)_C$ group to be unaffected by transformations in $SU(3)$.

The type of fields corresponding to SM particles can be seen as based on their spin. Matter particles (fermions) have spin $s = \frac{1}{2}$ and are represented by a spinor field Ψ and its conjugate $\bar{\Psi}$; gauge bosons have spin $s = 1$ and are represented by a vector field A; the higgs boson has spin $s = 0$ and is represented by a scalar field Φ. The simplest Lorentz-invariant Lagrangian one could write for each of these fields are

$$\mathscr{L}_0 = \frac{1}{2}\partial_\mu \Phi \partial^\mu \Phi + \frac{1}{2}(\frac{mc}{\hbar})^2 \Phi^2, \;\; \mathscr{L}_{\frac{1}{2}} = (\hbar c)\bar{\Psi}\gamma^\mu \partial_\mu \Psi + mc^2 \bar{\Psi}\Psi, \;\; \mathscr{L}_1 = \frac{1}{16\pi} F_{\nu\mu} F^{\nu\mu} + \frac{1}{8\pi}(\frac{mc}{\hbar})^2 A^\mu A_\mu \tag{2.1}$$

where γ^μ are the Dirac matrices, and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. These are known as $free$ Lagrangians because they do not include terms to account for interactions between fields (interactions between particles), and thus don't provide an accurate description of what we observe in nature.

2.1.3 Electromagnetism

One of the more elegant features of the SM is the way that the interaction terms arise. We know that these free Lagrangians are invariant with respect to global transformations, having the general form:

$$\Psi \to \Psi' = e^{iq\phi}\Psi \tag{2.2}$$

where q and ϕ are constants. The interaction terms come about when we require our Lagrangian to not only observe symmetry under global transformations, but local ones as well. Local transformations are those which can take a different value at different points in spacetime. This is reflected in the transformation parameter ϕ,

which we now consider to be dependent on the spacetime coordinates X^μ when applying a transformation:

$$\Psi \to \Psi' = e^{iq\phi(X^\mu)}\Psi \tag{2.3}$$

In order for the SM Lagrangian to remain invariant under local transformations, extra terms must be added to $\mathscr{L}_{\frac{1}{2}}$ in eqn. 2.1. The extra terms that are put in place to ensure gauge invariance naturally give rise to new fields, and interactions between our matter fields Ψ and those new fields. The transformation presented above in eqn. 2.3 above happens to be local in the $U(1)$ gauge group. The new $U(1)$ gauge invariant Lagrangian is

$$\mathscr{L}_{\frac{1}{2}} = (\hbar c)\bar{\Psi}\gamma^\mu D_\mu \Psi + mc^2\bar{\Psi}\Psi + \frac{1}{16\pi}F_{\nu\mu}F^{\nu\mu} \tag{2.4}$$

with the primary differences from $\mathscr{L}_{\frac{1}{2}}$ in eqn. 2.1 being the inclusion of the "covariant derivative" $D = \partial_\mu + iqA_\mu$ and the term $\frac{1}{16\pi}F_{\nu\mu}F^{\nu\mu}$. One can see that the covariant derivative introduces a new vector field A_μ that interacts with Ψ, and from $\mathscr{L}_1$ in eqn. 2.1 one can see that the last term corresponds to the kinetic portion of a *massless* spin-1 field Lagrangian. This can be summarized as follows: starting with the free Lagrangian of a spin $\frac{1}{2}$ (matter) field, the requirement of local gauge invariance under U(1) transformations gives rise to a field A_μ which must be massless. This A_μ gauge field is interpreted to be the field that gives rise to the photon. The coupling g in D represents the strength of the interaction between the photons and fermions. Since experiment shows that the photon is massless, we are not alarmed by the lack of a mass term from the spin-1 field, and this $U(1)$ symmetric Lagrangian seems to give a complete description of electromagnetic interactions.

2.1.4 The Strong Force

The strong force, described by the theory of Quantum Chromodynamics (QCD), acts only on the six quarks (and their antiparticles): u ($\bar{u}$), d ($\bar{d}$), c ($\bar{c}$), s ($\bar{s}$), t ($\bar{t}$), and b ($\bar{b}$). The charge driving the strong force is the *color* charge, and unlike electromagnetism, there are three distinct color charges labeled red R, green G, and blue B. The motivation behind the naming scheme is that all particles directly observed in nature are color neutral, i.e. have either a net color charge of RGB, $\overline{RGB}$, or any combination of color and anti-color. This is a phenomenon of QCD known as $color\ \ confinement$. With the additional types of charges, it is expected that the $SU(3)$ case is more complex than in $U(1)$. In fact, the $SU(3)$ group has eight degrees of freedom compared to the single degree of freedom in $U(1)$. In group theory language, $SU(3)$ has eight *generators* and is a $non-abelian$ group. meaning that the group elements (transformations) don't necessarily commute. Because quarks are also spin $\frac{1}{2}$ particles, we can again start with $\mathscr{L}_{\frac{1}{2}}$ from eqn. 2.1, though this time keeping

in mind that there are separate components in the internal vector space corresponding to $SU(3)$ that represent the different color quarks:

$$\Psi = \begin{pmatrix} \Psi_R \\ \Psi_G \\ \Psi_B \end{pmatrix}$$

This time with the additional degrees of freedom and non-abelian group elements, our global transformations are described by

$$\Psi \to \Psi' = e^{-i\frac{q}{\hbar c}\lambda\cdot\phi}\Psi \tag{2.5}$$

where λ is the vector of $SU(3)$ generators, and ϕ is now a *vector* of parameters. The factor $q/\hbar c$ establishes the coupling strength for all colors. As in the $U(1)$ case, the global symmetry is promoted to a local symmetry,

$$\Psi \to \Psi' = e^{-i\frac{q}{\hbar c}\lambda\cdot\phi(X^\mu)}\Psi \tag{2.6}$$

Also as in the $U(1)$ case, the requirement of local gauge invariance gives rise to vector fields A_μ which must be massless, as well as terms in the Lagrangian of the form $ig\hbar c\bar{\Psi}\gamma^\mu\lambda\cdot A_\mu\Psi$ which describe the interactions between the quark fields Ψ and the new vector fields A_μ. One can see that in the $SU(3)$ case, A_μ must also have eight components; these are the gluon fields. Unlike $U(1)$, requiring gauge invariance in $SU(3)$ also give rise to terms where the components of A_μ interact with each other, indicating that gluons can interact amongst themselves as well as with the quarks. After accounting for kinetic terms of the new gauge fields and letting them propogate, the QCD Lagrangian becomes

$$\begin{aligned}\mathscr{L} = {} & \hbar c\bar{\Psi}\gamma^\mu D_\mu\Psi + mc^2\bar{\Psi}\Psi + \frac{1}{16\pi}(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu)(\partial^\mu A_{\nu a} - \partial^\nu A^{\mu a}) \\ & - \frac{g}{16\pi}f^{ade}A^d_\mu A^e_\nu(\partial^\mu A_{\nu a} - \partial^\nu A^{\mu a}) - \frac{g}{16\pi}f^{abc}A^{\mu b}A^{\nu c}(\partial_\mu A^a_\nu - \partial^\nu A^a_\mu) \\ & + \frac{g^2}{16\pi}f^{abc}f^{ade}A^b_\mu A^c_\nu A^{\mu d}A^{\nu e}\end{aligned} \tag{2.7}$$

where f^{ijk} are structure constants of $SU(3)$. Here the covariant derivative takes the form $D_\mu = \partial_\mu + ig\lambda\cdot A_\mu$.

2.1.5 The Electroweak Interactions

The symmetry group that describes the unified electroweak theory is $SU(2)_L \times U(1)_Y$. The $SU(2)$ group alone cannot by itself fully describe the weak interactions as we observe them in nature. The current SM theory of weak interactions is thus understood to be a unified symmetry of these two symmetry groups that at low energies undergoes what is known as *spontaneous symmetry breaking* and becomes the two forces we experience separately as the weak and electromagnetic force. The mechanism responsible for the

phenomenon of symmetry breaking in $SU(2)_L \times U(1)_Y$ is the *Higgs mechanism*, and will be described in the following subsection.

A Lagrangian for fermions can be written in terms of right and left-handed fields:

$$\mathscr{L} = \hbar c \bar{\Psi}_L \gamma^\mu \partial_\mu \Psi_L + \hbar c \bar{\Psi}_R \gamma^\mu \partial_\mu \Psi_R + mc^2 (\bar{\Psi}_L \Psi_R + \bar{\Psi}_R \Psi_L) \tag{2.8}$$

This is useful because, from experiment, we know that left and right-handed particles are treated differently in weak interactions. On the other hand, because of the different transformations of left and right chiral fields in the SM, the last term which gives these fields mass is not mathematically consistent and this Lagrangian can only represent a theory of *massless* fermions.

As suggested by the L subscript, $SU(2)_L$ acts only on left-handed matter particles, and $U(1)_Y$ acts on particles with *hypercharge* Y. Furthermore $SU(2)$ acts on two-component objects known as *doublets*, and $U(1)_Y$ on singlets. The representations of the SM fermions in the chiral basis of $SU(2)_L \times U(1)_Y$ groups are displayed below.

$$\chi_L = \begin{pmatrix} u \\ d \end{pmatrix}_L \quad \begin{pmatrix} c \\ s \end{pmatrix}_L \quad \begin{pmatrix} t \\ b \end{pmatrix}_L \quad \begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \quad \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}_L \quad \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}_L \tag{2.9}$$

$$\chi_R = u_R \quad d_R \quad c_R \quad s_R \quad t_R \quad b_R \quad e_R \quad \mu_R \quad \tau_R \tag{2.10}$$

the group transformations in terms of these representations for $SU(2)_L$ are

$$\chi_L \to \chi_L' = e^{-ig\vec{\sigma}\cdot\vec{\theta}} \chi_L \tag{2.11}$$

and for $U(1)_Y$

$$\chi_L \to \chi_L' = e^{-ig' Y_{\chi_L} \phi} \chi_L \qquad \chi_R \to \chi_R' = e^{-ig' Y_{\chi_R} \phi} \chi_L \tag{2.12}$$

where the generators of the $SU(2)$ group, $\vec{\sigma}$, are the Pauli Matrices; Y_χ are in place to dictate how much of g' corresponds to each χ; and, as in the previously discussed cases, θ and ϕ are parameters that we will promote to become spacetime-dependent and require our Lagrangian to satisfy local gauge invariance. This process again introduces appropriate covariant derivatives

$$\begin{aligned} \partial_\mu \chi_L \to D_\mu \chi_L &= \partial_\mu \chi_L + ig\vec{\sigma}\cdot\vec{W}_\mu \chi_L + ig' Y_{\chi_L} B_\mu \chi_L \\ \partial_\mu \chi_R \to D_\mu \chi_R &= \partial_\mu \chi_R + ig' Y_{\chi_R} B_\mu \chi_R \end{aligned} \tag{2.13}$$

where $\vec{W}_\mu$ are the three gauge fields corresponding to the three $SU(2)_L$ degrees of freedom, and B_μ is the field

corresponding to $U(1)_Y$. We then add kinetic terms for each of these fields in a similar fashion to the $SU(3)$ case with terms of the form

$$F_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \qquad F^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g\varepsilon^{abc} W^b_\mu W^c_\nu \tag{2.14}$$

As with the electromagnetism and strong force cases, the second part of the kinetic term for these gauge fields ($\mathscr{L}_1$ in eqn. 2.1) are not gauge invariant. Thus, the electroweak theory as an $SU(2)_L \times U(1)_Y$ group symmetry can only exist as a theory with massless gauge bosons as well as massless fermions (per the discussion earlier in this subsection). We know from experiment that this is not what is observed in nature. Through the higgs mechanism, $SU(2)_L \times U(1)_Y$ undergoes spontaneous symmetry breaking through the Higgs mechanism, by which the weak force gauge bosons and fermions gain mass. In the process, the B_μ field mixes with the certain component of the W_μ field to create the familiar photon (γ) and neutral Z boson:

$$\begin{aligned} A_\mu &= B_\mu cos\theta_W + W^Z_\mu sin\theta_W \\ Z^0_\mu &= -B_\mu sin\theta_W + W^Z_\mu cos\theta_W \end{aligned} \tag{2.15}$$

where θ_W is the Weinberg angle that we get from experimental measurement, which also relates the couplings related to the weak force gauge bosons to each other, and to the coupling g of the unbroken symmetry:

$$g_\gamma = g\, sin\theta_W \qquad g_{W^\pm} = g \qquad g_{Z^0} = \frac{g}{cos\theta_W} \tag{2.16}$$

2.1.6 The Higgs Mechanism

The subsections above describe the mechanism by which interactions between particles arise in the SM given free-field Lagrangians, namely, by requiring that they are invariant in local gauge transformations. Though successful in providing interaction terms, this technique falls short in providing a mechanism by which mass can be ascribed to the SM fermions, and the bosons associated with these new gauge fields. We know from experiment that the charged fermions, and the Z^0 and $W^\pm$ bosons in fact have mass (though recent experiments observe the neutrinos to have mass [28], the SM was formulated under the assumption that they are massless). The Higgs mechanism introduces a new scalar field with particular properties that interacts with the fermion and boson fields in a way that mass terms arise. The (gauge invariant) Higgs field is a complex scalar field whose Lagrangian is

$$\mathscr{L} = (D^\mu \Phi)^\dagger (D_\mu \phi) - V(\Phi) \tag{2.17}$$

where

$$V(\Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.18}$$

and the covariant derivative

$$D_\mu \Phi = \partial_\mu \Phi + ig\vec{\sigma} \cdot \vec{W}_\mu \Phi + \frac{i}{2} g' B_\mu \Phi \tag{2.19}$$

contains the $SU(2)_L \times U(1)_Y$ gauge boson fields $\vec{W}_\mu$ and B_μ.

The SM makes predictions via calculations that are perturbative in nature [29], thus we expand about the ground states of fields, and the fields we have discussed previous to this subsection happen to have vacuum expectation values (VEV) of zero. The key to the Higgs Mechanism and spontaneous symmetry breaking (in this context) is that this new field we introduce has a non-zero VEV, and upon choosing an arbitrary minimum to expand on we say that the symmetry is spontaneously broken. Our potential, eqn. 2.18, depends on two parameters: λ which is required to be positive for vacuum stability, and μ. The shape of the Higgs potential is thought to have evolved from the universe to its current configuration, where $\mu^2 > 0$. The resulting potential is displayed in fig. 2.1. It is evident that $\Phi = 0$ (point A) in this system is an unstable minimum, thus the VEV

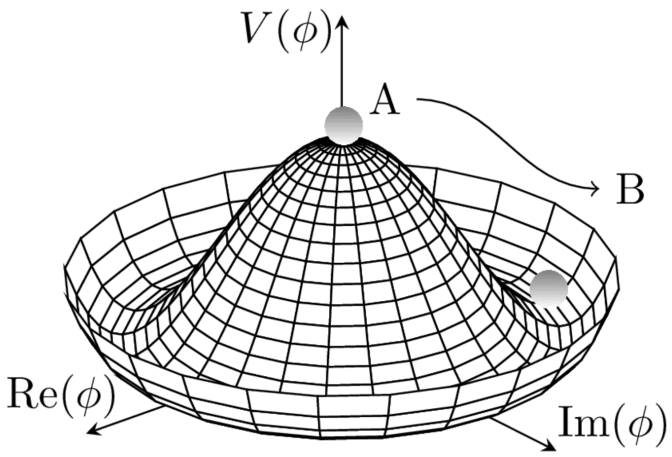

Figure 2.1: The Higgs Potential

is expected to lie at values of the field satisfying $\Phi^\dagger \Phi = \frac{v^2}{2}$ (point B), where $v = \sqrt{\frac{\mu}{2\lambda^2}}$. These values of Φ define a circle of radius v, as seen in fig. 2.1. Though there are an infinite number of minima in the subspace defined by this circle, we must choose a particular value to expand our Lagrangian about. After strategically picking the direction of the field, the ground state in the $SU(2)$ doublet representation becomes

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.20}$$

and because our field is a fluctuation about this ground state, it becomes

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix} \tag{2.21}$$

After rewriting the Lagrangian for the Higgs field in this new form, the following term arises:

$$(D^{\mu}\Phi)^{\dagger}(D_{\mu}\Phi) = \frac{v^2}{8}[g^2((W_{\mu}^1)^2 + (W_{\mu}^2)^2) + (gW_{\mu}^3 - g'B_{\mu})^2] \tag{2.22}$$

Because the vector bosons in the charge basis are represented as

$$W_{\mu}^{\pm} = \frac{1}{\sqrt{2}}(W_{\mu}^1 + W_{\mu}^2), \quad A_{\mu} = B_{\mu}cos\theta_W + W_{\mu}^Z sin\theta_W, \quad Z_{\mu}^0 = -B_{\mu}sin\theta_W + W_{\mu}^Z cos\theta_W \tag{2.23}$$

one can see that eqn. 2.22 represents a mass term for the gauge bosons with a value dependent on the higgs VEV v. It is also worth noting that the SM lagrangian also acquires terms there the Higgs field interacts with the gauge bosons.

The SM fermions are also massive, of course. The way we ascribe them a mass is via the addition of a *Yukawa* Lagrangian, roughly of the form

$$\mathscr{L} = \Gamma\chi_L\Phi\chi_R \tag{2.24}$$

where χ are the chiral fermion fields from eqns. 2.9 and 2.10. Similarly to the case with the gauge fields, given the component $v+h$ of the Higgs $SU(2)$ doublet, one can see that this gives rise to mass terms for the fermions as well as interaction terms between the fermion fields and the higgs field. The "strength" of interactions between the higgs field and the fermions is dictated by the Yukawa couplings contained in matrix Γ. The spin-0 particle associated with the Higgs field as described above is the Higgs boson.

2.2 Brief Discussion on Regularization and Renormalization

Though the picture laid out thus far gives a clean description of the SM in terms of its symmetries, some of the nuanced details involving formal calculations using the SM (these are beyond the scope of this dissertation) give rise to obstacles and nuisances. One of the main issues that arises in perturbative expansions of SM calculations involve integrals over momentum space of the form

$$\int d^4q \frac{1}{(\vec{q}^2 + m^2)^2} \tag{2.25}$$

which are divergent. A way around this is to instead let the upper limit of integration be a finite value Λ. The integral is then some function of Λ which goes to infinity as $\Lambda \to \infty$. Divergent integrals may also occur when integrating as momentum goes to zero. Similarly for this case, we let Λ be some finite, lower limit of integration. Cases where we have divergences at infinity are known as *ultraviolet* divergent, and the latter case involving integrals over zero momentum is known as an *infrared* divergence. This process is known as *regularization*. The new terms now have a dependence on Λ, which can be removed from divergent terms by essentially making a change of variables that cancel Λ but introduce "counterterms." This process is called renormalization, and it can be illustrated by a toy model with Lagrangian

$$\mathscr{L} = \frac{1}{2}[(\partial_\mu \Phi)^2 - m^2\Phi^2] - \frac{1}{4!}\lambda\Phi^4 \tag{2.26}$$

which after renormalization, i.e. reparametrization of the field and couplings we expand about, becomes

$$\mathscr{L} = \frac{1}{2}[(1+\delta Z)(\partial_\mu \Phi_r)^2 - (m_r + \delta m)^2\Phi_r^2] - \frac{1}{4!}(\lambda_r + \delta\lambda)\Phi_r^4 \tag{2.27}$$

where $\Phi_r = (1+\delta Z)^{-1/2}\Phi$, and δZ, δm, and $\delta\lambda$ are parameters of the counterterms which are strategically picked to remove the divergent integrals' dependence on Λ, and consequently give our perturbative methods predictive power. It will be important in what follows to notice that this is in essence the same Lagrangian, but with slightly modified mass and coupling terms. We can now interpret the mass one would see in experiment, m, as being composed of a bare mass m_r plus a "correction" term δm that depends on the how the field interacts. The correction term can be said to depend on the scale Λ as the effects that contribute to the experimentally observable mass can change with the energy (distance) that we experimentally probe these mechanisms in.

2.3 Shortcomings of the Standard Model

The SM in its described form is still widely used in high-energy physics, and its predictions have seen remarkable agreement in experiment over the decades. Nonetheless, as physics and astronomy move forward, and experiments become more precise and wide in scope, we are seeing mounting evidence that the SM is not the complete picture in describing the fundamental constituents of the universe. Issues with the SM can generally be seen as falling into two categories: (1.) an inability to explain something about the nature of the universe, and (2.) a disagreement between a quantity that is predictable within the SM framework and the experimentally observed quantity. From the first category, some of the issues are as follows.

- **Gravity** - Gravity is beyond the scope of the SM and is explained by Einstein's General Relativity.

There is no current proven theory in physics unifying these two schools of thought. [30]

- **Neutrino Masses** - The SM considered the neutrinos to be massless particles, whereas recently obsereved phenomena indicates that neutrinos undergo flavor oscillation, which suggests neutrinos are massive particles [28].

- **Dark Matter** - Astronomical measurements suggest that ordinary matter, as described by the SM, makes up only $\sim 5\%$ of the universe. An evasive form of matter known as dark matter (DM) that interacts primarily gravitationally makes up a considerably larger $\sim 25\%$ of the universe. The SM has no DM particle candidate, and no mechanism for describing this phenomenon [26].

- **Dark Energy** - The remaining $\sim 70\%$ of the content of the universe is observed to be related to the energy density of the vacuum, and is described as dark energy. The SM also lacks any description of dark energy [26].

- **Hierarchy and Naturalness Problems** - There are aspects of the SM (some related to the renormalization methods discussed previously) that seem to rely on a fine-tuned balance of parameters with no substantial physical explanation to yield meaningful results [31]. One manifestation of this is the large difference in scales between the fundamental forces. This will be discussed in some detail in the following chapter.

In contrast, some of the issues related to SM predictions that differ from current experimental results are below.

- **Rare B-Meson Decays** - Various experiments (Bebar, Belle, LHCb) have observed a collective 3σ deviation from the SM prediction in the rate that these decays should involve the three families of SM leptons. The τ lepton seems to be involved in $10-20\%$ more decays that predicted by the SM.

- **Muon Magnetic Dipole Moment** - Experiments at Brookehaven National labs and Fermilab study the behaviour of muons in a magnetic field. A value that can be calculated from these experiments that is related to the muon's fundamental interactions, the muons anomalous magnetic moment, is yielding different results to that predicted by the SM [32]

As will be described in the following chapter, the new physics we experimentally search for in what regards this dissertation provides promising and elegant physical explanations to address the hierarchy problem, while simultaneously providing a DM candidate and possible explanations for the discrepancies between SM predictions and measurments in both the rare B-meson decays and muon dipole moment. As such, an

extended but brief expansion of some of the details of these particular issues, in the context of this work, will be provided.

2.4 The Hierarchy Problem and the Higgs Mass

From the discussion in section 2.2, it is evident that the experimentally measured mass of a particle in a renormalized theory is somewhat of a composite quantity. This holds true for the SM particles, including the Higgs boson:

$$m_H^2 = m_{0,H}^2 + \delta m_H^2 \tag{2.28}$$

The contributions to the δM_H term in the above equation are referred to as the quantum corrections to the Higgs mass. Every particle that couples to the Higgs field contributes to this term. In other words, the field of every particle that has mass has a contribution to the quantum corrections of the Higgs mass. A point of contention in particle physics arising from this aspect of the SM is the question of how it is that these corrections balance so finely to yield the relatively small Higgs mass that we observe. Assuming there are undiscovered heavier particles as energy scales become larger, we should see larger and larger quantum corrections in δM_H which somehow would delicately cancel to yield the observed mass. One is then forced to either assume 1. that there is no new physics between the currently accessible energies ($\Lambda \sim 10^3\ GeV$) and the Planck scale ($\Lambda \sim 10^{18}\ GeV$), where the SM is predicted to break down due to gravitational effects it cannot explain, or 2. that the contributions from each of these terms somehow balances in a fine-tuned manner to yield the observed Higgs mass. This is one manifestation of what is known as the *Heirarchy Problem* in particle physics.

2.5 Dark Matter and Dark Matter Relic Density

2.5.1 Dark Matter Discovery and Identity

Evidence of DM has come primarily by astronomical observations that seem at odds with what would be expected in the presence of only baryonic (visible) matter. The earliest of these came from the observation of the unexpected velocities of galaxy in the Coma cluster [33], then from further observations suggesting the need for additional matter to explain the stability of disk galaxies as well as the motion of gases in their vicinity [34, 35]. Later work involving gravitational lensing further confirmed the existence of a large amount of non-visible matter around galaxies [36]. Finally, analysis of the cosmic microwave background (CMB) by the WMAP team provides precise measurements of the density of the universe associated with each type of matter. The total density parameter of the universe Ω is defined as the ratio of the observed density to the critical density, which is an important parameter in determining the geometry of the universe. So $\Omega = \rho/\rho_c$, and it has been experimentally determined that $\Omega = 1$ with a small margin of error [37]. From WMAP

we know that the contributing factors to the total density are matter (baryonic and dark) and dark energy, contributing $\sim 25\%$ and $\sim 75\%$, respectively; the density parameter for total matter is $\Omega_m \approx 0.25$, where the density related to baryonic matter is $\Omega_b \approx 0.05$ [37]. The difference in these suggests that, by large, DM is the dominating type of matter in the universe. The "dark" in DM of course refers to the fact that this type of matter is not seen to interact electromagnetically, only gravitationally. As will be described in more detail, current searches, such as ours, place hope in that DM interactis with traditional SM matter via the weak force.

2.5.2 Relic Abundance of Dark Matter

The most widely accepted theory in the scientific community explaining the origin and development of the universe is the standard model of cosmology, also known as the "Big Bang" theory, or more technically, the ΛCDM model. The in-depth details to the Big Bang theory are beyond the scope of this dissertation, but the main points as related to DM and its presence in the universe will be summarized here in a similar fashion to that in [38]. The main features of Λ-CDM are that the universe started in a hot, dense state and has been expanding since. The initial stages of the universe are characterized by high temperatures and matter in close enough proximity that particle interactions are frequent and highly energetic. The rate at which a particular species of particles is produced in an environment is expressed in the reaction rate,

$$\Gamma = n < \sigma v > \tag{2.29}$$

where n is the number density of the particle type, σ is the cross-section, and v is the relative velocity of the particles. As one would expect, in the context of the early universe, n will be primarily affected by the size and expansion rate of the universe. The same holds true for v as a smaller, denser universe carries a higher temperature T, hence more kinetic energy distributed amongst its constituents. The expansion rate and size of the universe are explained by the Friedmann equations, the most relevant of which, for our purposes, relates the expansion rate of the universe H (known as the Hubble constant), to its energy density ρ:

$$H^2 + \frac{k^2}{a^2} = \frac{8\pi G}{3}\rho \tag{2.30}$$

Here, k is a parameter that takes on the values ± 1 or 0 corresponding to the curvature of the universe, G is the familiar Newton's constant of gravitation, and a is the $scale\ factor$ that gives the relative size of the universe, $a = 0$ being the Big Bang and $a = 1$ the scale at present time. H is in fact defined as a function of a and its time derivative $H = \dot{a}/a$. As the universe is taken to expand adiabatically since its origin, a can be shown to

be proportional to time,

$$a(t) \propto (\frac{t}{t_0})^{\frac{2}{3(1+w)}} \tag{2.31}$$

where t_0 is taken to be the present time. The quantity w related to matter pressure and energy density in the universe which takes on different values depending on whether the content of the universe is dominated by radiation vs non-relativistic matter.

The era during which DM was produced in the early universe was dominated by radiation, thus $a \propto t^{1/2}$. Because production rates of heavier non-relativistic particles during this time are found to be significant relative to the expansion rate of the universe, we consider this period of the Big Bang to be at thermal equilibrium and describable by thermal state variables. The most enlightening of these for the purpose of describing particle production is temperature T, as it quantifies the available energy in the universe to potentially create matter (it's proportional to the v in eqn. 2.29). It also carries inverse proportionality to time, as the universe cools upon expansion. For non-relativistic particles, the number density can be expresses as

$$n_{NR} = g_i(\frac{mT}{2\pi})^{\frac{3}{2}} e^{-m/T}, \tag{2.32}$$

where m is the invariant mass of the particle species, and g_i is related to the degrees of freedom a particle carries. One can see from this expressions that the number density carries an exponentially decreasing factor as the universe expands and cools. From this we understand that eventually our decreased particle number density causes interactions involving these non-relativistic particles to decrease significantly. An equally important effect brought about by a lowered temperature is that, to create a particle of particular rest mass from smaller, lighter particles, the temperature of the universe must be such that $k_b T > mc^2$. Once T dips below this threshold, which is understood to happen when $H \sim \Gamma$, the heavy particle at hand will cease to interact and/or be created by lighter particles and its density in the universe will remain essentially the same thereafter in what is known as a "$freeze\text{-}out$".

The mechanisms described above are relevant to the study of DM as follows. It is believed that DM was created in the early universe and, during the phase of thermal equilibrium, interactions of the type $DM + DM \leftrightarrow SM + SM$ happened at equal rates in both directions as the lighter SM particles carried enough energy to create the more massive DM particles. As the effects described in the previous paragraph took hold, the rate of interactions involving production of the heavy DM particles decreased, and eventually ceased altogether. This left behind a constant density of DM in the universe, known as the dark matter relic abundance, or *relic density*. The current measured value is $\Omega_{DM}h^2 = 0.120 \pm 0.001$. As previously mentioned, the SM does not provide a candidate for DM. In fact, the issue with DM relic density is that the leading BSM theories with

DM candidates do not in their basic forms (with annihilation interactions as mentioned above) yield a correct predicted value in agreement with what is measured. New mechanisms must be introduced to account for this deviation, as will be described in the subsequent chapter.

2.6 B-meson Anomalies

The SM leptons appear in three generations (or families), with each generation having identical quantum numbers and only differing in mass. As such, the couplings corresponding to each SM symmetry are regarded as *universal* across generations, behaving equivalently after accounting for kinematic effects from the differences in mass. There is evidence coming from observations of rare B-meson decays via the BaBar, Belle, and LHCb experiments that nature may not respect the lepton universality put forward by the SM [39, 40, 41]. Mesons are composite particles made from a pair of quarks, and an example of a semi-leptonic B-meson decay is illustrated in Figure 2.2. The possible violation of universality is seen in these semi-leptonic

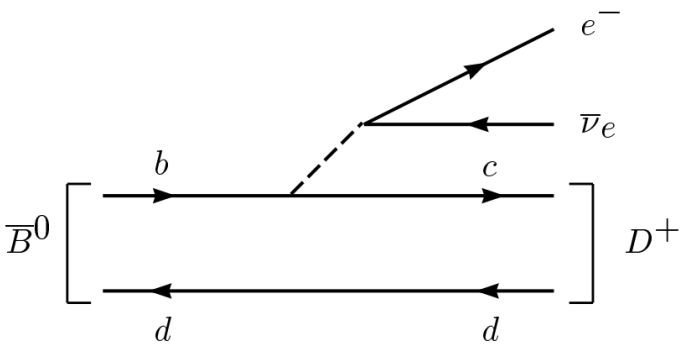

Figure 2.2: Neutral B-meson semi-leptonic decay [5]

decays by comparing the fraction of the number of times the decay yields a particular lepton family and comparing these to those predicted by the standard model. These decay probabilities are known as *branching fractions* or *branching ratios* (*BR*). The current deviation from SM prediction is seen in the ratio of BR of 3rd generation charged leptons τ as compared to the lighter families. This is quantified in terms of the values R_D and R_{D^*} as

$$R_{D^{(*)}} = \frac{BR(\bar{B} \to D^{(*)}\tau^-\bar{\nu}_\tau)}{BR(\bar{B} \to D^{(*)}\ell^-\bar{\nu}_\ell)} \tag{2.33}$$

where D is the D-meson, and $\ell = \mu, \tau$. The experimental results and the SM predictions are displayed in Figure 2.3, and the deviation of the averaged experimental results from the SM prediction currently stand at $\sim 3.2\sigma$.

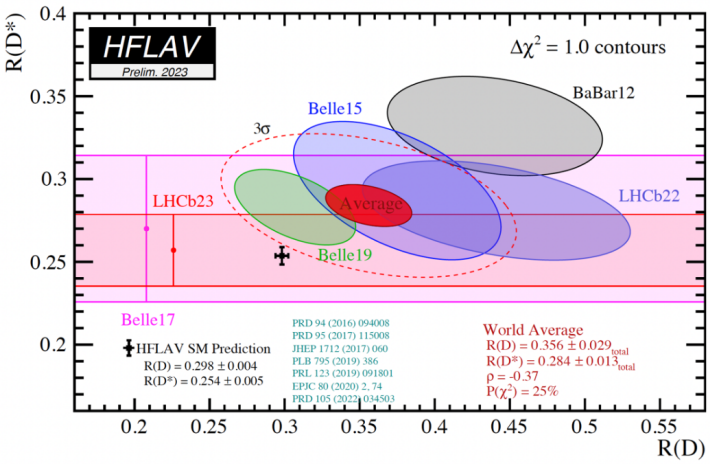

Figure 2.3: Latest Measurements of $R_{D^{(*)}}$ and SM prediction.[6]

CHAPTER 3

Beyond the Standard Model

To attempt to reconcile the shortcomings of the SM and provide an explanation for what we see in nature, many theories have been proposed. These are collectively known as Beyond Standard Model (BSM) theories, as many times they take the SM as a foundation to begin with and expand upon. One of the most promising BSM theories, which simultaneously resolves multiple of the above issues, is supersymmetry (SUSY). The main part of work described in this dissertation is a search for experimental evidence of the existence of SUSY. The primary appeal of SUSY is that it provides a physical explanation to the issues raised in the hierarchy problem, and that it provides a particle candidate for DM. Additionally, SUSY models could provide explanations to the source of disagreement in SM predictions and experiment in both the B-meson and the muon dipole moment cases. A secondary focus of this work is an interpretation of our experimental search in the context of BSM physics postulating the existence of a new particle named the leptoquark (LQ).

3.1 Supersymmetry

3.1.1 SUSY and the Hierarchy Problem

For the sake of simplicity, we can consider the quantum corrections to the Higgs boson mass corresponding to specific loop diagrams of a general fermion (left) or vector boson (right) which couple to the Higgs field with strengths λ_f and λ_s respectively

$$\delta m_H^2 = -\frac{|\lambda_f|^2}{8\pi^2}\Lambda_{UV}^2 + ... \qquad\qquad \delta m_H^2 = \frac{|\lambda_s|^2}{16\pi^2}\Lambda_{UV}^2 - ... \tag{3.1}$$

where Λ_{UV} is the momentum cutoff upper limit used in dealing with the divergent integrals related to loop-diagrams, as described in the regularization and renormalization procedure outlined in the previous subsection. Since the mass of a particle is dictated by how strongly it couples to the Higgs field, one can see how the λ^2 term for potentially new, heavy particles could have a large influence on the quantum corrections of the Higgs mass. Observing that the correction terms from fermions and bosons carry opposite signs, it becomes apparent that if boson-fermion pairs with similar couplings to the Higgs field appeared, we would see this cancellation effect we are looking for. Supersymmetry (SUSY) introduces a new symmetry to the SM that relates SM fermions to bosons and vise-versa, thus giving a more elegant and justified solution to the Hierarchy Problem without the need of ad hoc fine-tuning of parameters.

Stated in more technical terms, SUSY introduces a bosonic field for each SM fermionic field, and con-

versely a fermionic field for SM bosonic fields. This resulting structures are SUSY representations known as *supermultiplets* containing the SM particles along with their SUSY *superpartners* which carry the same quantum numbers, with the exception of a spin offset by one half. The correspondence becomes a bit muddled as there is mixing amongst fields due to symmetry breaking which gives rise to mass eigenstates that are a superposition of fields, as will be described. An additional important feature of SUSY is that is extends the SM Higgs sector to include two Higgs doublets as opposed to one, where each of the doublets contain a charged and neutral component: (H_u^+, H_u^0) and (H_d^0, H_d^-). This extension is necessary for SUSY to remain internally consistent and to yield the correct cancellation of terms we seek [2]. Though there is a spectrum of more intricate versions of SUSY with additional features, the version described above includes only the necessary SUSY extensions to the SM to make a working unit, and it is appropriately known as the Minimal Supersymmetric Standard Model (MSSM). Table 3.1 lists the field content of MSSM,

FIELDS OF MSSM

Supermultiplets	Bosonic Fields	Fermionic Partners
gluon/gluino	g	$\tilde{g}$
gauge/gaugino	$W^{\pm}, W^0$	$\tilde{W}^{\pm}, \tilde{W}^0$
	B	$\tilde{B}$
slepton/lepton	$(\tilde{\nu}_L, \tilde{\ell}_L)$	(ν_L, ℓ_L)
	$\tilde{\ell}_R$	ℓ_R
squark/quark	$(\tilde{q}_{u,L}, \tilde{q}_{d,L})$	$(q_{u,L}, q_{d,L})$
	$\tilde{q}_u, R$	$q_{u,R}$
	$\tilde{q}_d, R$	$q_{d,R}$
Higgs/Higgsino	(H_d^0, H_d^-)	$(\tilde{H}_d^0, \tilde{H}_d^-)$
	(H_u^+, H_u^0)	$(\tilde{H}_u^+, \tilde{H}_u^0)$

Table 3.1: Field content of MSSM. [2]

where $\ell = e, \mu, \tau$ are the charged lepton fields, $\nu = \nu_e, \nu_\mu, \nu_\tau$ the neutrino fields, $q_u = u, c, t$ the fields associated with up-type quarks, and $q_d = d, s, b$ the fields associated with bottom-type quarks. The convention to denote a SUSY superpartner fields is to use a tilde over the corresponding SM field. The L and R subscripts in the SM case represent left and right chiral fields as usual, though in the SUSY case it is mainly a label to indicate the corresponding SM counterpart, as SUSY sleptons and quarks (superpartners to SM leptons and quarks) have spin $s = 0$ and thus are not chiral fields. The "s" added to the front of the name is meant to indicate that these fields are now scalar. The convention for superpartners to SM bosons is then to add an "ino" at the end of the name, and these are now spin $s = 1/2$ fields. The left and right column entries in Table 3.1 combine to form the mentioned *supermultiplets* that transform into each other under SUSY transformations.

From electroweak symmetry breaking (EWSB), we know that the SM B^0 and W^0 components of the corresponding gauge fields mix to form the photon (γ) and neutral Z boson. If SUSY were an unbroken symmetry, the equivalent case for the SUSY partner fields would hold, forming what are know as the *zino*

($\tilde{Z}^0$) and *photino* ($\tilde{\gamma}$). If SUSY were an unbroken symmetry, we would also expect the masses of the SUSY superpartners to have the same mass as their SM counterparts, and we know from experiment that it is not the case. SUSY is therefor understood to be a *broken* symmetry. The mechanism by which we arrive at MSSM as a broken symmetry is beyond the scope of this document, though one of the implications is that there is also mixing between Higgsino and gaugino fields. The neutral components mix to form the *neutralinos* $\tilde{\chi}_1^0$, $\tilde{\chi}_2^0$, $\tilde{\chi}_3^0$, and $\tilde{\chi}_4^0$, and the charged components mix to form the *charginos* $\tilde{\chi}_1^\pm$, $\tilde{\chi}_2^\pm$, $\tilde{\chi}_3^\pm$, and $\tilde{\chi}_4^\pm$ where a larger subscript corresponds to a larger mass. The "amount" of each fundamental field in each of the neurtalino and chargino mass eigenstates is a parameter that requires experimental confirmation, similar to the Weinberg angle of EWSB. A mass eigenstate with a mix dominated by $\tilde{H}$, $\tilde{W}$, or $\tilde{B}$ content is referred to as higgsino-, wino-, or bino-like, respectively.

