Search for Supersymmetry with a Compressed Mass Spectrum in Events with a Soft τ Lepton, a Highly Energetic Jet, and Large Missing Transverse Momentum in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV

By

Savanna Rae Starko

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

May 8, 2020

Nashville, Tennessee

Approved:

Alfredo Gurrola, Ph.D.
Will Johns, Ph.D.
Paul Sheldon, Ph.D.
Kelly Holley-Bockelmann, Ph.D.
Shane Hutson, Ph.D.

To my parents, Frank and Nanette Starko, extraordinarily caring and wise

and

To my brother, Anthony Starko, like whom I strive to be more and more each day

ACKNOWLEDGEMENTS

Less than two weeks before I began my PhD program, I was diagnosed for the second time with thyroid cancer. Prior to this, an apartment was secured, and boxes were packed. I had everything planned, for plans make me feel secure. More quickly than I could imagine though, the chaos of a second diagnosis infiltrated those plans. I was fearful for what was to come, but I had a choice to make as to whether to continue embarking on the PhD program. Thanks to the support of many along the way, my goal of a PhD in Physics from Vanderbilt University could come to fruition.

First and foremost, I have to thank God, my ultimate protector. Personally, I believe that spirituality is a deep-rooted connection that all beings have to each other, forged by something much greater than us. That something much greater, in my eyes, is God. Over the course of my time in Nashville, I was baptized in the United Methodist Church, accepting Jesus Christ as my Lord and Savior. I came to realize that I thrive most on connection with others. Connection makes triumphs much sweeter and trials more bearable. Connection allows me to share myself in meaningful ways with others, which I believe to be my God-given purpose for my Earthly time.

Second, I must thank my immediate family for the unwavering support they have always shown me. My parents, Frank and Nanette, have never led me to believe that there's anything that I can't do. We've experienced deep adversities together, including my fragile medical condition, but I've always known that we'll be okay because we approach life as a team. "Team Starko" is so much stronger together than each member alone. One of my biggest role models is my brother, Anthony. Though he's younger, he has always shown me care, courage, and wisdom that extends far beyond his years. He's been there to listen throughout this entire process, and for that, I am deeply grateful. He and my future sister-in-law, Katie, are caring and compassionate individuals who extend a hand to those in need without ever giving it a second thought.

My research group has been an integral part of my life for the last several years. I am grateful for Will Johns, Paul Sheldon, and my advisor, Alfredo Gurrola, for their backing of my developing career. What's most important to me are the countless opportunities they've afforded me to mentor other students, building those all-important connections. To Alfredo, I'm most grateful for the times where he's held the space for me to grow as a person, not just a physicist. In particular, his encouragement to apply for the NSF Graduate Research Fellowship allowed me the chance to remember who I am- someone who strives to be a good person and care meaningfully for others. My fellow graduate students have contributed to my growth as a young adult, too. Our times spent outside of the office, whether enjoying dinner or engaging with the community, are some of the happiest memories I'll have of the past several years. My students have enriched my life ten-fold, as I've been fortunate to see them grow in search of each their own purpose, too.

At various times over the course of my PhD program, I've experienced fulfillment in mind, body, and spirit that has ultimately led me to remember my core values and unmatched desire to experience all this life has to offer with those closest to me.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vii

LIST OF FIGURES ix

Chapter

I Introduction 1
The Standard Model of Particle Physics 1
Symmetry Breaking 3
The Higgs Mechanism 4

II Motivation 7
Connecting Small Scale to Large Scale 7
Employing Supersymmetry 7
Motivating the Search for Compressed Mass Spectra 8
Parameterizing the DM Relic Density 8

III The Large Hadron Collider 17

IV The CMS Experiment 20
CMS Geometry 20
The Superconducting Solenoid 21
The Tracker System 21
The Pixel Detector 22
The Silicon Strip Detector 22
The Calorimeters 23
The Electromagnetic Calorimeter 23
The Hadronic Calorimeter 24
The Muon System 24
Data Acquisition and Triggering 25

V Analysis Strategy 26

VI Trigger Efficiency Studies 30

Trigger Efficiency Measurements 30
Tag and Probe Method 30
Muon Efficiencies and Scale Factors 31
Tau Efficiencies and Scale Factors 31
Trigger Efficiency for this Analysis 32

VII Optimization Studies 35

VIII Particle Identification and Event Reconstruction 40
Jet Reconstruction 40
Jet Identification 41
Electron Reconstruction 42
Electron Identification 42
Muon Reconstruction 43
Muon Identification 43
Tau Reconstruction 44
Tau Identification 46
E_T^{miss} Reconstruction 48

IX Data and Monte Carlo Samples 50

X Background Estimation: Boost and Recoil Studies 53
The Importance of $Z(\rightarrow \mu\mu)$+ISR 53
Shifting Focus to Jet Resolution and ISR Modeling for $Z(\rightarrow \mu\mu)$+ISR 55
Conclusions from Boost and Recoil Studies on $Z(\rightarrow \mu\mu)$+ISR 58
Validating ISR Weights on $W(\rightarrow \mu\nu)$+ISR 61

XI Background Estimation: τ_h Identification Studies 66
The Importance of $Z(\rightarrow \tau\tau)$+ISR 66
Estimating the QCD Contribution to the $Z(\rightarrow \tau\tau)$+ISR Control Region 68
Conclusions to Draw from $Z(\rightarrow \tau\tau)$+ISR 69
Emulating Hadronic Tau Kinematics Using Muons 71

XII Background Estimation: $t\bar{t}$ Contribution 75

XIII Background Estimation: QCD Contribution 81

XIV Systematics 90

XV Results and Conclusions 96
Information for Reinterpretation of Results 101

XVI Discussion of Limits 104

REFERENCES 108

LIST OF TABLES

Table		Page
1	The SM Quarks and Leptons [27]	2
2	The SM Bosons [27]	3
3	First Generation Fermionic Charges [27]	4
4	Signal Region Selections	27
5	Contributions to the isolated single-μ trigger efficiency in 2015 data, integrated over $p_T > 22$ GeV.	31
6	Data-to-simulation scale factors for different working points of the MVA-based isolation discriminant, using highly boosted Z/γ^* events decaying to τ lepton pairs [38].	32
7	Event selection criteria for trigger efficiency studies.	33
8	Base event selections for optimization studies	35
9	Optimized event selection criteria	39
10	Loose Jet-ID Selections.	41
11	Tight Jet-ID Selections.	41
12	Electron ID Selections	43
13	μ Identification Criteria	44
14	Reconstructed Tau Decay Modes	46
15	MC Samples (2016)	50
16	Data Samples (2016)	50
17	MC Samples (2017)	51
18	Data Samples (2017)	51
19	$Z(\rightarrow \mu\mu)$+ISR Event Selections	54
20	Event Weights by Z-Boost (2016 & 2017)	58
21	$W(\rightarrow \mu\nu)$+ISR Event Selections	61
22	Background and data yields in the W+jets control regions (2016)	64
23	$Z(\rightarrow \tau\tau)$+ISR Event Selections.	67
24	Background and data yields in the $Z(\rightarrow \tau\tau)$+ISR control region (2016).	71
25	Background and data yields in the $Z(\rightarrow \tau\tau)$+ISR control region (2017).	71
26	Event selection differences between signal region and $t\bar{t}$ control regions.	75
27	$t\bar{t}$ Control Region Event Selections	76

28 Background and data yields in the $t\bar{t}$ control regions (2016). 78
29 Background and data yields in the $t\bar{t}$ control regions (2017). 80

30 Summary of the cuts used for the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ and $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CRs. 83
31 Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the Z region (2016). 85
32 Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the Z region (2017). 85
33 Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W region (2016)..... 87
34 Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W region (2017)..... 87

35 Event Weight Uncertainties by Z-Boost 90
36 Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W+jets and Z+jets regions and their relative differences. 95
37 Systematics values given in percent- "s" indicates "shape" uncertainties. 95

38 Example of some correlated and uncorrelated nuisance parameters. 106

LIST OF FIGURES

Figure Page

1 Higgs potential [11]. 5

2 High potential sensitivity when searching for compressed mass spectra (SUS-17-003). 9

3 Schematic of current LHC set-up [49]. 19

4 CMS detector schematic [25]. 20
5 CMS coordinate geometry [33]. 21
6 CMS pixel detector [44]. 22
7 CMS ECAL detector diagram [14]. 24

8 ISR topology for this study. 27
9 $N(\tau_h)$ for four representative sets of SUSY masses ($m(\tilde{\chi}_1^{\pm})$, $m(\tilde{\tau})$, and $m(\tilde{\chi}_1^0)$) for $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV. 28

10 Trigger correlation between ISR jet and the E_T^{miss}: 50 to 300 GeV (top) and 300 to 600 GeV (bottom). 34

11 Optimization of the $p_T(\tau_h) > X$ criterion, using signal points with LSP masses of 270 GeV, 360 GeV, and 450 GeV. 36
12 Optimization of the $20 < p_T(\tau_h) < X$ criterion, using signal points with LSP masses of 270 GeV, 360 GeV, and 450 GeV. 37
13 Optimization of the $p_T^{\text{lead}}(j)$ criterion, using signal points with LSP masses of 270 GeV and 360 GeV. 38

14 Sketch depicting tau lepton isolation cones. 45
15 Relative τ_h E_T resolution for PF based reconstruction (solid line), calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections. 46
16 Relative p_T resolution of reconstructed τ_h candidates. 47
17 E_T^{miss} resolution for PF-based reconstruction, calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections. 49

18 Z-Boost [GeV] (left) and H_T [GeV] (right) distributions for $Z(\rightarrow \mu\mu)$+ISR (2016). .. 55
19 Jet recoil parallel to Z-Boost [GeV] (2016). 56
20 Profile of $\frac{|r_{\mathrm{T}}|}{|\vec{Z}_{\mathrm{T}}|}$ (2016). 56