3.1.2 SUSY and Dark Matter

It is tempting to look to the SM neutrino as a potential candidate for DM, as they are neutral, and only interact via the weak force. However, the requirement for neutrinos to be in thermal equilibrium with the rest of early universe constituents coupled with the fact that they have very small masses means they were (and mostly still are) moving at relativistic speeds [42]. They contribute to what is called *hot* DM. For the development of observed astronomical structures, as understood in Big Bang cosmology, we know the majority of DM should be non-relativistic, *cold* DM, thus disqualifying neutrinos as the candidate. The theorized SUSY neutralino χ_1^0, on the other hand, fits the description of cold DM. It is electrically neutral, does not have color charge, and interacts only weakly. The last piece of the puzzle comes from a required symmetry in MSSM that conserves a quantity known as $R-parity$:

$$P_R = (-1)^{3(B-L)+2s} \tag{3.2}$$

where B and L are baryon and lepton number, respectively, and s is spin. B and L are conserved quantum numbers in the SM defined as $B = \frac{1}{3}(n_{quarks} - n_{antiquarks})$ and $L = n_{lepton} - n_{antilepton})$. MSSM does not explicitly conserve B and L, though it is a consequence of the introduction of R-parity. Another important effect of R-parity conservation is that it requires the number of SUSY particles in any interaction to be conserved as SM particles carry $P_R = +1$ and SUSY particles $P_R = -1$. It is a consequence of this symmetry that the lightest neutralino χ_1^0 (also known as the lightest supersymmetric particle LSP) becomes a stable particle that cannot decay to lighter SM particles, thus solidifying it as a DM particle candidate. It should be noted that it is not known for a fact that DM interacts via the weak force as we know it, only hypothesized.

3.1.3 SUSY and DM Relic Density

Per the discussion in Section 2.5.2 we know that the composition of matter in the current universe is a result of complex dynamics involving competing effects in the early universe; particles interacted in a hot, dense thermal environment but faced a rapidly expanding universe that diluted energy densities and slowed interaction rates. The Boltzmann equation quantifies the rate of change of particle number density n for a particular species due to the various effects. If we consider an R-parity conserving model where the heavier particles decay at a rate much faster that the age of the universe, i.e. our MSSM, the Bolztmann equation reads

$$\frac{dn}{dt} = -3Hn - <\sigma_{eff} v> (n^2 - n_{eq}^2) \tag{3.3}$$

where v is again the relative velocity, n_{eq} is the number density at equilibrium, the angle brackets denote a thermal average, and σ_{eff} is the effective cross section of processes that annihilate dark matter, such as $DM + DM \rightarrow SM + SM$. As one would expect, the DM relic density is dependent on the cross-section of interactions that annihilated DM in the early universe before freeze-out conditions were met. A detailed derivation of this relation is described in [7] and the general result relevant to this analysis can be simply stated as

$$\Omega_{DM} h^2 \propto \frac{1}{<\sigma_{eff} v>} \tag{3.4}$$

The effect on varying $<\sigma_{eff} v>$ values on the DM relic abundance can be seen in Figure 3.1 below, where a larger value results in a lower $\Omega_{DM} h^2$. Under current SUSY models, a free parameter of the theory is the composition of the LSP, χ_1^0, with respect to the underlying fundamental fields (as described in Sec. 3.1.1). When taking χ_1^0 to be bino-like, the theoretically predicted values of $\Omega_{DM} h^2$ reflect an overabundance of DM compared to astronomical observations. In contrast, taking χ_1^0 to be wino-like yields an underabundance. The processes typically dominating σ_{eff} that we have considered ($DM + DM \rightarrow SM + SM$) are those primarily taken into account in various models as the reduction mechanism of DM, though there are other processes that become important when certain conditions are met. Such a process is that of the form $DM + X \rightarrow SM + SM$, known as *coannihilation*, where our DM candidate now annihilates with an unknown partner X that is not necessarily another DM or even SUSY particle. As described in detail in [43, 44], the contributions to σ_{eff} from these different processes can be calculated via the Boltzmann equations, yielding

$$\sigma_{eff} = \frac{g_1^2}{g_{eff}^2} (\sigma_{11} + 2\sigma_{12} \frac{g_2}{g_1} (1+\Delta)^{\frac{3}{2}} e^{-x\Delta}) \tag{3.5}$$

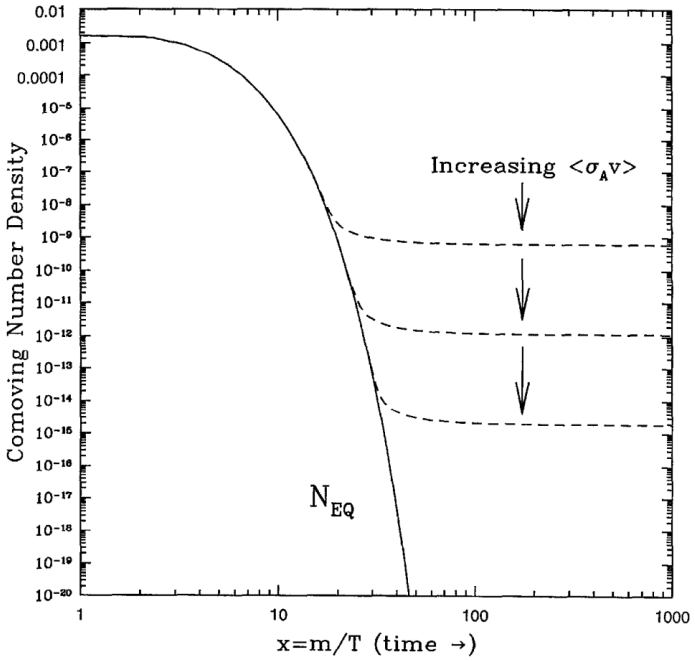

Figure 3.1: Comoving Number Density with Varying $< \sigma v >$ [7]

where g_n are particle degrees of freedom, g_{eff} is the effective degree of freedom driven by

$$g_{eff} = g_1 + g_2(1+\Delta)^{3/2} e^{-x\Delta}. \tag{3.6}$$

x is the parameter $x = \frac{m_{DM}}{T}$ (T being the temperature of the thermal bath), and Δ are the mass splittings

$$\Delta = \frac{m_X - m_{DM}}{m_{DM}} \tag{3.7}$$

where m_X and m_{DM} are the masses of the coannihilation partner X and DM particle, respectively. The two terms in Eqn. 3.5 contain the cross-sections σ_{11} and σ_{12}, where the first corresponds to processes initiated by two DM particles ($DM + DM \rightarrow SM + SM$) and the second to the coannihilation scenarios we are interested in ($DM + X \rightarrow SM + SM$). One can see from both Eqns. 3.5 and 3.6 that the coannihilation term is exponentially suppressed, but becomes a significant contribution when we have small Δ. This suggests that inclusion of coannihilation mechanisms (via small Δ) to the reduction of DM in the universe could explain the discrepancy in current models with respect to observation. This fact is a primary motivation for the parameter space we search in this analysis. As previously discussed, the role of the DM candidate in SUSY is filled by χ_1^0. A viable candidate for the role of X is taken to be the SUSY $\tilde{\tau}$ (stau), and it is this particle that we consider in coanihilation scenarios in this analysis. This choice is motivated by the fact that, as the lightest of the

SUSY sleptons, the stau is a possible next-to-lightest-supersymmetric particle (NLSP) in MSSM which would naturally decay to our LSP.

3.2 Leptoquark-Portal Dark Matter

The leptoquark (LQ), is a theorized bosonic particle that, as the name implies, couples to a lepton and a quark. The LQ appears in many BSM theories, including Grand Unified Theories (GUTs) [45], and Technicolor theories [46]. Though traditional LQ models only consider couplings of LQ to SM leptons and quarks, more novel models, such as that we consider in this study, can also pair with DM. As such, the LQ can act as a mediator in the coannihilation mechanism that hopes to explain discrepancies in DM relic density as visualized in Fig. 3.2.

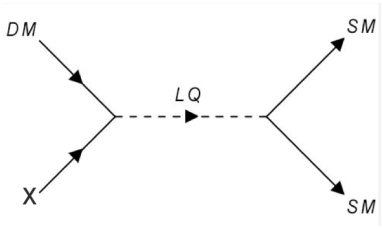

Figure 3.2: LQ-mediated DM Coannihilation

We adopt here a *simplified* model, where the minimum model content for the LQ to contribute to the coannihilation process $DM + X \rightarrow SM + SM$, and to yield agreement with DM relic density, are applied as described in [8]. To achieve this, we introduce three new fields: a field that gives rise to our DM particle, "*DM*," a field X for our coannihilation partner, and a field M_s for our mediating particle (in this case, the LQ). The particles associated with each of these new fields are a majorana fermion (one who is indistinguishable from its antiparticle), a dirac fermion, and a scalar boson, respectively. The Lagrangian can be written as

$$\begin{aligned}\mathscr{L} = \frac{i}{2}\overline{DM}\partial DM + i\overline{X}DX + |D_\mu M_s|^2 - \frac{m_{DM}}{2}\overline{DM}DM - m_X\overline{X}X - V(M_s,H) \\ -(y_D\overline{X}M_sDM + y_{Ql}\overline{Q_L}M_s l_R + y_{Lu}\overline{L_L}M_s^c u_R + h.c.)\end{aligned} \tag{3.8}$$

where

$$V(M_s,H) = V(H) + m_{M_s}^2 M_s^\dagger M + \frac{1}{4}\lambda_{M_s}(M_s^\dagger M_s)^2 + \varepsilon_{M_s}M_s^\dagger M_s(H^\dagger H - \frac{v^2}{2}). \tag{3.9}$$

M_s is the mediator field associated with our LQ, M_s^c is the charge-conjugate field, and $V(H)$ is the SM Higgs potential. y_D is the dark sector Yukawa coupling, y_{Ql} and y_{Lu} are the Yukawa coupling matrices associated with LQ $\rightarrow l + q$ vertices. As what we are describing here is a model capable of providing a coannihilation method equally described by the Boltzmann equations in section 3.1.3 (Eqns. 3.5, 3.6, 3.7), the need for a particular Δ value for coannihilation to contribute significantly to DM reducion applies as well. The required

Δ for a spectrum of DM and X masses is displayed in Figure 3.3 below.

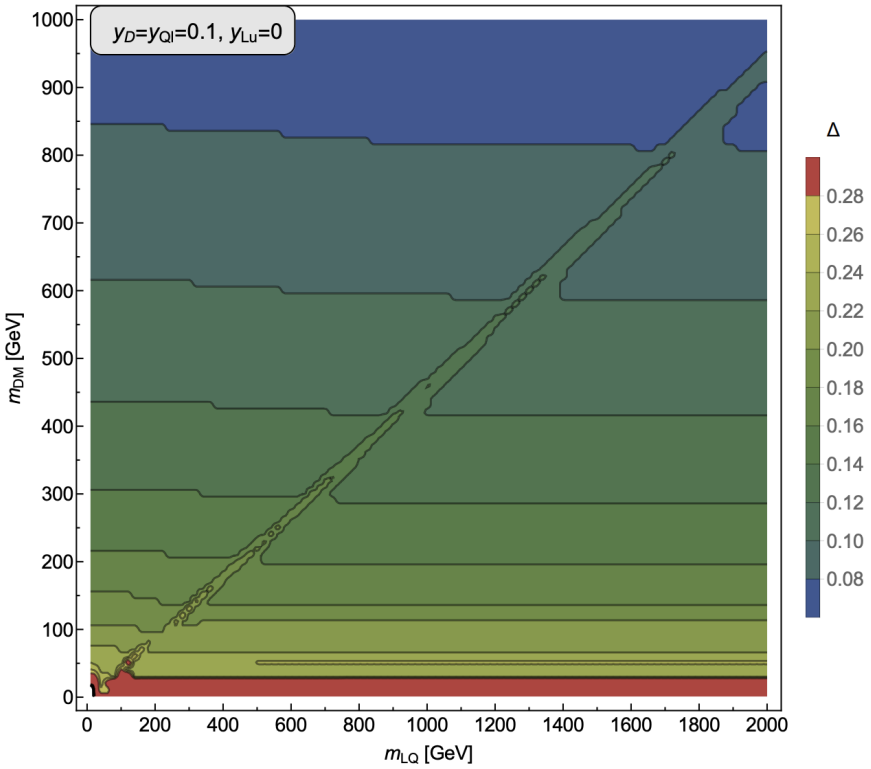

Figure 3.3: Δ value required for agreement with measured $\Omega_{DM}h^2$ [8]

3.3 Brief discussion on SUSY, LQ-portal DM, and the B-Meson Anomalies

SUSY and LQ models have the potential to explain the B-meson anomalies [47, 9, 48]. In the SUSY case, lepton flavor violation (LFV) interactions are introduced in slepton/guagino exchanges, as well as higgs-mediated processes [49]. An example of some MSSM process diagrams that could contribute to the anomaly are displayed in Figure 3.4. The process by which the LQ could mediate the B-meson decays is displayed in Figure 3.5. LFV processes can be introduced in various ways to LQ models (not necessarily in conflict with what we search for in this analysis). An example of such a model with LFV introduced via a coupling structure is described in detail in ref. [48].

3.4 SUSY - Experimental Status

The search presented in this document is a search for SUSY which would simultaneously provide a DM candidate, and, with it, some insight into the nature of DM. Searches for SUSY are vast as the unconstrained parameter space to cover is large. Following from the theoretical motivations discussed in this chapter, we focus our search on R-parity conserving MSSM models with a small mass difference between the LSP and its

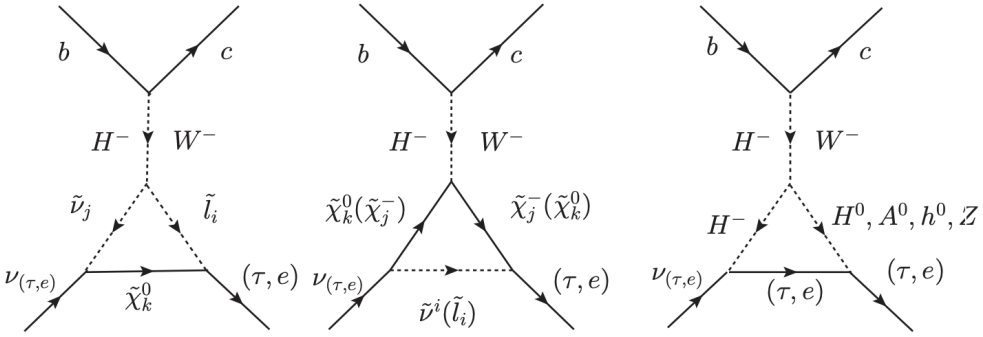

Figure 3.4: Pengiun diagrams contributing to B-meson decays [9].

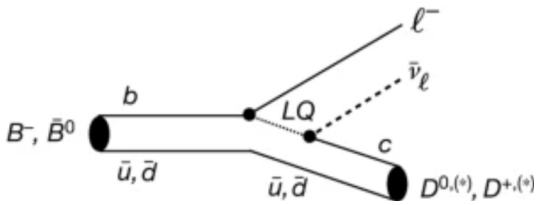

Figure 3.5: Feynman diagram of LQ-mediated process potentially contributing to B-meson decays. [10]

coannihilation partner, $\Delta m = m_{\tilde{\tau}} - m_{\chi_1^0}$. This parameter space comprises what is known as the *compressed mass spectrum*, and it satisfies requirements to correctly predict DM relic density. We also look for 3rd generations SM decay products in our final state, as the disagreements in predicted and observed values of $R_{D^{(*)}}$ suggest there may be enhanced couplings related to these. SUSY searches at colliders rely on SUSY particle production via interactions with SM particles. The produced SUSY particles may then go on to decay to our DM candidate (LSP). As SM particles are less massive than theorized SUSY (including our DM candidate) particles in many BSM theories, the energy required to produce such events is large (on par with collision energies in the early universe). Particle colliders are uniquely suited to host these sorts of high-energy interactions in hopes of finding evidence of new physics. As the most powerful collider in history, the Large Hadron Collider (LHC) is the most promising modern experimental vessel for DM production searches.

Results of the two most recent SUSY searches that serve as relevant comparisons to our study are presented here. The first of these is SUS-19-012, a search for electroweak production of charginos and neutralinos in proton-proton collisions at the in the CMS detector at the LHC [50]. The diagrams of the processes searched for are displayed in Fig. 3.6, where partons from incoming protons with a center-of-mass energy $\sqrt{s} = 13\ GeV$ produce a lightest chargino $\chi_1^\pm$ and next-to-lightest neutralino χ_2^0. The resulting limits set by this search are displayed in Fig. 3.7. There was no significant indication of new physics found in this search,

though one can see from the black contours representing excluded masses at 95% confidence level that this search doesn't have sensitivity in the compressed mass spectrum region we are interested in.

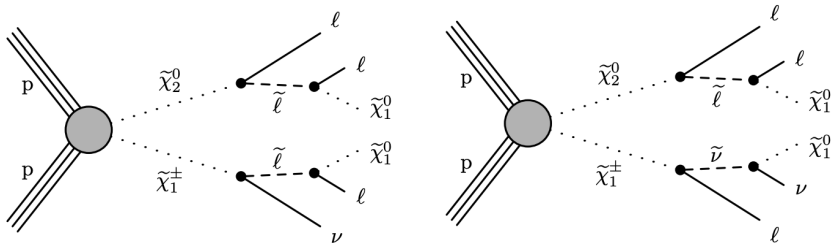

Figure 3.6: Primary processes considered in SUS-19-012

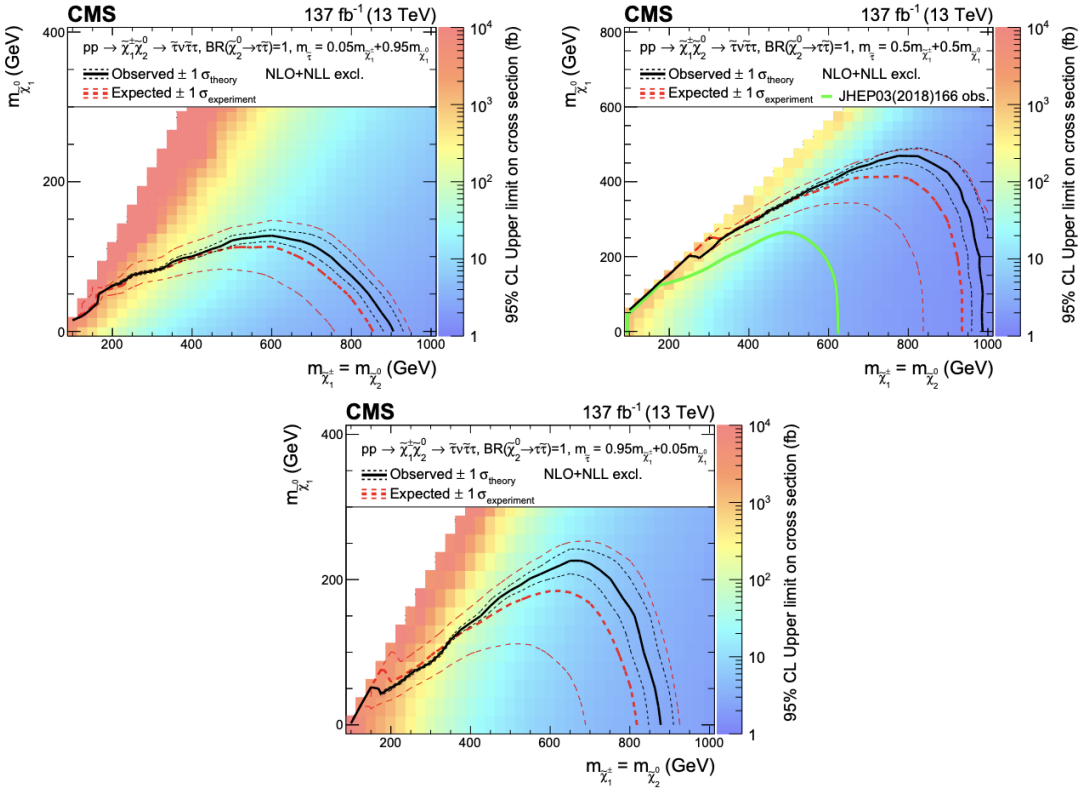

Figure 3.7: Interpretation of results for τ-dominated decays in SUS-19-012 for three different mass-splitting parameters

The next work we consider is a search for direct pair production of the SUSY stau ($\tilde{\tau}$), SUS-21-001 [51]. The primary production process considered in this analysis is displayed in Fig. 3.8, and the resulting exclusion limits can be seen in Fig. 3.9. We again note that no evidence of new physics was found in this study, and also the lack of sensitivity in compressed mass spectra scenarios.

A common feature of these two analyses is the lack of sensitivity in the compressed mass realm. This is

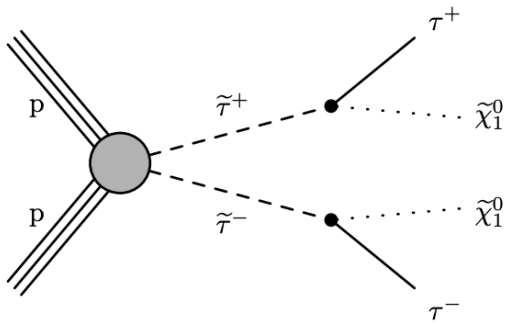

Figure 3.8: Primary processes considered in SUS-21-001

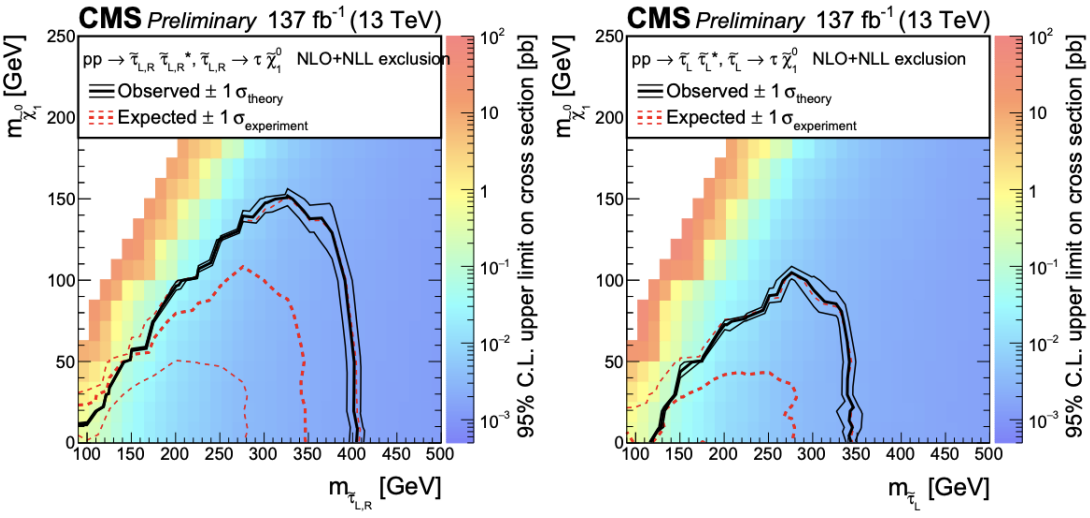

Figure 3.9: Interpretation of SUS-21-001 results for degenerate (left) and left-handed (right) stau sleptons.

a crucial feature if one hopes to search for MSSM models that explain the DM relic density disagreement. One of the strengths of this analysis, as will be developed and described in detail in subsequent chapters, is the novel methodology adopted to gain sensitivity in compressed mass spectra regions.

3.5 Leptoquarks - Experimental Status

LQs have been the subject of numerous searches at the LHC, with no direct evidence found as of yet. However, these searches almost exclusively consider a model where the LQ solely couples to a SM lepton and quark. The model we consider, which is extended by including a coupling to DM, is novel and searches for it at LHC are a limited. The most recent of these is a search for DM in events with a LQ and missing transverse momentum, CMS-EXO-17-015 [52]. As pictured in Fig. 3.10, the lepton in the final state of this search is a SM μ, where our search looks in final states with hadronically-decaying τs, as theoretical motivations suggest enhanced couplings in this scenario.

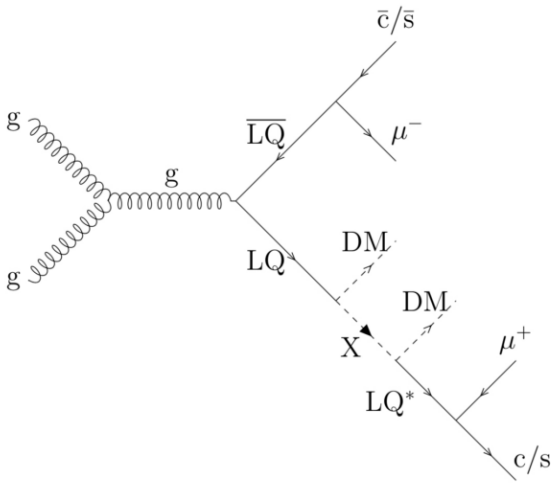

Figure 3.10: Example diagram of processes considered in CMS-EXO-17-015

CHAPTER 4

The Large Hadron Collider

The Large Hadron Collider (LHC), located at the European Center for Nuclear Research (CERN) on the French-Swiss border, is the world's largest and most powerful particle accelerator and collider. By the time the LHC was completed, direct evidence for many SM particles had been found. This is even the case for heavier particles such as the top quark, which was detected at collisions in the Collider Detector at Fermilab (CDF) via the Tevatron accelerator in 1995 [53]. At this point, the final piece of the puzzle to confirm the longstanding theoretical support for the SM was to search for confirmation of the existence of the Higgs boson. Thus, the main purpose the LHC was built was to investigate the nature of the weak force and symmetry breaking via the Higgs, which was expected to be evident at the planned LHC energies. Upon discovery of this Higgs in 2012 [54], work on investigating the nature of the weak force has continued, though there has been increased focus on searching for DM and promising BSM theories, such as SUSY.

The LHC design incorporates a large, magnet-accelerated superconducting ring, and while the tunnel geometry was originally intended for the LEP electron-positron machine, adjustments were made to accommodate the LHC [55]. The official approval to build the LHC was given by the CERN council in 1994 and was driven by cost-saving advantages gained through the reuse of the LEP tunnel. The tunnel is 27km long and lies between 45 m and 170 m under the surface depending on location. The beampipe tubes are kept at ultrahigh vacuum levels, and at a temperature of $-271.3^{\circ}C$ in order for the superconducting properties of the accelerator magnets to maintain functionality [56]. The LHC accelerates proton bunches in opposite directions in order to create proton-proton collisions.

The center-of-mass (CM) energy in LHC collisions has varied over the various run stages. For the data considered in this analysis (Run II), collisions were conducted at a COM energy of $\sqrt{s} = 13$ TeV, though, at the time of writing, the LHC is conducting Run III at $\sqrt{s} = 13.6$ TeV. To acquire this energy and speeds near the speed of light, the protons are passed through a series of systems. The first step is Linac4, a linear accelerator, which accelerates hydrogen ions to an energy of 160 MeV to then be fed to the Proton Synchrotron Booster where the ions are stripped of their electron, leaving single proton nuclei. At this stage, they are successively passed to the Proton Synchrotron and Super Proton Synchrotron where they aqcuire 450 GeV of energy before being injected into the main ring to be brought to full speed [57].

The aim of a collider is to maximize its *luminosity* $\mathscr{L}$, which is a measure of the number of collisions it can provide per unit time, per unit area. The number of events a particular process produces can be expressed at $N = \sigma \mathscr{L}_{int}$, where σ is the cross-section of interest, and $\mathscr{L}$ is the instantaneous luminosity integrated over

a period of time. At its full capacity, the LHC is expected to deliver an integrated luminosity of 100 fb^{-1} per year, corresponding to a rate of $\sim 10^9$ inelastic events per second at the all-purpose detectors.

The LHC has four main collision points located at the four main detectors: The Compact Muon Soldenoid CMS, ATLAS, ALICE, and LHCb. CMS and ATLAS are general purpose detectors, ALICE focuses on heavy-ion collisions, and LHCb focuses on experiments involving b quarks [58, 59, 60, 61]. There are also five other highly-specialized detectors operating at the LHC: FASER, SND, LHCf, MoEDAL, and TOTEM. A diagram of the layout of the LHC with its main detectors and features can be seen in Fig. 4.1.

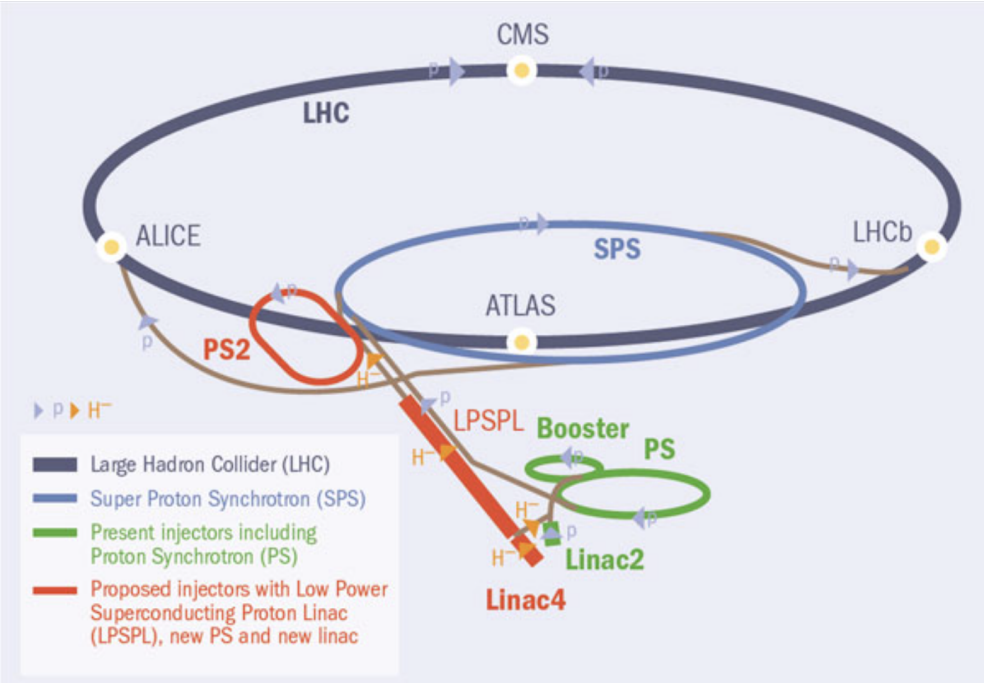

Figure 4.1: Schematic of the LHC Accelerator Complex at CERN [11]

CHAPTER 5

The CMS Detector

The CMS detector, along with ATLAS, is one of the general-purpose detectors at the LHC. It is located at Point 5 of the LHC on the French side of the facility. The CMS detector stands out with its unique construction method compared to other large detectors at the LHC. Rather than being built on-site, it was assembled in 15 sections at ground level and later lowered into an underground cavern near Cessy, France. There, it was reassembled to form a complete detector measuring 28.7 meters in length, 15 meters in width, and 15 meters in height [58]. The CMS experiment is one of the most extensive international scientific collaborations in history. It brings together approximately 5500 particle physicists, engineers, technicians, students, and support staff from 241 institutes across 54 countries [58].

The goals of the LHC program that the CMS detector was built to attain are as follows [13]:

- Capabilities for identifying and accurately measuring the momentum of muons across a broad spectrum of momenta and angles, and accurate determination of muon charges for momenta below 1 TeV.
- An inner tracker with good charged-particle momentum resolution and reconstruction efficiency, and efficient triggering and offline identification of tau leptons and b-jets.
- Good resolution of electromagetic energy and diphoton and dielectron mass. Wide geometric coverage, good rejection of neutral pions (π^0), and efficient photon and lepton isolation at high luminosities.
- Hadron calorimeters with a large hermetic geometric coverage and with fine lateral segmentation in order to achieve good missing-transverse-energy (E_T^{miss} or p_T^{miss}) and dijet-mass resolution.

A big-picture rendering and a transverse slice of CMS and its components are shown in Figures 5.1 and 5.2, respectively.

The CMS detector adopts a coordinate system with its origin centered at the primary collision point within the experiment. The y-axis is oriented vertically upward, while the x-axis points radially inward towards the center of the LHC. The z-axis aligns with the beam direction. ϕ is the azimuthal angle measured from the x-axis in x-y plane, and the radial coordinate in this plane is r. θ is the polar angle measured from the z-axis. Because of its approximate Lorentz invariace, the more widely used coordinate in the θ direction is the pseudorapidity η, defined as

$$\eta = -ln\, tan(\theta/2). \tag{5.1}$$

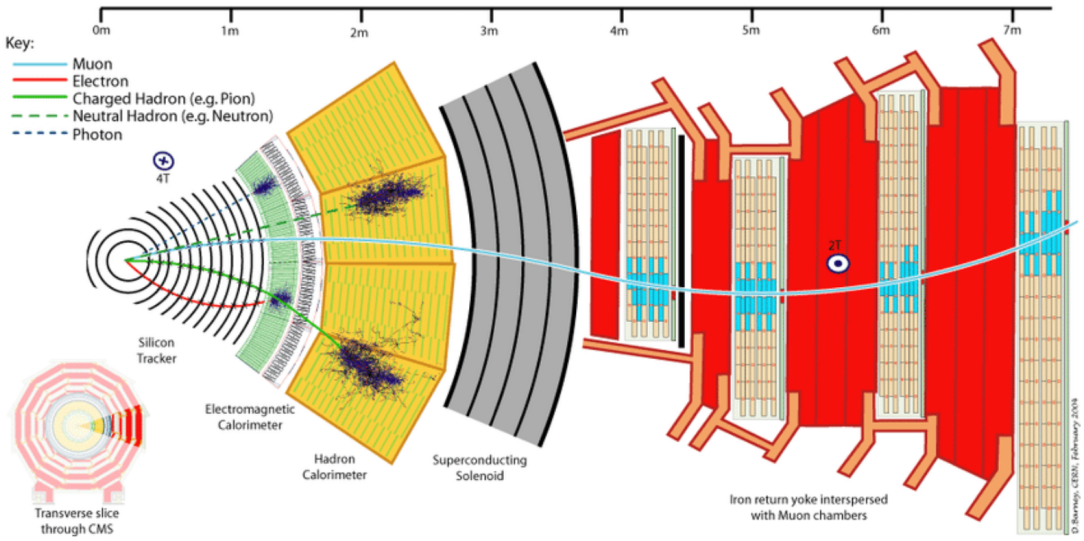

Figure 5.1: Length-wise cross-sectional view of the CMS detector. [12]

The transverse momentum (p_T) and transverse energy (E_T) are calculated using the x and y components of the corresponding four-vectors. Because the pre-collision momentum in the detector is effectively carried across the beamline, momentum conservation dictates that the total transverse momentum (in the x-y plane in CMS) should be zero. Thus, a useful quantity to study is the missing transverse momentum p_T^{miss}, also sometimes denoted as missing transverse energy E_T^{miss}.

As can be seen from Figures 5.1 and 5.2, the central feature of a the detector is the large superconducting solenoid, measuring 13m in length and 6m in diameter. Within it are contained are the tracking system, electromagnetic calorimeters, and hadronic calorimeters. Surrounding the solenoid are the muon systems. It's worth noting that the CMS detector has undergone upgrades since its initial construction and installation. The planned updates are grouped into Phase I and Phase II, and changes from both of these have been implemented in the detector at different times. This description follows only the features and changes as relevant to the period of data-taking used in this analysis, namely, the 2016-2017 Run II period.

5.1 The Superconducting Solenoid

The superconducting magnet in the CMS detector provides a magnetic field of 3.8 T for the purpose of bending charged particles via a Lorentz force. From the curvature of the path provided by the tracking system, we can deduce important properties of the particles involved such as momentum and charge. The CMS solenoid is the largest superconducting magnet ever built. It operates at a temperature of -268.5°C. It is wound in four coils and stores an energy of 2.6 GJ at full current. A rendering of the housing and modules

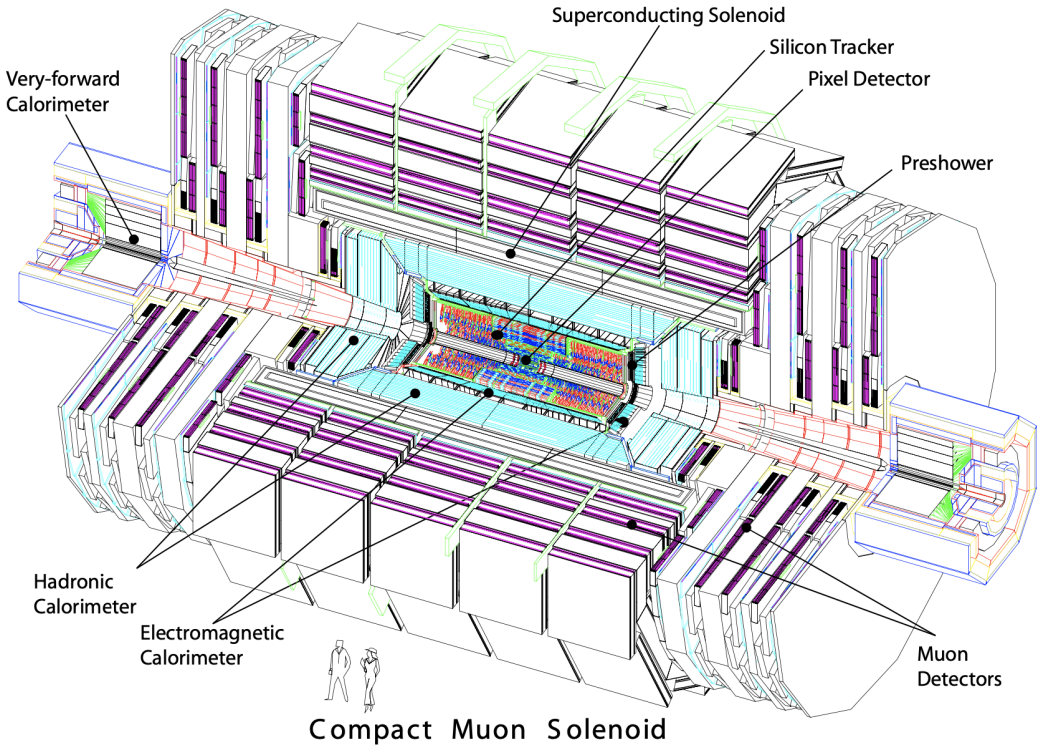

Figure 5.2: Big-picture view of the CMS detector. [13]

composing the solenoid magnet can be seen in Fig. 5.3.

5.2 The Inner Tracker

The inner tracking system (IT) is the first part of the detector which objects from the collision will encounter. The tracker was designed to accurately detect the trajectories of charged particles while minimally disturbing them by making strategic position measurements [13]. It is also capable of reconstructing secondary interaction vertices. As a whole, it measures 5.8 m in length with a diameter of 2.8 m, and the magnetic field provided by the solenoid fully spans its volume. The detection elements of the tracker are silicon-based, and it is split into two main parts: a pixel detector and a strip detector. Each detector also features endcap disks which extend the reach of the detector to $|\eta| < 2.5$. A cross-section showing the components and arrangement of the tracker is displayed in Fig. 5.4 (pre-2017 upgrade). The design and development team for the IT was made up of 500 physicists and engineers from 51 institutions who worked for 12 to 15 years to construct the unit.

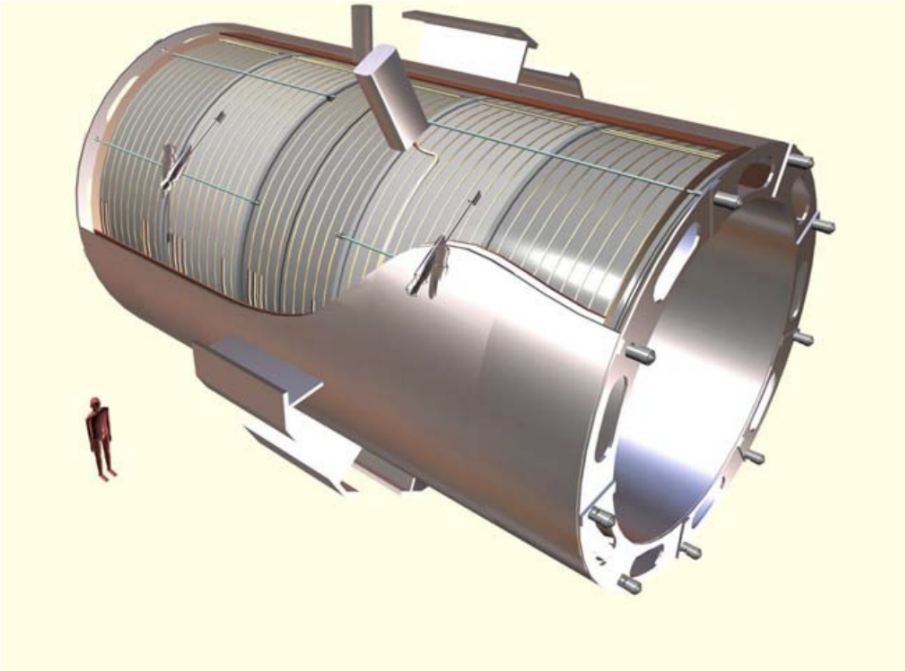

Figure 5.3: Superconducting Solenoid Housing and Modules [13]

5.2.1 The Pixel Detector

The pixel detector lies closest to the interaction region. It is made up of modules with tile-like silicon sensors, each of which measures 100 $\mu m \times 150\ \mu m$ [14]. These are arranged into a barrel surrounding the beamline, and two endcaps in the forward regions. The electrons ejected from the sensors due to interaction with charged particles are collected by application of a small voltage, as illustrated in Fig. 5.5 (right), and the signal is then amplified by readout chips (ROCs). Each ROC serves a grid of 52×80 pixels. An illustration of the components of the pixel detector modules is displayed in Fig. 5.5 (left). Between the data taking periods of 2016 and 2017, the pixel detector underwent a set of upgrades resulting in improved ROC chips, an additional layer in both the barrel and endcaps, an inner layer of the barrel at a closer radius to the interaction region, and a lighter and more efficiently cooled carbon fiber support system [62]. The upgraded pixel detector has a first layer at a radial distance of 2.9 cm, compared to the pre-ugrade distance of 4.4 cm. Its additional endcap layers also allow for multiple-hit coverage even at high $|\eta|$ values, resulting in more robust trajectory reconstructing capabilities.