21 Z-Boost [GeV] distribution without boost weights (left) and with weights (right) for events in the $Z(\rightarrow \mu\mu)$ + ISR jet control region (2016). 57
22 Z-Boost [GeV] distribution without boost weights for events in the $Z(\rightarrow \mu\mu)$ + ISR jet control region (2017)..... 58
23 Parallel Recoil r_T [GeV] without boost weights (left) and with weights (right) (2016). 59
24 Parallel Recoil r_T [GeV] with weights (2017)..... 59
25 Top Row: Muon p_T without boost weights (left) and with weights (right); Second Row: First leading jet p_T without boost weights (left) and with weights (right); Third Row: H_T without boost weights (left) and with weights (right); Bottom Row: $m_T(\tau_h, E_T^{miss})$ without boost weights (left) and with weights (right) (2016)..... 60
26 Top Row: Muon p_T without boost weights (left) and with weights (right); Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) (2016). 62
27 Top Row: Muon p_T without boost weights (left) and with weights (right) using 2017 data ($E_T^{miss} > 230$ GeV cut). Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) using 2017 data ($E_T^{miss} > 50$ GeV)..... 63
28 Top Row: Muon p_T without boost weights (left) and with weights (right); Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) (2016). 64
29 Results for $W(\rightarrow \mu\nu)$+ISR control region using 2017 data and MC..... 65

30 Mass sideband used to find the QCD OS-to-SS ratio for OS (SS) [left] ([right])..... 69
31 Top Row: $p_T(\tau_h)$ and $m(\tau_h, \tau_h)$; Second Row: $N(j)$ and $p_T(j)$; Third Row: $p_T^{\text{lead}}(j)$ and E_T^{miss} (2016)..... 70
32 $m_T(\tau_h, E_T^{miss})$ in the $Z(\rightarrow \tau\tau)$+ISR control region (2016) 71
33 $m_T(\tau_h, E_T^{miss})$ in the $Z(\rightarrow \tau\tau)$+ISR control region (2017). 72
34 Fraction of momentum carried away by τ_h in τ lepton hadronic decays..... 73
35 Emulated E_T^{miss} and $m_T(\tau_h, E_T^{miss})$ distributions, using muon events to emulate the τ_h kinematics (2016)..... 74

36 $p_T(\tau_h)$ (left) and $m_T(\tau_h, E_T^{miss})$ (right) for events with exactly 1 b-jet and "Tight" ID for τ_h (2016)..... 76
37 Top Row: $p_T(\tau_h)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Middle Row: $N(j)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Next Row: $m_T(\tau_h, E_T^{miss})$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Bottom Row: E_T^{miss} with exactly 1 b-jet (left) and for exactly 2 b-jets (right) for events with "Tight" τ_h ID (2016). 77

38 Top Row: $p_T(\tau_h)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Second Row: $N(j)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Third Row: E_T^{miss} with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Bottom Row: m_T with exactly 1 b-jet (left) and for exactly 2 b-jets (right) for events with "VTight" ID for τ_h (2016). 79
39 Results for $t\bar{t}$+ISR control regions using 2017 data. The top left plot is for CR1, top right for CR2, bottom left for CR3, and bottom right for CR4. 80

40 Sketch for ABCD methodology for the QCD background estimation. 81
41 Feynman diagram of $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ and $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$. 82
42 Top (Bottom) row: $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ CR w/ (1) "Tight" isolation and (2) passing "Loose" but failing "Tight" for 2016 (2017). 84
43 $p_T(\tau_h)$ dependence of the tight-to-loose ratio in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ region for 2016 (2017), left (right). 84
44 Top (Bottom) row: $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CR w/ (1) "Tight" isolation and (2) passing "Loose" but failing "Tight" for 2016 (2017). 86
45 $p_T(\tau_h)$ dependence of the tight-to-loose ratio in the $W(\rightarrow \mu\nu) + \tau_h^{fake}$ region for 2016 (2017), left (right) 86
46 2D histogram representing the predicted yield of QCD multijet events in CR C (left) and CR D (right), as a function of m_T and $p_T(\tau_h)$. 88
47 Final yield of QCD estimate in CR D. Response 2D histogram from CR C is re-weighted for Z CR (left) and W CR (right) for 2016 (top row) and 2017 (bottom row). 88
48 CR D yield from QCD prediction using the shape extracted from MC in CR C in 2016 (2017), left (right). 89

49 $m_T(\tau_h, E_T^{miss})$ for Z+jets (top) and for W+jets (bottom) in the SR. 91
50 FastSim MET correction for signal 93
51 Data/MC ratios, as a function of m_T, in the $Z(\rightarrow \tau\tau)$ and $t\bar{t}$ control regions that are used to assign shape-based systematics. 94

52 Expected signal and background m_T distributions in the search region with 2016 data (top) & 2017 data (bottom) iterations. 97
53 ISR topology for this study. 97
54 Signal acceptance for cascading production of $\tilde{\tau}$s as a function of $m(\tilde{\chi}_1^{\pm})$, where $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV and $m(\tilde{\tau}) = \frac{m(\tilde{\chi}_1^{\pm}) + m(\tilde{\chi}_1^0)}{2}$ 98

55 The 95% confidence level upper limits on the production cross sections ($\sigma_{95\%\,CL}$) as a function of $m(\tilde{\chi}_1^{\pm})$. The solid blue line shows the theoretical cross section, and the dashed blue line its uncertainty. The observed limit is shown with the solid black line, while the expected limit is shown with the dashed black line. The green (yellow) band corresponds to the one (two) standard deviation range about the central value of the expected limit. 98
56 2016 SR yields for BG and signal, as a function of m_T bin. Two benchmark signal samples are considered: (i) $m(\tilde{\chi}_1^{\pm}) = 200$ GeV, $m(\tilde{\chi}_1^0) = 150$ GeV; (ii) $m(\tilde{\chi}_1^{\pm}) = 300$ GeV, $m(\tilde{\chi}_1^0) = 250$ GeV. 99
57 2017 SR yields for BG and signal, as a function of m_T bin. Two benchmark signal samples are considered: (i) $m(\tilde{\chi}_1^{\pm}) = 200$ GeV, $m(\tilde{\chi}_1^0) = 150$ GeV; (ii) $m(\tilde{\chi}_1^{\pm}) = 300$ GeV, $m(\tilde{\chi}_1^0) = 250$ GeV. 99
58 Signal acceptance for direct $\tilde{\tau}$ production of as a function of $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$ and specified values of $m(\tilde{\tau})$. 100
59 95% C.L. upper limits on the direct $\tilde{\tau}$ pair production signal cross sections, as a function of $m(\tilde{\tau})$ and $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$. 101
60 95% CL upper limits on the signal cross-section (for the electroweakino model) for each m_T bin and for two benchmark signal samples: (i) $m(\tilde{\chi}_1^{\pm}) = 200$ GeV (top); (ii) $m(\tilde{\chi}_1^{\pm}) = 300$ GeV (bottom). 102

61 Pulls and impact plots of the nuisance parameters used for 2016 limit calculation. 105
62 Goodness of fit test for the combination: The black histogram represents the probability density function for the test statistic, obtained using 500 toys, assuming the "saturated" model as prescribed by the CMS statistics committee. The blue arrow represents the observed value of the goodness of fit indicator (i.e. using the data). The p-value is the integral above the observed value, which represents the probability of obtaining the observed data distributions assuming a null hypothesis. 107

Chapter I

Introduction

The Standard Model of Particle Physics

Particle Physics is a subdiscipline of Physics that bridges connection between the largest and smallest scale aspects of the universe. On the large scale, particle physicists consider, for instance, the formation and rotation of galaxies. In contrast, particle physicists examine the fundamental particles, the building blocks of which the universe is comprised, on the small scale. These particles are considered fundamental in that they presumably cannot be broken down into smaller constituent parts. The interaction of quarks as protons collide is an example of a small scale interaction. For decades, Particle Physics as a discipline has relied on the Standard Model (SM) to describe the fundamental particles of the universe and the ways in which those particles interact [34]. The SM is divided into two main subcategories: fermions and bosons. Fermions are particles with half-integer spin and include quarks and leptons, whereas bosons are particles with integer spin. The four fundamental forces through which these particles may interact are the strong, weak, electromagnetic, and gravitational forces.

We begin by considering the properties of the quarks. There are six known quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). These quarks are arranged in three generations of couplets: u and d, c and s, and t and b, where each subsequent generation is higher in mass than the previous. Quarks have electric charge and color charge. The color charge dictates the strong force interactions in which a quark can participate. Since the quarks possess electric charge, they may interact through the electromagnetic force. A quark's anti-quark has similar properties but opposite electric charge. In addition, quarks may participate in weak force interactions. A bound state of two quarks is known as a meson. An example of a meson is a π^0, which could be a bound state of an u and an anti-up ($\bar{u}$) quark. When three quarks are bound, this is a baryon. An example of a baryon is the proton, a bound state of two u quarks and a d quark. Mesons and baryons collectively comprise the hadrons. According to the concept of asymptotic freedom, we are unable to observe a single unbound quark. Quarks exist in bound states because the strength of the force between them increases with the distance by which they are separated.

Next, we consider the properties of the leptons. There are six known leptons: electron (e), electron neutrino (ν_e), muon (μ), muon neutrino (ν_μ), tau (τ), and tau neutrino (ν_τ). Like the quarks, these leptons are arranged in three generations of couplets: e and ν_e, μ and ν_μ, and τ and ν_τ, where each subsequent generation again represents a higher mass state than the previous. For example, the mass of the τ lepton is nearly four orders of magnitude larger than the mass of

the e. Leptons have electric charge, except for the neutrinos, which are chargeless. Hence, the leptons interact primarily via the weak and electromagnetic forces. Key properties of the quarks and leptons are described to follow in Table 1, where e is the magnitude of the charge on a single electron: 1.6×10^{-19} C.