5.2.2 The Strip Detector

As a whole unit, the strip detector covers the radial distance from 20 cm to 120 cm from the beamline. It is composed of four layers in the inner barrel (TIB), arranged in shells, and two inner endcaps (TID) composed of three small discs each. Surrounding both the TIB and TID is the outer barrel (TOB), consisting of six

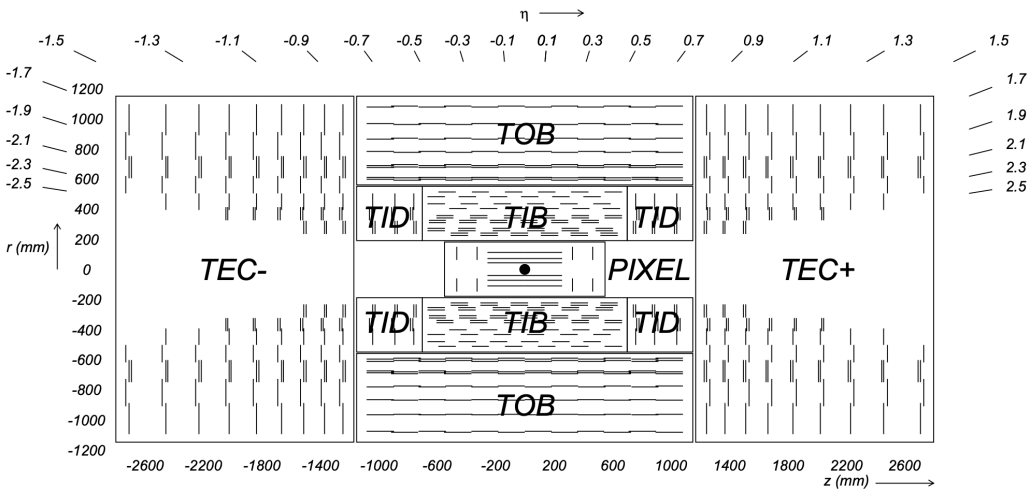

Figure 5.4: Schematic of cross-section of CMS inner tracking system. Acronyms: TIB - tracker inner barrel; TID - tracker inner disks; TOB - tracker outter barrel; TEC - tracker end caps. [13]

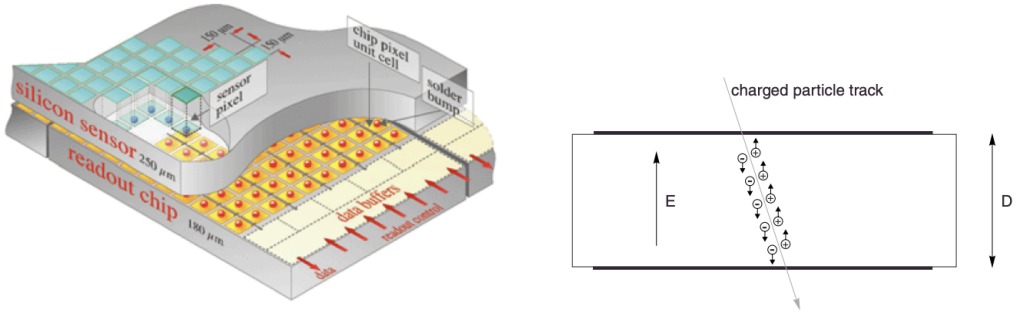

Figure 5.5: Pixel Detector Module (left) [14], and Diagram of charged particle moving through silicon detector (right) [15]

concentric layers. The tracker is completed with two endcaps (TEC), sealing off each end. The arrangement of each strip module in the inner barrel can be seen in Fig. 5.6. Each section of the tracker utilizes silicon modules specifically optimized for their respective locations within the detector [63]. It is made of silicon strip sensors embedded in 15,200 modules that function in a manner similar to the pixel detector. The amplification in the case of the strips is handled by Analogue Pipeline Voltage (APV25) chips.

In the 2016 data-taking period of Run II, a faulty setting caused the pre-amplification system of the APV chips in the strip detector to become oversaturated causing a drop in efficiency at high instantaneous luminosities, as shown in Fig. 5.7 [17]. Due to this effect, the 2016 data set and Monte Carlo samples we use in this analysis had to undergo special post-processing to account for these effects. This issue affected part of the data in 2016, but was fixed for the remainder of Run II.

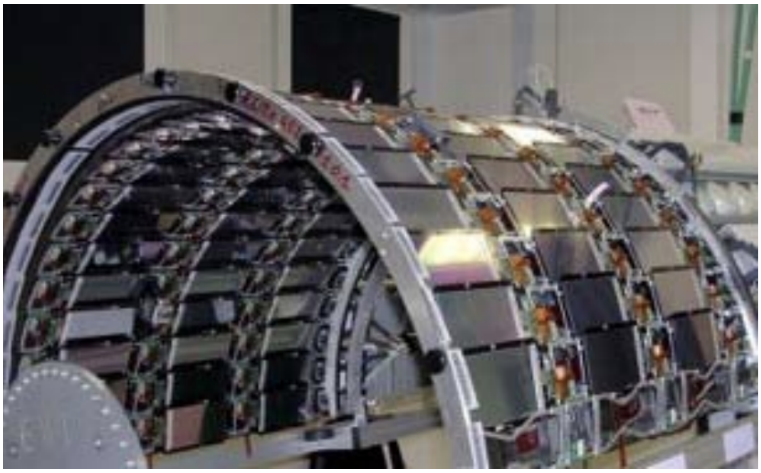

Figure 5.6: Half of the strip detector inner barrel (TIB). [16]

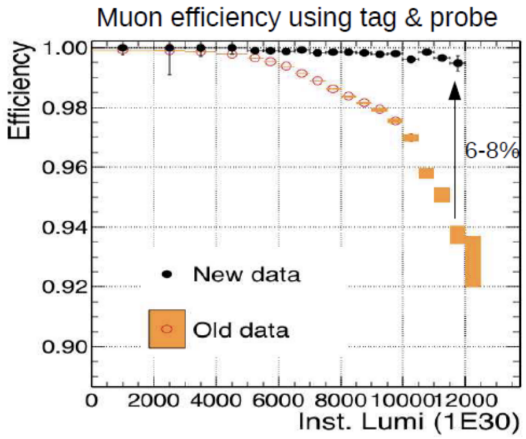

Figure 5.7: Loss of efficiency due to APV pre-amplification issue. ”Old data” corresponds to period 0f 2016 issue, and ”New data,” is in 2017, post-fix. [17]

5.3 The Electromagnetic Calorimeter

The electron calorimeter (ECAL) is just beyond the inner tracker, and it is in place to measure the energy of electrons and photons. It is composed of a barrel covering the range $|\eta| < 1.479$, and two endcap sections covering the range $1.479 < |\eta| < 3.0$ [13]. The barrel is made of 61,200 scintillating lead tungsten crystals divided into 36 modules. The gap between crystals within a module is 0.35 mm, and 0.5 mm between modules. The endcaps are made with 7,324 crystals each, divided into units of 5×5 crystals. Preshower detectors are located in front of each of the endcaps, spanning an angular region $1.653 < |\eta| < 2.6$, for the purpose of improving determination of electron and photon positions, helping identify electrons vs. minimum ionizing particles, and to identify neutral pions (π^0). A layout of the ECAL and its parts can can be seen in Fig. 5.8. Because the number of scintillation photons produced by the crystals is temperature dependent, it is

kept at a constant nominal temperature of 18 °C. The cooling systems supplies water flow independently to each module to maintain temperature.

The crystals longitudinally transmit light with wavelengths above $\sim$ 350 nm, and the resulting scintillated light is mainly in the blue-green visible range of $\sim$420-430 nm as can be seen in Fig. 5.9. The photodetectors tasked with detecting scintillated photons are photodetavalanche photodiodes in the barrel (APDs) and vacuum phototriodes (VFTs) in the endcaps. These must be radiation tolerant and be capable of operating in the strong magnetic field supplied by the solenoid. The varying conditions in the barrel vs endcaps in both those variables were motivating factors in choosing different photodetectors [13].

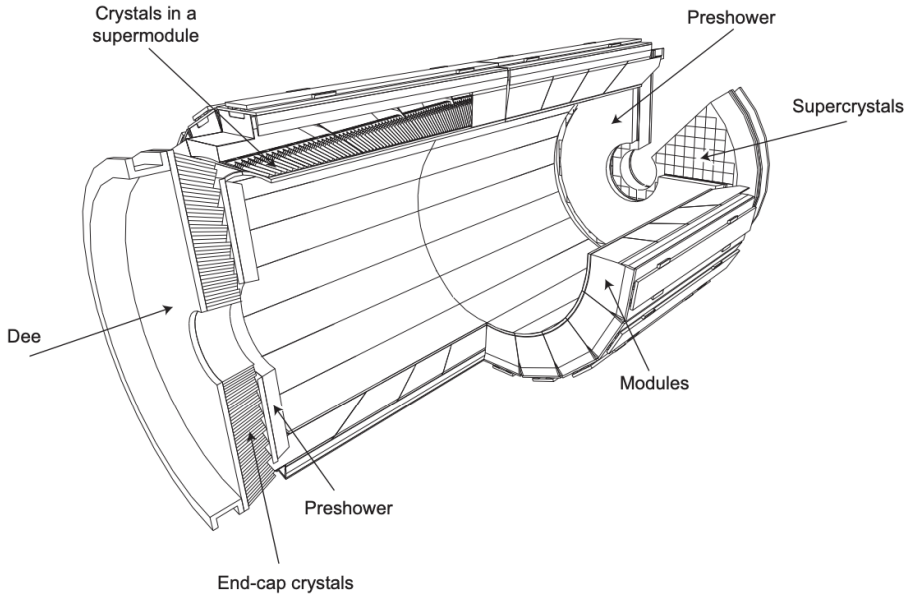

Figure 5.8: CMS Electron Calorimeter. (*Dee* is a name used to refer to each half of an endcap). [13]

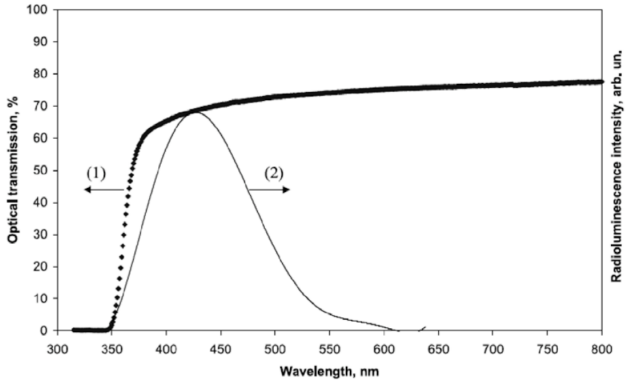

Figure 5.9: Longitudinal optical transmission and radioluminescence of $PbWO_4$ crystals. [13]

5.4 Hadron Calorimeter

The Hadron Calorimeter (HCAL) is the next layer of the detector, lying between the ECAL and the superconducting solenoid at $R = 1.77$ m to $R = 2.95$ m. The purpose of the HCAL is to measure the energy and position of hadrons while simultaneously contributing to accurate missing transverse energy measurements to aid inference of the properties of invisible object such as neutrinos.The HCAL is composed of four parts: a barrel (HB, seperated into halfs HB+ and HB-), two endcaps (HE), two forward calorimeters (HF), and the outer calirmeters (HO); the HO is the only part that lies beyond the solenoid magnet. With this full configuration the HCAL is able to cover a wide angular span of incoming material up to $|\eta| < 5.2$ and is designed to be hermetic, thus capturing nearly every collision remnant. A longitudinal cross-section view shows the location and angular coverage of each part in Fig. 5.10. The HB is made of 36 azimuthal wedges, with a minimal distance of 2mm between each wedge.

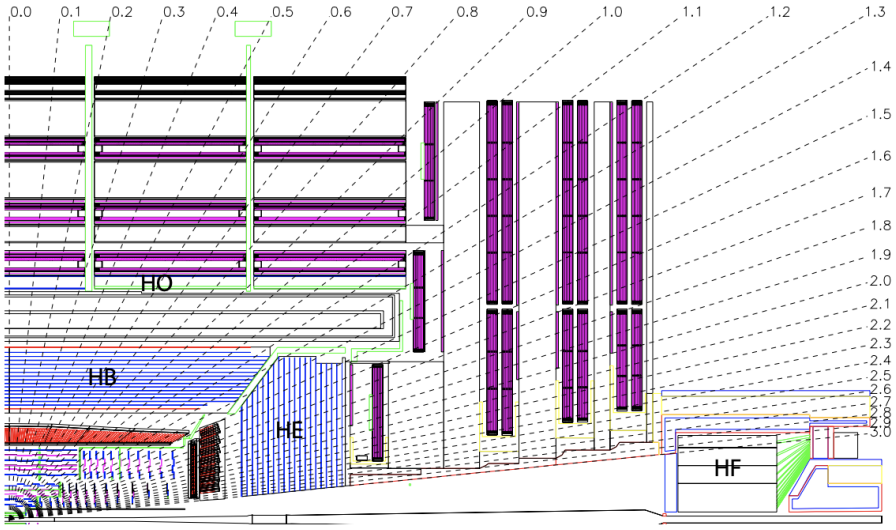

Figure 5.10: HCAL cross-sectional view. Key: hadron barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. [13]

The HB and HE are composed of alternating scintillating and absorbing material. The absorbing material is composed of brass (and some steel) plates aligned parallel to the beam axis, and the scintillator plates are mostly Kuraray SCSN81 plastic chosen for their stability and radiation tolerance [13]. The readout from the scintillators is carried by 0.94 mm-diameter plastic fiber, and each tray of plates also contains 1 mm steel tube and a quartz fiber for calibration purposes. The photodetector agents in HB, HE, and HO were originally hybrid photo diodes (HPD) but were replaced with silicon photomultiplier devices (SiPMs) in HO in 2014, in HE in 2017, and in HB in 2019 [64] (the data used in this analysis spans 2016-2018). The purpose of the HO is to catch additional material that is not stopped in the ECAL and HCAL barrel sections. It uses the same readout and scintillator material as HB and HE, but instead uses steel from the solenoid magnet as an absorber.

The HF covers the high $|\eta|$ region of the detector and instead collects Cherenkov light through quartz fibres. The HCAL is a "sampling detector." The electronic elements integrate the signal from photodetectors every 25 ns.

5.5 The Muon System

As implied by the detector name, the muon system of the CMS detector is one of its most powerful and important features. The design goal for the muon system was to provide precise and robust muon measurements, and it has three functions: identifying muons, measuring muon momentum, and providing a triggering mechanism. It was designed to reconstruct muon momentum and charge over the entire kinematic range of the LHC, and to accomplish this it uses three types of gaseous detectors [13]. Similar to the other detector components that have been discussed, it is composed of a barrel region and two endcaps. The barrel region covers $|\eta| < 1.2$, and the endcaps covers $0.9 < |\eta| < 2.4$.

The three gas ionization chambers are the drift tube chambers (DT), chathode strip chambers (CSC), and resistive plate chambers (RPC). The DTs are mostly contained in the barrel, CSCs in the endcaps, and RPCs in both. The layout of of each type within the detector can be seen in Fig. 5.11. DTs are 4 cm wide, and a wire is contained in the gas volume. When muons ionize the gas in the DT, the resulting electrons drift towards the positively charged wire, where they are then amplified and produce a signal. By measuring the time it takes for the electron to drift to the wire, accurate position measurment can be made. CSCs are arrays of anode wires crossed with copper cathode strips in the gas volume. This configuration allows for the measurement of a two-dimensional coordinate when the gas is ionized and products of ionization move towards the corresponding wires. CSCs were chosen for the endcaps due to their hardiness in uneven magnetic fields and their ability to handle a large influx of muons. The closely spaces wires also provide fast response, resulting in efficient and speedy triggering. The purpose of the RPCs is to provide a redundancy in the triggering system to ensure good efficiency and coverage of the full kinematic region. They are composed of a high-resistivity plastic arranged in parallel anode and cathode plates separated by a thin volume of gas. Upon a muon hit, the signal is picked up by external metalic strips, allowing a fast momentum measurement and fast trigger decisions. The cumulative effort of these detectors results in muon identification efficiencies $> 95\%$ [18].

5.6 Trigger

The peak interaction rate at CMS is 40 MHz, with each interaction producing a large number of particles. Full reconstruction of the information for each event is not only a near technical impossibility, but also an inefficient way to do physics, as most interactions will not be type of events of interest in CMS studies. The

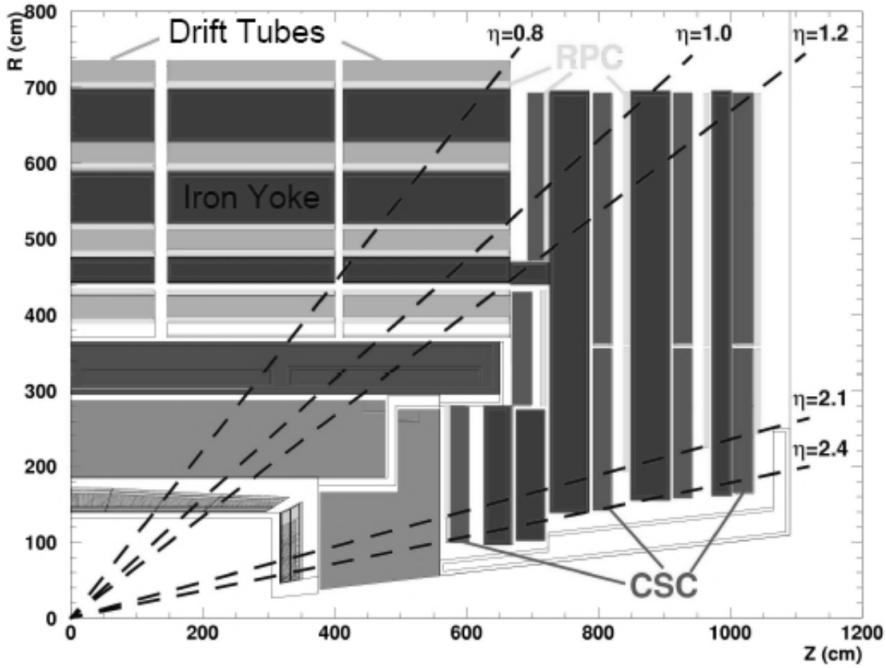

Figure 5.11: Cross-sectional area showing the muon system layout and corresponding detector type. [18]

purpose of the CMS trigger system is to provide near real-time and offline means to gather potential events of interest and further filter them.

The functionality of the trigger system is accomplished in two parts. The level 1 (L1) trigger system is a purely hardware-based implementation which takes raw detector information primarily from the muon system (as described in the previous section), the ECAL, and HCAL to collect an effective rate of 100 kHz events. The level 2 (L2) and level 3 (L3) processing, collectively known as the high-level triggers (HLT), then collects the filtered events and further processes the dataset in a software-based approach that is capable of looking at further properties of each event in different parts of the detector. Due to the offline nature of the HLT, it can take advantage of full detector information, apply complex algorithms, and undergo updates noninvasive to the detector hardware when necessary in order to improve identification and reconstruction of event objects. After HLT processing, the data we are left with reflects an effective interaction rate of 1 kHz [65].

CHAPTER 6

Event Reconstruction and Particle Identification

6.1 Particle Flow

Particle Flow (PF) is the technique used to provide event descriptions at CMS by using information from multiple parts of the detector. The PF concept was initially developed by the ALEPH collaboration at the Large Electron-Positron Collider (LEP) [66]. Historically, energy of objects such as jets (to be described in more detail later in this chapter), electrons, and photons, was measured purely by the showers they left behind in the calorimeters with no reference to information from the tracker. PF incorporates elements from both tracking and caloremeters to achieve higher efficiencies, higher resolutions, and lower misidentification rates. The CMS detector has demonstrated great suitability for implementing the PF algorithm due to several advantageous features. It benefits from a powerful magnetic field that effectively distinguishes neutral and charged hadrons. Additionally, the detector's tracker exhibits fine granularity, enabling precise measurements of jets with transverse momenta reaching up to 1 TeV. CMS also features a highly segmented electromagnetic calorimeter (ECAL), which accurately determines energy fractions within jets. Complementing the ECAL is a comprehensive hadronic calorimeter (HCAL) that ensures complete coverage, and an exceptional system for identifying and tracking muons.

The PF approach is based on the concept of a linking algorithm. A particular signature in a single subdetector may constitute a PF element, and the elements may then be linked to form what is called a PF block. For example, a particular trajectory in the tracker constitutes a particle flow element, and, seperately, an energy deposit in the ECAL constitutes a PF element. Given the right conditions, these two elements may be linked and constitue a PF block correspnding to an electron to jointly aid with its reconstruction and identification. Overall, PF results in more accurate energy measurements, better efficiency, and better resolution of all physics objects [19]. An effect of these improvements, which is important to this analysis, is a better performance in reconstruction of missing transverse energy (p_T^{miss} or MET) compared to a calorimeter-only approach, as seen in Fig. 6.1.

6.2 Jet Reconstruction and Identification

Other than p_T^{miss}, the two other objects central to this analysis are jets and τ leptons. The identification of a τ begins with the identification of a jet, thus we begin with a discussion on jet reconstruction and identification. Jets are a collimated "spray" of particles resulting from the hadronzation of isolated, colored particles. In the LHC, these can originate from gluons and u, d, s, b and c quarks, where u d s are referred to as the light

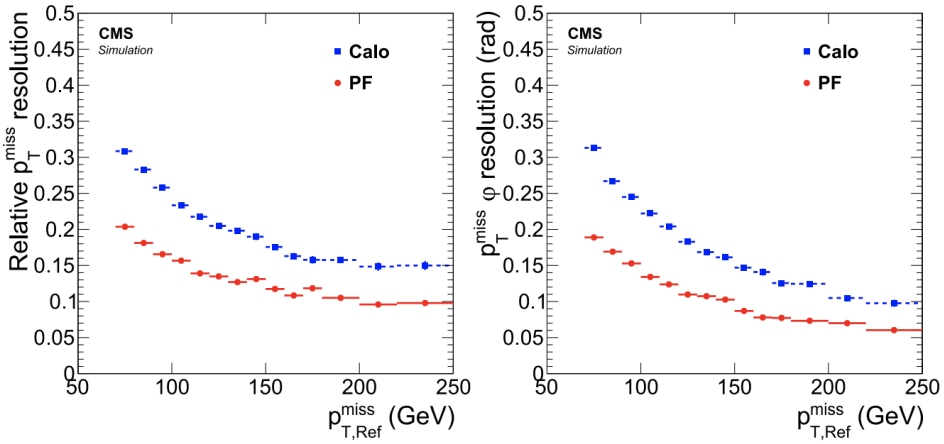

Figure 6.1: Relative resolution and angular resolution of PF p_T^{miss}[19]

flavors, and b c the heavy flavors. Though mostly made of hadronic material, jets can also carry photons coming from π^0 decays, and leptons coming from quark decays. Generally, the composition of a jet is based on its momentum and originating object. There are three different supported reconstruction methods for jets: the calorimeter approach, a "jet-plus-tracks" approach, and the PF approach [67]. The PF technique, as discussed in the previous section, reconstructs the individual particles within a jet from multiple subdetector signatures and these can then be clustered into a jet according to a range of parameters, such as the angular distance between objects.

The clustering of jets is handled by the anti-kT algorithm, resulting in what are known as AK4 jets. Anti-kT is a sequential recombination algorithm that clusters PF objects according to a weighted distance in the (η,ϕ) plane, where particles with proximity to higher p_T objects will cluster first, and separate objects are clustered in together iteratively until a threshold radius parameter R value between clusters is reached, typically resulting in a final hard jet cone with radius R [20]. Resulting clusters from a simulated event are displayed in Fig. 6.2 for an R value of 1.

6.2.1 Correcting Jet Atributes

The clustering of objects is the first step in the factorized approached taken in CMS to accurately reconstruction jet properties. Conditions such as the presence of pileup and varying response in the detector as a function of jet p_T, η, and flavor require us to apply a range of corrections to data in order to account for these effects and achieve accurate measurements. Jets used in Monte Carlo simulation (MC) must also be "smeared" for accurate comparisons to data. The processing can be summarized as follows:

- **Pile-up Removal(L1)** - Pile-up are interactions occurring in the same bunch crossing as our event of in-

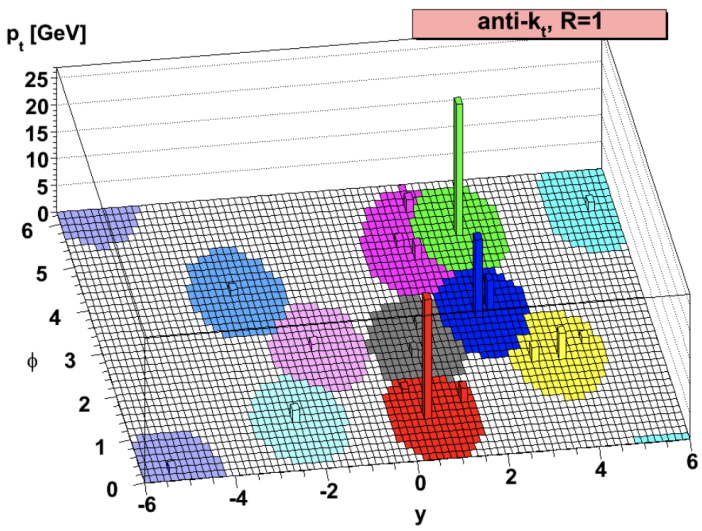

Figure 6.2: Jet clusters from anti-kT algorithm in simulated partonic event. [20]

terest. Using tracker information, we can track the vertex corresponding the pile-up events and remove the particles associated with it. As neutral particles escape the tracker undetected, a different approach is necessary to remove their energy contribution in the calorimeters. This is done by calculating the probability that a nuetral particle originates from pileup, and then scaling the jet energy accordingly. The latter technique is known as pileup per particle identification (PUPPI).

- **Jet Energy Corrections (L2L3 MC-Truth Corrections)** - To correct for the varying response of the detector, a simulated particle response factor is derived from QCD multijet samples simulated to detector-level with the frameworks GEANT4 and PYTHIA 6.4 [68]. This is done by comparing the reconstructed p_T of jet objects to the original parton p_T. These are derived as a function of p_T and η and applied similarly, individually to each jet. The corrections are also cross-checked with data driven methods [67]. The simulated particle response is defined as:

$$R_{ptcl}(\langle p_T \rangle, \eta) = \frac{\langle p_t \rangle}{\langle p_{T,ptcl} \rangle}[p_{T,ptcl}, \eta] \tag{6.1}$$

where $p_T, ptcl$ is the transverse momentum of the original object, and p_T is that of the reconstructed jet. η-dependent detector response corrections for jets in all 2016-2018 can be seen in Fig. 6.3.

- **Jet Energy Resolution Corrections** - After the main effects on the overall jet energy scale (JES) have been accounted for, we focus on jet p_T resolution, defined as the width of an estimated Gaussian fit to

the distribution $p_{T,reco}/p_{T,ptcl}$. These are intended to not have an effect on JES, but only on the width of this distribution. Because JER is worse in data than MC, we apply smearing to MC by either 1. a scaling method that makes use of generator level partons matched to reconstructed jets to derive a correction factor, or 2. a stochastic method that doesn't need the gen level information, thus is applied to pile-up jets.

- **Residual Corrections** - Additional corrections for small effects (% level) on JES are applied in what is called the L3L3Residual phase. These will not be discussed here as they have minimal effects, though further discussion can be found in ref. [68].

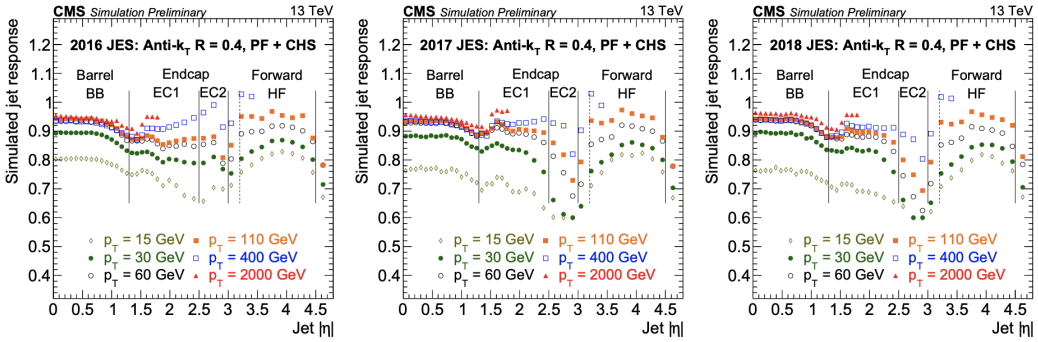

Figure 6.3: Jet response corrections as a function of $\eta^{j}et$ for various p_T values for 2016-2018 (left to right) [21]. Note the different significant inconsistency in response with increasing $|\eta|$

6.2.2 Jet Identification

To reject unwanted jets caused by noise, PF defines two main working points for jets: loose and tight [3]. These work by limiting the percent composition jets can have of particular objects. Because the sources of noise are primarily the calorimeters, the tighter working point mainly limits neutral content coming from the HCAL and ECAL, as these can't be validated with a corresponding track, as is the case with the charged hadronic content. The "tight" working point is used in this analysis, and the detailed requirements are shown in Table 6.1.

6.3 Tau Reconstruction and Identification

As motivated in the first few chapters, the τ lepton plays a central role in our search for SUSY and potential explanations of experimental disagreements with SM predictions. The τ is the heaviest of the leptons with a mass of 1.77 GeV. It is the only lepton that has both hadronic and leptonic decay modes. The branching ratios and specific decay channels are details in Table 6.2. One can see that the decay of the τ always includes at least one neutrino.

Selection	Cut
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
Additionally applied for $\|\eta\| < 2.4$:	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Table 6.1: "Tight" PF Jet ID Requirements [3]

	Decay Mode	Branching Ratio (%)
LEPTONIC	$\tau^- \to e^- \bar{\nu}_e \nu_\tau$	17.8
	$\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$	17.4
HADRONIC	$\tau^- \to h^- \nu_\tau \gamma$	11.5
	$\tau^- \to h^- \pi^0 \nu_\tau$	26.0
	$\tau^- \to h^- \pi^0 \pi^0 \nu_\tau$	9.5
	$\tau^- \to h^- h^+ h^- \nu_\tau$	9.8
	$\tau^- \to h^- h^+ h^- \pi^0 \nu_\tau$	4.8
	other hadronic decays	3.2
	Total Hadronic	64.8

Table 6.2: τ lepton decay modes and branching fractions.

Because of their larger overall branching ratios, CMS analyses often target the hadronic decays modes of the τ (as do we), denoted by τ_h . 11.5 % of the time we see decays to one charged hadron, 35.5 % to one charged hadron plus neutral hadrons, and 15% to three charged hadrons and neutral hadrons. The algorithms for reconstructing τs is the *hadron-plus-strips* (HPS) algorithm. It takes reconstructed AK4 jets with parameters $R = 4$, $p_T > 14$ GeV, and $|\eta|$ <2.3 as input [69], and then identifies the particular decay mode by looking at the content of the jet. The decay modes are identified by the number of charged hadrons leaving signatures in the tracker and HCAL, and the number of "strips". π^0s are common in the τ decay, these don't leave a mark in the tracker, but they mainly decay into two photons which can also produce electron-positron pairs and leave a signature in the ECAL within the jet radius. These are the "strips" in HPS. Fig. 6.4 shows a diagram of the different τ decay modes and their interactions with subdetectors. One can see in this schematic that certain decay modes contain intermediate resonances of ρ and a mesons, which go on to decay to the πs. To enhance the accuracy of τ_h reconstruction, we can take advantage of these resonances and require the resulting $\pi^\pm$ in the final state to have an invariant mass equal to that of the ρ (770 MeV) or a (1260 MeV) depending on the number of *prongs* (charged hadrons) in the reconstructed jet [4]. Another strategy for increasing tau reconstruction is to require the PF candidates constituents of the jet to have a total charge equal to the τ or $\bar{\tau}$, i.e.± 1. The HPS algorithm classifies the decay modes as shown in Table 6.3

HPS decay mode	Hadronic decay(s)
1 prong + 0 strips	$h^{\pm}\nu_{\tau}$, $h^{\pm}\pi^{0}\nu_{\tau}$ (low energy π^{0})
1 prong + 1 strip 1 prong + 2 strips	$h^{\pm}\pi^{0}\nu_{\tau}$, $h^{\pm}\pi^{0}\pi^{0}\nu_{\tau}$
2 prong + 0 strip 2 prong + 1 strip	Experimental modes to recover 3-prong τ_h's, where 1 track is lost (not included in this analysis)
3 prong	$h^{\pm}h^{\mp}h^{\pm}\nu_{\tau}$

Table 6.3: Description of the HPS decay modes [4]

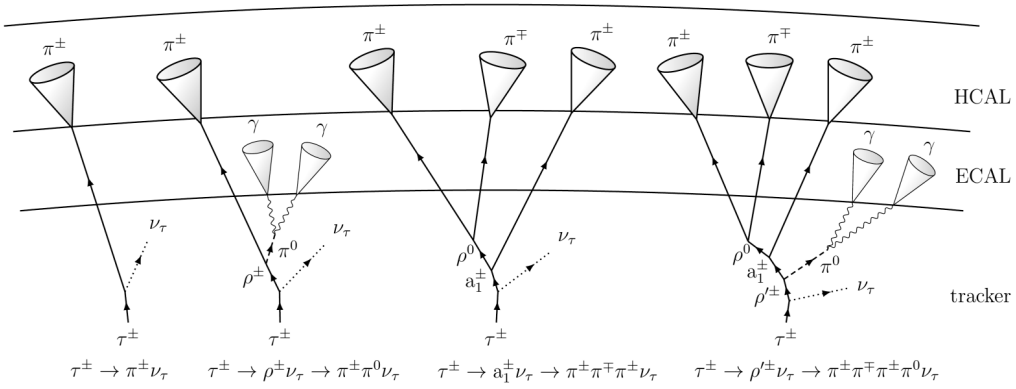

Figure 6.4: Hadronic decay of τs as seen in sublayers of CMS. [22]

The HPS algorithm results in a τ_h candidate with a corresponding decay mode. The p_T range yielding accurate reconstruction and decay mode tagging is taken to be $p_T(\tau_h) > 20$ GeV. The $h^-h^+h^-\pi^0\nu_\tau$ decay mode is not explicitly considered due to a larger probability of misidentification. The primary objects misidentified as τ_h in CMS are [4]:

- **Non-τ_h Jets:** These carry highly collimated quark and gluon material that can be reconstructed as a τ_h.
- **Muons:** They carry a charge of 1 and aren't stopped by the ECAL, therefore they can immitate 1-prong τ_hs.
- **Electrons:** Emission of Bremsstrahlung radiation by electrons can lead to photon signatures in the ecal similar to those from a ρ decay.

Dedicated algorithms are needed in order to discriminate against these backgrounds. For analyses conducted during the first part of Run II, including the previous iteration of this analysis SUS-19-002 [23], the canonical algorithm made use of discriminants based on multi-variate analysis (MVA). MVA uses a machine learning (ML) technique named boosted decision trees (BDT) to derive the discriminators [4]. Later in Run II, a new identification algorithm, *DeepTauv2p1*, was developed by the Tau Physics Object Group (POG)

at CMS based on a multi-classifier using deep neural networks. DeepTau takes particle-level inputs such as track quality and particle type, and also high-level input corresponding mainly to those used also in MVA, such as τ_h 4-momentum and charge, decay mode, relative isolation, and energy distribution of strips. Fig.6.4 shows efficiency and misidentification probability comparisons for MVA and Deep Tau for 2016. For all of Run II, superior performance is seen in DeepTau, and it is the the algorithm recomended by the Tau POG to be used in all current analyses.

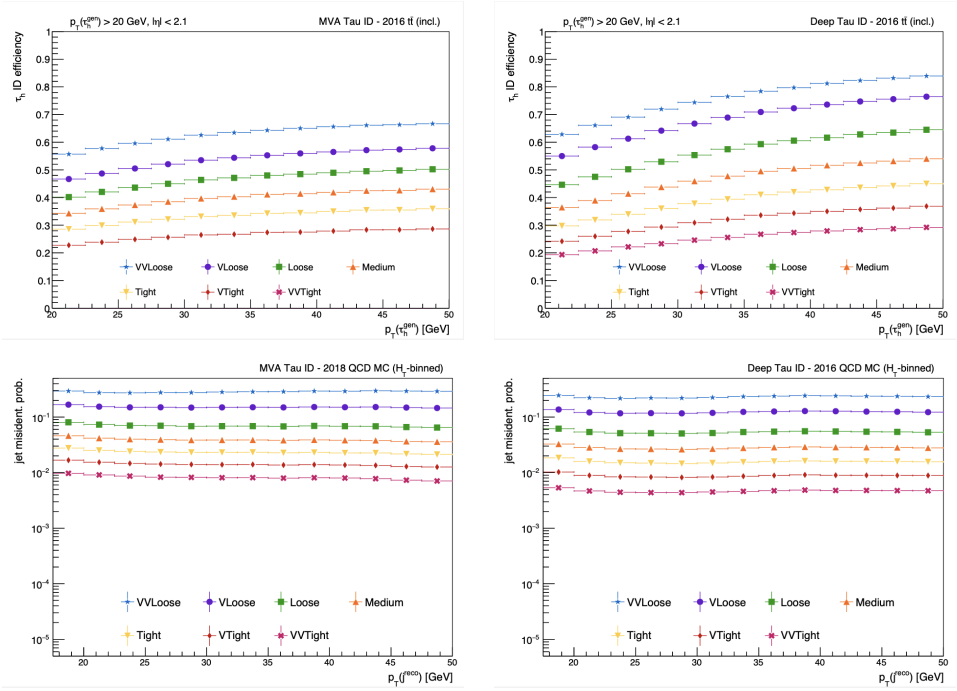

Table 6.4: Comparison of MVA (left) and DeepTau (right) algorithm performance in τ_h ID efficiency (top) and misidentification (bottom).

DeepTau derives discriminators for use in rejecting jets, muons, and electrons reconstructed at τ_h's. Values of discriminators providing particular target efficiencies are classified by *working points* (WP). For anti-jet and anti-electron discriminators, these are: *VVVLoose, VVLoose, VLoose, Loose, Medium, Tight, VTight*, and *VVTight*. For anti-muon discriminators the WPs are: *VLoose, Loose, Medium*, and *Tight*. As one would expect, tighter WPs are associated with lower misidentification probabilities, but also lower efficiencies.The exact discriminator names and working points utilized for τ_h identification in the signal region are shown on Table 6.5

6.4 Electron and Photon Reconstruction and Identification

Electron and photon reconstruction is integrated with the PF technique previously described. The two main detector components associated with their reconstruction are the tracker (mainly aiding electron reconstruc-

Discriminator name	Working point
DeepTau2017v2p1VSjet	VTight
DeepTau2017v2p1VSe	Tight
DeepTau2017v2p1VSmu	Tight

Table 6.5: Deep Tau ID discriminator working points used in this analysis.

tion) and the ECAL. The ECAL is generally the last layer reached where they will deposit all their energy. As an electron moves through the tracker it tends to produce bremsstrahlung radiation (photons), thus it appears in the ECAL as a multiple-particle shower as opposed to well-contained particle. A photon moving through the traker may produce elctro-positron pairs, thus also causing a shower effect in the ECAL as opposed to a clean singular signature. In a similar fashion to hadronic tau reconstruction, algorithms are in place to cluster the various objects into a single one. To correct for radiative effects of electrons seen in the tracker, the Gaussian sum filter (GSM) algorithm is used. The general features of electron and photon reconstruction are discussed here and more detailed information can be found in ref. [70].