QUARK	Mass	Charge	Doublet Partner
u	~2.4 MeV/c^2	$+\frac{2}{3}$e	d
d	~4.8 MeV/c^2	$-\frac{1}{3}$e	u
c	~1.3 GeV/c^2	$+\frac{2}{3}$e	s
s	~95 GeV/c^2	$-\frac{1}{3}$e	c
t	~172.4 GeV/c^2	$+\frac{2}{3}$e	b
b	~4.2 GeV/c^2	$-\frac{1}{3}$e	t
LEPTON	Mass	Charge	Doublet Partner
e	~0.511 MeV/c^2	$-e$	ν_e
ν_e	$<$ 2.2 eV/c^2	0	e
μ	~105.7 MeV/c^2	$-e$	ν_μ
ν_μ	$<$ 1.7 MeV/c^2	0	μ
τ	~1.8 GeV/c^2	$-e$	ν_τ
ν_τ	$<$ 15.5 MeV/c^2	0	τ

Table 1: The SM Quarks and Leptons [27]

The final group of SM particles we consider is the bosons. There are five known integer-spin bosons incorporated into the SM: the photon (γ), the gluon (g), the Z^0, the $W^\pm$, and the Higgs boson (H). The bosons act as mediators for the types of interactions in which SM particles can participate. The underlying physics of these interactions is governed by gauge groups of the SM: $SU(3)_C \times SU(2)_L \times U(1)_Y$. The g is the mediator of the strong force interaction, an example of which being what holds together the constituent quarks of a proton. The strong interaction is governed by the gauge group $SU(3)_C$. The Z^0 and $W^\pm$ bosons are the mediators of the weak force interactions. When interactions between subatomic particles or the mechanism of radioactive decay are considered, as examples, weak force interactions are at play. The γ mediates electromagnetic interactions. Electromagnetic interactions may be associated with both moving and stationary charged particles. The electroweak interactions are governed collectively by the gauge groups of $SU(2)_L$ and $U(1)_Y$. Lefthanded helicity states of fermion couplets are dictated according to $SU(2)_L$. The $U(1)_Y$ gauge group is necessary in describing interactions between particles with hypercharge. Associated to the Higgs field is the H, which is deemed responsible for endowing particles with mass. The SM fails to incorporate the gravitational force at this time through a mediator or otherwise. Key properties of the bosons described are included in Table 2.

BOSON	Mass	Charge
g	0	0
Z^0	91.1876 $\pm$ 0.0021 GeV	0
$W^{\pm}$	80.398 $\pm$ 0.025 GeV	± 1
γ	1.8x10^8 eV	$< 5\text{x}10^{-30}$
H	125.18 GeV	0

Table 2: The SM Bosons [27]

Despite its success as being the canonical means through which particle physicists describe the fundamental particles and the ways those particles interact for decades, the SM is inherently incomplete. As mentioned previously, the SM fails to incorporate gravitational interactions. There exist other broad-based questions that remain unanswered by the SM. We question why there is an imbalance of matter and antimatter in the universe, favoring matter. Unknown is the origin of the masses of the chargeless neutrinos. We can consider why gravity is so much weaker than all of the other forces, a question known as the hierarchy problem. As a final example, left under investigation is the nature of astronomical dark matter (DM). Particle physicists are taxed with probing deeply into the potential answers to these broad-based and unresolved questions.

Symmetry Breaking

Symmetries constitute the backbone of many theories in Particle Physics. When considering symmetries, one aspect at the forefront is the invariance of a theory under certain group transformations [50]. Classical electromagnetism is wrapped into the skew rank two field tensor $F_{\mu\nu}$. This field tensor may be parametrized in terms of a vector potential A_μ as follows:

$$F_{\mu\nu} = \partial_\nu A_\mu - \partial_\mu A_\nu. \tag{I.1}$$

The possibility exists then to modify A_μ without affecting $F_{\mu\nu}$. As such, A'_μ may be considered:

$$A'_\mu = A_\mu + ie\partial_\mu \Lambda, \tag{I.2}$$

where $\Lambda(x)$ is a scalar field, and e is the strength of interaction with this field [50]. While not necessary in classical theory, consideration of A_μ is necessary in the more modern quantum theory.

Within the SM are intrinsic symmetries that stem from the unitary groups: $U(1)$, $SU(2)$, and $SU(3)$. Symmetry under U(1) means the fields are invariant under phase transformations $\phi' = \phi e^{i\theta}$ in order that physical properties arising from the SM are unaffected. This idea arises from the gauge freedom in A_μ. Symmetry under SU(2) and SU(3) means a physical state is unaffected by rotations in the two-dimensional and three-dimensional complex space, respectively.

Each of the groups $U(1)_Y$ and $SU(2)_L$ influence the electric charge of a fermion. The electric charge (Q) of a fermion is parametrized as follows:

$$Q = T_L^3 + \frac{Y}{2}, \tag{I.3}$$

where T is the weak isospin from $SU(2)_L$, and Y is the hypercharge of $U(1)_Y$ [27]. The values of Q, T_L^3, and Y for each of the for each of the first generation lefthanded (L) and righthanded (R) fermions are listed in Table 3.

Particle	Q	T_L^3	Y
ν_e	0	$\frac{1}{2}$	-1
e_L	-1	$\frac{-1}{2}$	-1
e_R	-1	0	-2
u_L	$\frac{2}{3}$	$\frac{1}{2}$	$\frac{1}{3}$
d_L	$-\frac{1}{3}$	$-\frac{1}{2}$	$\frac{1}{3}$
u_R	$\frac{2}{3}$	0	$\frac{4}{3}$
d_R	$-\frac{1}{3}$	0	$-\frac{2}{3}$

Table 3: First Generation Fermionic Charges [27]

A surprising facet of the SM is that the Z^0 and $W^{\pm}$ bosons have mass. This is because, under the assumption that a physical state remains invariant under a spacetime phase transformation, the mediators of the fundamental particles' interactions are massless. With this in mind, the origin of the masses of these particles is tied to the Higgs mechanism, which is described in a subsequent section. In connection with the Higgs mechanism is the idea of spontaneous symmetry breaking. Spontaneous symmetry breaking results when there are degenerate ground state solutions to a mathematical theory. Once a choice of solution is made, the symmetry of the system is "broken." This resolves the issue with observed massive gauge bosons because a "broken symmetry" is postulated to occur in the vaccuum state where the Higgs field allows for these mass eigenstates. We note that the choice of model to describe the interactions of the fundamental particles must follow local gauge invariance to avoid creating physical impossibilities.

The Higgs Mechanism

One of the intriguing aspects of the SM, as mentioned previously, is the fact that the photon is a massless boson, while the W and Z bosons have mass. This can be explained in relation to the Higgs particle, a massive scalar boson discovered at the Large Hadron Collider (LHC). The Higgs field is introduced to the SM because of the broken symmetry in $SU(2)_L \times U(1)_Y$ with a Lagrangian density of

$$l = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi), \tag{I.4}$$

where the potential $V(\phi)$ is given by

$$V(\phi) = \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2, \tag{I.5}$$

and $\lambda > 0$ (μ and λ are dimensionless) to maintain bounded potentials as $\phi \to \infty$. The SM Higgs field is expressed as

$$\phi = \begin{pmatrix} \phi^\dagger \\ \phi^0 \end{pmatrix}, \tag{I.6}$$

where

$$\phi^\dagger = \frac{\phi_1 + i\phi_2}{\sqrt{2}}, \phi^0 = \frac{\phi_3 + i\phi_4}{\sqrt{2}}. \tag{I.7}$$

The Lagrangian can be minimized under the assumption that the field has one scalar and one imaginary component in order to determine the vacuum state. Minimization of this form results in the following:

$$\phi^\dagger\phi = -\frac{\mu^2}{2\lambda} = \frac{\nu^2}{2}. \tag{I.8}$$

When $\mu^2 < 0$, there are two stable ground states: $+\nu$ and $-\nu$. This is referred to as the "Mexican hat potential" and is depicted in Figure 1.

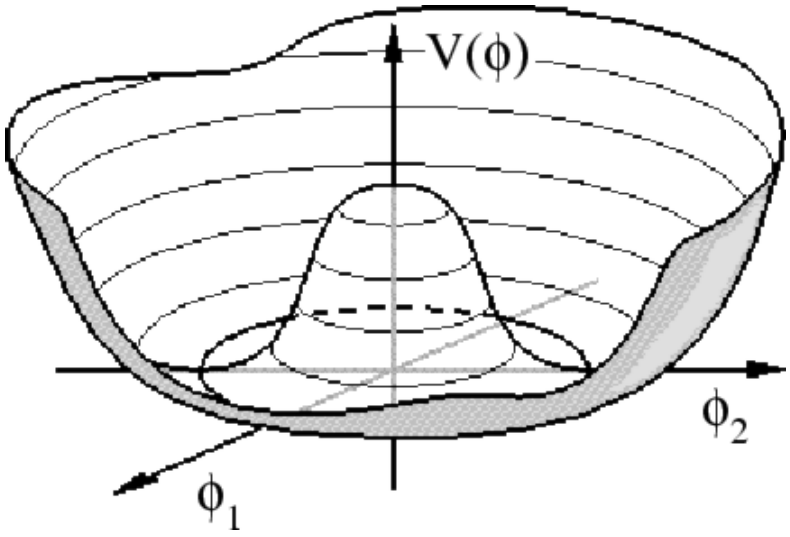

Figure 1: Higgs potential [11].