The reconstruction of electrons starts with the formation of clusters from crystals in the ECAL that exceed energy thresholds of ~ 80 MeV in the EB, and ~ 300 MeV in the EE. Of the clusters appearing in an event, the one with the highest energy and exceeding a transverse energy $E_T > 1$ GeV is defined as the *seed cluster* (where E_T is defined as $E_T = \sqrt{m^2 + p_T^2}$). In order to account for additional objects due to the effects described above, the seed cluster is then added to other clusters within a defined area around it to form a *super cluster* (SC). When seeds in the tracker compatible with the properties of the SC are found, they are fed to the GSM algorithm. Electron-positron pairs produced photons may also leave tracks not associated with the primary vertex. A dedicated algorithm seeded from these *generic tracks* with $p_T > 2$ GeV in the tracker identifies those actually created by photons producing pairs.

The PF elements that are linked to create blocks are those we have now found via their dedicated algorithms: ECAL clusters, SCs, GSF track, and generic tracks. These are tagged as a general object that at this point doesn't differentiate photons and electrons. These are known as *refined superclusters*, and are subject to a set of requirements. Those with a GSF track are labelled an electron, and those without, a photon. If a refined cluster fails to meet the criteria defined for an electron or photon, its elements may be considered in hadronic material ID algorithms. The reconstruction efficiency resulting from this approach is $> 95\%$ for electrons of $p_T > 20$ GeV.

6.5 Muon Reconstruction and Identification

The CMS detector meets its original goal of providing optimal muon reconstruction at an efficiency of 99%. This is accomplished in part from the high purity of muons present in the muon systems, as other visible material from collisions is absorbed in the calorimeters. There are three types of reconstructed muons given the detector activity they display. *Standalone-muons* are defined as those whose trajectory can be fully reconstructed in the muon system from activity in the relevant chambers. *Tracker muons* are reconstructed IT tracks with p_T >0.5 GeV and with corresponding activity in at least one muon system. *Global muons* are standalone muons that can also be matched to an IT track.

Muon ID working points are also defined to fit use cases with varying focus on low misidentification rates vs. high efficiency. For the muon these are *loose*, *medium*, and *tight*. The performance of muon identification under loose and tight WPs as a function of detector η are displayed in Fig. 6.5. Because the tight muon ID maintains an efficiency between 95-99%, this is the WP considered for the purposes of this analysis. We can further distinguish muons as coming from a primary vertex or a heavy quark decay by implementing isolation criteria. This is done by PF by analyzing summed p_T of hadronic activity in geometrical cones around the reconstructed muon with suitable $\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ values.

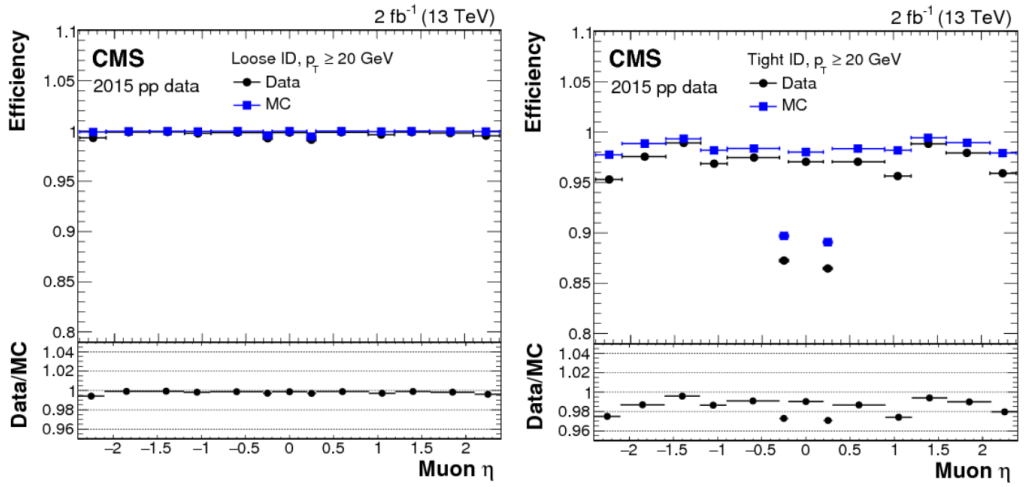

Figure 6.5: Muon ID efficiencies for loose (left) and tight (right) in data and MC as a function of detector η for $p_T\mu > 20$ GeV. [22]

6.6 Missing Transverse Momentum Reconstruction

Missing transverse momentum p_T^{miss}, sometimes interchangeably referred to as missing transverse energy E_T^{miss} (or MET), is a reflection of of the presence of particles that do not interact with the detector. These include neutrinos and potentially new physics object, such as the DM candidates this analysis is concerned

with. It is based on the concept that the momentum carried by the incoming protons is restricted to the z-axis, and therefore the total transverse momentum (in the x-y plane) measured post-collision should also be zero. Because the CMS detector is designed to be hermetic, any deviation from a null p_T must indicate material not visible to the detector. This can be quantified as

$$\vec{p}_{T,PF}^{miss}(raw) = -\sum_{i=1}^{N_{Particles}} \vec{p}_{T,i}, \tag{6.2}$$

which uses PF information as denoted by the T subscript. As discussed in Section 6.2, some PF jet attributes, such as their energy scale (JES) and resolution (JER) must be corrected to account for detector effects. This effect must be propagated to the calculation of p_T^{miss} for accurate reconstruction [66]:

$$\vec{p}_{T,PF}^{miss}(raw) = \sum_{i=1}^{N_{Particles}} \vec{p}_{T,i} - \sum_{j=1}^{N_{PF\,jets}} (\vec{p}_{T,j}^{corr} - \vec{p}_{T,j}) \tag{6.3}$$

6.7 b-Jet Identification

b-jets are jets originating from b quarks. Hadronization of the q-quark is known to produce hadrons with a relatively long lifetime, such that a secondary vertex can often be reconstructed displaced from the main interaction point (primary vertex).

The DeepCSV algorithm is used to identify jets as originating from the hadronization of a b-quark. DeepCSV builds on the CSVv2 algorithm, which combines reconstructed secondary vertex and track-based lifetime information to build a discriminator to distinguish between jets from b-quarks and those from charm or light quarks and gluons. After adding additional variables, extended CVSv2 information is then used as input to train a deep neural network.

The minimum thresholds on these discriminators define loose, medium, and tight operating points with a misidentification probability of $\sim$10%, 1%, and 0.1%, respectively, for an average jet p_T of $\sim$80 GeV. The medium operating point with an efficiency $\sim$ 60% is used in this analysis. [71].

CHAPTER 7

Analysis Strategy

Examples of traditional SUSY and LQ searches at CMS were given in Chapter 3. These entail direct pair production of stau's or chargino-neutralino pairs ($pp \to \tilde{\tau}^+\tilde{\tau}^-$, and $pp \to \tilde{\chi}_1^\pm\tilde{\chi}_2^0$) as can be seen in diagrams in Figures 3.6 and 3.8. The $\tilde{\tau}$'s in the former case will immediately decay to the LSP ($\tilde{\chi}_1^0$) and a SM τ, and the chargino-neutralino in the latter case will undergo cascade decays resulting in multiple LSPs along with SM neutrinos and τ's. The reason these analyses lack sensitivity in the compressed mass spectra regions, which we target to satisfy DM relic density motivations, is that with small Δm between our DM particle and coannihilation partner, i.e. $\Delta m(\tilde{\tau},\tilde{\chi}_1^0) = m_{\tilde{\tau}} - m_{\tilde{\chi}_1^0}$ for the MSSM we consider, result in very low p_T (soft) final decay products. These generally will lie below the p_T threshold of the reconstruction ability of the detector as described in the previous chapter. The same argument applies in processes considered within the LQ-portal DM model in previous CMS searches (Fig. 3.10). The LQ can be produced in pairs at the LHC through gluon-gluon fusion or quark-gluon scattering. Searches have been carried out, though mainly under the assumption of traditional LQ models where the only decay channel is LQ $\to l + q$. As in the SUSY case, these searches predict high production rates, but lack sensitivity in compressed mass scenarios where the $\Delta m(X,LQ) = m_X - m_{LQ}$ we are interested in targeting is also a small value. It is worth noting that in the LQ-portal DM case, the requirement for this ΔM is not necessarily to be small, but to take on particular values that will yield a Δ value satisfying the measured DM relic density values as shown in Fig. 3.3. Our motivation for additionally requiring small $\Delta m(X,LQ)$ is that larger values are the subject of current CMS searches and these analysis are not sensitive in compressed mass regions.

To provide a kinematic boost to our system, we require a high-p_T jet in our selections. One can see from conservation of momentum principles that this will result in a recoil effect and provide additional momentum to our final-state products and improve detectability. High p_T jets in the processes we consider are understood to come from initial state radiation (ISR) of incoming quarks/gluons. The primary production mechanisms at the LHC we consider in this analyses are pictured in Fig. 7.1. For SUSY, though the dominating process is the pictured production of $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$, we also see production of $\tilde{\chi}_1^\pm\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0\tilde{\chi}_2^0$.

The effects of ISR on a topology with pair production of LQ's that goes on to cascading decay can be seen in Fig. 7.2. The x-axis is the E_T^{miss} in two generator-level MC samples, one produced without an ISR jet (blue) and one with (black) an ISR jet required to have $p_T > 100$ GeV. There is a clear increase in E_T^{miss} in the $N(jet) = 1$ case, corresponding to a p_T boost to final state DM particles from the LQ decay. The same effect is propagated to our final state τ's, increasing the likelihood of detector reconstruction and providing

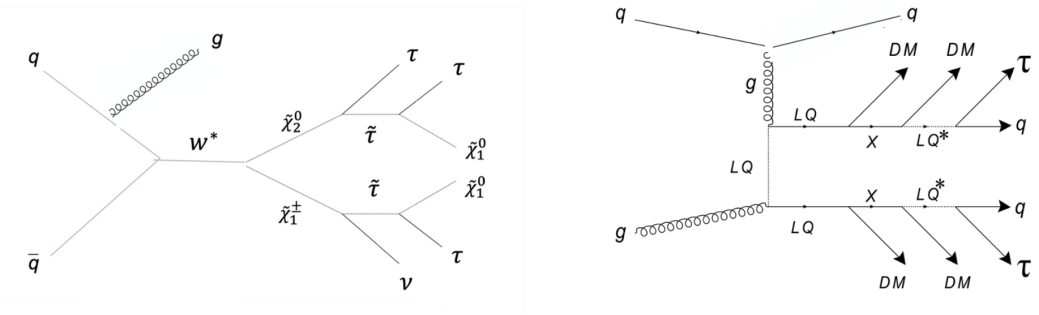

Figure 7.1: Topology of primary production mechanisms in the SUSY (left) and LQ-portal DM (right) models considered in this work.

sensitivity to our study. This is the main motivator for the requirement of a high-p_T ISR jet and large E_T^{miss}, in this analysis. Because of the larger branching compared to decays with lighter leptons, our study sees increased sensitivity when we require our τ decay hadronically (τ_h). Though τ_h's receive the kinematic boost from ISR jet, these remain relatively soft, thus we require only one τ with low p_T. This is quantified and plotted in Fig. 7.3, where we see very low acceptance of any τ_hs above the first for various mass scenarios in the SUSY case. A similar results is seen the LQ case. This will be quantified in what follows of this chapter. The full set of selections comprising our signal region (SR) is listed in Table 7.1, and the optimization process arriving to these values will be quantified in the chapter to follow.

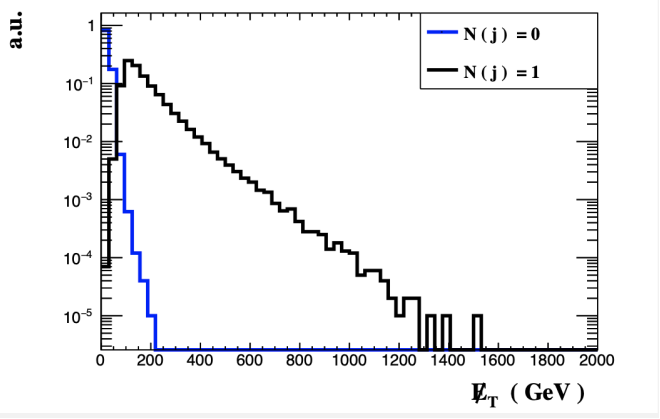

Figure 7.2: Generator-level E_T^{miss} of final state DM particles produced via LQ decays in 1-jet and 0-jet scenarios

As mentioned, the choice of free parameters in this analysis is motivated by DM relic density considerations, B-meson anomaly considerations, and also by analysis of phase space lacking sensitivity in current searches. In the SUSY case we consider a simplified wino-bino MSSM category with $\tilde{\tau}$-dominated decays. Thus we have a zino-like $\tilde{\chi}_1^0$, wino-like $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ with the degenerate masses, and a $\tilde{\tau}$ mass

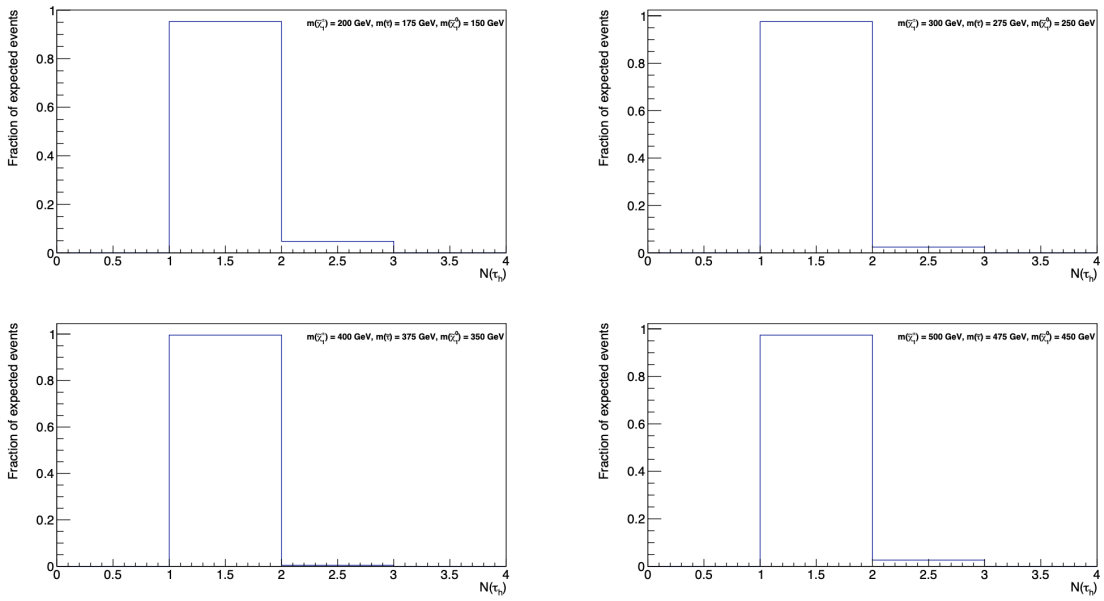

Figure 7.3: Number of signal MC τ_hs in SR for various mass scenarios and a $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV [23]

Object	Selection cuts
Trigger	HLT_PFMET120_PFMHT120_IDTight
τ_h	$N(\tau_h) = 1$, 20< $p_T(\tau_h)$ <40 GeV, $\lvert\eta(\tau_h)\rvert < 2.1$, DeepTau Vtight, anti-$e/\mu$ discr., prong: 1hps
e veto	$N(e) = 0$, $p_T(e)$ >10 GeV, $\lvert\eta(e)\rvert < 2.1$, I_{rel} <0.15, Medium CBID
μ veto	$N(\mu) = 0$, $p_T(\mu)$ >10 GeV, $\lvert\eta(\mu)\rvert < 2.1$, I_{rel} <0.15, Tight PF ID
Jets	≥ 1 with $p_T \geq 30$ GeV, $\lvert\eta\rvert < 2.4$, "Tight" jet ID
Leading Jet	$p_T > 100$ GeV
b-jet veto	N(b) =0, p_T(b) >30 GeV, $\lvert\eta(\text{b})\rvert < 2.4$, medium DeepCSV WP
E_T^{miss}	E_T^{miss} >230 GeV
QCD Rejection	$\lvert\Delta\phi(j, E_T^{miss})\rvert_{\text{min}} > 0.7$

Table 7.1: Signal Region Selections

$m_{\tilde{\tau}} = 0.5(m_{\tilde{\chi}_1^{\pm}} + m_{\tilde{\chi}_1^0})$. Similarly, for the LQ-portal DM model we consider $m_X = 0.5(m_{LQ} + m_{DM})$, and we assume maximal branching fraction to DM and X in the LQ$\rightarrow$ DM+X decay (as usual in CMS searches).

The big-picture approach in this analysis is of course to compare prediction to data for evidence of new physics. The role of our SM prediction is filled by large samples of MC events produced via Feynman rules under the assumption of SM physics. In the end, SM predictions via MC are statistically compared to CMS data for deviations potentially representing new physics, or limits are set if no excess is observed. The distribution, or fit variable, where we perform a binned likelihood fit to look for potential signs of signal events is the transverse mass between E_T^{miss} and our visible τ_h:

$$m_T = \sqrt{2E_T^{miss} p_T(\tau_h)(1 - cos\Delta\phi(E_T^{miss}, \tau_h))}. \tag{7.1}$$

Though the mass of the original $\tilde{\chi}_2^0$, $\tilde{\chi}_1^\pm$ or LQ aren't fully reconstructable due to E_T^{miss} and soft objects that we expect to not be reconstructed in the detector, we still expect a large presence of signal in the tail of the $m_T(E_T^{miss}, \tau_h)$ distribution, as well as good background reduction in this range, thus enhancing the sensitivity of our study.

Before a physically accurate and statistically sound comparison can be made, both our MC and data must undergo a series of validations and corrections. In the MC case, this is due in part to the perturbative nature of QFT calculations and the numerical implementation of these via computational methods and algorithms. In the case of data, as mentioned in the previous chapter, a number of corrections and calibrations are needed to account for detector effects in reconstructing event objects; Some of these are due to expected, standard operation of the detector, others from unexpected issues such as the APV chip pre-amplifier issue in 2016. The required corrections to data have been mostly covered in the previous chapter, and further necessary treatments will be covered as needed moving forward. For MC, some corrections are derived by CMS groups specializing in the relevant physics objects and have a detailed, required prescription. Each of the studies resulting in the prescribed corrections and treatments are done under particular conditions and particular topologies that cannot be expected to cover the full range of scenarios studied in each analysis. Thus, after applying what is required and recommended by the expert groups, a dedicated background estimation is conducted to study their validity in our analysis, and to derive further corrections where necessary to achieve a high-level of confidence in the final result. This analysis aims to extend the work done in the CMS search SUS-19-002 [23] and include a new interpretation correspinding to the LQ-portal DM model. Though most of the work is novel, including the addition of 78% more CMS data, updated detector elements, and updated reconstruction and identification algorithms, some of the work done in SUS-19-002 remains relevant to this work and will be included here and cited where applicable.

Our approach to background (BG) estimation uses both data and MC. The strategy is to construct dedicated, BG-enriched control regions (CR) orthogonal to our SR (with minimal contamination from signal events), each with the purpose of studying a particular aspect of our SR selections to ensure any mismodeling is accounted for. In these CRs, we aim for a high purity of the targeted BG. After subtracting MC from non-targeted BGs, this generally allows for extraction of scale factors (SFs) to be applied to our BG MC in the SR to yield a correct estimation. This method can can be expressed as follow. The number of events expected for signal and BG are a result of the total number of events of the relevant process produced at the LHC times the efficiencies ε associated with each SR selection cut:

$$N_{SR}^{BG,data} = \sigma_{BG} \times \mathscr{L}_{int} \times \varepsilon_{\tau_h ID}^{true} \times \varepsilon_{MET\,cut}^{true} \times \varepsilon_{ISR}^{true} \times \varepsilon_{other}^{true} \tag{7.2}$$

$$N_{SR}^{BG,MC} = \sigma_{BG} \times \mathscr{L}_{int} \times \varepsilon_{\tau_h ID}^{MC} \times \varepsilon_{MET cut}^{MC} \times \varepsilon_{ISR}^{MC} \times \varepsilon_{other}^{MC} \tag{7.3}$$

Where ε_{other}^{MC} accounts for efficiencies of all other SR selections. And, in general, for a given CR (assuming application of relevant recommended SFs):

$$\frac{N_{SR}^{BG,data}}{N_{SR}^{BG,MC}} = \frac{N_{CR}^{BG,data}}{N_{CR}^{BG,MC}} = SF_{CR} \rightarrow N_{SR}^{BG,data} = SF_{CR} \times N_{SR}^{MC} \tag{7.4}$$

To minimize uncertainties due to the non-targeted, subtracted backgrounds, we strategically order our background estimation approach to extract corrections and SFs from highest purity CRs (where uncertainties due to non-targeted BG are negligible) first and apply these to subsequent CRs as follows. (*i*) We begin with a 97% pure Drell-Yan ($Z(\rightarrow \mu\mu)$) CR to extract corrections necessary for mismodeling in processes arising from boosted *Z*/*W* bosons, as well k-factor corrections to cross-sections necessary for particular DY+jets (Drell-Yan processes with associated jets) and W+jets (processes where a W boson is produced in association with jets) MC samples; (*ii*) We then construct a $t\bar{t}$ (processes producing a top and antitop pair) CR that is unaffected by these issues and is 97% pure in $t\bar{t}$ + single top MC to extract data-to-MC SFs to apply in SR and subsequent CRs (we correct for both these backgrounds similary since mismodeling of both is known to be related to boosted top quarks); (*iii*) The *Z*/*W* boost weights, HT-binned W+jets and DY+jets k-factor corrections, and $t\bar{t}$/single top data-to-MC SFs are applied in a W+Jets CR that is 84% pure in W+Jets, but has 12% composition of $t\bar{t}$, single top, and DY+jets MC, which has been appropriately corrected; (*iv*) Additional studies regarding τ_h ID are conducted in a 95% pure $Z(\rightarrow \tau\tau)$ CR; (*v*) QCD studies are conducted in a data-driven manner, where all MC in use has been corrected and scaled as suggested by all previously mentioned studies. It is also worth noting that validation studies of diboson MC in a dedicated CR suggest that MC is generally well modeled and not in need of additional scaling. This approach gives us a high degree of confidence that the conclusions of our background estimation studies hold, and that the standard methods of accounting for systematic uncertainty in the corrections and SFs acquired are appropriate.

The BG estimation studies of this analysis and their purposes are broken-down in what follows. Primary backgrounds to the signal region are W+jets, QCD multijet (QCD), and $t\bar{t}$:

Drell-Yan (with μs): This CR focuses its attention on studying the effects of the ISR jet selection in our SR. To accomplish this, we take advantage of lepton universality by constructing a DY+jets, namely $Z(\rightarrow \mu\mu)$ + ISR jet, CR. The motivations behind studying this CR are as follow:

- In the eyes of the SM, the physics of this process should be similar to that of an equivalent process with hadronic taus.
- The reconstruction of muons in the CMS detector is far cleaner, with generally lower uncertainties and

unwanted effects which allows us to isolate ISR effects.

- There is no real E_T^{miss} produced in this process, reducing the contributing effects to the efficiency of this selection even more.

TTbar ($t\bar{t}$): After analyzing a full set of MC in our SR, we determine that this background composes $\sim 13\%$ of our SR BG's. We study its modeling in a dataset that is orthogonal to SR while retaining its main important attributes. We obtain a $t\bar{t}$-enriched CR by taking our SR selections and changing our $N(b-jet)$ selection from requiring exactly 0, to exactly 2 b-jets. We use the data-to-MC SF to correct the $t\bar{t}$ prediction in SR MC.

W+Jets: The purpose of the W+Jets CR are the following:

- As the source of the mismodeling in events with boosted Z bosons (as extracted in the DY+jets CR with μ's) is the same as that present in boosted W bosons, we use this CR to validate the Z/W boost weights extracted in the DY CR.
- A similar argument applies for the necessary corrections to the cross-sections of W+Jets and DY+jets samples which are binned in HT (the sum scalar sum of all jet p_T in an event).
- W+jets events contain real E_T^{miss}, thus, once ISR corrections are applied, the effects of E_T^{miss} are the isolated and corrections related to it can be extracted.
- Similarly to the $t\bar{t}$ case, we analyze the amount of W+jets BG by running a full set of MC in our SR. W+jets is by far the largest BG in SR. Our W+jets CR allows for a detailed study of the expected $m_T(E_T^{miss}, \tau_h)$ distribution in SR and any needed corrections.

Drell-Yan (with τ_hs): Since the main lepton in our final state is a τ_h, we construct a dedicated $Z(\rightarrow \tau\tau \rightarrow \tau_h\tau_h)$ + ISR CR to understand the effects of τ_h ID in a topology similar to our SR. These effects are nicely isolated in this CR after the corrections of ISR effects and k-factors derived from previous CR's, as well as the lack of real E_T^{miss}.

Drell-Yan (with one τ_h and one μ): Due to the lack of a trigger efficient in the low $p_T(\tau_h)$ range, our τ_h ID studies via the Drell-Yan CR with two τ_h's is conducted with $p_T(\tau_h) > 60$ GeV. To validate that the conclusions we draw from these studies are applicable to τ_h"s in our SR, we construct a CR where instead of two hadronically decaying τs, one of the τs decays leptonically into a μ, ν_τ, and ν_μ. We then require the τ_h to fall in the range of our SR τ_hs ($p_T = [20, 40]$ GeV), and instead require events to fire a single-muon trigger that is efficient in our selected $p_T(\mu)$.

QCD: QCD background sneaks into our SR when QCD multijet processes common in proton-proton collisions produce a jet with conditions that cause it to be identified at a τ_h (commonly referred to as just a $fake$). After seeing a QCD MC yield of $\sim 20\%$ in an initial SR analysis, we conclude that QCD is an important BG to understand. QCD produced via MC generators lacks reliability. This is because the high-order processes that are not uncommon to contribute in collisions at CMS are computationally difficult. Because of this, we used a purely data-driven technique known as a $fake\text{-}factor$ method to estimate QCD in the signal region. This entails extracting the shape of the m_T distribution from a CR similar to SR, except with flipped τ_h isolation requirements. We then extract transfer factors from separate CRs designed to emulate the properties of QCD jets faking τ_h's in nominal and inverted τ_h ID conditions to be applied to our m_T distribution from the flipped isolation CR.

CHAPTER 8

Trigger Studies

Event triggering in data in CMS was described in section 5.6. Beyond online and offline triggering in the detector, effects of triggers must also be simulated to obtain MC that accurately describes the SM background we expect to see in the detector. The trigger selection directly impacts the event selection and subsequent data analysis, thus we conduct a dedicated study. Our SR selections generally consist of exactly one soft τ_h, a high-p_T jet from ISR, and substantial E_{miss}. The existing τ_h triggers at CMS exhibit inefficiency in the low-p_T(τ_h range relevant to this analysis. Therefore, an alternative approach is adopted, instead looking to the large amount of E_T^{miss} and ISR activity in SR to trigger our events of interest. We use the $HLT\_PFMET120\_PFMHT120\_IDTight$ trigger. We conduct efficiency studies on our trigger with the goals of ensuring high efficiency in selecting events in our SR, and to understand how the behavior differs in data versus MC. To accomplish this, we make use of a single-muon trigger which is known to perform optimally in its domain, $HLT\_IsoMu24$. We construct a high-purity W+Jets CR with one high-quality muon, which we will trigger on. Since there is no E_T^{miss} requirement in this CR and the muon trigger collects events based solely on the muon, we can look at the full E_T^{miss} spectrum in this CR. We then compare this E_T^{miss} distribution to that in the same CR, but of events that have also passed our E_T^{miss} trigger to study the thresholds where our E_T^{miss} trigger becomes fully efficient in data and MC. The efficency is found as:

$$\varepsilon_{E_T^{miss}trig} = \frac{N(pass\ CR\ cuts\ \&\ pass\ E_T^{miss}\ trig.\ \&\ pass\ \mu\ trig.)}{N(pass\ CR\ cuts\ \&\ pass\ \mu\ trig.)} \tag{8.1}$$

where N denotes the total number of events passing the given requirements. One can see that this method effectively factors out the muon trigger effects and allows us to study our E_T^{miss} trigger efficiency accurately. The full W+Jets CR seletion used are displayed in Table 8.1, and the resulting efficiency as a function of E_T^{miss} can be seen in our results in Fig. 8.1

The resulting efficiency curves show that our trigger reaches its full efficiency of $\sim 97\%$ at $E_T^{miss} \sim 230$ GeV and that the data and MC turn on curves see good agreement at $E_T^{miss} \geq 200$ GeV, as seen in the middle ratio pad of our results. This motivates our SR selection of $E_T^{miss} \geq 230$ GeV. To gauge the systematic uncertainty in our trigger efficiencies, we fit a function to our data and analyze the error on the fit. The function used is

$$\frac{\mathbf{a}}{2}\left(1+erf\left(\frac{x-\mathbf{b}}{\mathbf{c}\sqrt{2}}\right)\right). \tag{8.2}$$

Trigger(denominator)	HLT_IsoMu24_v
$N(\mu)$	≥ 1
$p_T(\mu)$	$[20,40]GeV$
$\lvert\eta(\mu)\rvert$	< 2.1
$ID(\mu)$	Tight
$N(j)$	≥ 1
$p_T^{lead}(j)$	> 100 GeV
$\lvert\eta^{lead}(j)\rvert$	< 2.4
$N(b-jets)$	0
Overlaps removal	$\Delta R > 0.4$
Trigger(numerator)	HLT_PFMET120_PFMHT120_IDTight

Table 8.1: Selection criteria for W+Jets CR used for trigger studies.

where **a**, **b**, and **c** are the fit parameters, and erf is the standard Gaussian error function. Though the resulting errors on the fit where our E_T^{miss} selection lies are $< 1\%$, we see in Figure 8.1 an ε_{data}-to-ε_{MC} ratio of 0.974, so we assign the deviation from unity of this ratio (2.6%) as our systematic uncertainty on the trigger efficiency.

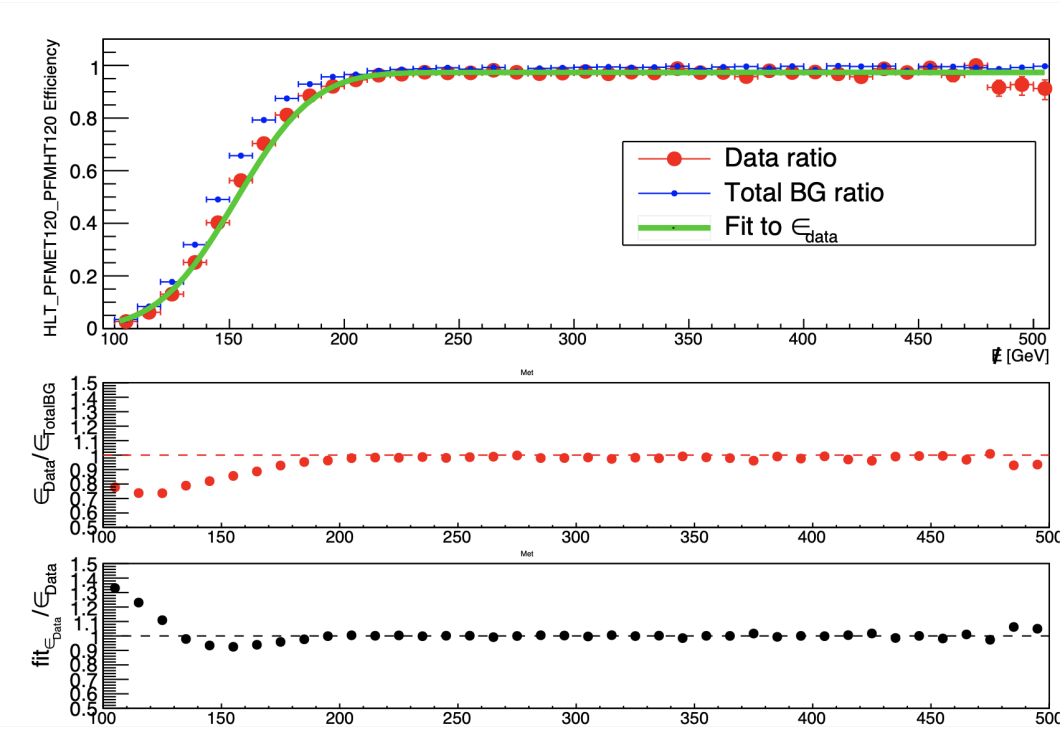

Figure 8.1: HLT_PFMET120_PFMHT120_IDTight trigger efficiency for data and total BG MC (top). Ratio of efficiency in data to efficiency in total BG MC (middle). Ratio of fit to data to efficiency in data (bottom)

CHAPTER 9

Signal Optimization

To ensure optimal discovery potential in our analysis, it's necessary to identify key variable windows, in the form of our cuts (selection criteria), that yield the best balance of enhanced signal and reduced background. This is done by varying our cuts in the windows allowed by important criteria, such as trigger efficiencies and object reconstruction thresholds, and evaluating the following figure of merit:

$$\frac{S}{\sqrt{S+B}} \tag{9.1}$$

where S is our total signal yield and B is our total BG yield passing the given set of cuts. This figure of merit is often referred to as the *significance*. $\sqrt{S+B}$ represents the statistical uncertainty on our yields, therefore this figure of merit provides a (rough) measure of how many standard deviations away from SM background we expect our signal yield to be. The resulting optimal cuts used in this analysis are partially a legacy of those studied in SUS-19-002 [23], and the corresponding results are for the first two variables, $p_T(\tau_h)$ and $p_T(jet^{lead})$, presented here. The base criteria upon which the optimization process builds upon is summarized in Table 9.1.

Selection Criteria
$N(\tau_h) \geq 1$
$p_T(\tau_h) > 20\,\text{GeV}$
$\lvert\eta(\tau_h)\rvert < 2.1$
$N(j) \geq 1$
$P_{\text{lead}}(j) > 100\,\text{GeV}$
$\lvert\eta(\text{lead } j)\rvert < 2.4$
$N(b\text{ - jets}) = 0$
$E_{\text{miss}} > 230\,\text{GeV}$
Overlaps removal $\sim \Delta R(\tau_h, j) > 0.3$

Table 9.1: Optimization Study Base Selections

The optimization of the $p_T(\tau_h)$ selection is approached in two different ways, first by investigating a lower threshold only which is varied in increments of 10 GeV from 20-60 GeV. The results are shown in Fig. 9.1 for three $M(\tilde{\chi}_1^0)$ scenarios. The next approach is to look at performance under a lower *and* upper threshold. The lower threshold is held at 20 GeV due to τ_h detector reconstruction considerations, and the upper threshold is varied from 30-70 GeV again in increments of 10 GeV. The results are displayed in Fig. 9.2. The optimal significance for the $p_T(\tau_h)$ selection is found with a lower threshold of 20 GeV and an upper threshold of 40 GeV. Thus motivating and justifying the corresponding selection in SR.

A lower p_T threshold is also studies for the ISR jet selection. A range from 100 to 150 GeV is studied with signal corresponding to $m(\tilde{\chi}_1^0) = 270$ GeV, and from 110 to 160 GeV for signal with $m(\tilde{\chi}_1^0) = 360$ GeV. The results are shown in Fig. 9.3. Though there is a minute increase in significance for $p_T(jet^{lead} = 110$, 120 GeV, the lower threshold is kept at 100 GeV to improve the efficiency of the selection and minimize overall statistical uncertainties.

To validate and enhance the optimization studied in SUS-19-002 and discussed thus far, we further study selections related to our τ_h candidates to ensure optimal significance in SR. We first look at the impact of looking for an additional τ in our signal region in hopes that another may be reconstructed and further enhance signal and/or reduce background. We first consider a scenario where the τ with second largest p_T goes on to decay hadronically and is reconstructed in our detector (total of $2\tau_h$'s). We then consider the possibility that the second τ went on to decay leptonically, producing a muon ($\tau_h + \mu$). Because the p_T of a reconstructed electron would be similar to that of the μ, we use the μ case to quantify both scenarios and omit an explicit electron requirement. The results of these studies are presented in Table 9.2. The significance is at least an order of magnitude higher when selecting exactly 1 τ_h and no other leptons, thus motivating this selection in our SR. This is well expected given the low number of additional τ_h's present in our signal region, as seen in Fig. 7.3.

As discussed in Section 6.3, τ_h decays happen in various ways and leave different signatures in the detector that can most often be differentiated upon reconstruction. Our nominal SR selections include τ_h's reconstructed as having exactly one charged track in the seeded jet (1-Prong, or 1HPS). We study the effects on signal significance if we add τ_h's reconstructed with three charged tracks in their corresponding jet (3-prong τ_h's) to our current selection, and also the scenario where we *only* consider 3-prong τ_h's . The results are seen in the 9.2. The significance is considerably lower for every scenario considered, with the exception of the final state with 1- and 3-prong τ_h's in our selection, where the significance remains about the same as our nominal selection. We ultimately keep our nominal 1-prong τ_h requirement because admitting 3-prong τ_h's is known to increase fake τ_h's coming from QCD processes, adding higher uncertainties to our SR.

		$m(\tilde{\chi}_1^{\pm}) = 200, m(\tilde{\chi}_1^0) = 175$	$m(\tilde{\chi}_1^{\pm}) = 300, m(\tilde{\chi}_1^0) = 275$
1 τ_h (1-prong)	S	1668 ± 23	504 ± 19
	B	36067 ± 126	36067 ± 126
	$\mathbf{S}\sqrt{\mathbf{S}+\mathbf{B}}$	**7.17**	**2.19**
2 τ_h (1-prong)	S	2.3 ± 0.9	0.9 ± 0.8
	B	49 ± 5	49 ± 5
	$\mathbf{S}/\sqrt{\mathbf{S}+\mathbf{B}}$	**0.26**	**0.1**
1 τ_h (1-prong) + 1 μ	S	20 ± 2.5	6.7 ± 2.2
	B	1371 ± 14	1371 ± 14
	$\mathbf{S}\sqrt{\mathbf{S}+\mathbf{B}}$	**0.5**	**0.17**
1 τ_h (1 or 3-prong inclusive)	S	1852 ± 24	559 ± 20
	B	42662 ± 136	42662 ± 136
	$\mathbf{S}\sqrt{\mathbf{S}+\mathbf{B}}$	**7.34**	**2.24**
1 τ_h (3-prong exclusive)	S	184 ± 8	55 ± 6
	B	6595 ± 66	6595 ± 66
	$\mathbf{S}\sqrt{\mathbf{S}+\mathbf{B}}$	**1.75**	**0.52**

Table 9.2: Significance values for additional lepton and different τ_h decay mode scenarios.

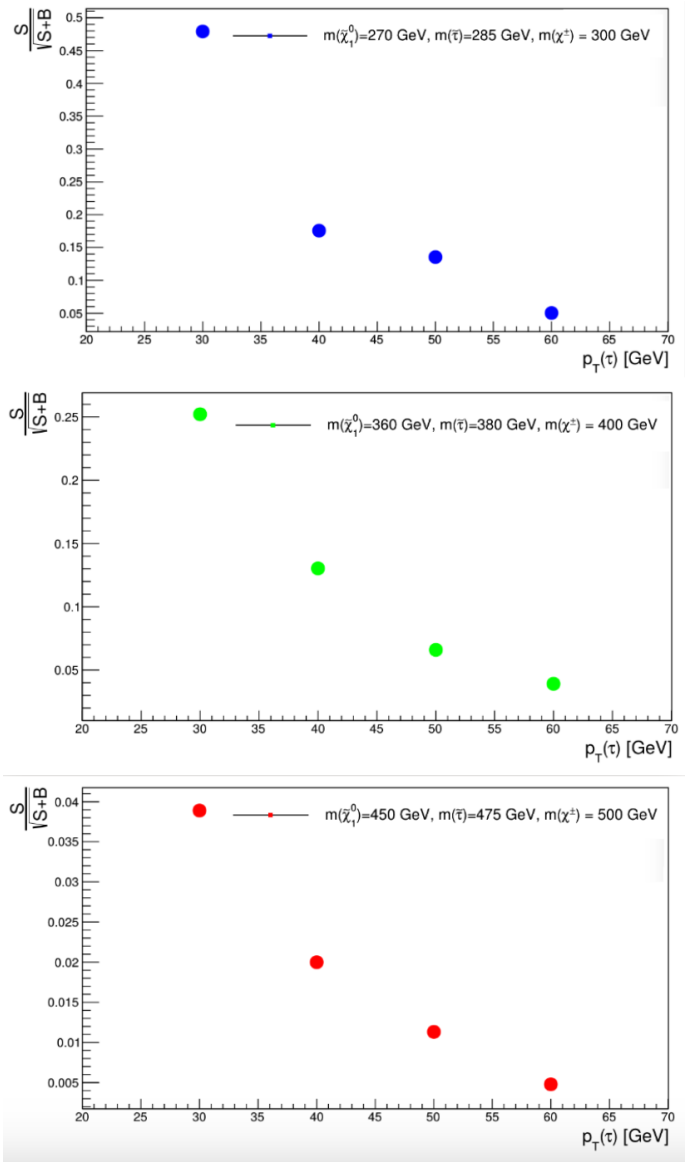

Figure 9.1: $p_T(\tau_h)$ lower threshold optimization for $m(\tilde{\chi}_1^0) = 270, 360, 450$ GeV [23]

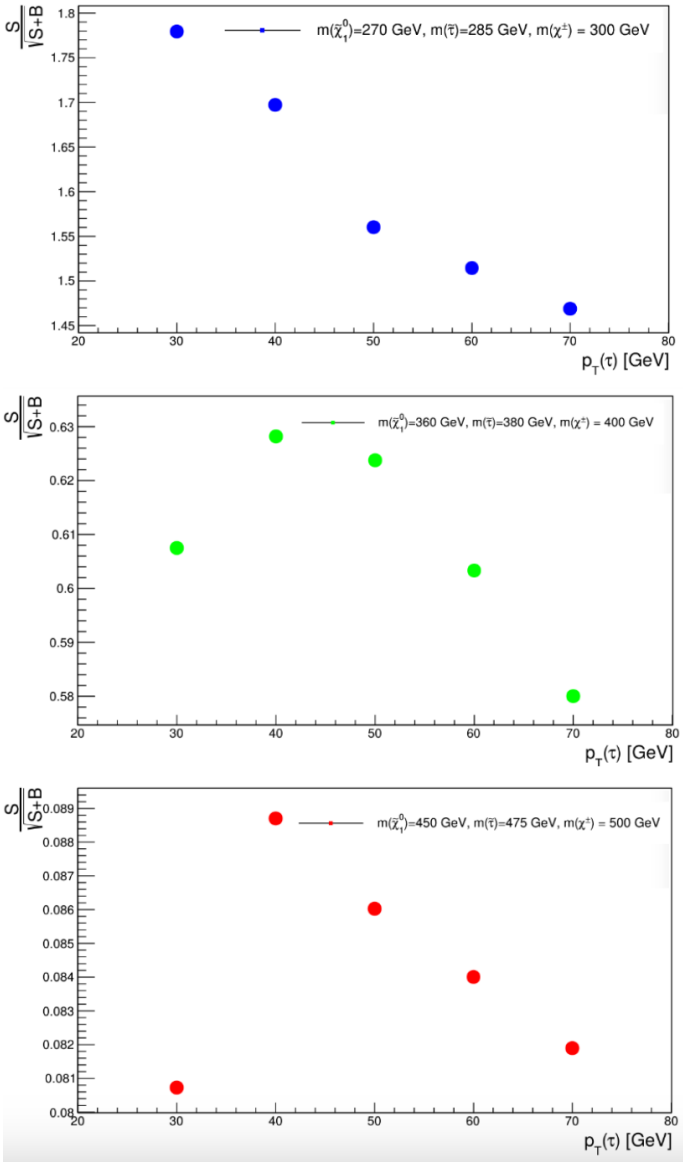

Figure 9.2: $p_T(\tau_h)$ upper threshold optimization with a fixed lower threshold at 20 GeV for $m(\tilde{\chi}_1^0) = 270, 360, 450$ GeV [23]

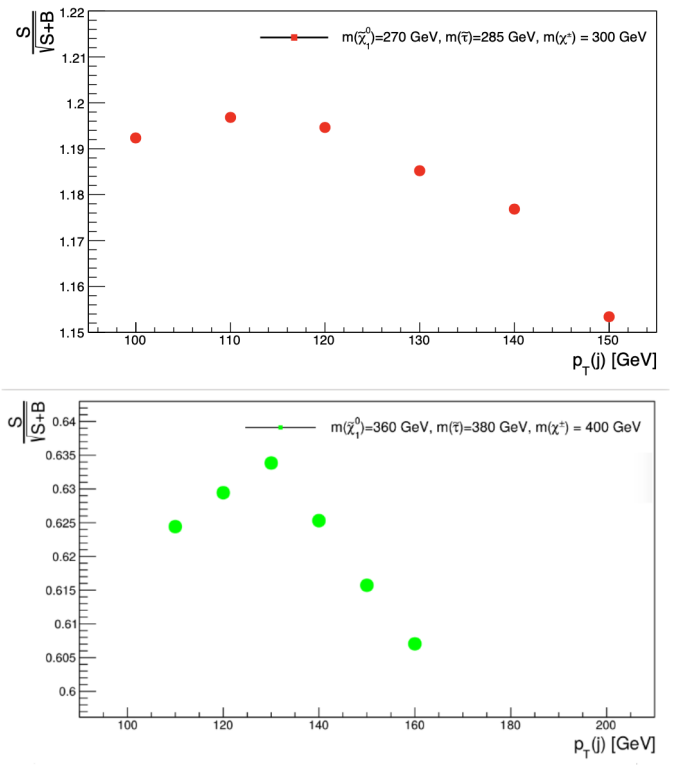

Figure 9.3: $p_T(jet^{lead})$ minimal threshold optimization for $m(\tilde{\chi}_1^0) = 270, 360$ GeV [23]

CHAPTER 10

Data and Monte Carlo Samples

10.1 Data Samples

This analysis is based on samples centrally processes by CMS in a campaign labelled Ultra Legacy (UL). UL is a reprocessing of the "pre-legacy" version and it features better calibration of detector elements, as well as fixes related to unwanted detector effects. These are in the NanoAODv8 format for both data and simulation, which is a compressed version of more complete collection, meant to be lighter and more efficient to use. The samples make use of the data collected by the CMS detector for proton-proton collisions at 13 TeV in 2016, 2017, and 2018 with corresponding integrated luminosities of 36.33 fb^{-1}, 41.48 fb^{-1} and 59.83 fb^{-1}, respectively. A new feature of the UL campaign is that the data for 2016 is split into APV and non-APV (also labelled pre-VFP and post-VFP) era, where the year was processed separately to correct for the APV chip issues described in Section 5.2.2. These eras carry integrated luminosities of 19.52 fb^{-1} 16.81 fb^{-1}, respectively.

The samples are grouped into primary datasets (PD) typically dictated by the object used to trigger the events. In order to reject run ranges or events known to have faulty data after quality monitoring certification, official files in the JSON format are used. These are listed in Table 10.1. The MET PD's are used for the SR and are shown in Tab. 10.2. In additon, muon (Tab. 10.3) and tau (Tab. 10.4 PDs are utilized in order to understand the modeling of different selection efficiencies as well as perform background estimations.

Era	JSON file	Luminosity (fb^{-1})
Legay 2016	Cert_271036-284044_13TeV_Legacy2016_Collisions16_JSON.txt	35.92
Legacy 2017	Cert_294927-306462_13TeV_UL2017_Collisions17_GoldenJSON.txt	41.53
Legacy 2018	Cert_314472-325175_13TeV_Legacy2018_Collisions18_JSON.txt	59.74

Table 10.1: Good run and luminosity section files for Run II collision data.

10.2 Monte Carlo Samples

The MC event samples used across CMS analyses are produced via the MC generators MadGraph, PYTHIA, and POWHEG [72]. The production begins by generating parton-level interactions resulting in outgoing SM leptons, bosons and quarks. Hadronization of the final state quarks, and sometimes lepton decays, are then handled by "showering" algorithms. This step is usually done with PYTHIA. Finally, detector effects must be simulated. This is handled by internal CMS software based on GEANT4 and PYTHIA frameworks. MC samples used for CMS analyses are required to be centrally produced by the CMS MC group. The samples

Era	Physics sample	Official CMS datasets
2016 (pre-VFP)	Run 2016Bv1	/MET/Run2016B-ver1_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016Bv2	/MET/Run2016B-ver2_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016C	/MET/Run2016C-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016D	/MET/Run2016D-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016E	/MET/Run2016E-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016F	/MET/Run2016F-HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2016 (post-VFP	Run 2016F	/MET/Run2016F-UL2016_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2017G	/MET/Run2016G-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017H	/MET/Run2016H-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2017	Run 2017B	/MET/Run2017B-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017C	/MET/Run2017C-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017D	/MET/Run2017D-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017E	/MET/Run2017E-UL2017_MiniAODv1_NanoAODv2-v3/NANOAOD
	Run 2017F	/MET/Run2017F-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
2018	Run 2018A	/MET/Run2018A-UL2018_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2018B	/MET/Run2018B-UL2018_MiniAODv1_NanoAODv2-v5/NANOAOD
	Run 2018C	/MET/Run2018C-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2018D	/MET/Run2018D-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD

Table 10.2: Run II collision data samples: MET primary datasets (UL NanoAODv8)

we used have been produced at LO with NLO generators (MadGraph and POWHEG), then the yields have been scaled with NLO, or NNLO cross-sections where available. The W+Jets and DY+jets samples used are binned in HT, as these HT-binned samples provide better statistics in the boosted, ISR topology we consider. Because these samples are only available for HT $>$ 100 GeV, the low-HT contribution is taken from inclusive DY+jets and W+Jets samples and then stitched together with the HT-binned samples via an HT filtering method. For auxiliary studies in this analysis, QCD MC samples were occasionally used where the importance or contribution of QCD was considered to be minimal. When in use, these were also HT-binned samples due to statistical considerations similar to the DY+jets and W+jets case.

In addition to those mentioned up to this point, further corrections to MC must be done in order to achieve a proper comparison with data. These include shortcomings in the production due to computational considerations and the state of the detector during particular runs. A general description of some of these additional corrections is listed here.

- **Pileup Weights**: Pileup distributions in MC are assigned to best match what is seen in data, though varying detector conditions and statistical limitations at the time pileup is assigned to MC require us to assign weights to account for differences. These weights are found as:

$$w_{PU}(n) = \frac{P_{data}(n)}{P_{MC}(n)} \tag{10.1}$$

  where $w_{PU}(n)$ are our weights, and $P_{data}(n)$ and $P_{MC}(n)$ are pileup districutions in data and MC.

- **"MET" Filters**: Non-targeted activity such as detector noise and cosmic rays can induce large E_T^{miss}

Era	Physics sample	Official CMS datasets
2016 (pre-VFP)	Run 2016Bv1	/SingleMuon/Run2016B-ver1_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016Bv2	/SingleMuon/Run2016B-ver2_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016C	/SingleMuon/Run2016C-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016D	/SingleMuon/Run2016D-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016E	/SingleMuon/Run2016E-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016F	/SingleMuon/Run2016F-HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2016 (post-VFP)	Run 2016F	/SingleMuon/Run2016F-UL2016_MiniAODv1_NanoAODv2-v4/NANOAOD
	Run 2016G	/SingleMuon/Run2016G-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016H	/SingleMuon/Run2016H-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2017	Run 2017B	/SingleMuon/Run2017B-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017C	/SingleMuon/Run2017C-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017D	/SingleMuon/Run2017D-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017E	/SingleMuon/Run2017E-UL2017_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2017F	/SingleMuon/Run2017F-UL2017_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2017G	/SingleMuon/Run2017G-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
2018	Run 2018A	/SingleMuon/Run2018A-UL2018_MiniAODv1_NanoAODv2-v3/NANOAOD
	Run 2018B	/SingleMuon/Run2018B-UL2018_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2018C	/SingleMuon/Run2018C-UL2018_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2018D	/SingleMuon/Run2018D-UL2018_MiniAODv1_NanoAODv2-v2/NANOAOD

Table 10.3: Run II collision data samples: muon primary datasets (UL NanoAODv8).

to be seen in the detector. This referred to as $false\ MET$ and CMS has filters in place to remove these events.

- **Jet Energy Scale and Jet Resolution Corrections**: Jet-by-jet application of corrections in both data and MC to account for detector response in jet reconstruction. See Section 6.2 for more detail.

- τ_h **ID Scale Factors**: Known corrections associated with τ_h ID must be applied for the anti-jet, anti-muon, and anti-electron case. We implement the recipe with the set of recommended τ_h SF's provided by the Tau POG [4]. As mentioned in the analysis strategy, we also conduct further studies to understand τ_h ID as it applies specifically to our analysis.

- **b-tagging Scale Factors**: As in the τ_h ID case, there is an associated recommended set of SF's to be applied by the corresponding CMS expert group, the JetMET group in this case. We apply the recommnded SFs corresponding to our b-tagging WP [71].

- **L1-Prefiring Weights**: In 2016 and 2017 runs, wear in the ECAL crystals due to radiation caused triggering effects with incorrect assignments of bunch crossings to events, known as "pre-firing." The effect of pre-firing from the ECAL resulted in a loss of efficiency in the $2.0 < |\eta| < 3.0$ region which must be accounted for in MC. We implement the EGamma POG recipe in MC in the two affected years [73].

- **2018 HEM veto**: There was a loss of power to the hadronic endcap calorimeter (HEM) in 2018. To reflect this detector condition in 2018 MC, we veto jets in the affected area ($-3 \leq \eta \leq -1.65$ and

Era	Physics sample	Official CMS datasets
2016 (pre-VFP)	Run 2016Bv1	/Tau/Run2016B-ver1_HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016Bv2	/Tau/Run2016B-ver2_HIPM_UL2016_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2016C	/Tau/Run2016C-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016D	/Tau/Run2016D-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016E	/Tau/Run2016E-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016F	/Tau/Run2016F-HIPM_UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2016 (post-VFP)	Run 2016F	/Tau/Run2016F-UL2016_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2016G	/Tau/Run2016G-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2016H	/Tau/Run2016H-UL2016_MiniAODv1_NanoAODv2-v1/NANOAOD
2017	Run 2017B	/Tau/Run2017B-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017C	/Tau/Run2017C-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017D	/Tau/Run2017D-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2017E	/Tau/Run2017E-UL2017_MiniAODv1_NanoAODv2-v2/NANOAOD
	Run 2017F	/Tau/Run2017F-UL2017_MiniAODv1_NanoAODv2-v1/NANOAOD
2018	Run 2018A	/Tau/Run2018A-UL201_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2018B	/Tau/Run2018B-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2018C	/Tau/Run2018C-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD
	Run 2018D	/Tau/Run2018D-UL2018_MiniAODv1_NanoAODv2-v1/NANOAOD

Table 10.4: Run II collision data samples: tau primary datasets (UL NanoAODv8).

$-1.57 \leq \phi \leq -0.87$).

Though some of the technical details of the MC samples used are not relevant to this document, a set of lists is provided in Tables 10.5, 10.6,10.7, and 10.8 with the purpose of showing the cross-sections used corresponding to each process and to what order they are calculated.

Process	Official CMS dataset	Cross section [pb]
TTBar	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	365.34
Single top	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[**]/NANOAODSIM	134.2
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[**]/NANOAODSIM	80.0
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[***]/NANOAODSIM	39.65
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[***]/NANOAODSIM	39.65
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	6.893
Z+jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[**]/NANOAODSIM	6025.2 (NLO)
Z+jets H_T-binned ($m(\ell\ell) \geq$50 GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.0028
Diboson	/WW_TuneCP5_13TeV-pythia8/[**]/NANOAODSIM	118.3
	/WZ_TuneCP5_13TeV-pythia8/[**]/NANOAODSIM	47.13
	/ZZ_TuneCP5_13TeV-pythia/[**]/NANOAODSIM	16.523
W+jets (H_T-incl.)	/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/[**]/NANOAODSIM	61334.0
W+jets (H_T-binned)	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[**]/NANOAODSIM	0.0248
QCD (H_T-binned)	QCD_HT50to100_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	246300000.0
	QCD_HT100to200_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	27990000.0
	QCD_HT200to300_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	1559000.0
	QCD_HT300to500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	351900.0
	QCD_HT500to700_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	29070.0
	QCD_HT700to1000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	5962.0
	QCD_HT1000to1500_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	1005.0
	QCD_HT1500to2000_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	101.8
	QCD_HT2000toInf_TuneCP5_PSWeights_13TeV-madgraph-pythia8/[**]/NANOAODSIM	20.5
Higgs	/ZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	0.311
	/ggZH_HToBB_ZToBB_M-125_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	0.07784
	ggZH_HToBB_ZToLL_M-125_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	0.006954
	ggZH_HToBB_ZToNuNu_M-125_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	0.006954
VBS/VBF W/Z+Jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[**]/NANOAODSIM	32.05
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[**]/NANOAODSIM	39.05
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[**]/NANOAODSIM	6.215
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[**]/NANOAODSIM	10.66
TT+X	/TTWJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[**]/NANOAODSIM	0.2043
	/TTWJetsToQQ_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[**]/NANOAODSIM	0.4062
	/TTZToLLNuNu_M-10_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.2529
	/TTZToQQ_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.5297
	/TTGJets_TuneCP5_13TeV-amcatnloFXFX-madspin-pythia8/[**]/NANOAODSIM	3.6970
	/TTTT_TuneCP5_13TeV-amcatnlo-pythia8/[**]/NANOAODSIM	0.0090
V+jets	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[**]/NANOAODSIM	51.1

Table 10.5: List of background simulation samples for 2016 pre-VFP in the UL NanoAODv8 data format and its corresponding cross sections in pb, where [*] = RunIISummer20UL16NanoAODAPVv2-106X_mcRun2_asymptotic_preVFP_v8-v1, and [**] = RunIISummer20UL16NanoAODAPVv2-106X_mcRun2_asymptotic_preVFP_v9-v1, and [***] = RunIISummer20UL16NanoAODAPVv2-106X_mcRun2_asymptotic_preVFP_v9-v2

Process	Official CMS dataset	Cross section [pb]
TTBar	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	365.34
Single top	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	134.2
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	80.0
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	6.893
Z+jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	6025.2 (NLO)
Z+jets H_T-binned ($m(\ell\ell) \geq$50 GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WW_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	118.3
	/WZ_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	47.13
	/ZZ_TuneCP5_13TeV-pythia/[*]/NANOAODSIM	16.523
W+jets (H_T-incl.)	/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	61334.0
W+jets (H_T-binned)	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0248
QCD (-binned)	/QCD_Pt_15to30_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	1244000000.0
	/QCD_Pt_30to50_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	106500000.0
	/QCD_Pt_50to80_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	15700000.0
	/QCD_Pt_80to120_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	2346000.0
	/QCD_Pt_120to170_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	407700.0
	/QCD_Pt_170to300_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	103700.00
	/QCD_Pt_300to470_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	6826.0
	/QCD_Pt_470to600_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	551.2
	/QCD_Pt_600to800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	156.7
	/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	26.25
	/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	7.465
	/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.6487
	/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.08734
	/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.005237
	/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.0001352
VBS/VBF W/Z+Jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	32.05
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	39.05
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	6.215
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.66
V+jets	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	51.1

Table 10.6: List of background simulation samples for 2016 post-VFP in the UL NanoAODv8 data format and its corresponding cross sections in pb, where [*] = RunIISummer20UL16NanoAODv2-106X_mcRun2_asymptotic_v15-v1

Process	Official CMS dataset	Cross section [pb]
TTBar	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	365.34
Single top	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	134.2
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	80.0
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	6.893
Z+jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	6025.2 (NLO)
Z+jets H_T-binned ($m(\ell\ell) \geq$50 GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WW_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	118.3
	/WZ_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	47.13
	/ZZ_TuneCP5_13TeV-pythia/[*]/NANOAODSIM	16.523
W+jets (H_T-incl.)	/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	61334.0
W+jets (H_T-binned)	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0248
QCD (-binned)	/QCD_Pt_15to30_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	1244000000.0
	/QCD_Pt_30to50_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	106500000.0
	/QCD_Pt_50to80_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	15700000.0
	/QCD_Pt_80to120_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	2346000.0
	/QCD_Pt_120to170_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	407700.0
	/QCD_Pt_170to300_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	103700.00
	/QCD_Pt_300to470_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	6826.0
	/QCD_Pt_470to600_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	551.2
	/QCD_Pt_600to800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	156.7
	/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	26.25
	/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	7.465
	/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.6487
	/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.08734
	/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.005237
	/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.0001352
VBS/VBF W/Z+Jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	32.05
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	39.05
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	6.215
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.66
V+jets	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	51.1

Table 10.7: List of background simulation samples for 2017 in the UL NanoAODv8 data format and its corresponding cross sections in pb, where [*] = RunIISummer20UL17NanoAODv2-106X_mc2017_realistic_v8-v1

Process	Official CMS dataset	Cross section [pb]
TTBar	/TTTo2L2Nu_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	88.29
	/TTToHadronic_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	377.96
	/TTToSemiLeptonic_TuneCP5_13TeV-powheg-pythia8/[**]/NANOAODSIM	365.34
Single top	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	134.2
	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_13TeV-powheg-madspin-pythia8/[*]/NANOAODSIM	80.0
	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/[*]/NANOAODSIM	39.65
	/ST_s-channel_4f_leptonDecays_TuneCP5_13TeV-amcatnlo-pythia8/[*]/NANOAODSIM	6.893
Z+jets H_T-incl.	/DYJetsToLL_M-10to50_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	18610.0 (NNLO)
	/DYJetsToLL_M-50_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	6025.2 (NLO)
Z+jets H_T-binned ($m(\ell\ell)$ $\geq$50 GeV)	/DYJetsToLL_M-50_HT-100to200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	213.4
	/DYJetsToLL_M-50_HT-200to400_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	65.42
	/DYJetsToLL_M-50_HT-400to600_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	7.31
	/DYJetsToLL_M-50_HT-600to800_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.49
	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.661
	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.119
	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_PSweights_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0028
Diboson	/WW_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	118.3
	/WZ_TuneCP5_13TeV-pythia8/[*]/NANOAODSIM	47.13
	/ZZ_TuneCP5_13TeV-pythia/[*]/NANOAODSIM	16.523
W+jets (H_T-incl.)	/WJetsToLNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	61334.0
W+jets (H_T-binned)	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1695.0
	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	532.4
	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	61.6
	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	12.4
	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	5.77
	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	1.023
	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/[*]/NANOAODSIM	0.0248
QCD (-binned)	/QCD_Pt_15to30_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	1244000000.0
	/QCD_Pt_30to50_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	106500000.0
	/QCD_Pt_50to80_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	15700000.0
	/QCD_Pt_80to120_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	2346000.0
	/QCD_Pt_120to170_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	407700.0
	/QCD_Pt_170to300_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	103700.00
	/QCD_Pt_300to470_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	6826.0
	/QCD_Pt_470to600_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	551.2
	/QCD_Pt_600to800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	156.7
	/QCD_Pt_800to1000_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	26.25
	/QCD_Pt_1000to1400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	7.465
	/QCD_Pt_1400to1800_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.6487
	/QCD_Pt_1800to2400_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.08734
	/QCD_Pt_2400to3200_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.005237
	/QCD_Pt_3200toInf_TuneCP5_13TeV_pythia8/[*]/NANOAODSIM	0.0001352
VBS/VBF W/Z+Jets	/EWKWMinus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	32.05
	/EWKWPlus2Jets_WToLNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	39.05
	/EWKZ2Jets_ZToLL_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	6.215
	/EWKZ2Jets_ZToNuNu_M-50_TuneCP5_withDipoleRecoil_13TeV-madgraph-pythia8/[*]/NANOAODSIM	10.66
V+jets	/ZGToLLG_01J_5f_TuneCP5_13TeV-amcatnloFXFX-pythia8/[*]/NANOAODSIM	51.1

Table 10.8: List of background simulation samples for 2018 in the UL NanoAODv8 data format and its corresponding cross sections in pb, where [*] = RunIISummer20UL18NanoAODv2-106X_upgrade2018_realistic_v15_L1v1-v1, and [**] = RunIISummer20UL18NanoAOD-106X_upgrade2018_realistic_v11_L1v1-v1

CHAPTER 11

Semi-Data-Driven Background Estimation

11.1 $\mathbf{Z}(\rightarrow \mu\mu) + \mathbf{ISR}$

We construct this Drell-Yan + Jets CR with goals as outlined in our strategy section: to study the effects of high-p_T ISR on the BGs in our SR. This is made possible by the clean reconstruction of μ's in CMS and the high purity we can achieve of events where the required jet is in fact ISR. Additionally, after correcting for mismodeling related to boosted W/Z bosons, we can use this CR to study the validity of the cross-sections used in H_T-binned DY and W+Jets samples, extracting k-factor corrections where appropriate to then validate in a W+jets CR. The yield in our $Z \rightarrow \mu\mu$ + ISR CR can be parametrized as

$$N^{CR}_{Z\rightarrow\mu\mu} = \sigma_Z \cdot L_{\text{int}} \cdot \varepsilon^2_{\mu} \cdot \varepsilon_{\text{ISR}} \cdot \varepsilon_{\text{other}}, \tag{11.1}$$

where σ_Z is the best known cross-section for this process, L_{int} is the integrated luminosity in the data period considered, ε_{μ} is the efficiency for the selection of a muon, ε_{ISR} is the efficiency related to our ISR jet selection, and ε_{other} is the efficiency related to the remaining selections, which are expected to have a small effect (e.g. lepton and b-jet vetoes, and our QCD rejection cut). We can compare this to the DY+jets BG yield expected in our SR, which primarily comes from $Z \rightarrow \tau\tau$ processes where one τ decays hadronically and is identified, whereas the other τ is lost, showing up as E_T^{miss}. This yield is given by,

$$N^{SR}_{Z\rightarrow\tau\tau} = \sigma_Z \cdot L_{\text{int}} \cdot \varepsilon_{\tau_h} \cdot (1-\varepsilon_{\tau_h}) \cdot \varepsilon_{E_T^{miss}} \cdot \varepsilon_{\text{ISR}} \cdot \varepsilon_{\text{other}}, \tag{11.2}$$

where ε_{τ_h} is the efficiency related to selecting a τ_h ($1-\varepsilon_{\tau_h}$ is the efficiency related to the loss of the second τ), and $\varepsilon_{E_T^{miss}}$ is the efficiency associated with our E_T^{miss} cut. The advantages of first considering a CR with μ's are made clear in these expressions, as the efficiencies related to selecting τ_h's are known to carry more uncertainty and dependence on analysis topology than in the μ case. Furthermore, since these DY processes have no real E_T^{miss} and we don't implement a corresponding cut in this CR, this allows us to isolate and assess any differences in ε_{ISR} and or σ_Z between data and MC.

11.1.1 Boost Studies

The main selections comprising our $Z \rightarrow \mu\mu$ CR include our ISR jet with $p_T(j^{lead})$ >100 GeV and two tight-ID μ's, which we require to carry opposite signs and have a reconstructable mass $m_{\mu\mu} = [80, 100] GeV$

in order to ensure consistency with the Z-boson mass, i.e. to ensure they are its decay products. We also require the μ's to have an η value central in the detector as this is where reconstruction and ID is most reliable. To lower the possibility of including jets which have been erroneously reconstructed as μ's, we require $\Delta R(j,\mu) > 0.4$. We used 2018 MC and single-muon PDs for this study. The CR selections are summarized in Table 15.2. Because the main objects resulting from partonic interactions are the ISR jet and

Selection	Cut
$N(\mu)$	2
$p_T(\mu)$	> 30 GeV
$\lvert\eta(\mu)\rvert$	< 2.1
$Q_{\mu_1} * Q_{\mu_2}$	< 0 [OS]
$m_{\mu\mu}$	[80, 100] GeV
$\lvert\eta(j)\rvert$	< 2.4
$p_T(j)$	> 30 GeV
$N(j)$	≥ 1
$p_T(j_{lead})$	≥ 100 GeV
Trigger	HLT_IsoMu24

Table 11.1: Selections for $Z(\rightarrow \mu\mu)$+ISR CR

Z boson, we expect the Z to experience a recoiling effect from conservation of momentum effects due to high-energy nature of our ISR selection. This can be quantified by looking at the vector sum of the individual $p_T(\mu)$, $p_T(\mu\mu)$, which reflects the boost of the parent Z boson. This distribution is displayed in Fig. 11.2.

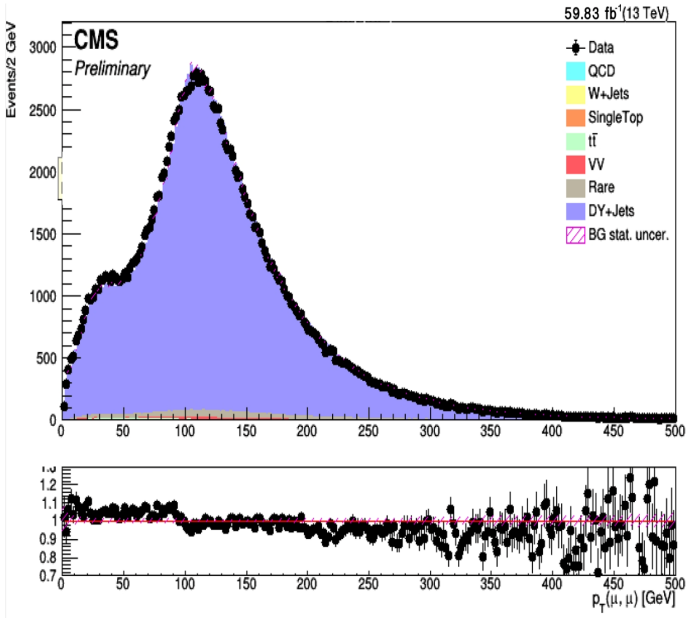

Figure 11.1: $Z \rightarrow \mu\mu$ CR $p_T(\mu,\mu)$ distibution reflecting the boost of the parent Z boson

As can be clearly seen in the ratio pad of Figure 11.2, the $p_T(\mu,\mu)$ exhibits significant disagreement

between data-and-MC. We take this distribution as our base to calculate boost weights meant to correct for this effect in processes with Z and W bosons (as the mismodeling in the W case is understood to come from similar sources). These are applied on an event-by-event in subsequent studies. The resulting weights are listed in Table 11.2. We compare the effects of other key distributions in this CR before and after the application of the derived boost weights to assess data-to-MC agreement. We study muon momentum and momentum of the leading jet as displayed in Figure 11.2, and improved modeling of MC is evident in both variables.

$p_t(\mu,\mu)$ Bin	Weight
1: 0-50 GeV	1.040 ± 0.006
2: 50-100 GeV	1.038 ± 0.005
3: 100-150 GeV	0.987 ± 0.003
4: 150-200 GeV	0.993 ± 0.005
5: 200-300 GeV	0.947 ± 0.012
6: 300-400 GeV	0.928 ± 0.02
7: 400-600 GeV	0.919 ± 0.020
8: 600+ GeV	0.769 ± 0.061

Table 11.2: Boost Weights by $p_T(\mu.\mu)$.

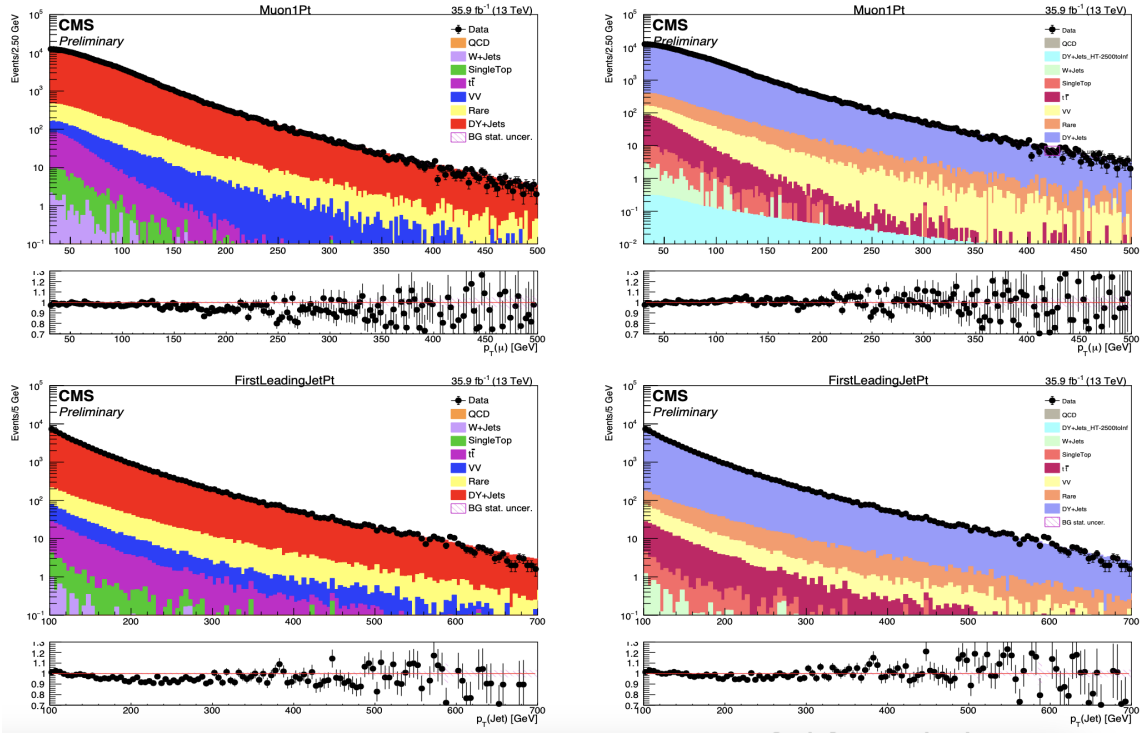

Figure 11.2: $p_T(\mu)$ (top) and $p_T(j^{lead})$ (bottom) distributions in $Z \to \mu\mu$ CR before (left) and after (right) application of boost weights.

11.1.2 K-Factor Corrections

The cross-sections used with our H_T-binned MC samples are found by taking the leading order (LO) calculation corresponding to the samples and multiplying by the appropriate k-factor. K-factors for processes with Zs and Ws produced in association with ≥ 1 jets have known H_T dependence studied in [74, 75]. Because of our ISR requirement and use of H_T-binned samples for gen-level $H_T > 100GeV$, the composition of DY + jets and W + jets in SR MC is dominated by the H_T-binned MC samples. We take advantage of the independence of this $Z \to \mu\mu$-enriched CR on $\varepsilon_{E_T^{miss}}$ and ε_{τ_h} to extract the necessary corrections to our cross-sections from the data-to-MC ratio of the H_T distribution in Fig. 11.3 (right). These are then applied as k-factors to our H_T-binned MC samples as a function of gen-H_T. The resulting "k-factor correction" SFs are listed in Table 11.3 and are consistent with what is seen in the data-to-simulation comparisons in the published CMS cross-section measurements papers [74, 75]. We note that, as illustrated in Fig. 11.3 (left), the $p_T(\mu,\mu)$ distribution reflecting the boost of the Z does not contain a notable dependence on H_T, which gives us confidence that our method of first extracting boost weights, and then k-factors is valid.

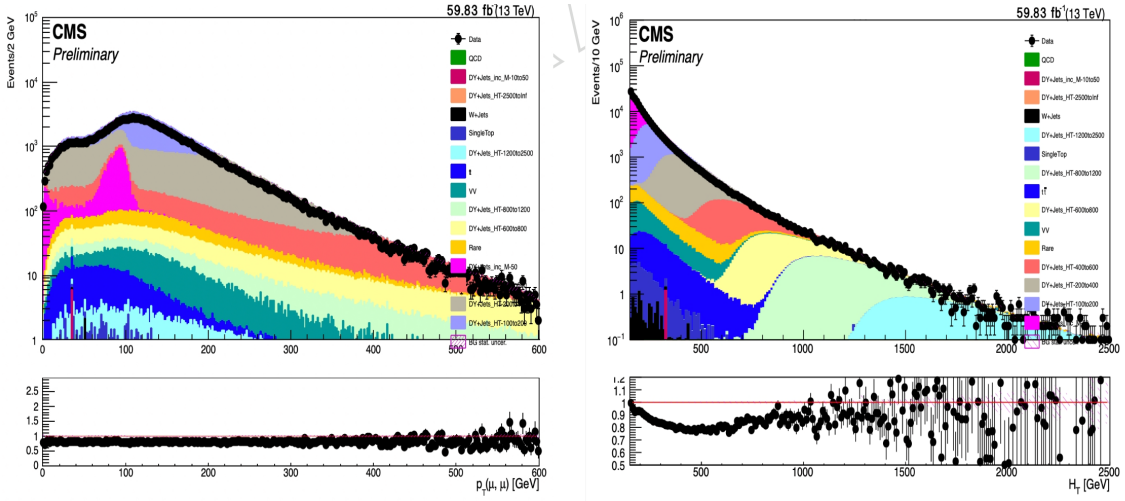

Figure 11.3: Z-Boost (left) and H_T (right) distributions for $Z(\to \mu\mu)$+ISR (2018) with different gen-H_T samples displayed separately.

Gen-H_T	Scale Factor
1: 100-200 GeV	0.7927 ± 0.019
2: 200-400 GeV	0.7927 ± 0.002
3: 400-600 GeV	0.7641 ± 0.003
4: 600-800 GeV	0.8048 ± 0.006
5: 800-1200 GeV	0.8629 ± 0.009
6: 1200-2500 GeV	0.9308 ± 0.020
7: 2500+ GeV	0.9887 ± 0.18

Table 11.3: K-factor Correction SFs by gen-H_T

11.2 $t\bar{t}$ Estimation

Top pairs are typically produced at the LHC via the quark-quark or gluon-gluon interactions producing a high-energy gluon, which then produces the $t\bar{t}$ pair. The top quark sometimes decays into a b-quark and W-boson, which can then go on to decay to leptonically. To reduce $t\bar{t}$ BG in our signal region, we implement a b-jet veto. Though significant reduction of $t\bar{t}$ is achieved in this way, there are still ways in which $t\bar{t}$ enters our SR, comprising $\sim 8\%$ of our total BG and $\sim 19\%$ of BG in the $m_T \geq 100$ GeV tail. To study the modeling of $t\bar{t}$ + jets BG in SR, we construct a control region with selections identical to those in SR, with the exception of the 2 b-jet requirement, in order to maintain the conditions as close to SR as possible while looking in an orthogonal region with no signal contamination.

The 2-bjet requirement is very effective in isolating $t\bar{t}$ MC, resulting in purities $> 97\%$ in this CR. Because the data-to-MC shape in various key distributions is well modelled, we extract a global $SF_{t\bar{t}}$ by subtracting non-$t\bar{t}$ BG from data and MC and taking the ratio of total data-to-MC, which we can then apply to our SR $t\bar{t}$ BG according to the method described by Equation 7.4. The restulting $SF_{t\bar{t}}$ are listed in Table 11.4. The low content of non-$t\bar{t}$ BG in this CR minimizes the systematic errors associated with $SF_{t\bar{t}}$. Because $t\bar{t}$ background is corrected in subsequent CRs, this uncertainty-limiting effect is propagated through the rest of the analysis and applied similarly to other BGs. Some key kinematic variables in the $t\bar{t}$ CR with 2018 data and MC are displayed in Figure 11.4, along with the χ^2/NDF value to a horizontal line fit in each case. Values of $\chi^2/NDF \leq 2.0$ indicate that the shapes of these distributions are well modelled and that the application of an overall $SF_{t\bar{t}}$ to the $t\bar{t}$ MC in SR is expected to yield an accurate estimate.

Year	2016	2017	2018
$SF_{t\bar{t}}$	1.12 ± 0.05	1.21 ± 0.04	0.97 ± 0.05

Table 11.4: $t\bar{t}$ control region scale factors.