Choosing $+\nu$ or $-\nu$ as the ground state defines the direction of the field. This yields the following:

$$\phi_3 = \nu, \phi_1 = \phi_2 = \phi_4 = 0 \to \phi_0 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ \nu \end{pmatrix}. \tag{I.9}$$

To determine the excited states, we expand around one of the ground states. With a perturbation $H(x)$, the field is as follows:

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + H(x) \end{pmatrix}. \tag{I.10}$$

The excited states are SM particles. Choosing the direction of the field breaks the symmetry. Even so, by introducing a covariant derivative, the $SU(2)_L \times U(1)_Y$ symmetry is preserved, and massive gauge bosons can be accommodated [24]. As mentioned in the introduction of this section, this accurately predicts the masses of the $W^{\pm}$ and Z bosons despite the massless nature of the photon. This is a significant and remarkable accomplishment in SM particle physics.

Chapter II

Motivation

Connecting Small Scale to Large Scale

Particle Physics is a unique discipline in that many open-ended questions have potential far-reaching consequences at small and large scales. One of these open-ended questions, among others, is the identity of astronomical DM and how that fits into the larger scope of SM Particle Physics. At the small scale, the particle identity of astronomical DM is unknown. At the large scale, there exists evidence from astronomers that DM is necessary for the formation, rotation, and overall evolution of galaxies. This evidence comes from collaborations such as Planck and WMAP [9, 41]. To address this and other open-ended questions, the use of hadron colliders may be employed (to be described in a subsequent section); however, question remains as to whether we have reached high enough energies yet to probe the particle interactions we seek. This may be an inhibitor to the complete connection of small scale and large scale phenomena.

Employing Supersymmetry

Even with this uncertainty in mind, we can at present employ specific models to have the best potential of probing desired particle interactions. Supersymmetry (SUSY) is a potential extension of the SM that aims to resolve several open-ended questions that remain. Under the framework of SUSY, every fermionic SM particle is matched with a bosonic superpartner and vice versa. The supersymmetric partners of the quarks are squarks, of the leptons are sleptons, and of the neutrinos are sneutrinos. The supersymmetric equivalents of the bosons are the gluinos, photinos, higgsinos, winos, and binos. The superposition of states of the bosonic superpartners creates an electrically charged set of particles (charginos) and an electrically neutral set of particles (neutralinos). The concept of SUSY supports the idea of the lighter Higgs boson that is observed at a mass of 125 GeV. Another reason to value the idea of SUSY is that it provides a DM candidate particle in the form of the lightest neutralino ($\tilde{\chi}_1^0$).

There are challenges to combat in the search for the $\tilde{\chi}_1^0$. The search for the $\tilde{\chi}_1^0$ at hadron colliders via direct production mechanisms is difficult due to lower production rates; thus, the search for $\tilde{\chi}_1^0$ is most often conducted targeting cascading decays of heavier supersymmetric particles. In the interactions in which $\tilde{\chi}_1^0$ is most likely to be produced, the particle is low-energy (soft) because it arises from these cascading decays of heavier particles. Thus, the requirement of initial state radiation (ISR) may be made. The ISR comes in the form of a quark or gluon jet (spray of particles) with charge stipulated in order to conserve charge in the interaction. This jet provides a natural

kinematic boost to the system that, by conservation of momentum, gives additional momentum to the soft decay products, including the $\tilde{\chi}_1^0$. This aids in the detection of these soft decay products.

Another way to combat the difficulty in searching for the $\tilde{\chi}_1^0$ is to introduce the idea of DM coannihilation (CA), where CA refers to the interaction of $\tilde{\chi}_1^0$ with another supersymmetric particle to produce SM particles [26]. This is important in the context of the DM relic density (to be parameterized in the next section), the amount of dark matter present in our universe today (Ωh^2). If we assume that the DM particle is mostly Bino (Z-like), then there exists an overabundance of DM in the universe with respect to the value quoted by astronomical experiment [9, 41]. On the other hand, if we assume that the DM particle is mostly Wino (W-like), then there exists an underabundance of DM in the universe with respect to the astronomical value. To address these discrepancies, we focus-in on the value of the DM coannihilation cross section ($< \sigma v >_A$) in order to modulate Ωh^2. The relationship is inverse; Ωh^2 decreases as $< \sigma v >_A$ goes up. An extra dependency built-in here is the value of the mass difference (Δm) between the coannihilating particles. When the mass gap is small, this drives up $< \sigma v >_A$. Thus, a small Δm between coannihilating particles increases the $< \sigma v >_A$, which in turn brings down the Ωh^2 to a value more consistent with the astronomical measurement.

Motivating the Search for Compressed Mass Spectra

In the analysis to follow, we target DM coannihilation where the $\tilde{\chi}_1^0$ coannihilates with the supersymmetric partner of the tau (τ) lepton, the stau ($\tilde{\tau}$). We target a small mass difference between the $\tilde{\tau}$ and the $\tilde{\chi}_1^0$ ($\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$) to raise the $< \sigma v >_A$ and to lower Ωh^2 for certain models. This is with the aim of achieving consistency with the DM relic density as quoted by astronomical measurement (WMAP, etc.). When the Δm is small, this is a scenario known as a compressed mass spectrum. We seek small $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$ to explore areas of the phase space with the highest potential sensitivities at the Compact Muon Solenoid (CMS) experiment (to be described later) of the LHC. This can be seen in Figure 2 to follow for the area shaded in red [6].

In the next subsection, we elaborate on the parameterization of Ωh^2 to further illustrate the interconnectedness of Ωh^2, $< \sigma v >_A$, and Δm.

Parameterizing the DM Relic Density

General Boltzmann Equation

We begin by developing the general form of the Boltzmann equation, which describes how the number density (number of particles per unit volume) of a particular particle species changes with time. The general form of the Boltzmann equation is as follows:

$$\frac{\mathrm{d}n_\mathrm{i}}{\mathrm{d}t} + 3\frac{\dot{a}}{a}n_\mathrm{i} = 0, \tag{II.1}$$

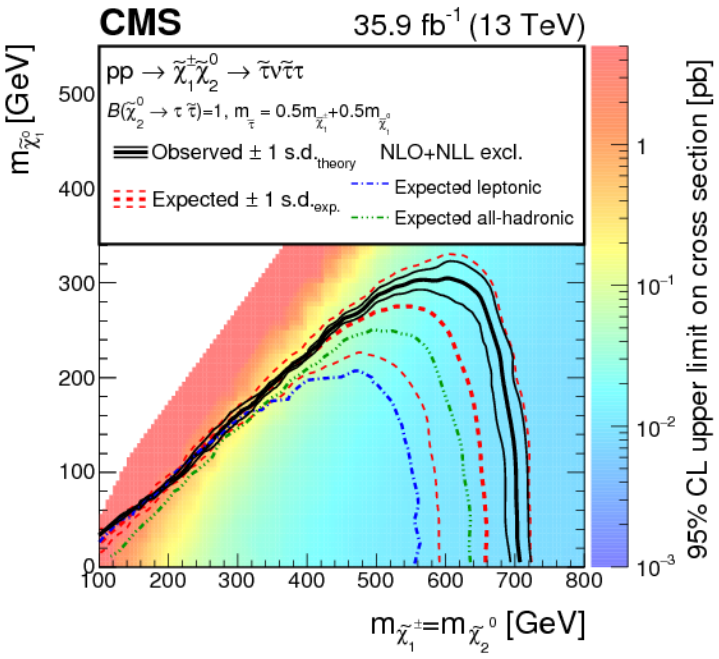

Figure 2: High potential sensitivity when searching for compressed mass spectra (SUS-17-003).

where n_{i} is the number density for particle species i, $\frac{\dot{a}}{a}$ is the relative expansion rate of the universe otherwise known as the Hubble constant, and t is time.

This formulation of the Boltzmann equation can be made more compact in the following way. For

$$\frac{\mathrm{d}n_{\mathrm{i}}}{\mathrm{d}t} + 3\frac{\dot{a}}{a}n_{\mathrm{i}} = 0, \tag{II.2}$$

multiply by $\frac{a^3}{a^3}$ such that

$$\frac{a^3}{a^3}\left[\frac{\mathrm{d}n_{\mathrm{i}}}{\mathrm{d}t} + 3\frac{\dot{a}}{a}n_{\mathrm{i}} = 0\right]. \tag{II.3}$$

Now, bringing a factor of a^3 inside the brackets gives

$$\frac{1}{a^3}\left[\frac{\mathrm{d}n_{\mathrm{i}}}{\mathrm{d}t}a^3 + 3a^2\dot{a}n_{\mathrm{i}}\right] = 0. \tag{II.4}$$

Recognition of the term within the brackets as a product rule for differentiation gives the following:

$$\frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3 n_{\mathrm{i}}\right] = 0. \tag{II.5}$$

To represent generally the interaction between species i and j without loss of generality, we utilize the term $C_i[n_j]$. To properly account for interactions between particles in the previous equation, we write instead that

$$\frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3 n_{\mathrm{i}}\right] = C_{\mathrm{i}}[n_{\mathrm{j}}]. \tag{II.6}$$

Before, the time derivative was 0 for no change with respect to time in $a^3 n_{\rm i}$. Now, however, there exist interactions between particles, and the time derivative can have nonzero value. To apply this information in context, let us consider the following reciprocal interaction:

$$p_1 + p_2 \Longleftrightarrow p_3 + p_4, \tag{II.7}$$

where p_1, p_2, p_3, and p_4 represent four species of particles that are not necessarily identical. Of the many parameters in this interaction, let's say we wish to track the number density of species 1 (n_1).

Particles of type 1 will be both created and destroyed in the process of this reciprocal interaction. The creation will be dependent on n_3 and n_4 since species 3 and 4 interact to produce species 1. The destruction will depend on n_1 and n_2 since species 1 and 2 interact to produce 3 and 4. Thus, in general,

$$\frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3 n_i\right] = -\alpha n_1 n_2 + \beta n_3 n_4, \tag{II.8}$$

where the creation and destruction terms are $\beta n_3 n_4$ and $-\alpha n_1 n_2$, respectively. By its placement in the previous equation, we know that α will be the time-averaged annihilation cross section for species 1 and 2. This cross section is represented by $< \sigma v >$. When the system of particles is in equilibrium, the time derivative in this equation will go to 0. Hence, in that case,

$$\begin{aligned} -\alpha n_1 n_2 + \beta n_3 n_4 &= 0 \\ \alpha n_1 n_2 &= \beta n_3 n_4. \end{aligned} \tag{II.9}$$

Solving this equation for β,

$$\begin{aligned} \beta &= \alpha \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq} \\ &= < \sigma v > \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq}, \end{aligned} \tag{II.10}$$

where the subscript eq is to remind that this is a relationship derived from equilibrium considerations. Substituting for α and β gives the following:

$$\frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}\left[a^3 n_{\rm i}\right] = - < \sigma v > \left[n_1 n_2 - \left(\frac{n_1 n_2}{n_3 n_4}\right)_{\rm eq} n_3 n_4\right]. \tag{II.11}$$

This derivation leading to the equation above helps to justify the dimensions of terms in the full Boltzmann equation to follow.

Full Boltzmann Equation

The full Boltzmann equation appears as follows:

$$\frac{\mathrm{d}n_{\mathrm{x}}}{\mathrm{d}t} + 3Hn_{\mathrm{x}} = -(<\sigma v>)_{\mathrm{T}} \left[n_{\mathrm{x}}^2 - n_{\mathrm{eq}}^2\right], \tag{II.12}$$

where n_{x} is the number density of particle species x, H is the Hubble constant, $(<\sigma v>)_{\mathrm{T}}$ is the thermally-averaged annihilation cross-section, and n_{eq} is the number density at the time of equilibrium, which is when the rates of creation and destruction were equal. Recall that this is much earlier on in the evolution of the universe when the normal matter still possessed enough kinetic energy to interact and produce dark matter, and the dark matter was still concentrated enough in the universe to interact to produce normal matter.

The first change of variables we will make is

$$Y = \frac{n}{T^3}, \tag{II.13}$$

where n is still a number density, and T is temperature. The reasoning behind this definition is as follows: The entropy density s for the system is given by the following:

$$s = \left(\frac{2\pi^2}{45}\right) g_* T^3, \tag{II.14}$$

where g_* is the number of degrees of freedom for the relativistic dark matter particle. In fundamental units, the entropy σ is defined as $\ln(g)$, where g is the number of accessible states of the system. Thus, s is a measure of a (a number)/(volume) in fundamental units, and so is n. Thus, dividing n by s, or, more generally, n by T^3 gives a dimensionless quantity to track in Y. We note from here that $n \propto T^3$.

We keep in mind as well that the quantity aT is time-independent for our purposes. Now, manipulating again the lefthand side of the Boltzmann equation,

$$\frac{1}{a^3}\frac{\mathrm{d}}{\mathrm{d}t}\left[na^3\right] = a^{-3}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{n(aT)^3}{T^3}\right). \tag{II.15}$$

Bringing $(aT)^3$ out of the time derivative since it is a constant,

$$\begin{aligned} a^{-3}\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{n(aT)^3}{T^3}\right) &= T^3\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{n}{T^3}\right) \\ &= T^3\frac{\mathrm{d}Y}{\mathrm{d}t}. \end{aligned} \tag{II.16}$$

Recall from previously that

$$\frac{\mathrm{d}n_{\mathrm{x}}}{\mathrm{d}t} + 3Hn_{\mathrm{x}} = -(<\sigma v>)_{\mathrm{T}} \left[n_{\mathrm{x}}^2 - n_{\mathrm{eq}}^2\right]. \tag{II.17}$$

Making full substitutions for n in terms of $Y = \frac{n}{T^3}$ in the above and with the subscript x suppressed,

$$\begin{aligned} T^3 \frac{\mathrm{d}Y}{\mathrm{d}t} &= (<\sigma v>)_{\mathrm{T}} \left[(Y_{\mathrm{eq}} T^3)^2 - (Y T^3)^2 \right] \\ &= (<\sigma v>)_{\mathrm{T}} T^6 \left[Y_{\mathrm{eq}}^2 - Y^2 \right]. \end{aligned} \tag{II.18}$$

Solving this equation for $\frac{\mathrm{d}Y}{\mathrm{d}t}$,

$$\begin{aligned} \frac{\mathrm{d}Y}{\mathrm{d}t} &= T^3 (<\sigma v>)_{\mathrm{T}} \left[Y_{\mathrm{eq}}^2 - Y^2 \right]. \\ &= -(<\sigma v>)_{\mathrm{T}} T^3 \left[Y^2 - Y_{\mathrm{eq}}^2 \right]. \end{aligned} \tag{II.19}$$

Next, we move on to a second change of variables. The second change of variables we make is as follows:

$$x = \frac{m}{T}, \tag{II.20}$$

where m is particle mass and T is temperature. The Hubble constant H can be expressed as $\frac{\mathrm{d}x}{\mathrm{d}t} = Hx$, which is the relationship known as Hubble's Law [12]. This originates as follows for $x = \frac{m}{T}$:

$$\begin{aligned} \frac{\mathrm{d}x}{\mathrm{d}t} &\approx -m \left(\frac{1}{T^2} \right) \frac{\mathrm{d}T}{\mathrm{d}t} \\ &\approx -\frac{m}{T} \left(\frac{\dot{T}}{T} \right) \\ &\approx Hx. \end{aligned}$$

Justification for this comes from the Cosmological Principle [13]. The Cosmological Principle, as developed by Einstein, says that on large scales, the universe is essentially isotropic and homogeneous. In the way that we use the Hubble constant $\frac{\dot{a}}{a}$ to describe the relative expansion of the universe, we use an approximately equivalent expression $-\frac{\dot{T}}{T}$, which is the negative of the relative change in temperature. This is acceptable since $\dot{T}$ is an inherently negative quantity for the expanding universe.

Utilizing the chain rule for derivatives, we know that

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = \frac{\mathrm{d}Y}{\mathrm{d}t} \frac{\mathrm{d}t}{\mathrm{d}x}. \tag{II.21}$$

Substituting in this equation for $\frac{\mathrm{d}Y}{\mathrm{d}t}$ and $\frac{1}{Hx}$ for $\frac{\mathrm{d}t}{\mathrm{d}x}$, we have

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = T^3 (<\sigma v>)_{\mathrm{T}} \left[Y_{\mathrm{eq}}^2 - Y^2 \right] \frac{1}{Hx}. \tag{II.22}$$

We make use of x since interesting dynamics are for when the temperature is on the order of the

mass of the particle. Under these conditions, we make a radiative consideration that

$$H = \frac{H(m)}{x^2}, \tag{II.23}$$

where $H(m)$ is a function of m for these dynamics.

We consider the radiation-dominated era of our universe, being one of many eras during the universe's evolution. In the radiation-dominated era, the radiation (energy) density (ϵ) for photons was proportional to T^4, as we can discern from the Stefan-Boltzmann Law. Dividing through by $c^2 = 1$ in our system of natural units, we treat this energy density as essentially the same as a mass density. This radiation density constituted some fraction of the critical density ρ_{critical} for the dark matter particles.

The critical density is a mass density. Hence, dividing two mass densities in taking the radiation density ϵ by the critical density ρ_{critical} gives a dimensionless quantity. In the definition of ρ_{critical} in what follows, it is the case that $\rho_{\text{critical}} \propto H^2$. Thus, through these relationships,

$$\begin{aligned} \frac{H^2}{T^4} &\propto \text{constant} \\ H^2 &\propto T^4 \\ H &\propto T^2 \\ H &\propto \frac{1}{x^2}, \end{aligned} \tag{II.24}$$

if we incorporate a functional dependence on the mass for H, as shown in the previous parametrization for H.

Another factor defined for ease of calculation is the following:

$$\lambda = \frac{m^3(<\sigma v>)_{\text{T}}}{H(m)}. \tag{II.25}$$

Next, we substitute for λ and H in the previous equation for $\frac{\mathrm{d}Y}{\mathrm{d}x}$. This yields, upon substitution,

$$\frac{\mathrm{d}Y}{\mathrm{d}x} = T^3 \left(\frac{H(m)\lambda}{m^3}\right) \left[Y_{\text{eq}}^2 - Y^2\right] \frac{1}{x} \left(\frac{x^2}{H(m)}\right). \tag{II.26}$$

Since $T^3 = \frac{m^3}{x^3}$,

$$\begin{aligned} \frac{\mathrm{d}Y}{\mathrm{d}x} &= \frac{m^3}{x^3} \left(\frac{H(m)\lambda}{m^3}\right) \left[Y_{\text{eq}}^2 - Y^2\right] \frac{1}{x} \left(\frac{x^2}{H(m)}\right) \\ &= \frac{\lambda}{x^2} \left[Y_{\text{eq}}^2 - Y^2\right] = \frac{-\lambda}{x^2} \left[Y^2 - Y_{\text{eq}}^2\right]. \end{aligned} \tag{II.27}$$

This formalism is to be applied as we move to integrating for the DM relic density.