11.3 W+Jets

The W+Jets CR serves two purposes: to validate the boost weights and k-factor corrections derived in the $Z \to \mu\mu$ CR, and to study the study the m_T distribution of W+Jets background in conditions similar to those in SR. W bosons can decay leptonically or hadronically. When decaying leptonically the resulting final state always includes a neutrino. It is understood that W+jets is our primary background because the W boson is produced in associated with a jet, which could mimic our signal ISR jet. Furthermore, the W decay could include a τ which goes on to decay hadronically and satisfy our τ_h selections. Finally, the neutrino from the leptonic W decay appears as E_T^{miss} in the detector and can carry enough energy to satisfy our high E_T^{miss} selection in SR. To keep our selections as close to SR as possible while looking in orthogonal data and MC

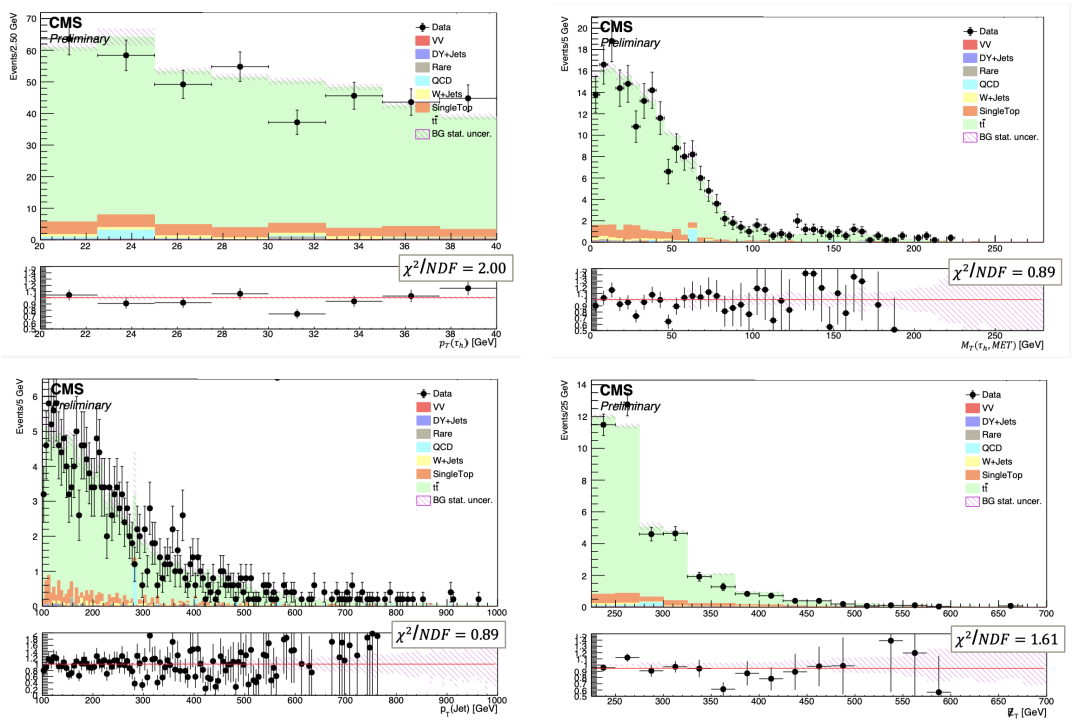

Figure 11.4: Top Row: $p_T(\tau_h)$ and $m_T(\tau_h, E_T^{miss})$ (2018); Second Row: $p_T(jet^{lead})$ and $\eta(jet^{lead})$ (2018)

subsets, we again take advantage of lepton universality and require a μ instead of a τ_h, while simultaneously vetoing any τ_h's. We also add an additional $\Delta|\phi|(E_T^{miss}, j^{lead})$ requirement to ensure a boosted topology. Every other selection is kept as it is in the SR. The list of selections our W+jets CR can be seen in Table 11.5.

11.3.1 Boost Weight Validation

The source of mismodeling in processes with boosted Z bosons, as those present in our SR BG MC, are understood to come from the limited ability of MC generators to produce events at high orders in QCD. Because our topology is boosted due to QCD (ISR) activity, we expect this to be a non-negligible effect in this analysis. For a W boson produced in association with jets, the same argument applies, thus we expect the boost weights derived from our $Z \to \mu\mu$ CR to be valid corrections for the same effect in our W+jets CR. We apply the boost weights on an event-by-event basis to our W+Jets BG (and any other BG process containing bosted W/Z's) according to the momentum W momentum at generator level. We then analyze the kinematic variable understood to be most reflective of the W boost. Because of our μ and large E_T^{miss} selections in this CR, we expect the decay products of the W boson in W+Jets MC to be a muon and a neutrino. Thus, we perform a goodness-of-fit test to the ratio pads of the $p_T(\mu)$ and E_T^{miss} distributions, these are shown in Figure

Table 11.5: $W(\rightarrow \mu\nu)$+ISR Event Selections

Selection	Cut
$N(\mu)$	1
$p_T(\mu)$	[20,40] GeV
$\lvert\eta(\mu)\rvert$	< 2.1
E_T^{miss}	> 230 GeV
$\lvert\eta(j)\rvert$	< 2.4
$N(j)$	≥ 1
$P_T^{lead}(j)$	≥ 100 GeV
QCD rejection	$\lvert\Delta\phi(j_{lead}, E_T^{miss})\rvert > 0.7$
Trigger	HLT_IsoMu24_

11.5. The corresponding χ^2/NDF are 0.79 and 1.27, respectively for $p_T(\mu)$ and E_T^{miss}, indicating excellent agreement in the shapes of MC compared to data.

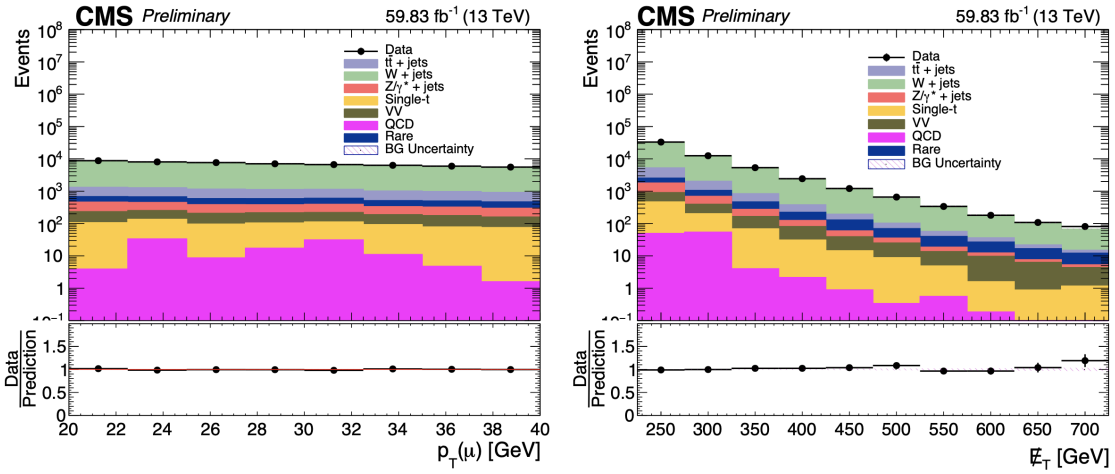

Figure 11.5: Muon p_T and E_T^{miss} with applied W/Z boost weights in W+Jets CR (2018).

11.3.2 Validating K-Factor Corrections

After necessary corrections related to the ISR activity, we turn our attention to validation of the k-factor corrections derived in the $Z \rightarrow \mu\mu$ section. Figure 11.6 shows the H_T distributions for our W + jets CR (with boost weights applied) without (left) and with (right) k-factor correction SFs. There is a notable improvement in the modeling of H_T, and a $\sim 27\%$ improvement in the global data-to-MC scale factor.

11.3.3 Extracting Shape of W+Jets BG in SR

The improved modeling of key variables displayed in previous sections gives us a high degree of confidence that effects related to ISR activity (including its effects on E_T^{miss}) and H_T dependence of k-factors have been

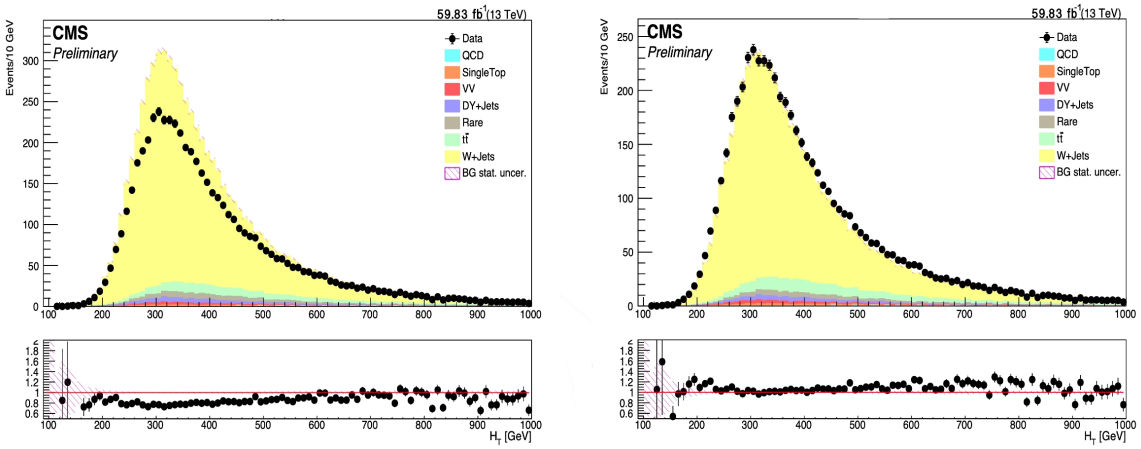

Figure 11.6: H_T distribution in W+jets CR without (left) and with (right) k-factor correction SFs (boost weights applied, 2018).

accounted for. We can now take a look at the $m_T(\mu, E_T^{miss})$ variable in our $W(\rightarrow \mu\nu)$+ISR CR, which, with the exception of effects related to τ ID, we expect to be an accurate representation of the shape of W + jets BG in our fit variable in the SR ($m_T(\tau, E_T^{miss})$. Fig. 11.7 shows the $m_T(\mu, E_T^{miss})$ distribution in this CR for the full RunII luminosity ($L_{2016} + L_{2017} + L_{2018} = 137.6\ fb^{-1}$). From [74, 75] we understand that the k-factors appropriate for our samples are not only a function of H_T, but vary by the number of jets present in the event. From Fig. 11.8 (left) it is apparent that our derived k-factor corrections have resulted in an accurate data-to-MC prediction where the majority of our W+Jets background lies, i.e. at $N(jets) = 1$. However, disagreement between data and MC in the tail of the $m_T(\mu, E_T^{miss})$ distribution (Fig. 11.7) indicates that further corrections to the shape of W + jets in our SR fit variable are required. To account for this effect, we apply a bin-by-bin, m_T-based scale factor in our prediction of W + jets in the SR. These SFs are listed in Table 11.6. Further evidence for the mismodeling of W + jets BG with $N(jets) \geq 1$ comes from the distribution of $\Delta\Phi_{min}(jet^{2nd}, E_T^{miss})$ in Fig. 11.8 (right), where there is a pronounced difference between data and MC in events with at least 2 jets. Additional kinematic plots for our leading ISR jet are displayed in Figure 11.9, showing excellent agreement between data and MC after the boost and k-factor corrections have been applied.

11.4 $Z \rightarrow \tau\tau$

Even though DY+jets only comprises 3.1% of overall BG in SR and 1.8% of BG in the high $m_T(\tau_h, E_T^{miss})$ range, we construct a $Z \rightarrow \tau\tau$ CR where both τ's go on to decay hadronically for the main purpose of studying the effects of τ ID. The strength of this CR is partly due to the fact that we now have confidence that we understand and have systematically corrected for any source of mismodeling and potential differences in data and MC efficiencies. In this CR we require two hadronically decaying taus (τ_h's), where the $p_T(\tau_h)$

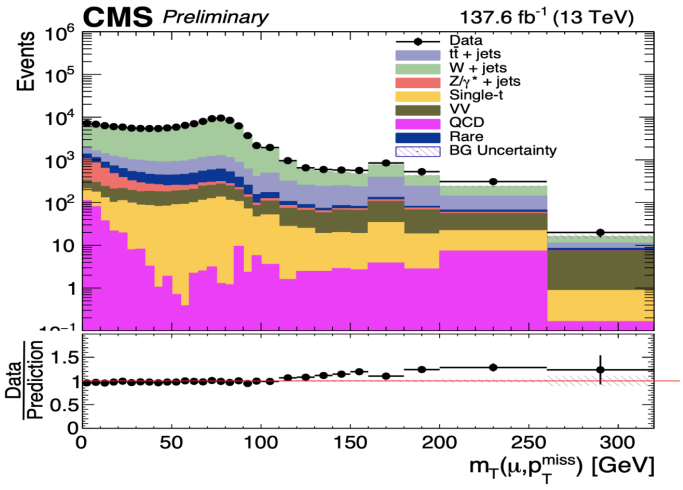

Figure 11.7: $m_T(\mu, E_T^{miss})$ in W+jets CR with full 2016-2018 data and MC. Mismodeling in the tail of this distribution is understood to be related to k-factor differences in $n(jet) > 1$ events.

for both is required to be > 60 GeV. The difference in $p_T(\tau_h)$ compared to signal region τ_h's is in place for two reasons: to ensure minimal signal contamination upon requiring two τ_h's as opposed to one with a different p_T range, and to allow the use of a $di-\tau$ trigger. Other than the p_T range, the τ_h are required to meet the conditions of our SR τ_h's: They are identified via the *DeepTauv*2 algorithm, passing the criteria for the "Vtight" WP, with "Tight" WP on the *DeepTauv*2 anti-electron and anti-muon discriminators. All POG recommended SF's with use of *DeepTauv*2 have been implemented. The reason we decide to use a new HLT trigger in this CR is that, since we aim to reconstruct both τ_h's coming from the Z boson decay, there is no real E_T^{miss} to be expected from this process, thus we do not make a E_T^{miss} selection. Along with using the listed di-tau trigger, we use tau PDs in this CR. The yield expected in this CR can be expressed as follows:

$$N^{CR}_{Z\to\tau\tau} = \sigma_Z \cdot L_{\text{int}} \cdot \varepsilon^2_{\tau_h} \cdot \varepsilon_{\text{ISR}} \cdot \varepsilon_{\text{other}}. \tag{11.3}$$

where σ_Z is the cross-section for this process, and ε_x are the efficiencies related to the associated selection. At this point in this analysis we have comprehensively studied, and corrected where necessary, any mismodeling related to our ISR selection. We have also validated those corrections via our W+jets CR. As in other CR's, uncertainties in σ_Z and L_{int} are small, and effects due to ε_{other} are also expected to be small. As a result, this CR provides a clean method for analyzing ε_{τ_h} ID and extracting corrections for our DY+jets BG in SR if necessary. As in the case where the Z boson decays to μ's, we require our τ_h's to be opposite sign and have a reconstructable mass $m_{\tau_h\tau_h}$ consistent with the mass of the Z boson. In this case, because reconstruction of τ_h's is seeded by jets, and because the reconstruction process is more susceptible to uncertainty and loss of material related to the original decayed τ compared to μ's, we require only $m_{\tau_h\tau_h} < 100$ GeV. As in the signal

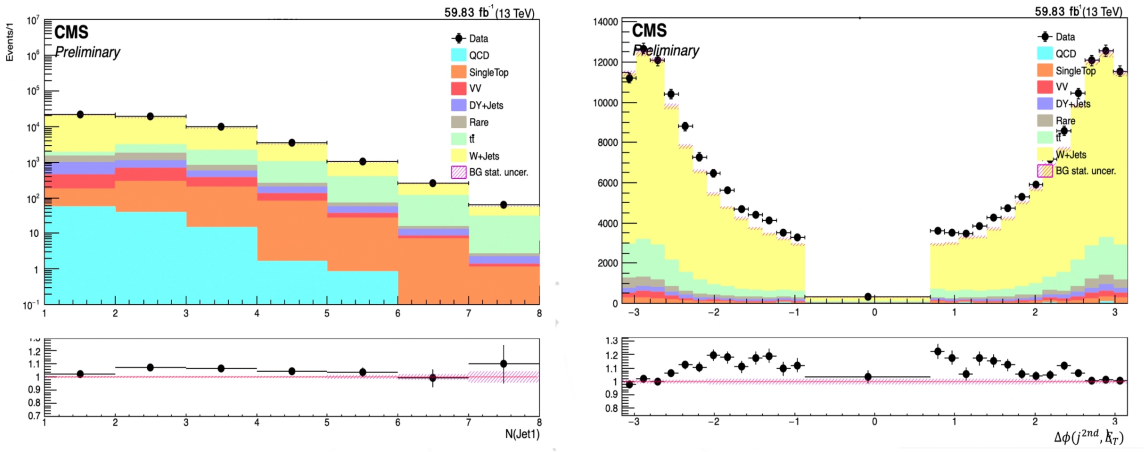

Figure 11.8: $N(jets)$ distribution (left) and $\Delta\Phi_{min}(jet^{2nd}, E_T^{miss})$ (right) in W+Jets CR (boost weights and k-factor correction SFs applied, 2018). Mismodeling in these distributions is understood to be related to k-factor differences in $n(jet) > 1$ events.

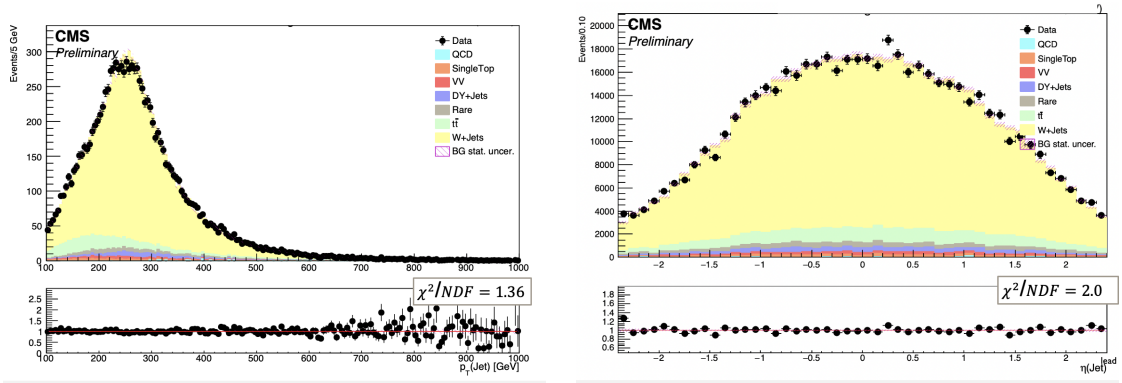

Figure 11.9: $p_T(j^{lead})$ (left) and $\eta(jet^{lead})$ (right) in W+jets CR (2018)

region, we implement the requirement for an ISR jet with $p_T(j^{lead}) > 100$ GeV with a "Tight" identification working point, and we include a similar separation between our jets and leptons to that in our DY+jets CR with μ's. The full selections of our $Z \to \tau_h\tau_h$ CR can be seen in Table 11.7.

The resulting kinematic distributions using can be seen in Figure 11.10 using 2018 MC and data, with the exception of the lower-right plot which uses the full 2016-2018 data for the purpose of analyzing the distribution from this CR that relates to our fit variable in the SR. Z/W boost weights and k-factor corrections have been applied to MC where applicable. For the sake of shape comparison, we have also applied to these plots the total data-to-MC scale factors derived in this CR. The full set of $SF_{Z\to\tau_h\tau_h}$ to be applied to our SR DY+jets yield can be seen in Table 11.8. We understand these SFs to be relatively large because this CR considers τ_hs with $p_T \geq 60GeV$, where the Tau POG recommends decay-mode-dependent $\tau_h ID$ SFs to be applied. We have withheld from applying these because (1) our SR τ_hs with $p_T = [20,40]GeV$ are below the

$\mathbf{m_T}$	**SF**
0-10	0.965 ± 0.010
10-20	0.966 ± 0.010
20-30	0.98 ± 0.011
30-40	0.98 ± 0.011
40-50	0.969 ± 0.011
50-60	0.99 ± 0.010
60-70	0.985 ± 0.009
70-80	0.999 ± 0.009
80-90	0.982 ± 0.010
90-100	0.964 ± 0.015
100-110	0.989 ± 0.030
110-120	1.065 ± 0.040
120-140	1.096 ± 0.040
140-160	1.168 ± 0.040
160-180	1.101 ± 0.044
180-200	1.24 ± 0.062
200-250	1.269 ± 0.088
250-300	1.466 ± 0.307

Table 11.6: m_T-dependent SFs extracted from W+jets CR to be applied to SR W+jets BG

Selection	Cut
$N(\tau_h)$	≥ 2
$p_T(\tau_h)$	> 60 GeV
$\lvert\eta(\tau_h)\rvert$	< 2.1
$Q_{\tau_{h1}}$*$Q_{\tau_{h2}}$	< 0 [OS]
$m_{\tau_h\tau_h}$	< 100 GeV
$\lvert\eta(j)\rvert$	< 2.4
$N(j)$	≥ 1
$p_T(j_{lead})$	≥ 100 GeV
N(b − jets)	= 0
Trigger	2016:HLT_DoubleMediumIsoPFTau*, HLT_DoubleMediumCombinedIsoPFTau*
	2017 & 2018:HLT_DoubleMediumChargedIsoPFTau(HPS)*

Table 11.7: $Z(\rightarrow \tau\tau)$+ISR CR Event Selections

range recommended for decay-mode-dependent SFs and (2) the decay-mode-dependent SFs are flat across $p_T(\tau_h)$, where the SFs we apply are p_T-based and yield better shape correction. As validation, the data-to-MC SFs for an identical CR, with decay-mode-dependent instead of p_T-based SFs applied, are also listed in Table 11.8. These show SFs close to unity. Since the m_T shape in this control region is well-modeled by MC, the *DY* + jets shape in the signal region is taken directly from simulation.

11.4.1 Validating Results With $Z \rightarrow \mu\tau_h$ Control Region

The choice of $p_T(\tau_h) \geq 60$ GeV in the $Z \rightarrow \tau_h\tau_h$ CR described above was partially motivated by the lack of a trigger that is efficient in the p_T range of our SR τ_hs. In the SR, the requirement of $E_T^{miss} \geq 230$ GeV allows

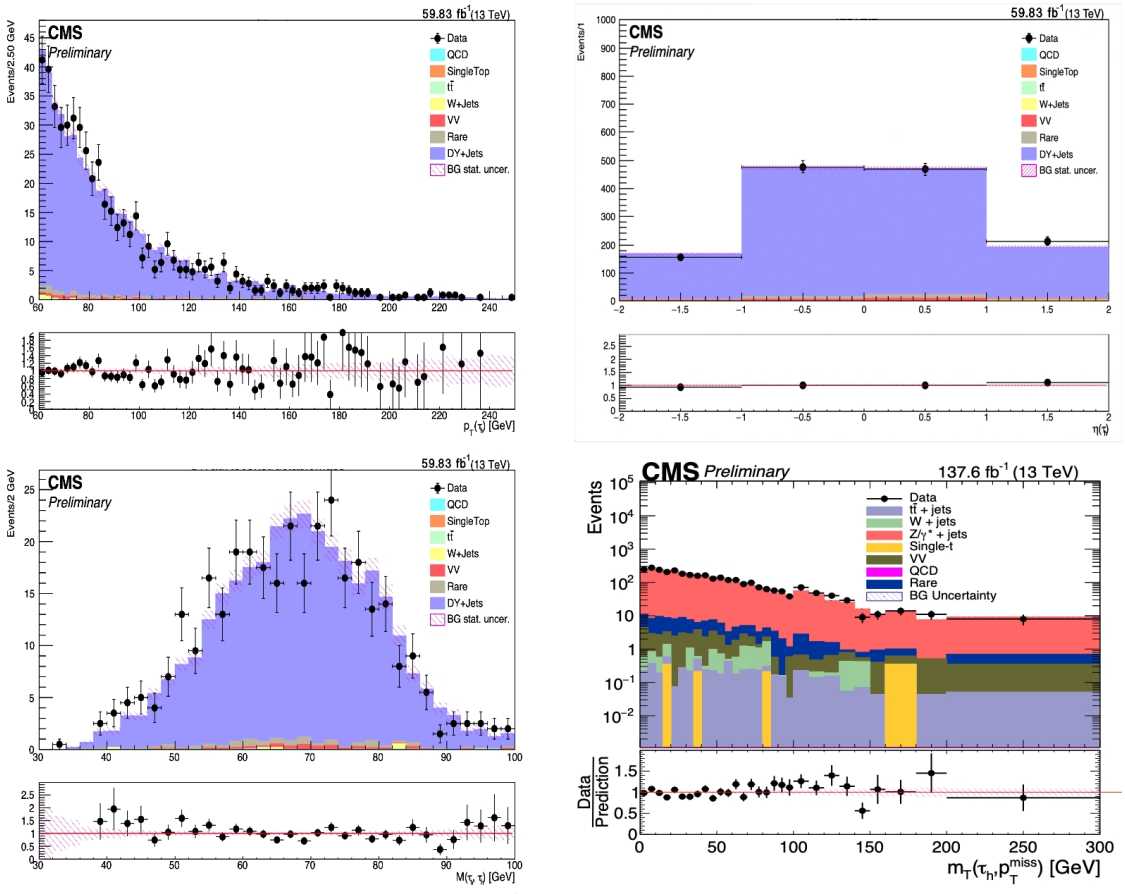

Figure 11.10: Top Row: Tau p_T and $m(\tau_h, \tau_h)$ (2018); Second Row: $\eta(jet)$ and $m_T(\tau_h, E_T^{miss})$ (2018 and all years, respectively)

Control Region	2016 SF	2017 SF	2018 SF
p_T-dep. SFs	1.38 ± 0.08	1.30 ± 0.07	1.13 ± 0.05
dm-dep. SFs	0.99 ± 0.06	0.91 ± 0.05	0.92 ± 0.05

Table 11.8: *DY* + jets data-to-MC SFs in $Z \to \tau_h\tau_h$ CR with p_T- and decay-mode-dependent τ_h ID SFs.

us to select events firing a E_T^{miss} trigger, but we make no such requirement in the $Z \to \tau_h\tau_h$ CR in order to focus on effects related to τ_h ID. To further study τ_h ID and confirm that the findings in the section above hold for our SR τ_h's, we construct a CR, using 2018 MC and data, where one of the τ's from the decay of the Z boson goes on to decay leptonically instead of hadronically. We implement a m_T requirement similar to the $Z \to \tau_h\tau_h$ CR, though this time on $m_T(\tau_h, \mu)$. These criteria allow us to select events with $p_T(\mu) > 30GeV$, for which we have an efficient trigger (HLT_IsoMu24), and we can now implement τ_h selections similar to those in our SR for a comparison. Further selection details are listed in Table 11.9.

As implied by Equation 11.3, the data-to-MC ratio in the $Z \to \tau_h\tau_h$ CR is associated with $\varepsilon^2_{\tau_h}$, namely

Table 11.9: $Z(\rightarrow \tau_h\mu)$+ISR Event Selections

Selection	Cut
$N(\tau_h)$	1
$p_T(\tau_h)$	[20,40] GeV
$\lvert\eta(\tau_h)\rvert$	< 2.1
$N(\mu)$	1
$p_T(\mu)$	≥ 30 GeV
$\lvert\eta(\mu)\rvert$	< 2.1
$Q_{\tau_h} * Q_\mu$	< 0 [OS]
$m_{\tau_h,\mu}$	< 100 GeV
$\lvert\eta(j)\rvert$	< 2.4
$N(j)$	≥ 1
$p_T(j_{lead})$	≥ 100 GeV
N(b – jets)	= 0
Trigger	HLT_IsoMu24

because we require exactly two τ_hs and expect the differences in other efficiencies involved to have been accounted for or equal when taking the ratio N^{data}/N^{MC}. It follows from the requirement of exactly one τ_h that the DY+jets data-to-MC ratio ($SF_{Z\rightarrow\tau_h\mu}$) in this CR is associated with ε_{τ_h}, or $\sqrt{SF_{Z\rightarrow\tau_h\tau_h}}$. The DY + jets data-to-MC ratio measured in this CR is 1.06 ± 0.10 which stands in agreement with the 2018 ratios found in the $Z \rightarrow \tau_h\tau_h$ section (1.13 ± 0.05). Furthermore, we see in Fig. 11.11 that the key variables associated with τ_hs in this CR are well modeled, as in the $Z \rightarrow \tau_h\tau_h$ case.

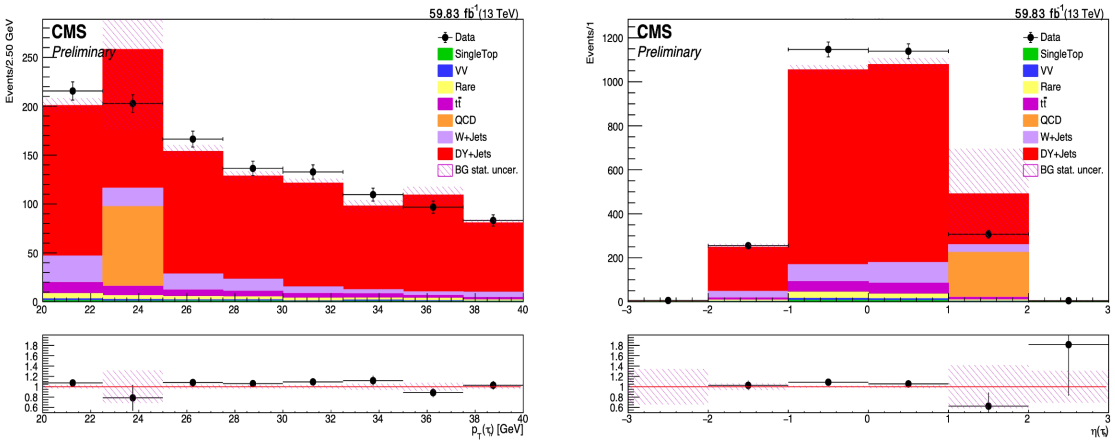

Figure 11.11: Tau p_T and η in the $Z \rightarrow \tau_h\mu$ CR (2018)

11.5 Diboson (VV)

Diboson background makes up only $\sim 1.5\%$ at $m_T(\tau_h, E_T^{miss}) < 100$ *GeV* and 4.0% at $m_T(\tau_h, E_T^{miss}) > 100$ *GeV* in SR. Though the contribution is small, we take the conservative approach of constructing a CR using

2018 data and MC to validate the use of diboson background directly from MC. To emulate the diboson content in our SR in an orthogonal CR, we require $N(\mu) \geq 3$ and cuts related to jets and E_T^{miss} similar to our SR. These are detailed in Table 11.10. The diboson data-to-MC ratio in this CR is 1.26 ± 0.34. Though the statistical uncertainty is large, this ratio is consistent with unity. Furthermore, we see in the plots of Fig. 11.12 that the kinematics of key variables are well modeled within uncertainties. Overall, these studies give us confidence in the use of MC to estimate diboson in the SR.

Selection	Cut
Trigger	HLT_PFMET120_PFMHT120
$N(\mu)$	≥ 3
$p_T(\mu)$	≥ 20 GeV
$\lvert\eta(\mu)\rvert$	< 2.1
$N(j)$	≥ 1
$p_T^{\text{lead}}(j)$	≥ 100 *GeV*
E_T^{miss}	≥ 230 *GeV*
Overlap removal	True ($\Delta R > 0.4$)

Table 11.10: Event selections motivating the diboson estimation

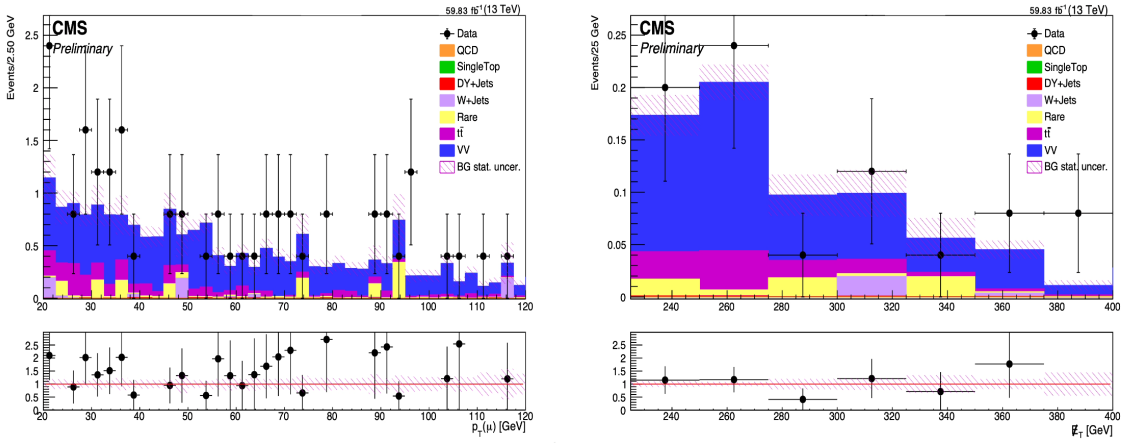

Figure 11.12: $p_T(\mu)$ (left) and E_T^{miss} (right) in a 2018 diboson CR.

CHAPTER 12

Data Driven Background Estimation: QCD

As previously mentioned, QCD multijet BG sneaks into our SR when a jet is misidentified as a τ_h. For the purposes of the CR's studied thus far, the MC samples suffices to give a description of the presence of QCD. For our SR estimation, on the other hand, we require a more precise, reliable method to estimate the QCD contribution which can account for higher-order QCD processes that evade production in MC samples. Figure 12.1 illustrates the motivation for the QCD estimation method for this analysis. This figure, produced with 2016 data and MC, shows the m_T and $\Delta\phi(j_{lead}, E_T^{miss})$ distributions for a sample of events with SR cuts but an inverted τ_h isolation requirement and an inverted $\Delta\phi_{min}$ cut ($|\Delta\phi_{min}| \leq 0.7$). The selection criteria are summarized in Table 12.1. An inverted τ_h isolation requirement means that the τ_h in the event passes the "VLoose" requirements but fails the "VTight" requirements for τ_hs. To note in Figure 12.1 is the significant discrepancy between Data and MC at high $m_T(\tau_h, E_T^{miss})$ and $\Delta\phi(j_{lead}, E_T^{miss})$ near zero ($|\Delta\phi(j_{lead}, E_T^{miss})| < 0.7$). Here we have excluded QCD MC and applied all relevant corrections to other BG MC, so this gives us a high degree of confidence that this discrepancy is the missing QCD multijet background. This is also the motivation for the $|\Delta\phi(j, E_T^{miss})|_{min} < 0.7$ cut used in the SR to reduce QCD BG.

Selection	Cut
Trigger	HLT_PFMET120_PFMHT120
$N(\tau_h)$	1
$p_T(\tau_h)$	[20, 40] GeV
$\lvert\eta(\tau_h)\rvert$	< 2.1
$N(j)$	≥ 1
$p_T^{\text{lead}}(j)$	≥ 100 GeV
$\lvert\eta_{lead}(j)\rvert$	< 2.4
E_T^{miss}	≥ 230 GeV
$\lvert\Delta\phi_{min}(j, E_T^{miss})\rvert$	< 0.7
$N(\text{b}-\text{jet})$	0
τ_h ID	pass "VLoose" but fail "VTight"
Overlap removal	$\Delta R > 0.4$

Table 12.1: Event selections motivating the QCD background estimation

The traditional method used for data-driven estimations is the "ABCD" method illustrated in Figure 12.2, where region B is our SR and B, C, and D are orthogonal CR's constructed with particular combinations of the *DeepTau* isolation requirement and the ($|\Delta\phi(j_{lead}, E_T^{miss})|$) variable. In our case, region B involves flipping the τ_h WP from events passing "VTight," to events passing VVLoose and failing "VTight." Region C carries a flipped τ_h ID isolation as well as a flipped $|\Delta\phi(j_{lead})|$ requirement to now include only events with $|\Delta\phi(j_{lead})| < 0.7$. Region D carries the SR τ_h ID criteria, but has the flipped $|\Delta\phi(j_{lead})| < 0.7$ requirement.

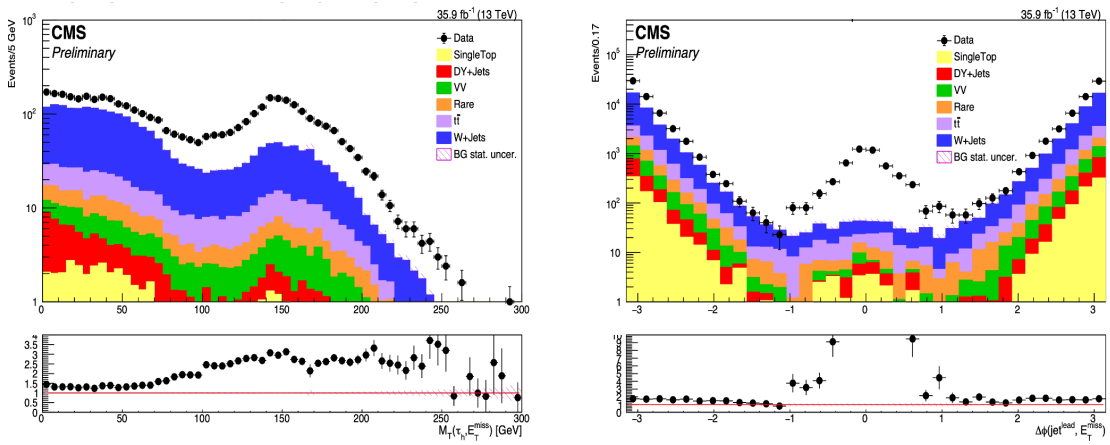

Figure 12.1: $m_T(\tau_h, E_T^{miss})$ (left) and $\Delta\phi(j_{lead}, E_T^{miss})$ (right) for events passing the "VLoose" τ_h-isolation requirement but failing "VTight" (2016)

The typical approach is to take data and subtract out non-QCD-MC in region B, and then scale this QCD yield by a $transfer\ factor$ that is meant to reflect how QCD would scales from the inverted τ_h ID working point to the nominal WP in the SR. Region C and D come into play to extract these transfer factors. Because of the similarity of selection to those in the SR (other than the inverted $|\Delta\phi|$ cut meant to increase QCD purity), a ratio of QCD events in region D to those in region C are reflective of how QCD scales when τ_h WP's change. This is in fact how the transfer factors are typically found, and then applied to the QCD yield in region B to normalize the QCD yield to what we expect to see in the SR. Because regions C and D in our analysis only reaches a QCD purity of $\sim 63\%$ in the high m_T tails, we can't be confident that uncertainties involved in non-QCD-multijet backgrounds will have an effect in the derivation of transfer factors (TF) from these regions.