The Path to the DM Relic Density After Variable Changes

We define the variable Δ such that

$$\Delta = Y - Y_{\rm eq}. \tag{II.28}$$

With that definition of Δ,

$$\frac{\mathrm{d}\Delta}{\mathrm{d}x} = \frac{\mathrm{d}Y}{\mathrm{d}x} - \frac{\mathrm{d}Y_{\rm eq}}{\mathrm{d}x}. \tag{II.29}$$

Substituting this into the equation for $\frac{\mathrm{d}Y}{\mathrm{d}x}$ yields the following:

$$\begin{aligned} \frac{\mathrm{d}\Delta}{\mathrm{d}x} &= \frac{-\lambda}{x^2}\left[Y^2 - Y_{\rm eq}^2\right] - \frac{\mathrm{d}Y_{\rm eq}}{\mathrm{d}x} \\ &= \frac{-\lambda}{x^2}\left[(Y + Y_{\rm eq})(Y - Y_{\rm eq})\right] - \frac{\mathrm{d}Y_{\rm eq}}{\mathrm{d}x}. \end{aligned} \tag{II.30}$$

Since $(Y + Y_{\rm eq}) = \Delta + 2Y_{\rm eq}$ and $(Y - Y_{\rm eq}) = \Delta$,

$$\begin{aligned} \frac{\mathrm{d}\Delta}{\mathrm{d}x} &= \frac{-\lambda}{x^2}(\Delta + 2Y_{\rm eq})\Delta - \frac{\mathrm{d}Y_{\rm eq}}{\mathrm{d}x} \\ &= \frac{-\lambda}{x^2}(\Delta^2 + 2Y_{\rm eq}\Delta) - \frac{\mathrm{d}Y_{\rm eq}}{\mathrm{d}x}. \end{aligned} \tag{II.31}$$

At equilibrium, when production and destruction rates were equal for the DM, the temperature was much warmer than times much later. Thus, since $Y = \frac{n}{T^3}$, we are able to say that $Y_{\rm eq} \ll Y$ for late, late times. With that, $\Delta \approx Y$, and

$$\begin{aligned} \frac{\mathrm{d}\Delta}{\mathrm{d}x} &\approx \frac{-\lambda}{x^2}\Delta^2 \\ \frac{\mathrm{d}Y}{\mathrm{d}x} &\approx \frac{-\lambda}{x^2}Y^2. \end{aligned} \tag{II.32}$$

We introduce the variable l as a step through time, where $l = 0$ is for freeze-out. We write the previous equation more generally as follows:

$$\frac{\mathrm{d}Y}{Y^2} = \frac{-\lambda}{x^{l+2}}\mathrm{d}x. \tag{II.33}$$

The bounds on the integration to get Y are from freeze-out to some late-late (∞) time later.

$$\int_0^\infty \frac{\mathrm{d}Y}{Y^2} = \int_0^\infty \frac{-\lambda}{x^{l+2}}\mathrm{d}x. \tag{II.34}$$

This means

$$-\frac{1}{Y_{l=\infty}} + \frac{1}{Y_{l=0}} \approx -\frac{\lambda_{l=0}}{x_{l=0}}. \tag{II.35}$$

To be factored in is a thermal suppression term for Y. This appears as follows:

$$Y \propto \left(\frac{m}{T}\right)^{\frac{3}{2}} \mathrm{e}^{-\frac{m}{T}}. \tag{II.36}$$

With this in mind, since the universe was much warmer at freeze-out than at some infinite time later, it is the case that $Y_{l=0} \gg Y_{l=\infty}$. This leads to the reformulation of the previous equation for $\frac{\lambda_{l=0}}{x_{l=0}}$ as follows:

$$\begin{aligned} \frac{1}{Y_{l=\infty}} &= \frac{\lambda_{l=0}}{x_{l=0}} \\ Y_{l=\infty} &= \frac{x_{l=0}}{\lambda_{l=0}}. \end{aligned} \tag{II.37}$$

Substituting with λ into this equation yields

$$Y_{l=\infty} = \frac{H(m)}{m^3 < \sigma v >_{l=0}} x_{l=0}. \tag{II.38}$$

In this equation, now that there exists an expression for Y at late-late times after freeze-out, we seek the number density n of the dark matter particles at $l = \infty$.

The Path to Dark Matter Relic Density: Finding Number Density

The product of the number density n and the mass of the DM particle m gives the mass density ρ of the DM in the universe. That is, since $Y = \frac{n}{T^3}$, we have

$$\rho = m Y_{l=\infty} T^3_{l=\infty}. \tag{II.39}$$

We have to make a correction to our equations to account for a cosmological process that involves the reheating of photons. Photons can be reheated when $\mathrm{e}^+\mathrm{e}^-$ annihilation occurs to produce energy at higher temperatures than the ambient photon temperature, thus raising the temperature of the surrounding photons. This temperature change means that there is a nonconstant aT during this time. Thus, we must scale the DM relic density accordingly with a correction factor. With this necessary correction, we have

$$\begin{aligned} \rho &\approx Y_{l=\infty} m T^3_{l=\infty} \left(\frac{a_{l=\infty}}{a_{l=0}}\right)^3 \left(\frac{T_{l=0}}{T_{l=0}}\right)^3 \\ &\approx \frac{m Y_{l=\infty} T^3_{l=0}}{30}, \end{aligned} \tag{II.40}$$

where the a terms still describe the radius of the universe as in the Hubble constant.

Next, we define the parameter Ω_χ, the fraction of the present day DM density coming from χ.

Mathematically, we define Ω_χ as follows:

$$\Omega_\chi = \frac{\rho}{\rho_{\text{critical}}}, \tag{II.41}$$

where the critical density ρ_{critical} is the DM density required for the universe to stop its expansion but only after an infinite time. The critical density is defined as

$$\rho_{\text{critical}} = \frac{3}{8\pi} H_0^2 \left(M_{\text{pl}}^2\right), \tag{II.42}$$

where the new term, M_{pl}^2, is the square of the Planck mass, defined by $\frac{1}{\sqrt{8\pi G}}$ for Newton's gravitational constant G. We emphasize again that the interesting dynamics occur in the so-called "radiative era," where mass is on the order of temperature. That is, the parameter $x = \frac{m}{T}$ is approximately 1.

The literature gives the following dependence for the average annihilation cross section $< \sigma v >$:

$$< \sigma v > \approx < \sigma v >_{l=0} \, x^{-l}, \tag{II.43}$$

such that for our purposes $< \sigma v >_{l=0} \approx < \sigma v >_p$ for $p \neq 0$. In the radiative era, we can also apply the following formula for the Hubble constant in terms of temperature T and the number of degrees of freedom g_* for the relativistic DM particle:

$$H(T) = T^2 \sqrt{\frac{4\pi^3 G g_*}{45}}. \tag{II.44}$$

Plugging the equations for ρ_{critical} and ρ into the equation for Ω_χ gives

$$\Omega_\chi \approx \frac{\rho}{\rho_{\text{critical}}} \approx \left(\frac{m Y_{l=\infty} T_{l=0}^3}{30}\right) \left(\frac{8\pi}{3}\right) \left(\frac{1}{H_0^2 M_{\text{pl}}^2}\right). \tag{II.45}$$

Using the equation for $Y_{l=\infty}$ yields

$$\Omega_\chi \approx \left(\frac{m T_{l=0}^3}{30}\right) \left(\frac{H(m)}{m^3 < \sigma v >}\right) (x_{l=0}) \left(\frac{8\pi}{3}\right) \left(\frac{1}{H_0^2 M_{\text{pl}}^2}\right). \tag{II.46}$$

Since $m \approx T$, and with the equation for $H(T)$ in the above,

$$\begin{aligned} \Omega_\chi &\approx \left(\frac{T_0^3}{30 T^2}\right) T^2 \sqrt{\frac{4\pi^3 G g_*}{45}} \left(\frac{x_{l=0}}{< \sigma v >}\right) \left(\frac{8\pi}{3}\right) \left(\frac{1}{H_0^2 M_{\text{pl}}^2}\right) \\ &\approx \sqrt{\frac{4\pi^3 G g_*}{45}} \left(\frac{8\pi}{90}\right) \left(\frac{T_0^3 x_{l=0}}{H_0^2 M_{\text{pl}}^2 < \sigma v >}\right). \end{aligned} \tag{II.47}$$

This illustrates the interconnectedness of the Ωh^2 with the $< \sigma v >$ and m through x, as desired.

Chapter III

The Large Hadron Collider

The Large Hadron Collider (LHC) is the world's largest particle accelerator. Built at the European Center for Nuclear Research (CERN), the LHC is a two-ring proton-proton collider with 27 km diameter spanning the border between France and Switzerland. The LHC operates at extremely high energies, exceeding those of even FermiLab's decommissioned Tevatron in the U.S. [37]. The LHC was built nearly 100 m underground to simulate nearly the conditions of the early universe in order to further understanding of the fundamental particles and the ways in which they interact. The potential exists for the LHC to shed light on electroweak symmetry breaking and unanswered questions associated with the SM at TeV scales.

The acceleration of protons at CERN is accomplished in a series of steps, ending in the circulation of proton beams in the LHC at nearly the speed of light and energies amounting at present to 6.5 TeV. The initial source of protons is hydrogen gas stripped of electrons. The first accelerator is Linac 2, which brings the protons to 50 MeV of energy. The proton beam is then passed to the Proton Synchrotron Booster (PSB), which brings the protons to 1.4 GeV. Next in the succession of accelerators is the Proton Synchrotron, which takes the beam to 25 GeV. After that, the Super Proton Synchrotron (SPS) brings the beam to 450 GeV [16]. Finally, the beams are transferred to the LHC where, after a short time, they reach 6.5 TeV and continue to circulate for many hours.

Once inside the LHC, two beams circulate in opposite directions in pipes kept at ultrahigh vacuum. Superconducting electromagnets guide the beams around each ring. Each electromagnet is built from electric cable that conducts electricity without resistance, which requires the magnets to be at a temperature of approximately -270°C. This cooling is accomplished for the LHC by liquid helium [17]. Focusing of the beams is accomplished by quadrupole magnets, which are each 5-7 m long. The protons are passed through the LHC in bunches in order to increase the probability of proton-proton interaction. In addition, the protons are "squeezed" by the quadrupole magnets to increase the chance that they will collide [17]. Thus, the system of magnets within the LHC is crucial to the proper circulation of the beams around each ring.

Luminosity quantifies the probability that a proton-proton interaction will occur. Focus is on achieving higher luminosities with the impending upgrades to the LHC. Most generally, instantaneous luminosity is a measure of the number of interactions occurring per unit area per unit time. Thus, an increase in luminosity can be achieved by increasing the number of protons in a bunch, minimizing the cross-sectional area of each bunch, or increasing the rate at which the bunches cross paths. The units of instantaneous luminosity are $\mathrm{b}^{-1}\mathrm{s}^{-1}$ or $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The unit of barn ($b$) corresponds to cm^2 via the following conversion: 1 b = 10^{-24} cm^2. Integrated luminosity is instan-

taneous luminosity integrated over a span of time. We express integrated luminosity in units of b^{-1}. The integrated luminosity at the LHC is directly proportional to the number of particles per bunch, the number of bunches per beam, the frequency of the beam revolution, a factor to account for the relativistic nature of the beams, and a factor to account for the crossing angle of the beams at the point of interaction. The integrated luminosity is inversely proportional to the transverse emittance and the amplitude (β^*) function for the beams. Emittance is a reflection of how the bunches were prepared. When emittance is low, the particles are more likely to interact because the particles within each beam are better confined and have nearly the same momentum. The β^* function is approximately the ratio of the width of the beam to the emittance, which reflects how well the beam is "squeezed" by the magnets [7]. Bunches are spaced every 25 ns. With a luminosity of 10^{34} $cm^{-2}s^{-1}$, the LHC withstands nearly one billion events occurring every second.

A conjecture as to why certain questions associated with the SM remain unresolved is because the LHC has yet to reach high enough energies necessary to produce hypothesized high mass particles. Another potential complication may come from the fact that the rate at which new physics events are produced is significantly smaller than the well-known SM background processes. The number of events observed at the LHC is directly proportional to the integrated luminosity and the scattering cross section of the particular process considered. The scattering cross section has units of area and quantifies the probability of the interaction to occur. Thus, the discovery of new physics is contingent on high enough luminosity and sufficient rejection of SM background processes.

Figure 3 shows a schematic of the current LHC set-up. The four main experiments present at the LHC are ATLAS, ALICE, CMS, and LHCb. For the analysis described in this document, we utilize 77.2 fb^{-1} of Run II data collected in 2016 and 2017 by the Compact Muon Solenoid (CMS) detector described in the section to follow.

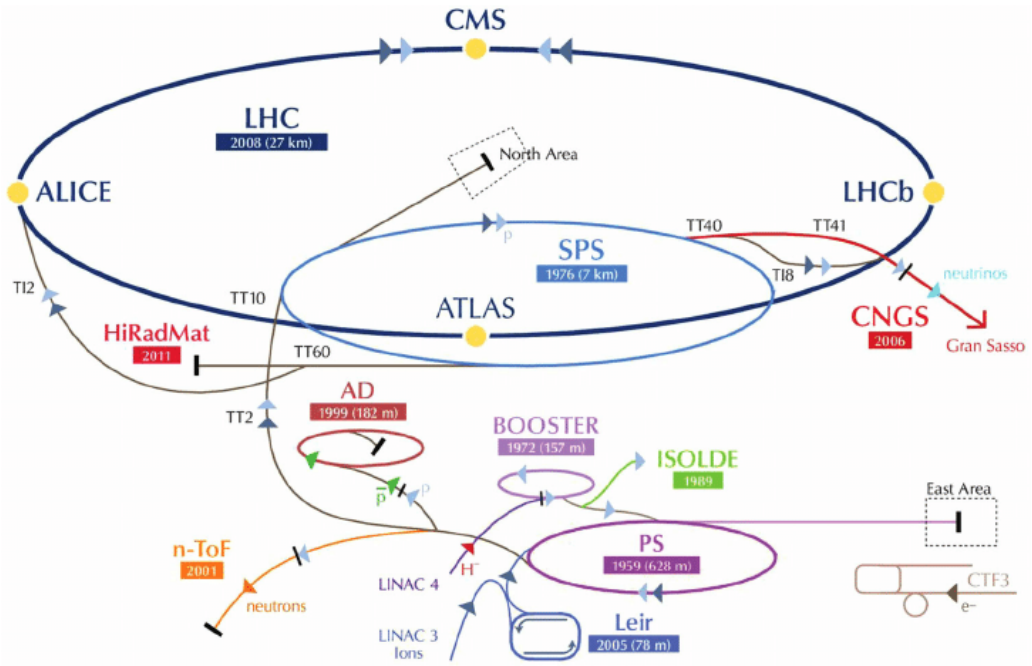

Figure 3: Schematic of current LHC set-up [49].

Chapter IV

The CMS Experiment

In this section, we describe the geometry of the CMS detector, the solenoid magnet, the tracker system, the pair of calorimeters, the muon system, and the system for data acquisition. A full cross-sectional schematic of the CMS detector is given in Figure 4.

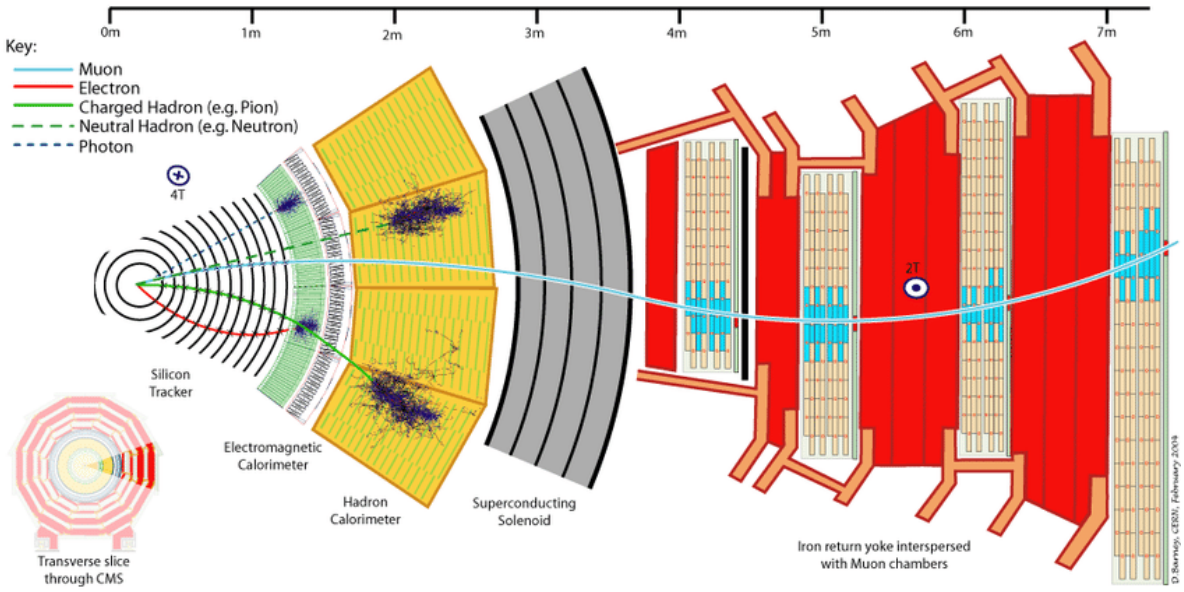

Figure 4: CMS detector schematic [25].

CMS Geometry

The CMS experiment utilizes a right-handed coordinate system. For this xyz-coordinate system, the x-axis points toward the center of the LHC ring. The y-axis points up in the transverse plane of the beampipe. The z-axis then points along the counterclockwise direction of the beampipe. In describing the CMS detector, a polar system of coordinates may also be employed. The polar angle (θ) is measured from the z-axis. The azimuthal angle ϕ is measured in the transverse (xy)-plane, and the radius r is measured out from the center of the beampipe. This coordinate system is shown in Figure 5.

The angular quantity more often utilized than θ is the pseudorapidity (η) of each outgoing particle. The η is measured as follows:

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{IV.1}$$

Particles that are more forward in the detector have η approaching ∞, whereas particles that are more central in the detector have η approaching 0.

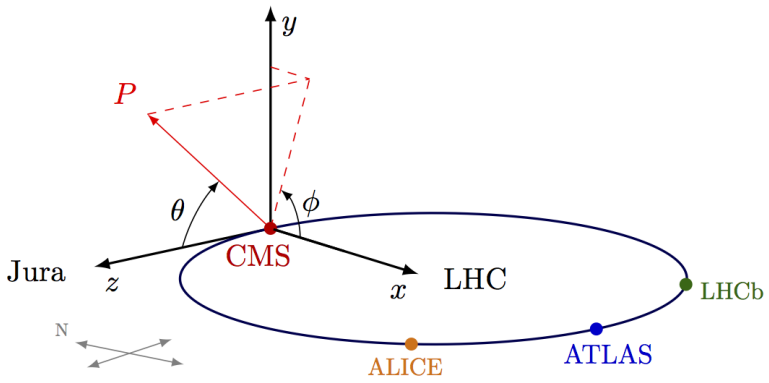

Figure 5: CMS coordinate geometry [33].

The Superconducting Solenoid

One of the most distinctive features of the CMS detector is the superconducting solenoid magnet. The magnet is superconducting in that it allows electricity to flow nearly without resistance. Contained within the steel return yoke of the detector, this magnet creates a field of 3.8 T. Since the field is created by a solenoid, the value is directly dependent on the number of turns in the coil and the current passed through the coil. In addition, the field is inversely proportional to the length of the coil. The purpose of the magnet is to bend the trajectories of outgoing particles [46]. A few key characteristics of an outgoing particle can be determined from its interaction with the magnetic field. The sign of the charge of the particle can be determined because the magnet will bend opposite charges in opposite directions. The relative magnitude of the particle's momentum can be determined because there will be less curve for a high-momentum track compared to a low-momentum track in the detector.