Thus, we turn to an alternative method to extract our "VLoose-to-VTight" TF's. Because QCD in the SR comes from jets faking τ_h's, we instead accomplish this by forming CR's where we achieve a high purity of events that we are confident include jets that have been misidentified as τ_h's. We construct a version of this CR where the fake τ_h's pass the "VTight" working points, and another where the fakes pass the "Vloose" and fail the "VTight" WP, then proceed to take a ratio of the yields to get our "VLoose-to-VTight" TF's. There are two sources of jets in QCD multijet BG: gluons and light quarks. For the sake of completeness, we study both scenarios by constructing two CR's, one with a high purity of jets originating from gluons and the other from light quarks. Considering the two different DeepTau WPs we need to consider to extract TF's, this gives us a total of 4 CR's. In both scenarios, we take advantage of the accurate and efficient reconstruction of μ's in CMS. For the case where the misidentified jet comes from a gluon, we construct a $Z(\rightarrow \mu\mu) + \tau_{h,fake}$ CR. The targeted Feynman diagram related to this process can be seen in Figure 12.3 (left). We accomplish

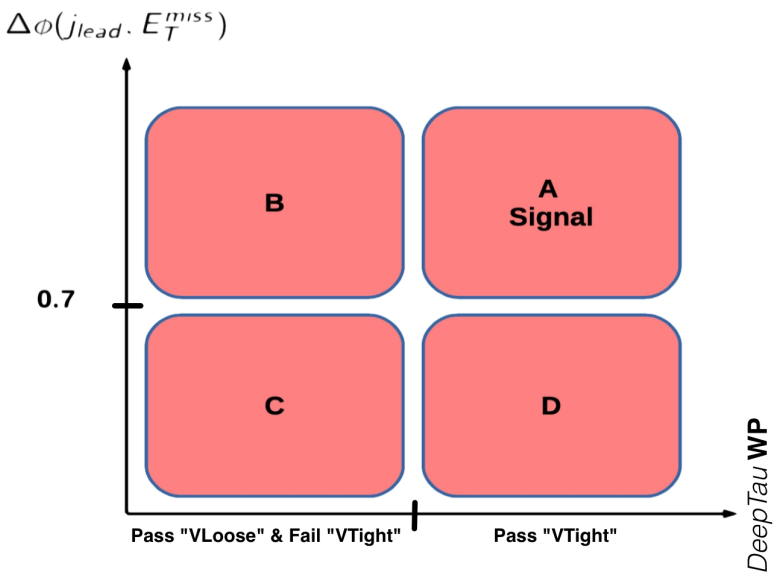

Figure 12.2: Classic ABCD method in data-driven QCD background estimation

the selection of these events by requiring the decaying muons to be consistent with those decaying from a Z boson (i.e. opposite charge and $m(\mu,\mu)$ reflecting $m(Z)$), then requiring an extra τ_h which can only come from an ISR jet. Because the incoming particles to create a Z boson must be a quark-antiquark pair (total charge of 0), we are confident that the source of the ISR jet in our selected events will primarily be a gluon. For the case of a light-quark jet misidentified as a τ_h, we construct a $W(\rightarrow \mu\nu) + \tau_h^{fake}$ CR. To capture the decay products of the W boson in the targeted process (right side diagram in Figure 12.3) we select exactly one μ while vetoing any additional muons that may come from DY processes. We also require $E_T^{miss} > 30$ GeV to account for the missing energy coming from the neutrino. The transverse mass between the muon and neutrino reflects the invariant mass of the originating W, so we require $m_T(\mu, E_T^{miss}) = [50, 120]$ to further ensure products that are W decays. After applying these cuts to ensure confidence that we have a high purity of the targeted process, we make the requirement for a τ_h that is understood to be a fake. The τ_h in both the Z and W CR cases is required to have $p_T\,\tau_h = [20-40]$ GeV to ensure similar conditions to those in the SR. It is worth noting that both CR's include a E_T^{miss} cut; in the W+$\tau_{h,fake}$ case it is due to the presence of the neutrino, and in the Z+$\tau_{h,fake}$ case it is in place to account for the mismeasurement of the associated jet. The full set of selections for our $\tau_{h,fake}$ CR's can be seen in Table 12.2.

We find the TF's as a function of $p_T(\tau_h)$ by taking the data in the $p_T(\tau_h)$ distributions and subtracting non-QCD-MC, then taking the bin-by-bin, nominal-to-inverted-τ_h ID ratio of the resulting distributions as

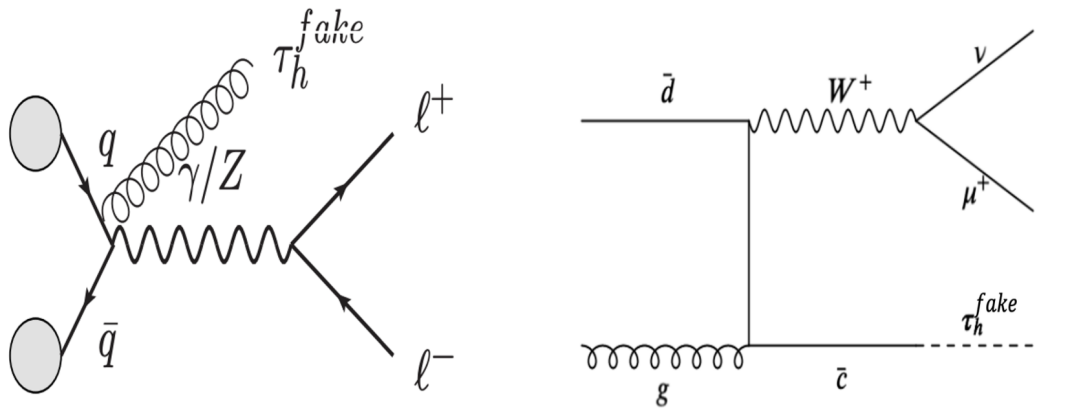

Figure 12.3: Feynman diagram of $Z(\rightarrow \mu\mu) + \tau_h^{fake}$ and $W(\rightarrow \mu\nu) + \tau_h^{fake}$

Central Selections	$Z(\rightarrow \mu\mu) + \tau_h^{fake}$	$W(\rightarrow \mu\nu) + \tau_h^{fake}$
Trigger	HLT_IsoMu24_	HLT_IsoMu24_
$N(\mu)$	2	1& Veto on other μ
$p_T(\mu)$	>30 GeV	>30 &>10 GeV
$\lvert\eta(\mu)\rvert$	<2.1	<2.1
$N(\tau_h)$	1	1
$p_T(\tau_h)$	>20 & <40 GeV	>20 & <40 GeV
$\lvert\eta(\tau_h)\rvert$	<2.1	<2.1
$Q(\mu_1)\times Q(\mu_2)$	-1	-
E_T^{miss}	>30 GeV	>30 GeV
$m(\mu,\mu)$	>70 & <110 GeV	-
$m_T(\mu, E_T^{miss})$	-	>50 & <120 GeV

Table 12.2: Selections defining the $Z(\rightarrow \mu\mu) + \tau_h^{fake}$ and $W(\rightarrow \mu\nu) + \tau_h^{fake}$ CR's

described by

$$Ratio^{VTight}_{VLoose-nonVTight} = \frac{N^{VTight}_{Z/W+Jets}}{N^{VLoose-nonVTight}_{Z/W+Jets}}. \tag{12.1}$$

where $N^{VTight}_{Z/W+Jets}$ is number of events in a bin passing the $VTight$ requirement, and $N^{VLoose-nonVTight}_{Z/W+Jets}$ is the number of events per bin failing the $VTight$ requirement but passing the $VLoose$ requirement. This is done separately for the gluon vs. soft quark jet cases and the TF's we obtain in both cases for 2018 are shown in Figure 12.4.

As mentioned previously, due to low QCD purity in Region C, we do not use region C/D as a way to extract transfer factors. Instead, we use regions C and D to perform a closure test by applying the transfer factors derived from our Z and W CR's to QCD (data - non-QCD MC) in region C, and check the data-to-MC agreement in region D. We perform this closure test using 2018 data and MC with TF's from both CR's as seen in Figure 12.4. The results are displayed in Fig. 12.5. We see better agreement when using the TF's

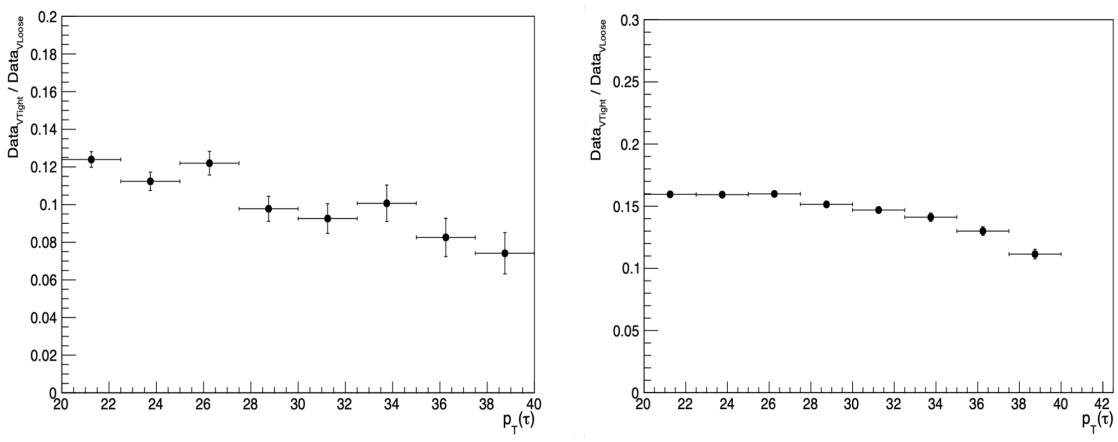

Figure 12.4: Transfer factors for $Z(\rightarrow \mu\mu) + \tau_h^{fake}$ (left) and $W(\rightarrow \mu\nu) + \tau_h^{fake}$ (right) CR's (2018)

extracted from the W CR, which is expected due to the dominating presence of $W + jets$ BG in the SR. Thus, we take the TF's from the W region as nominal and take the bin-by-bin difference in QCD yield in the $m_T(\tau_h, E_T^{mis})$ from Z vs W CR's as a systematic uncertainty. The W CR TF's for 2016-2018 are shown in Fig 12.6

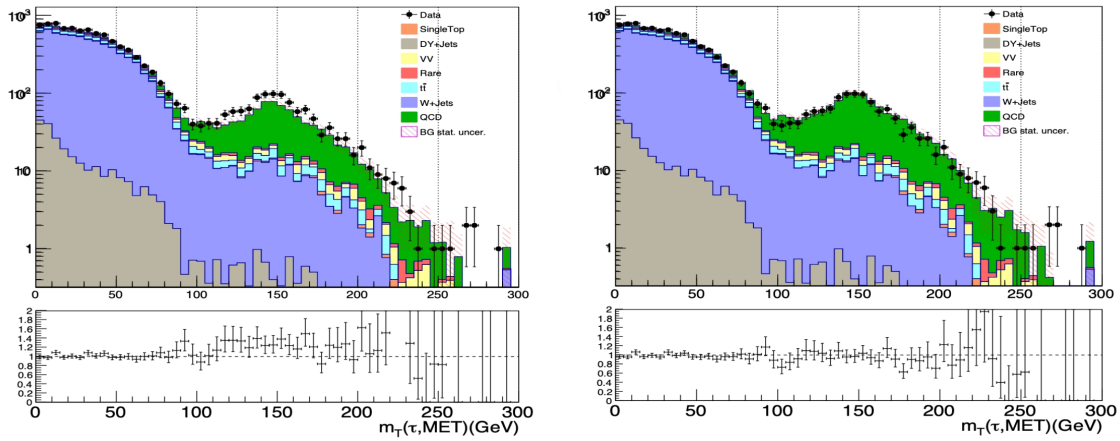

Figure 12.5: Closure test of data-driven QCD estimation method in inverted $|\Delta\phi|_{min}|$ CR's. $m_T(\tau_h, E_T^{mis})$ distributions with data-driven QCD estimates using $Z(\rightarrow \mu\mu) + \tau_h^{fake}$ TF's (left) and $W(\rightarrow \mu\nu) + \tau_h^{fake}$ TF's (right).

To apply our $p_T(\tau_h)$-based TF's and get the correct m_T shape in the SR, we construct a 2D histogram of $p_T(\tau_h)$ vs. $m_T(\tau_h, E_T^{mis})$ from our inverted τ_h ID CR (region B), subtract non-QCD-MC from data, apply the transfer factors as a function of pT, and integrate over the pT-axis. The 2D estimates of QCD in the SR (post transfer factor application) for years 2016-2018 are shown in Fig. 12.7.

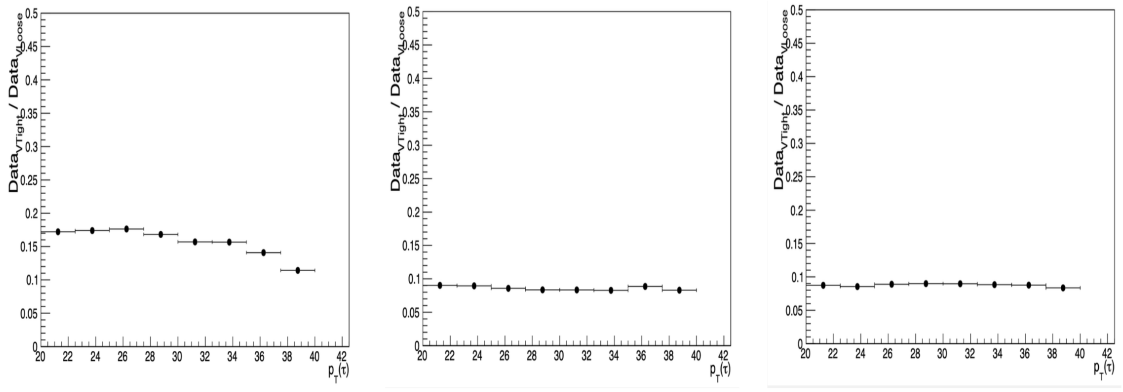

Figure 12.6: Transfer Factors from $W(\rightarrow \mu\nu) + \tau_h^{fake}$ CR for years 2016-2018 (from left to right).

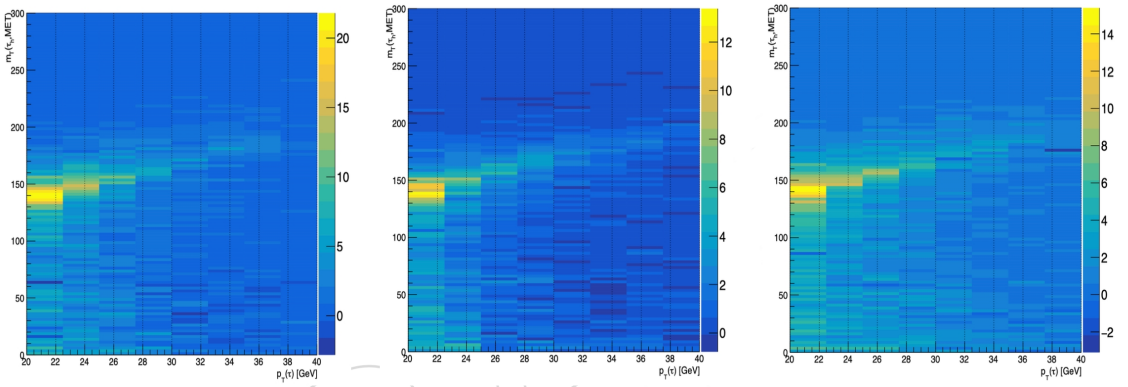

Figure 12.7: Signal region QCD estimations in 2D ($p_T(\tau_h)$ vs. $m_T(\tau_h, E_T^{mis})$) representation for years 2016-2018 (from left to right).

CHAPTER 13

Systematic Uncertainties

The uncertainties presented in the studies so far are by and large statistical. There are various sources which introduce systematic uncertainties into our analysis, which we will cover here. A comprehensive summary of all uncertainties we consider can be seen in Table 13.5, and the correlation of these across bins, processes, and years can be seen in Table 13.6

- **Luminosity**: There are many sources of uncertainty related to CMS luminosity measurements. The dominating contribution is from x-y non-factorization in Van Der Meer scans [76]. The Van Der Meer method is used to calibrate luminosity detectors by measuring the density of protons in an incoming beam. This method assumes that densities can be factorized as depending on x and y axes independently, which is not always physically the case. The full resulting uncertainties on measured luminosities are 1.3%, 2.3%, and 2.5% for 2016, 2017, and 2018 respectively.

- **Trigger**: To ascribe a systematic uncertainty to our trigger efficiencies, we perform a fit to the total MC efficiency using the function in Eq. 13.1:

$$f(x) = \frac{a}{2}(1 + erf(\frac{x-b}{c\sqrt{2}})) \tag{13.1}$$

  where *a, b,* and *c* are our fit parameters, and erf is the standard Gauss error function. The final fit can be seen in green in Fig. 8.1, and its parameters and associated errors are displayed in Table 13.1. To find the uncertainty in our fit, we propagate the errors as described in Eqn. 13.2 where $\delta_{a,b,c}$ are errors on our fit parameters, and $\delta_{ab,ac,bc}$ are elements of the associated covariant matrix. The uncertainty of our fit on the plateau, where our MET selection lies, is on the order of 10^{-3}%. We also perform a 0^{th} degree polynomial fit on the ratio of $\varepsilon_{MC}/\varepsilon_{data}$ on the plateau, which results in a function $p0 = 0.973$. We take the deviation from unity of this fit as our primary systematic uncertainty since the error on our fit is negligible (0.1%).

$$\partial f = \sqrt{(\frac{\partial f}{\partial a})^2\delta_a^2 + (\frac{\partial f}{\partial b})^2\delta_b^2 + (\frac{\partial f}{\partial c})^2\delta_c^2 + \frac{\partial f}{\partial a}\frac{\partial f}{\partial b}\delta_{ab} + \frac{\partial f}{\partial a}\frac{\partial f}{\partial c}\delta_{ac} + \frac{\partial f}{\partial b}\frac{\partial f}{\partial c}\delta_{bc}} \tag{13.2}$$

- ***Z/W* Boost Weights:** Because our binned likelihood fit is done in the $m_T(\tau_h, E_T^{miss})$ distribution, our systematic uncertainties are applied per $m_T(\tau_h, E_T^{miss})$ bin. The boost weights as a function of Z/W

Table 13.1: Fit Parameters and Errors

Parameter	Value
a	0.9736 ± 0.001
b	152.699 ± 1.648
c	26.582 ± 1.154

momentum boost are shown below in Table 13.2. To propagate this uncertainty to our fit variable, we

Table 13.2: Event Wgt Uncertainties by Z-Boost

Z-Boost Bin	Weight Uncertainty
1: 0-50 GeV	$< 1\%$
2: 50-100 GeV	$< 1\%$
3: 100-150 GeV	$< 1\%$
4: 150-200 GeV	$< 1\%$
5: 200-300 GeV	1.3%
6: 300-400 GeV	2.2%
7: 400-600 GeV	2.2%
8: 600+ GeV	7.9%

vary the boost weights by $\pm 1\sigma$. This is the error stated alongside the weights in Table 13.2 and is a result of the statistical uncertainties in data and MC. We then take the effect this up/down variation has on each bin in our $m_T(\tau_h, E_T^{miss})$ distribution as the systematic error associated with these weights. Though the weights carry uncertainties up to 8% at high p_T the effect in our fit variable bins is lower. This can be seen in Figure 13.1, where blue is the m_T distribution with the nominal weight applied, yellow corresponds to m_T when weights -1σ are applied to the relevant events, and green corresponds to weights $+1\sigma$ applied in a similar fashion. The resulting uncertainties can be seen in Table 13.3. We apply these as shape-based uncertainties and one can see they range from 1% at low values of m_T to 4% at high m_T.

- τ_h **Identification**: τ_h ID SFs are supplied by the Tau POG to correct for differences in DeepTau performance in data and MC. We apply these SFs using the POG-supplied tool, where the SF uncertainties are also available. We apply the τ_h ID SFs $\pm$ uncertainties to our signal region MC, and take the relative difference in yields from the nominal SFs in each case as our systematic uncertainty. These are observed to be independent of $m_T(\tau_h, E_T^{miss})$. The relative differences across BG's are $\sim 3\%$ and can be seen in detail in Table 13.5.

- b**-Tagging Efficiency:** Scale factors associated with b-tagging efficiencies are provided by the b-tagging and vertexing POG for the purpose of correcting for slight deviations in efficiencies in MC

$m_T(\tau_h, E_T^{miss})$ bin	DY+jets uncertainty (%)	W+jets uncertainty (%)
$0-10GeV$	1.0	1.0
$10-20GeV$	1.0	1.1
$20-30GeV$	1.1	1.0
$30-40GeV$	1.2	1.0
$40-50GeV$	1.3	1.0
$50-60GeV$	1.2	1.0
$60-70GeV$	1.1	0.9
$70-80GeV$	1.0	0.9
$80-90GeV$	0.7	0.8
$90-100GeV$	0.7	0.8
$100-130GeV$	0.4	0.8
$130-160GeV$	0.6	0.8
$160-190GeV$	0.5	0.9
$190-250GeV$	0.7	1.5
$250-400GeV$	2.4	3.9

Table 13.3: Systematic error on $m_T(\tau_h, E_T^{miss})$ in DY+Jets and W+Jets background due to *Z*/*W* boost weight uncertainties

and data. The uncertainties to these scale factors are accessible from POG resources, and are similarly dependent on the choice of working point, $p_T(jet)$, and $\eta(jet)$. We apply the SFs $\pm$ uncertainties and take the relative difference in event yields to be our systematic uncertainty. As expected, this effect is mostly apparent in $t\bar{t}$ and single top MC at 5.7% and 3.4%, respectively. The effect on yields of other backgrounds is $< 1\%$.

- **Electron Energy Scale:** Similar to the jet energy scale case described in section 6.2, electrons and muons carry small uncertainties related to reconstructed energy scale. For eletrons these are 1% and 2.5% respectively for the barrel and endcap regions of the ECAL. The systematic effect is $< 1\%$ in all backgrounds. This is expected as our use of electron ID is to veto electrons, and therefore has a small effect.

- **Muon Momentum Scale:** Muon momentum is also susceptible to the effects mentioned above. For these, we consider a 1% momentum scale uncertainty. We see less than %1 variation our fit variable bins related to this uncertainty.

- **Jet Energy Scale:** JEC and JER uncertainties are dominated by pileup effects on the low p_T end and by effects related to the use of simulation to extract response variables in the high p_T range [77]. These JEC with $\pm 1\sigma$ are provided by the JetMET group. As the JEC themselves, they range between 2%-5% and are a function of $p_T(jet)$ and $\eta(jet)$. As with our other uncertainties, we vary them $\pm 1\sigma$ and observe the effect in SR. As the deviations from nominal m_T when varying JEC uncertainties are seen to be consistent across the m_T spectrum (Figure 13.2), we apply these as overall systematic uncertainties

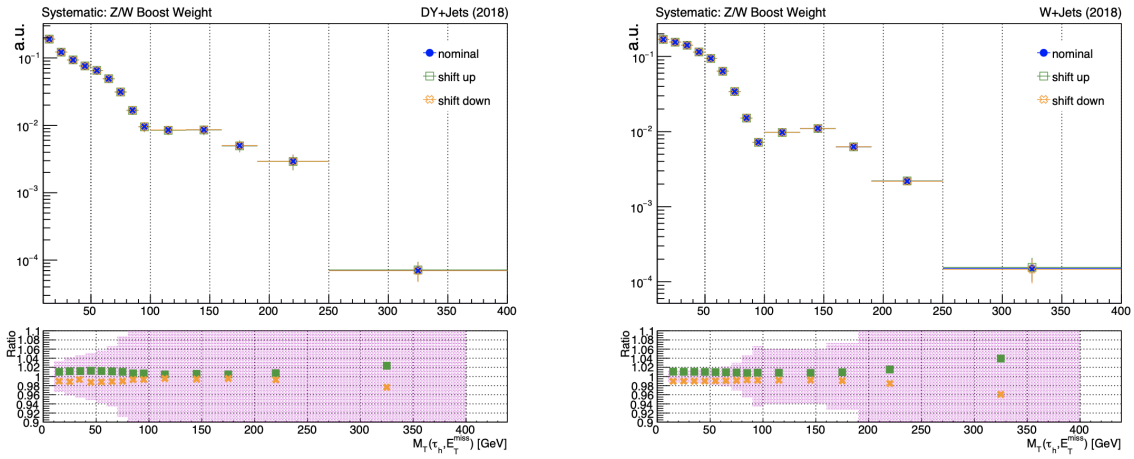

Figure 13.1: $\pm 1\sigma$ boost weight variation in $m_T(\tau_h, E_T^{miss})$ for Drell-Yan (left) and for W+Jets (right) in the Signal Region

in the yield of events per background in SR. It is worth noting that, even though the JEC uncertainties are in the order of 2-5%, most of the jets in our events from W+jets and DY+jets lie well above the lower $p_T(jet)$ threshold. Therefore, we see small deviations of overall yield in these backgrounds with varying JEC uncertainties. In backgrounds where our jets lie closer to the threshold of our jet cuts, such as $t\bar{t}$, we see bigger effects when varying JEC. The varied m_T plot for $t\bar{t}$ can be seen in Figure 13.2.

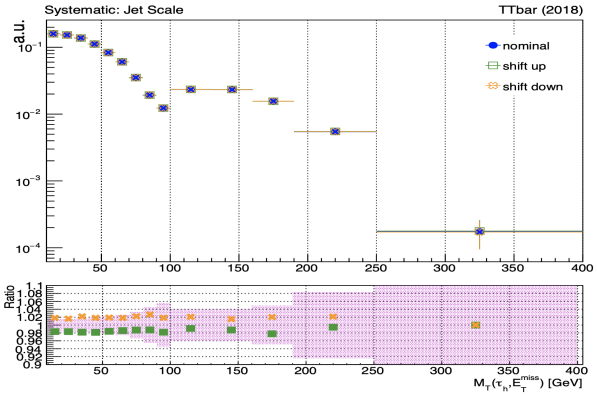

Figure 13.2: $m_T(\tau_h, E_T^{miss})$ distribution with varied uncertainties on JEC in $t\bar{t}$ MC events in SR.

- **Jet Energy Resolution:** Similar to JEC above, correction to jet energy resolution (JER) must be applied to MC to match that in data. The appropriate rescaling factors, along with their uncertainties, are provided by the JetMET group. We make use of these values to shift the JER factors by uncertainty up/down and observe the effect in our SR MC. The JER corrections are in general smaller in nature, and indeed we see an effect of $< 1\%$ across all backgrounds.

- **MET:** As has been discussed, the calculation of MET is dependent on the p_T of reconstructed PF objects. A consequence of this is that upon considering systematic uncertainties variations for various PF candidates, we effectively account for the corresponding uncertainty in MET. This is specifically the case in the JEC and JER case where the corrected jet properties are propagated to E_T^{miss}. The remaining uncertainty in MET to consider is that of unclustered MET. Jets below the unclustered energy threshold of 15 GeV are not corrected. Unclustered energy is subject to various detector effects and its uncertainties are also provided in our samples. We use the varied m_T templates as "shape variations" due to unclustered E_T^{miss}.

- **L1 Pre-Firing Weights:** As discussed, non-prefiring probabilities are applied as weights to 2016 and 2017 MC to account for prefiring issues in the detector reflected in data. These were derived by via measured pre-firing probabilities of the particular ECAL issues, they therefore carry corresponding uncertainties. The uncertainties are also provided by the appropriate CMS expert group. We apply the shifted weights to MC in our SR and apply the relative differences in event yield, per background process, as systematic uncertainties. These are summarized in Table 13.5

- **Semi-data-driven Scale Factors** Upon using the semi-data-driven technique of taking data-to-MC ratios to extract SFs, each of the yields we divide are susceptible to their own statistical errors and leave a corresponding error in the SF. The relative error on these SFs is our normalization systematic uncertainty in the process it's applied to.

- **Data-Driven QCD estimation:** Our data-driven QCD estimation strategy was based on the extraction of TF's from regions with high purity of fake τ_h. The statistics both these CR's were high enough that the resulting relative statistical uncertainties on our TF's are negligible. From our closure test, we see that the TF's from the W+Jets control region featuring light quark jets as the object faking our τ_h's yield the better data-to-MC agreement. To account for the fact that light quarks may not be solely responsible for our $\tau_{h,fake}$ content in SR, we take the relative difference in yields, per m_T bin, resulting from application of the W+jets and DY+jets TF's as our systematic uncertainty related to QCD closure/normalization. There p_T-binned TF's and their relative differences are listed in Table 13.4.

$p_T(\tau_h)[GeV]$	W+Jets $Ratio^{Tight}_{Loose}$	Z+jets $Ratio^{Tight}_{Loose}$	R. Difference
$[20.0, 22.5]$	0.087 ± 0.001	0.090 ± 0.001	3.3%
$[22.5, 25.0]$	0.085 ± 0.001	0.086 ± 0.001	1.2%
$[25.0, 27.5]$	0.089 ± 0.001	0.084 ± 0.001	5.6%
$[27.5, 30.0]$	0.090 ± 0.002	0.085 ± 0.001	5.5%
$[30.0, 32.5]$	0.090 ± 0.002	0.084 ± 0.001	6.6%
$[32.5, 35.0]$	0.088 ± 0.002	0.084 ± 0.002	2.3%
$[35.0, 37.5]$	0.088 ± 0.003	0.081 ± 0.002	8.0%
$[37.5, 40.0]$	0.083 ± 0.003	0.073 ± 0.002	7.2%

Table 13.4: QCD transfer factors per bin of $p_T(\tau_h)$ in the $W + Jets$ and $Z + Jets$ regions and their relative difference, This difference will be set a systematic uncertainty.

Source	W	DY	$t\bar{t}$	VV	QCD	ST	Rare
Lumi	[*]	[*]	[*]	[*]	–	[*]	[*]
μ ID	< 1	< 1	< 1	< 1	–	< 1	< 1
e ID	< 1	< 1	< 1	< 1	–	< 1	< 1
τ_h ID	3.3	2.7	3.3	2.7	–	3.3	2.9
Trigger	3.0	3.0	3.0	3.0	–	3.0	3.0
b ID	< 1	< 1	5.7	< 1	–	3.4	< 1
JES	< 1	< 1	1.8	< 1	–	1.8	< 1
JER	< 1	< 1	< 1	< 1	–	< 1	< 1
MMS	< 1	< 1	< 1	< 1	–	< 1	< 1
EES	< 1	< 1	< 1	< 1	–	< 1	< 1
Pileup	s	s	s	s	–	s	s
Unc. MET	s	s	s	s	–	s	s
L1 Prefiring	2.2	2.3	4.3	3.0	–	4.0	5.1
pdf	4.8	4.2	4.2	3.5	–	4.2	4.2
bin-by-bin stat.	s	s	s	s	–	s	s
Closure+Norm.	1.0	4.4	4.1	–	–	–	–
Z/W Boost	s	s	–	–	–	–	–
$Ratio^{Tight}_{Loose}$	–	–	–	–	s	–	–

Table 13.5: Summary of systematic uncertainties in % of m_T yield deviation, where s denotes a shape-based uncertainty, and [*] = 1.2, 2.3, and 2.5 for 2016-2018, respectively.

Source	Correlation across years	Correlation across processes	Correlation across bins
Lumi 2016 (1.2%)	1.0 uncorrelated, 0.6 correlated all years	100% for BGs/signal using MC	100%
Lumi 2017 (2.3%)	2.0 uncorrelated, 0.9 correlated all years, 0.6 correlated btw 2017/2018	100% for BGs/signal using MC	100%
Lumi 2018 (2.5%)	1.5 uncorrelated, 2.0 correlated all years, 0.2 correlated btw 2017/2018	100% for BGs/signal using MC	100%
Z/W boost	100%	100% btw Z/W+jets	100%
μ ID	stat 0%, syst 100%	100% for BGs/signal using MC	100%
e ID	100%	100% for BGs/signal using MC	100%
τ_h ID	0%	100% for BGs/signal using MC	100%
Trigger	0%	100% for BGs/signal using MC	100%
b ID	depends on the jet flavor and year	0% for non-top BGs/signal, 100% for top BGs using MC	100%
JEC	100%	100% for BGs/signal using MC	100%
JER	0%	100% for BGs/signal using MC	100%
TES	0%	100% for BGs/signal using MC	100%
MMS	stat 0%, syst 100%	100% for BGs/signal using MC	100%
MMR	stat 0%, syst 100%	100% for BGs/signal using MC	100%
EES	100%	100% for BGs/signal using MC	100%
MET uncl.	0%	100% for BGs/signal using MC	100%
Pileup	100%	100% for BGs/signal using MC	100%
pdf	100%	0% across processes	100%
scale	100%	0% across processes	100%
bin-by-bin stat.	0%	0% across processes	0%
Closure, Norm.	0%	0% across processes	100%

Table 13.6: Correlation across, years, processes, and bins for each systematic uncertainty

CHAPTER 14

Results and Discussion

14.1 Statistical Methods

Generally speaking, the interpretation of our results is approached as a hypothesis test aiming to quantify a potential discovery if significant deviation between data and SM predictions is seen; in the more common case where no significant excess is observed, we instead aim to set an upper limit on production cross-sections. This is done using a *profile likelihood ratio* test, described in detail in Ref. [78]. The base statistical tool underlying this technique is the likelihood function, which describes the probability that the configuration of our fit variable bins can be described by a given model. We can define a signal strength parameter μ, with a value of 0 for a BG-hypothesis and 1 for a nominal signal hypothesis. The variation in yields due to systematic uncertainties is incorporated in the form of nuisance parameters (NP) θ, which we allow to vary assuming log-normal priors for normalization parameters and gaussian in shape uncertainties. The test statistic used in our hypothesis testing is

$$t_\mu = -2\ln\lambda(\mu), \tag{14.1}$$

where $\lambda(\mu)$ is the *profile likelihood ratio*

$$\lambda(\mu) = \frac{L(\mu,\hat{\hat{\theta}})}{L(\hat{\mu},\hat{\theta})}. \tag{14.2}$$

$L(\mu,\hat{\hat{\theta}})$ is the *profile likelihood* for a given μ, with fixed NPs $\hat{\hat{\theta}}$ (though they are function of μ) which maximize the likelihood for the given μ we are testing. The denominator, $L(\hat{\mu},\hat{\theta})$, is the best-fit likelihood function with $\hat{\mu}$ and $\hat{\theta}$ found through log-likelihood minimization.

To quantify a potential discovery, our null hypothesis H_o is the BG-only hypothesis, i.e. $\mu = 0$, and we compare our test statistic to that representing a 5-sigma p-value. In the case of $\hat{\mu} < 0$ we take our $t_\mu = 0$ because the presence of signal is taken to be a positive event contribution.

In the case that no significant excess is observed, we use the LHC-standard CL_s technique to set limits [79], where CL_s is define as

$$CL_s = \frac{p_{s+b}}{1-p_b} = \frac{\int_{t_{obs}}^{\infty} f(t|s+b)dt}{1-\int_{-\infty}^{t_{obs}} f(t|b)dt} = \frac{CL_{s+b}}{CL_b}. \tag{14.3}$$

. Here, p_{s+b} is the p-value under a data+BG hypothesis, and p_b is a p-value under a BG-only hypothesis.

These are found by taking the appropriate integral of the probablity distribution function (PDF) for our test statistic under an $s+b$ or BG-only hypothesis with the bounds depending on the observed statistic t_{obs}. We then vary our signal strength parameter μ until the CL_s value is ≤ 0.5 and the corresponding cross-section is said to be excluded at a 95% confidence level. These are the *observed* limits of the study. To set *expected* limits where there is no observed test statistic, we use the test statistic corresponding to our signal sample MC yields in the SR.

As dictated by CMS collaboration guidelines, the data in our SR remains blinded, thus final results presented in this dissertation concerns the last of the scenarios mentioned: to set expected limits.

14.2 Results

The final BG yields with full Run II luminosity in the $m_T(\tau_h, E_T^{miss})$ distribution, along with a benchmark scenario for MSSM and LQ-portal DM signal, are plotted in Figure 14.1. For MSSM, the benchmark parameters are as follows: $m_{\tilde{\chi}_1^\pm} = 300$ GeV, $m_{\tilde{\chi}_1^0} = 250$ GeV, $m_{\tilde{\tau}} = 275$ GeV. For the LQ-portal DM benchmark scenario they are: $m_{LQ} = 750$ GeV, $m_X = 375$ GeV, $\Delta m = 25$ GeV. The uncertainties shown are purely statistical. A comprehensive analysis of systematic uncertainties has also been conducted as outlined in Section 13, and the resulting uncertainties are added as NP's and incorporated into our fitting framework for limit calculations. The expected 95% confidence level upper limits on production cross-sections are shown in Figure 14.2, and the year-by-year BG yields by process are listed in Tables 14.1, 14.2, and 14.3.

2016							
	Single Top	Diboson	DY+jets	Rare Processes	TTbar	W+jets	QCD (data-driven)
0-10	26.9 ± 2.9	36.7 ± 2.3	104.8 ± 4.8	106.7 ± 5.6	128.3 ± 2.6	1232.4 ± 21.4	87.8 ± 7.7
10-20	30.5 ± 3.0	32.3 ± 2.1	61.3 ± 3.6	89.7 ± 5.1	129.6 ± 2.6	1137.2 ± 20.5	61.9 ± 7.5
20-30	28.1 ± 2.8	29.9 ± 2.0	29.1 ± 1.9	83.5 ± 5.0	129.4 ± 2.6	1054.6 ± 19.9	45.1 ± 7.0
30-40	28.7 ± 2.8	28.9 ± 2.0	21.8 ± 1.7	67.9 ± 4.6	109.5 ± 2.4	957.8 ± 18.9	52.3 ± 7.2
40-50	23.5 ± 2.5	19.7 ± 1.6	16.9 ± 1.4	47.5 ± 3.8	85.6 ± 2.1	777.5 ± 17.1	54.9 ± 6.7
50-60	16.2 ± 2.1	15.7 ± 1.5	15.7 ± 1.4	28.6 ± 2.8	63.1 ± 1.8	604.4 ± 15.4	56.2 ± 6.4
60-70	10.1 ± 1.6	12.8 ± 1.32	12.5 ± 1.3	24.3 ± 2.6	38.7 ± 1.4	409.3 ± 12.5	31.9 ± 10.4
70-80	6.1 ± 1.2	6.3 ± 0.9	7.1 ± 1.0	9.9 ± 1.7	24.5 ± 1.1	219.2 ± 9.5	49.3 ± 5.0
80-90	3.1 ± 1.0	4.4 ± 0.8	3.7 ± 0.7	2.7 ± 0.8	13.9 ± 0.8	95.2 ± 6.1	43.3 ± 4.4
90-100	1.9 ± 0.7	3 ± 0.6	1.5 ± 0.4	2.0 ± 0.7	7.9 ± 0.6	36.7 ± 3.8	48.1 ± 4.2
100-110	0.8 ± 0.5	2.1 ± 0.5	1.3 ± 0.4	1.7 ± 0.7	5.7 ± 0.5	27.8 ± 3.4	47.6 ± 4.3
110-120	0.9 ± 0.6	3 ± 0.8	0.7 ± 0.2	2.0 ± 0.8	5.8 ± 0.5	27.2 ± 4.1	59.5 ± 4.5
120-140	1 ± 0.5	4.7 ± 0.8	1.8 ± 0.5	3.5 ± 0.9	10.8 ± 0.7	54.5 ± 5.4	198.9 ± 8.0
140-160	0.4 ± 0.1	5.5 ± 1.1	1.3 ± 0.4	5.3 ± 1.2	13.9 ± 0.8	73.6 ± 6.7	300.7 ± 9.6
160-180	1.2 ± 0.5	4.4 ± 0.8	0.8 ± 0.3	3.6 ± 1.1	9.4 ± 0.7	42.8 ± 4.6	175.2 ± 7.6
180-200	0.7 ± 0.4	2.4 ± 0.6	0.8 ± 0.5	0.8 ± 0.5	5.4 ± 0.5	27.7 ± 4.0	103.3 ± 5.5
200-500	0.6 ± 0.4	2.3 ± 0.6	0.2 ± 0.1	0.6 ± 0.4	2.9 ± 0.4	15.3 ± 2.4	55.7 ± 4.0

Table 14.1: 2016 Background MC yields with corresponding statistical uncertainties.

14.3 Discussion

We have studied the expected limits in a search for new physics at the LHC reflecting $L_{int} = 137.6 fb^{-1}$ of data collected in the CMS detector, encompassing all of Run II. A signal region topology including one soft

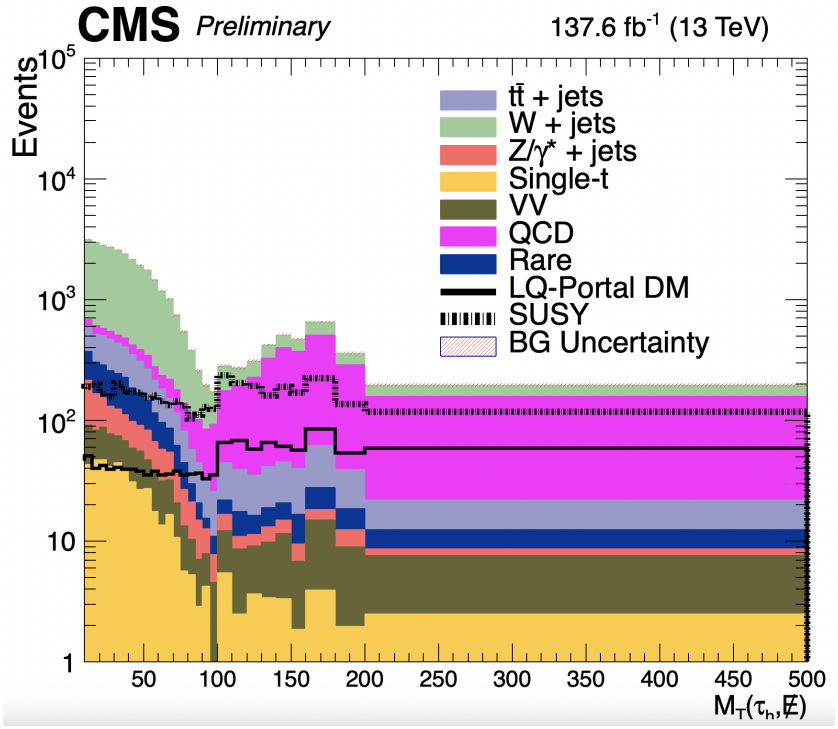

Figure 14.1: Expected BG and signal yields in SR with full Run II luminosity. LQ model parameters: $m_{LQ} = 750$ GeV, $m_X = 375$ GeV, $\Delta m = 25$ GeV. MSSM parameters: $m_{\tilde{\chi}_1^\pm} = 300$ GeV, $m_{\tilde{\chi}_1^0} = 250$ Gev, $m_{\tilde{\tau}} = 275$ GeV. This is not the post-fitting final yield, and only statistical uncertainties are displayed in error bands.

τ_h, a high-p_T ISR jet, and significant E_T^{miss} has been optimized to provide previously unachieved sensitivity in compressed mass spectrum regions. This phase space is crucial for searches involving BSM models featuring coannihilation mechanisms that can significantly contribute to the DM reduction cross-section and yield agreement with astronomical observations. We interpret our results in the context of two models. The first is an R-parity conserving MSSM, where our coannihilation partners are the SUSY LSP $\tilde{\chi}_1^0$ and the $\tilde{\tau}$. The heavier $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$ are wino-like, and $\tilde{\chi}_1^0$ is bino-like. We additionally consider a simplified LQ-portal DM model, where a generic, majorana fermion DM particle coannihilates with a dirac fermion partner X via a LQ mediator. Both scenarios involve production mechanisms with cascade decays to our DM candidate and soft τ_h's, which motivates the use of an ISR jet to boost our system. We have used updated, state-of-the art object reconstruction and identification algorithms, conducted a robust background estimation, and included a complete set of systematic uncertainties. Though the data remains blinded and observed limits remain to be seen, we have computed expected yields as displayed in Figure 14.2. These indicate expected exclusion power for masses up to ~ 340 GeV. Considering that the most recent SUSY searches at the LHC looking for 3rd generation decay products don't surpass the bounds set by the LEP experiment in compressed mass

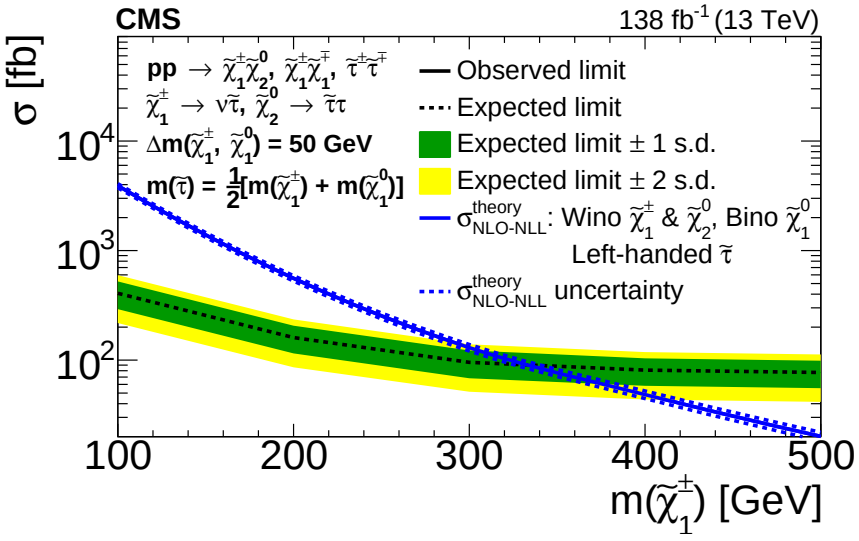

Figure 14.2: 95% confidence level upper limits on production cross-sections as a function of $m_{\tilde{\chi}_1^\pm}$

regions (see section 3.4), this is a significant improvement and demonstrates the power of looking in boosted topologies when targeting this parameter space. We also improve the limits set by the previous iteration of this analysis, SUS-19-002 (Figure 14.3). The expected upper limit on the cross-section is understood to be mainly driven by signal acceptance and the kinematics in the fit variable. Furthermore, our signal acceptance is mainly driven by the mass difference between coannihilating partners, ΔM, since this value constrains the momentum of the final state particles. The expected signal yield for a benchmark scenario under our LQ-portal DM model is displayed in Figure 14.1. As expected, this shows a similar distribution to that of our MSSM model in $m_T(E_T^{miss}, \tau_h)$ in SR. This is an indicator that we should expect to see similar 95% upper limits on cross-sections in the LQ-portal DM case. Because of this fact, and because the LQ production cross-sections are larger than those of chargino/neutrinalino production, we expect to set similar or improved limits on the LQ mass within the DM context we consider. This analysis is the first to consider the LQ-portal DM model with enhanced couplings to 3rd generation leptons, thus the results to be evaluated once official CMS-produced MC samples are made available will be the first of their kind.

2017							
	Single Top	Diboson	DY+jets	Rare Processes	TTbar	W+jets	QCD (data-driven)
0-10	29.3 ± 3.1	37.2 ± 3.1	113.9 ± 3.7	75.9 ± 3.8	139.8 ± 2.5	1442.2 ± 20.7	64.4 ± 4.6
10-20	34.2 ± 3.1	39.0 ± 3.1	59.9 ± 2.6	74.4 ± 3.9	149.3 ± 2.6	1362.4 ± 20.1	45.8 ± 4.9
20-30	34.6 ± 3.0	35.6 ± 3.0	36.1 ± 1.9	66.6 ± 3.7	141.1 ± 2.5	1244.7 ± 19.3	42.9 ± 4.2
30-40	31.6 ± 2.7	27.9 ± 2.7	27.3 ± 1.6	59.3 ± 3.5	123.2 ± 2.3	1123.1 ± 18.4	42.3 ± 4.1
40-50	19.7 ± 2.8	30.0 ± 2.8	23.0 ± 1.6	46.6 ± 3.0	102.8 ± 2.1	920.1 ± 16.5	35.0 ± 3.9
50-60	15.1 ± 2.2	19.0 ± 2.2	19.8 ± 1.5	31.0 ± 2.5	77.0 ± 1.8	739.4 ± 15.0	32.8 ± 3.6
60-70	12.2 ± 1.8	12.7 ± 1.8	13.9 ± 1.3	14.7 ± 1.7	51.6 ± 1.5	485.1 ± 12.4	42.2 ± 3.4
70-80	6.4 ± 1.6	9.7 ± 1.6	9.7 ± 1.0	11.9 ± 1.5	29.9 ± 1.2	292.0 ± 9.7	30.7 ± 3.0
80-90	4.3 ± 0.8	2.4 ± 0.8	7.1 ± 1.0	4.6 ± 0.9	16.8 ± 0.8	121.7 ± 6.4	30.7 ± 2.7
90-100	2.1 ± 0.8	2.9 ± 0.8	2.0 ± 0.4	2.1 ± 0.6	10.1 ± 0.7	56.1 ± 4.4	29.0 ± 2.8
100-110	2.5 ± 1.0	3.8 ± 1.0	1.2 ± 0.3	1.6 ± 0.6	8.6 ± 0.6	27.1 ± 3.0	37.0 ± 4.2
110-120	0.9 ± 0.9	3.3 ± 0.9	0.4 ± 0.1	1.8 ± 0.6	6.5 ± 0.5	22.5 ± 2.7	36.9 ± 2.7
120-140	3.5 ± 1.2	5.6 ± 1.2	1.9 ± 0.5	3.5 ± 0.8	13.6 ± 0.8	47.2 ± 4.3	127.3 ± 4.9
140-160	3.0 ± 1.2	5.6 ± 1.2	2.0 ± 0.6	3.6 ± 0.8	15.5 ± 0.8	56.2 ± 4.9	161.9 ± 5.5
160-180	0.5 ± 1.2	6.0 ± 1.2	1.1 ± 0.3	3.1 ± 0.8	11.7 ± 0.7	44.4 ± 4.2	118.4 ± 4.4
180-200	0.0 ± 1.0	3.9 ± 1.0	1.2 ± 0.4	3.0 ± 0.8	6.1 ± 0.5	19.6 ± 2.8	60.2 ± 3.2
200-500	0.6 ± 0.8	2.5 ± 0.8	0.2 ± 0.1	1.1 ± 0.5	3.0 ± 0.4	12.3 ± 2.0	33.9 ± 2.4

Table 14.2: 2017 Background MC yields with corresponding statistical uncertainties.

2018							
	Single Top	Diboson	DY+jets	Rare Processes	TTbar	W+jets	QCD (data-driven)
0-10	36.9 ± 4.8	59.9 ± 4.8	137.2 ± 12.9	125.0 ± 6.2	154.1 ± 2.5	1954.7 ± 25.0	76.3 ± 5.3
10-20	37.8 ± 4.8	58.9 ± 4.8	77.9 ± 2.8	105.2 ± 5.6	157.1 ± 2.5	1866.0 ± 24.4	66.3 ± 4.9
20-30	35.2 ± 4.5	52.5 ± 4.5	47.3 ± 2.2	90.2 ± 5.3	151.9 ± 2.5	1733.6 ± 23.8	57.1 ± 4.8
30-40	34.1 ± 3.8	37.5 ± 3.8	37.2 ± 1.9	88.7 ± 5.2	137.0 ± 2.3	1584.7 ± 22.7	52.3 ± 4.7
40-50	22.1 ± 3.9	40.7 ± 3.9	29.7 ± 1.7	60.1 ± 4.3	109.8 ± 2.1	1277.7 ± 20.2	53.6 ± 4.7
50-60	19.9 ± 3.1	24.3 ± 3.1	23.2 ± 1.5	42.7 ± 3.7	81.9 ± 1.8	1066.4 ± 19.0	45.6 ± 4.2
60-70	11.9 ± 3.1	24.7 ± 3.1	18.0 ± 1.3	28.2 ± 3.0	58.4 ± 1.5	717.1 ± 15.8	45.4 ± 3.8
70-80	6.2 ± 2.1	12.3 ± 2.1	11.1 ± 1.1	18.2 ± 2.4	35.2 ± 1.2	388.6 ± 11.8	39.4 ± 3.4
80-90	2.8 ± 1.8	8.2 ± 1.8	6.5 ± 0.8	7.1 ± 1.5	18.7 ± 0.9	168.2 ± 7.8	38.3 ± 5.2
90-100	1.7 ± 1.4	5.1 ± 1.4	4.0 ± 0.7	2.6 ± 0.9	11.9 ± 0.7	75.7 ± 5.1	42.8 ± 2.9
100-110	2.3 ± 1.2	3.6 ± 1.2	1.3 ± 0.4	1.2 ± 0.6	7.7 ± 0.5	42.6 ± 4.3	47.8 ± 2.9
110-120	1.0 ± 0.8	1.9 ± 0.8	1.4 ± 0.4	2.4 ± 0.8	7.8 ± 0.5	27.0 ± 3.1	56.1 ± 3.1
120-140	2.7 ± 1.6	7.0 ± 1.6	1.7 ± 0.4	4.0 ± 1.1	14.7 ± 0.8	79.8 ± 6.3	148.6 ± 5.2
140-160	2.9 ± 1.7	8.2 ± 1.7	2.8 ± 0.6	5.4 ± 1.3	15.2 ± 0.8	96.0 ± 7.0	223.7 ± 6.1
160-180	2.1 ± 1.6	6.3 ± 1.6	1.3 ± 0.4	4.4 ± 1.0	11.2 ± 0.6	53.7 ± 4.8	155.3 ± 7.7
180-200	1.1 ± 1.3	5.0 ± 1.3	1.2 ± 0.3	3.0 ± 0.9	7.1 ± 0.5	36.1 ± 4.0	83.1 ± 3.7
200-500	0.6 ± 0.9	2.3 ± 0.9	0.4 ± 0.1	1.7 ± 0.7	2.9 ± 0.3	16.3 ± 2.4	47.4 ± 2.8

Table 14.3: 2018 Background MC yields with corresponding statistical uncertainties.

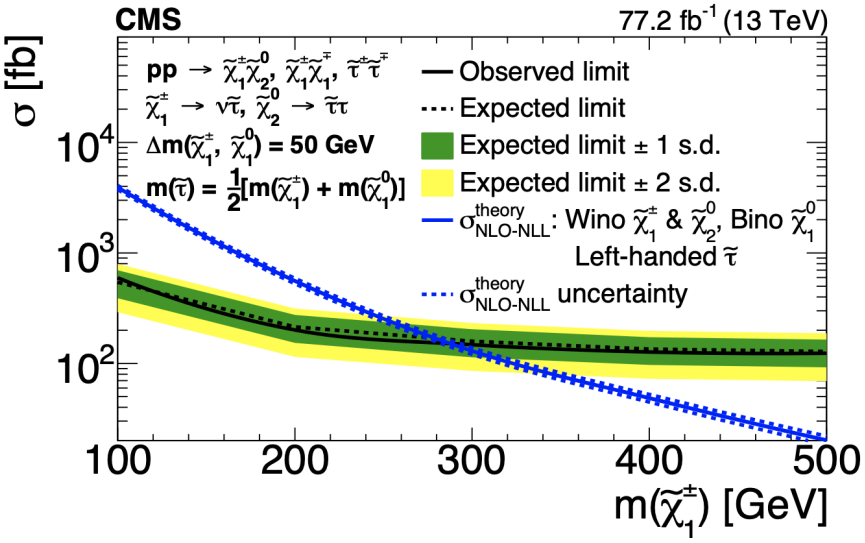

Figure 14.3: SUS-19-002 95% confidence level upper limits on production cross-sections as a function of $m_{\tilde{\chi}_1^{\pm}}$.

CHAPTER 15

Phenomenology at the LHC: Composite Bosons Leptoquarks from strongly-interacting Standard Model fermions via four-fermion operators of NJL type

15.1 Theory and Motivations

The Nambu–Jona-Lasinio (NJL) model was inspired by the Bardeen-Cooper-Schrieffer theory of superconductivity in condensed matter physics, whereby electrons form cooper pairs, resulting in superconductivity phenomena. NJL models have proposed solutions to the fermion mass heirarchy problem and explain the existence of the higgs as a tight composite state of two fermions [80]. In this study we consider four-fermion operators of the NJL type,

$$\sum_{f=1,2,3} G\left[\bar{\psi}_L^f \psi_R^f \bar{\psi}_R^f \psi_L^f\right]_{q=0,-1,2/3,-1/3}, \tag{15.1}$$

where q denotes the charge sector and f the family. The coupling G has an infrared (IR) fixed point at the weak coupling and a ultraviolet (UV) fixed point at the strong coupling. At the IR fixed point, these operators give rise to a $t\bar{t}$ composite particle candidate to the Higgs. At the strong coupling fix point, this model predicts the formation of composite fermions and bosons from bound states of two and three SM fermions, respectivley, which couple to their constituents via effective, contact interactions as displayed in Figure. 15.3

The operators described by Equation 15.1 allow for the formation of many quark-lepton composite boson bound states, which we also here refer to as leptoquarks (LQ). These are listed in Table 15.1. Is should be noted that many other non-LQ composite boson and fermion states are possible under this model, but they

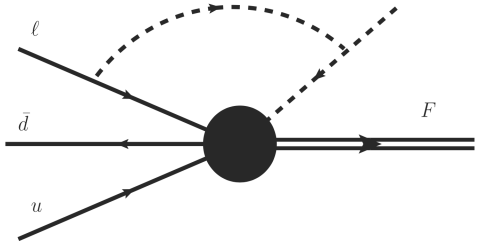

Figure 15.1: Composite Fermion Production

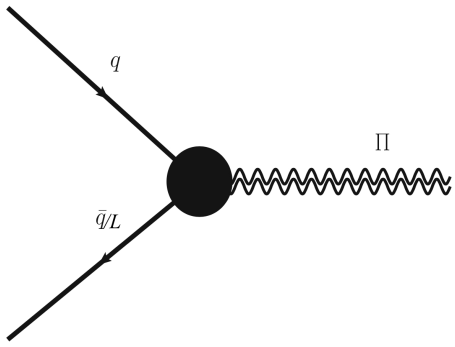

Figure 15.2: Composite Boson Production

Figure 15.3: Contact interactions described by NJL-type operators [24]

LQ bosons Π_a^Q	Charge $Q_i = Y + t_{3L}^i$,	$SU_L(2)$ 3-isospin t_{3L}^i,	$U_Y(1)$ charge Y,	$SU_c(3)$ color a
$\Pi_a^{+5/3} \propto \bar{e}_R u_{La}$	$+5/3$	$+1/2$	$+7/6$	3
$\Pi_a^{-1/3} \propto \bar{\nu}_R^e d_{La}$	$-1/3$	$-1/2$	$+1/6$	3
$\Pi_{u_a}^{+2/3} \propto \bar{\nu}_R^e u_{La}$	$+2/3$	$+1/2$	$+1/6$	3
$\Pi_{d_a}^{+2/3} \propto \bar{e}_R d_{La}$	$+2/3$	$-1/2$	$+7/6$	3
$\Pi_a^{-5/3} \propto \bar{u}_{Ra} e_L$	$-5/3$	$-1/2$	$-7/6$	3
$\Pi_a^{+1/3} \propto \bar{d}_{Ra} \nu_L^e$	$+1/3$	$+1/2$	$-1/6$	3
$\Pi_{u_a}^{-2/3} \propto \bar{u}_{Ra} \nu_L^e$	$-2/3$	$+1/2$	$-7/6$	3
$\Pi_{d_a}^{-2/3} \propto \bar{d}_{Ra} e_L$	$-2/3$	$-1/2$	$-1/6$	3

Table 15.1: Quark-lepton composite bosons identified by their constitute fermions and SM gauge charges.

are not the focus of our work. A previous paper done by many of the same collaborators working on this updated study explored the composite fermion case, where many new final states currently not searched for at the LHC were identified [24].

15.2 Experimental Status

In contrast with the novel LQ-portal DM model considered in the main body of this dissertation, the properties of the LQ in the context of this phenomenological study that we expect to see at colliders are in line with the traditional LQ, that is, with coupling to SM leptons and quarks and with no DM element. The LQ with properties associated with this model has been extensively searched for at the LHC. The primary production mechanisms targeted in LHC searches are single resonant production, pair production, and t-channel production [81]. Feynman diagrams depicting these can be seen in Figure 15.4. Current scalar LQ mass limits for Run II at the LHC have been set at ~ 1.7 TeV [82].

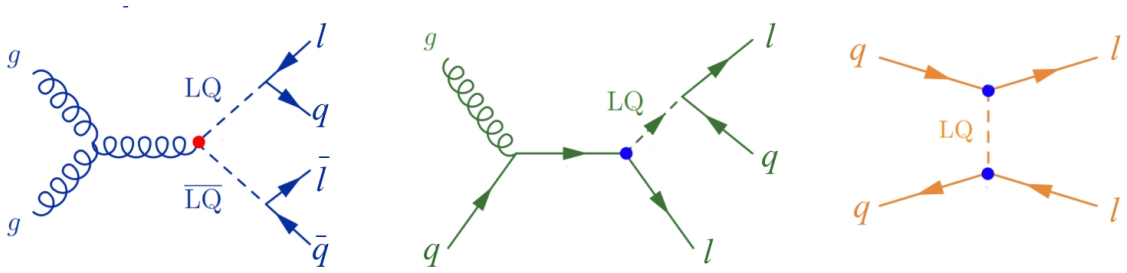

Figure 15.4: LQ production mechanisms subject to current LHC searches: pair production (left), single production (center), non-resonant t-channel (right).

15.3 Strategy

The signal events considered in traditional searches are designed to specifically target the processes displayed in the diagrams in Figure 15.4 by the explicit requirement of a resonant LQ (in the resonant cases) or the

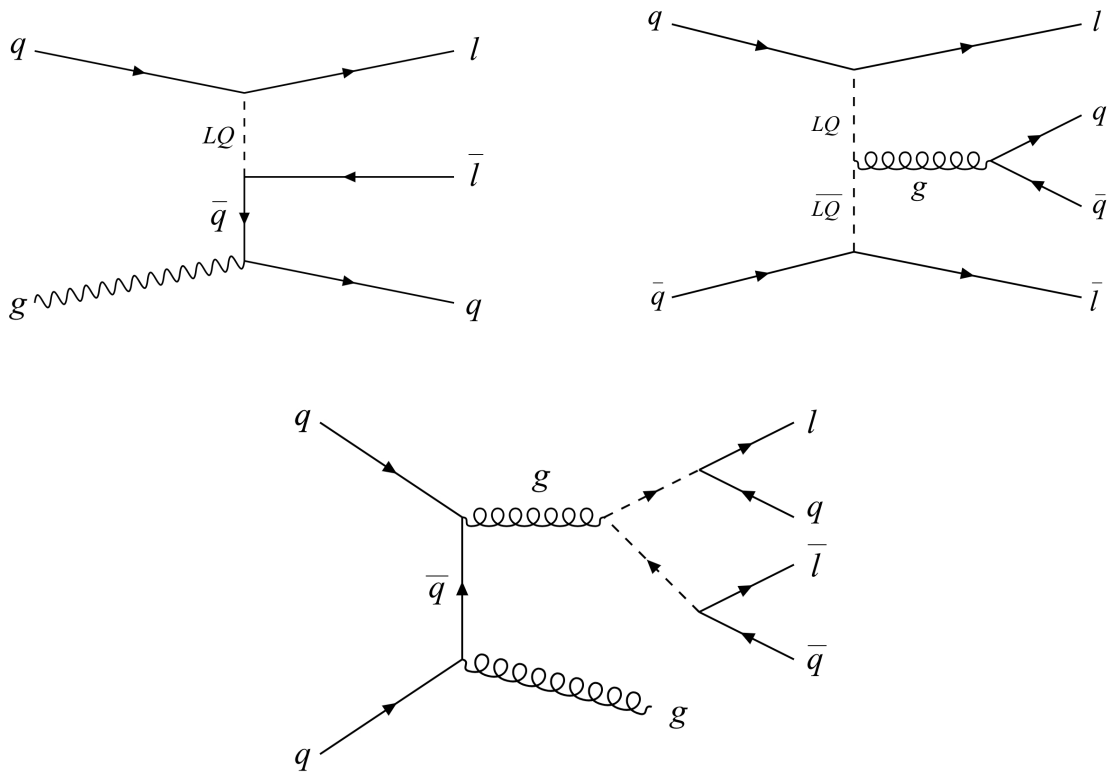

Figure 15.5: Single and pair production of LQs in association with 1 jet (top-left), 2 jets (top-right) and 3 jets (bottom) not currently considered in LHC searches.

presence of only two leptons at generator-level (in the t-channel case). It is possible to take a wider, more inclusive approach to studying LQ production mechanisms at the LHC by generating all Feynman diagrams containing LQs produced in association with zero, one, two, and three jets. When taking this approach, we find that there are processes different from those considered in current LHC searches that could contribute to the cross-section in a non-negligible way. An example set of these new diagrams are displayed in Figure 15.5. If there exists a kinematic variable with a signal signature that provides good distinguishability from BG processes, this could provide a notable enhancement in sensitivity compared to those of current LQ searches. We use the MadGraph5 MC generator to calculate cross-sections and produce large samples of MC events under this signal hypothesis, first studying the $\Pi_a^{+5/3}$ composite boson. This LQ is the bound state of the $e_R u_{La}$ field quanta, as described in Table 15.1. After generating parton-level events, these are showered via the Pythia8 generator, and detector effects are added via the Delphes simulator. The scalar sum of jet transverse momenta, S_T, is a kinematic variable often used in new physics searches as a fit variable. Because new physics searches typically involve new particles with masses $\mathcal{O}(TeV)$, the events producing these tend to be higher in energy than those producing SM objects. High jet activity is often present in higher-energy

events, thus, signal is expected to have a dominating presence at high S_T along with lower SM BG content. Another kinematic variable expected to provide good signal/BG separation is $\chi = exp(|\eta_{e1} - \eta_{e2}|$. We use a 2-dimensional S_T vs χ distribution as our fit variable in this study. The selections comprising our SR in this study are summarized in Table **??**. We implement a $E_T^{miss} < 50$ GeV requirement due to the lack or real E_T^{miss} in our signal process, and we veto b-jets because, since we are looking at a LQ composed of first generation fermions, we don't expect to see third generation decay products.

Selection	Cut
$N(e)$	2
$p_T(e)$	> 20 GeV
$\|\eta(e)\|$	< 2.5
$p_T(j)$	> 20 GeV
$\|\eta(j)\|$	< 5
E_T^{miss}	< 50 GeV
$N(b-jet)$	0

Table 15.2: Basic SR selections for inclusive $\Pi_a^{+5/3}$ production study.

15.4 Discussion and Future Work

We analyze the expected sensitivity of our study given a complete set of relevant SM BG. The 5σ exclusion curve corresponding to inclusive $\Pi_a^{+5/3}$ production, at an LHC $L_{int} = 300\ fb^{-1}$ (expected after RunIII), and $\sqrt{s} = 13$ TeV is shown in Figure 15.6. The LQ masses we consider range from 25 GeV to 30 TeV, and the range of couplings G from 0.5 to 2.5. It is important to point out that these results are preliminary and require further validation and development, but it shows promising exclusion of masses up to ~ 4 TeV for our lowest coupling, surpassing those currently quoted from Run II LQ searches at the LHC.

Current work in this project also involves the inclusion of processes initiated by leptons produced within incoming protons in the LHC beam. Direct production of the LQ in this manner, as shown in Figure 15.7, has recently been studied and found to be a valuable addition to currently studied production mechanisms at the LHC [25]. We aim to extend this work by considering the production of LQ via quark-lepton interactions in the jet-inclusive context of our study.

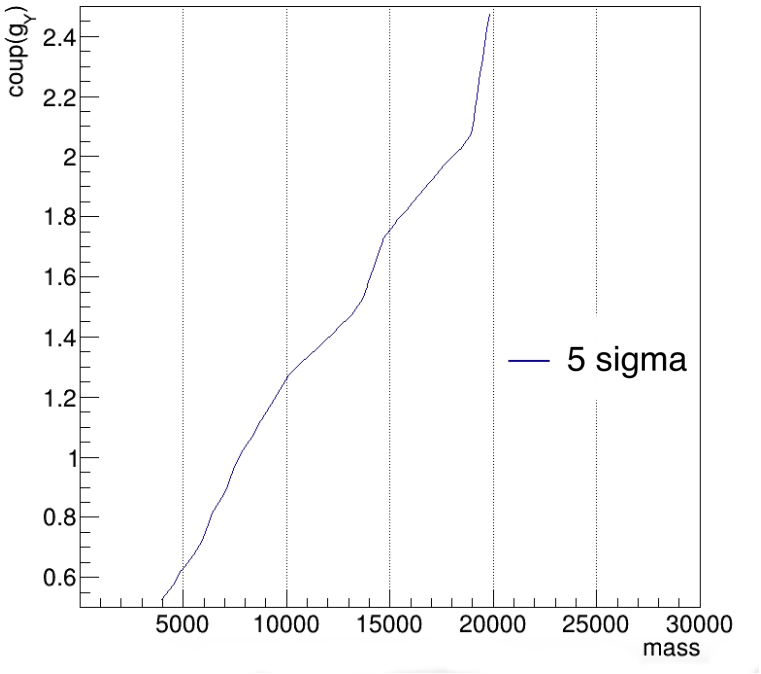

Figure 15.6: 5-sigma exclusion limit for jet-inclusive production of $\Pi_a^{+5/3}$.

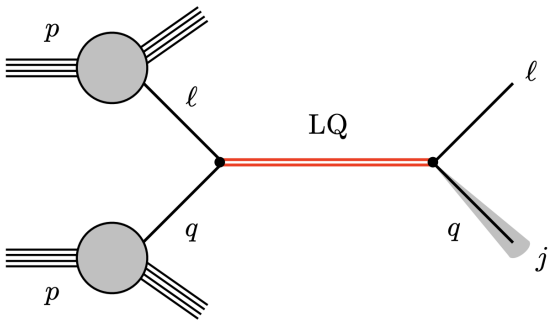

Figure 15.7: Direct LQ production via photon-lepton interactions at the LHC [25].

References

[1] Rainer Hauser Alex Flournoy. Introduction to the standard model of particle physics – part 1. https://inside.mines.edu/ aflourno/Particle/423.shtml, 2019.

[2] Howard E. Haber. Supersymmetry, part i (theory). October 2013.

[3] The CMS Collaboration. Jet algorithms performance in 13 TeV data. Technical report, CERN, Geneva, 2017.

[4] The CMS Collaboration. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13TeV$. *Journal of Instrumentation*, 13(10):P10005–P10005, Oct. 2018.

[5] Matthias Neubert. B Decays and the Heavy-Quark Expansion. In *Heavy Flavours II*, pages 239–293. World Scientific, June 1998.

[6] CERN Heavy Flavor Averaging Group (HFLAV). Preliminary average of r(d) and r(d*) for winter 2023. Presented at the CERN LHC seminar 21/Mar/2023.

[7] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Physics Reports*, 267(5):195–373, 1996.

[8] Michael J. Baker, Joachim Brod, Sonia El Hedri, Anna Kaminska, Joachim Kopp, Jia Liu, Andrea Thamm, Maikel de Vries, Xiao-Ping Wang, Felix Yu, and José Zurita. The coannihilation codex. *Journal of High Energy Physics*, 2015(12):1–86, Dec. 2015.

[9] Dris Boubaa, Shaaban Khalil, and Stefano Moretti. Explaining b decays anomalies in SUSY models. *Journal of Physics: Conference Series*, 1766(1):012018, Jan. 2021.

[10] Gregory Ciezarek, Manuel Franco Sevilla, Brian Hamilton, and et al. A challenge to lepton universality in b-meson decays. *Nature*, 546:227–233, 2017.

[11] Lyn Evans, Lucie Linssen, and CERN. The super-lhc is on the starting blocks. *CERN Courier*, July 2008.

[12] Stefanos Dris. Performance of the cms tracker optical links and future upgrade using bandwidth efficient digital modulation. CERN-THESIS-2007-015, 2010.

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

[14] CERN, CMS. Tracking. https://cms.cern/detector/identifying-tracks.

[15] M. Friedl. *The CMS silicon strip tracker and its electronic readout*. PhD thesis, Technische U. Wien, 2001.

[16] Paolo Azzurri. The CMS silicon strip tracker. *Journal of Physics: Conference Series*, 41:127–134, May 2006.

[17] F. Fiori and CMS Tracker Collaboration. CMS Tracker Operational Experience. In *Proceedings of VERTEX2016*, La Biodola, Elba (Italy), Sept. 2016.

[18] Pierluigi Paolucci. The CMS Muon system. In *9th ICATPP Conference on Astroparticle, Particle, Space Physics, Detectors and Medical Physics Applications*, pages 605–615, 2005.

[19] Milos Dordevic. The cms particle flow algorithm. *EPJ Web Conf.*, 191:02016, 2018.

[20] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. The anti-$k_t$jet clustering algorithm. *Journal of High Energy Physics*, 2008(04):063–063, Apr. 2008.

[21] The CMS Collaboration. Jet energy scale and resolution performance with 13 tev data collected by cms in 2016-2018. 2020.

[22] Andrea Cardini. Tau identification exploiting deep learning techniques. *PoS*, ICHEP2020:723, 2021.

[23] The CMS Collaboration. Search for supersymmetry with a compressed mass spectrum in events with a soft τ lepton, a highly energetic jet, and large missing transverso momentum in proton-proton collisions at $\sqrt{s} = 13tev$. *Physical Review Letters*, 124(4), Jan. 2020.

[24] R. Leonardi, O. Panella, F. Romeo, A. Gurrola, H. Sun, and S.-S. Xue. Phenomenology at the LHC of composite particles from strongly interacting standard model fermions via four-fermion operators of NJL type. *The European Physical Journal C*, 80(4), Apr. 2020.

[25] Luca Buonocore, Ulrich Haisch, Paolo Nason, Francesco Tramontano, and Giulia Zanderighi. Lepton-quark collisions at the large hadron collider. *Physical Review Letters*, 125(23), Dec. 2020.

[26] Fermilab. Dark energy survey releases most precise look at the universe's evolution. May 2021.

[27] Gino Isidori. B anomalies: hopes, disillusions, and future prospects.

[28] M.C. Gonzalez-Garcia and Michele Maltoni. Phenomenology with massive neutrinos. *Physics Reports*, 460(1-3):1–129, Apr. 2008.

[29] D. Griffiths. *Introduction to Elementary Particles*. John Wiley Sons, New York, USA, 1987.

[30] Institute of Physics. Do we know how the universe works? https://www.iop.org/explore-physics/big-ideas-physics/standard-modelgref.

[31] Madeleine O'Keefe. Fine-tuning versus naturalness. *Symmetry Magazine*, Jan. 2020.

[32] Muon $g-2$ Collaboration. Measurement of the positive muon anomalous magnetic moment to 0.46 ppm. *Phys. Rev. Lett.*, 126:141801, Apr. 2021.

[33] F. Zwicky. Die Rotverschiebung von extragalaktischen Nebeln. *Helvetica Physica Acta*, 6:110–127, Jan. 1933.

[34] J. Einasto, A. Kaasik, and E. Saar. Dynamic evidence on massive coronas of galaxies. *Nature*, 250:309–310, 1974. Received on April 10, 1974.

[35] Vera C. Rubin and Jr. Ford, W. Kent. Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions. *The Astrophysical Journal*, 159:379, Feb. 1970.

[36] Erin S. Sheldon, David E. Johnston, Morad Masjedi, Timothy A. McKay, Michael R. Blanton, Ryan Scranton, Risa H. Wechsler, Benjamin P. Koester, Sarah M. Hansen, Joshua A. Frieman, and James Annis. Cross-correlation weak lensing of sdss galaxy clusters. iii. mass-to-light ratios. *The Astrophysical Journal*, 703(2):2232, September 2009. Published 2009 September 17 • © 2009. The American Astronomical Society. All rights reserved.

[37] Planck Collaboration. Planck 2018 results. *Astronomy and Astrophysics*, 641:A6, Sept. 2020.

[38] Katherine Garrett and Gintaras Duda. Dark matter: A primer. *Advances in Astronomy*, 2011:1–22, 2011.

[39] BaBar Collaboration. Evidence for an excess of $\bar{B} \to d^{(*)}\tau^{-}\bar{\nu}_{\tau}$ decays. *Physical Review Letters*, 109(10), Sept. 2012.

[40] Belle Collaboration. Measurement of $\mathscr{R}(d)$ and $\mathscr{R}(d^*)$ with a semileptonic tagging method. *Physical Review Letters*, 124(16), Apr. 2020.

[41] LHCb Collaboration. Measurement of the ratios of branching fractions $\mathscr{R}(d^*)$ and $\mathscr{R}(d^0)$. https://arxiv.org/abs/2302.02886.

[42] Brent Follin, Lloyd Knox, Marius Millea, and Zhen Pan. First detection of the acoustic oscillation phase shift expected from the cosmic neutrino background. *Physical Review Letters*, 115(9), Aug. 2015.

[43] Kim Griest and David Seckel. Three exceptions in the calculation of relic abundances. *Phys. Rev. D*, 43:3191–3203, May 1991.

[44] Jinhui Guo, Yuxuan He, Jia Liu, and Xiao-Ping Wang. Heavy long-lived coannihilation partner from inelastic dark matter model and its signatures at the LHC. *Journal of High Energy Physics*, 2022(4), Apr. 2022.

[45] Ilja Doršner, Svjetlana Fajfer, Darius A. Faroughy, and Nejc Košnik. The role of the s 3 GUT leptoquark in flavor universality and collider searches. *Journal of High Energy Physics*, 2017(10), Oct. 2017.

[46] E. Eichten, I. Hinchliffe, K. D. Lane, and C. Quigg. Signatures for technicolor. *Phys. Rev. D*, 34:1547–1566, Sept. 1986.

[47] P. S. Bhupal Dev, Amarjit Soni, and Fang Xu. Hints of natural supersymmetry in flavor anomalies? *Physical Review D*, 106(1), July 2022.

[48] Hyun Min Lee. Leptoquark option for b-meson anomalies and leptonic signatures. *Physical Review D*, 104(1), Jul. 2021.

[49] A Masiero, P Paradisi, and R Petronzio. Anatomy and phenomenology of the lepton flavor universality in SUSY theories. *Journal of High Energy Physics*, 2008(11):042–042, Nov. 2008.

[50] The CMS collaboration. Search for electroweak production of charginos and neutralinos in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of High Energy Physics*, 2022:147, 2022. Article number: 147. Published on 26 April 2022.

[51] CMS Collaboration. Search for direct pair production of supersymmetric partners of τ leptons in the final state with two hadronically decaying τ leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s}$ = 13 tev, 2022.

[52] The CMS Collaboration. Search for dark matter in events with a leptoquark and missing transverse momentum in proton-proton collisions at 13 TeV. *Physics Letters B*, 795:76–99, Aug. 2019.

[53] David Harris. Standard model discoveries. *Symmetry Magazine*.

[54] The ATLAS Collaboration. Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, Sept. 2012.

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

[56] CERN. The large hadron collider. https://home.cern/science/accelerators/large-hadron-collider.

[57] CERN. Cern's accelerator complex. https://home.cern/science/accelerators/accelerator-complex.

[58] CERN. CMS. https://home.cern/science/experiments/cms.

[59] CERN. ATLAS. https://home.cern/science/experiments/atlas.

[60] CERN. ALICE. https://home.cern/science/experiments/alice.

[61] CERN. LHCb. https://home.cern/science/experiments/lhcb.

[62] W. Adam et al. The cms phase-1 pixel detector upgrade, Feb. 2021.

[63] CERN. Silicon Strips. https://cms.cern/detector/identifying-tracks/silicon-strips.

[64] Candan Isik. Phase 1 upgrade of the cms hadron barrel calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1042:167389, 2022.

[65] The CMS Collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. Dec. 2000.

[66] The CMS Collaboration. Particle-flow reconstruction and global event description with the CMS detector. *Journal of Instrumentation*, 12(10):P10003–P10003, Oct. 2017.

[67] CERN. Jets at CMS and the Determination of Their Energy Scale. https://cms.cern/news/jets-cms-and-determination-their-energy-scale.

[68] V. Khachatryan et al 2017. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12(02):P02014–P02014, Feb. 2017.

[69] Izaak Neutelings. Hadronic tau reconstruction and identification performance in ATLAS and CMS. *PoS*, LHCP2020:045, 2021.

[70] The CMS Collaboration. Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *Journal of Instrumentation*, 16(05):P05014, May 2021.

[71] The CMS Collaboration. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *Journal of Instrumentation*, 13(05):P05011–P05011, May 2018.

[72] CMS Monte Carlo production overview. https://opendata.cern.ch/docs/cms-mc-production-overview. Accessed: July 14, 2023.

[73] The CMS Collaboration. Performance of the CMS level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ tev. *Journal of Instrumentation*, 15(10):P10017–P10017, oct 2020.

[74] CMS Collaboration. Measurement of the differential cross sections for the associated production of a w boson and jets in proton-proton collistion at sqrt(s) = 13 tev. *Physical Review D*, 96(7), Oct. 2017.

[75] CMS Collaboration. Measurement of differential cross sections for the production of a z boson in association with jets in proton-proton collisions at $\sqrt{s}$ = 13 tev, 2022.

[76] The CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13$ TeV. Technical report, CERN, Geneva, 2019.

[77] CMS collaboration. Jet uncertainty;cms open data guide. https://cms-opendata-guide.web.cern.ch/analysis/systematics/objectsuncertain/jetmetuncertain/.

[78] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *The European Physical Journal C*, 71(2), Feb. 2011.

[79] A L Read. Presentation of search results: the cls technique. *Journal of Physics G: Nuclear and Particle Physics*, 28(10):2693, Sept. 2002.

[80] Y. Nambu and G. Jona-Lasinio. Dynamical model of elementary particles based on an analogy with superconductivity. i. *Phys. Rev.*, 122:345–358, Apr. 1961.

[81] Francesco Romeo. Search for leptoquark-like signatures with the atlas and cms detectors. *Nuclear and Particle Physics Proceedings*, 273-275:638–643, 2016. 37th International Conference on High Energy Physics (ICHEP).

[82] Nishita Desai and Amartya Sengupta. Status of leptoquark models after lhc run-2 and discovery prospects at future colliders, 2023.