The Tracker System

The momenta of outgoing particles are crucial to piecing together how a proton-proton collision occurred. A charged particle will interact with a magnetic field. According to Newton's second

law, the force on the particle is as follows:

$$F = ma = m\frac{v^2}{R} = qvB\sin(\theta), \tag{IV.2}$$

where m is the particle mass, v is the velocity, R is the radius of its trajectory, q is the charge, B is the value of the magnetic field, and θ is the angle between the particle's momentum and the direction of the magnetic field. The CMS tracker is responsible for reconstructing the paths of charged particles as they move through the magnetic field of the detector [47]. Accurate measurement of momentum is contingent on the tracker material disturbing the outgoing particles as little as possible. In the material that follows, we describe the two subcomponents of the tracker system: the pixel detector and the silicon strip detector.

The Pixel Detector

The CMS pixel detector is positioned as the closest detector to the beamline. As such, the pixel detector has the ability to withstand millions of particles per square-centimeter per second traveling through the detector. When a charged particle passes through the pixel detector, it leaves behind electron-hole pairs in the silicon material. The ejected charges are collected for amplification and readout via readout chips (ROCs). This is with the goal in mind of tracking the paths of particles emerging from a collision with extreme precision. The total coverage of the pixel detector extends out to approximately $|\eta| < 2.5$ [44]. The layout of the pixel detector is depicted in Figure 6.

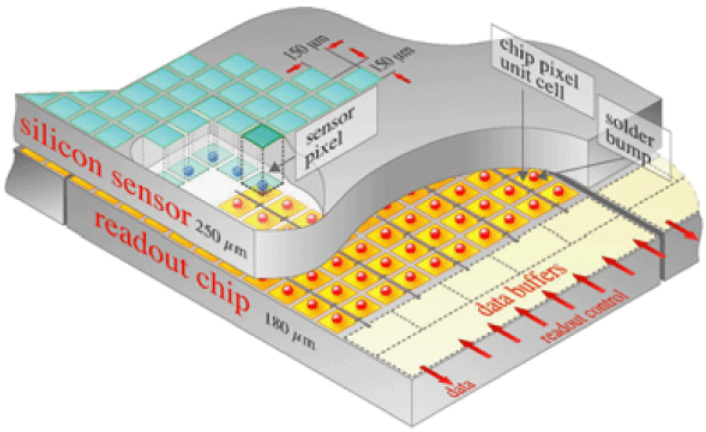

Figure 6: CMS pixel detector [44].

The Silicon Strip Detector

The next line of defense for outgoing particles in the CMS detector is the silicon strip detector. The silicon strip detector is comprised of a total of ten layers reaching 130 centimeters out from the center of the detector. Within these layers are silicon sensors that have fast response rates and

good spatial resolution. The operation and readout of these sensors is accomplished in much the same way as the cells of the pixel detector. As a charged particle crosses the material, it ejects electrons from the atoms, and these knocked-off charges give a very small current. This current is gathered, amplified, and readout by Analogue Pipeline Voltage (APV) chips [45]. This is with the goal in mind of tracking beyond the pixel detector the paths of particles emerging from a collision with extreme precision.

The Calorimeters

Calorimeters are utilized at CMS in order to measure the energy deposits of outgoing particles. Electrons and photons can generate electromagnetic showers with sufficient energy. Electromagnetic showers are often defined by a radiation length (X_0), defined as follows:

$$X_0 = 180\frac{A}{Z^2}, \tag{IV.3}$$

where A is the atomic weight and Z is the atomic number for the atom type of which the calorimeter material is made. A X_0 is defined as the distance a particle must travel through the calorimeter material so that the energy loss to photons is such that the particle energy decreases by a factor of $\frac{1}{e}$. The depth of the electromagnetic shower has some dependence on the energy of the incident particle [28]. For a photon or electron of ~ 50 GeV, $\sim 20X_0$ is necessary to accommodate the full extent of the electromagnetic shower.

Sufficiently energetic hadrons will generate hadronic showers based on the interaction between impinging particles and atomic nuclei. Hadronic showers are defined by absorption lengths (λ), defined as follows [30]:

$$\lambda = \frac{A}{N_A\sigma_{abs}}, \tag{IV.4}$$

where A is the atomic weight, N_A is Avogadro's number, and σ_{abs} is the absorption cross section. In the material that follows, we describe the two main calorimeters utilized for the CMS detector: the electromagnetic calorimeter (ECAL) and the hadronic calorimeter (HCAL).

The Electromagnetic Calorimeter

The ECAL is necessary for the reconstruction and identification of electrons and photons. The structure is composed of lead tungstate crystals that scintillate when electrons or photons pass through. This passing through of a charged particle produces light in short, fast, and well-defined pulses. This detected light is converted into electrical signals for amplification and analysis by photodetectors. The ECAL has a barrel section and two endcaps. The cylindrical barrel has 61,200 crystals, and the endcaps that seal off the barrel at either end have another nearly 15,000 crystals. A diagram of the CMS ECAL is shown in Figure 7.

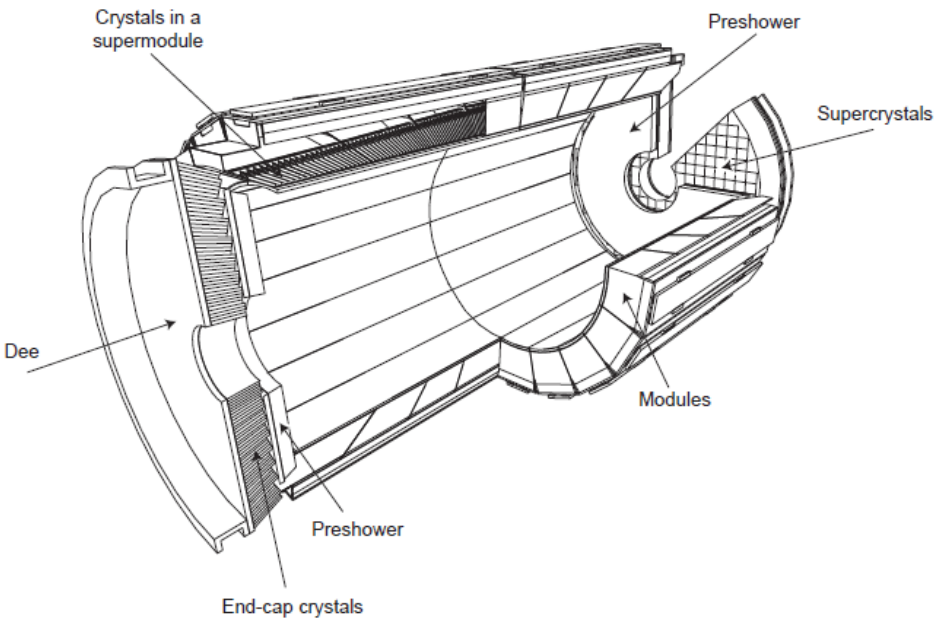

Figure 7: CMS ECAL detector diagram [14].

The Hadronic Calorimeter

The HCAL is necessary for the reconstruction and identification of hadrons. For this component of the CMS detector, there are alternating layers of absorber and scintillator material to prevent the escape of undetected particles. The HCAL is "hermetic" in nature, allowing for the capture of nearly every particle emerging from a collision. As a particle passes through the HCAL, it interacts with the scintillator material to produce blue-violet light. Optical fibers shift the wavelengths to green so that the light can be carried by optical cables to readout boxes. The outgoing particle deposits energy in a "tower." This total deposited energy is determined by Hybrid Photodiodes (HPDs) [42]. Again, the HCAL is necessary to determine the energies associated with both baryons and mesons.

The Muon System

The muon system is necessary to register signals from muons in the outermost regions of the detector. There exist four muon stations outside of the magnetic coil and interleaved within the plates of the steel return yoke. The robust structure of chambers aids in the filtering of background noise. Subcomponents of the muon system include drift tubes (DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). DTs measure the positions of muons in the barrel of the detector. In the endcap regions, there exist CSCs to account for the facts that the magnetic field is uneven and that the rates of particle interactions are high. The RPCs, fast and gaseous detectors, are included to complement the DTs and CSCs for muon detection. Muon reconstruction efficiency using only information from the muon system is typically $> 80\%$, except in the region of

$|\eta|$=1.2, where the efficiency is much lower due to the transition between the DTs and CSCs [43].

Data Acquisition and Triggering

A complex issue to consider at CMS is one of information storage. As previously mentioned, CMS contends with nearly one billion events occurring every second. Moreover, each proton-proton collision could produce an upwards of one thousand particles. Thus, there must exist a means by which to decide for which events to store the output information to search for signs of new and interesting physics.

This selection of events of interest is accomplished by a trigger. A trigger has some dependence on the spatial and time resolution of the detector, as it is necessary that each outgoing particle is identified and associated with the correct event. The first round of triggering is known as Level 1 (L1) and is purely hardware based. The L1 trigger looks at very basic requirements on an event like total energies, reducing the number of events considered every second from 10^9 to around 10^5. The High Level Trigger (HLT) makes more complex considerations in the search for new physics that reduces the number of events considered further from 10^5 to approximately 10^2 [48]. Specific triggers utilized in this analysis are described in the sections to follow.

Chapter V

Analysis Strategy

The ISR topology is characterized by the presence of one very energetic jet in the central region of the detector. As noted in the introduction, the use of an ISR jet with high p_T in the event topology creates a recoil effect that facilitates the detection of missing transverse momentum and provides a natural kinematic boost to the event to aid in the detection of otherwise low p_T (soft) leptons.

At the LHC, the SUSY $\tilde{\tau}$ can be produced directly ($pp \to \tilde{\tau}\tilde{\tau}$) or through cascading decays of the lightest chargino ($\tilde{\chi}_1^{\pm}$) and the next-to-lightest neutralino ($\tilde{\chi}_2^0$) in processes like $pp \to \tilde{\chi}_1^{\pm}\tilde{\chi}_1^{\mp} \to \tilde{\tau}\nu_\tau\tilde{\tau}\nu_\tau \to \tau\tilde{\chi}_1^0\nu_\tau\tau\tilde{\chi}_1^0\nu_\tau$ or $pp \to \tilde{\chi}_1^{\pm}\tilde{\chi}_2^0 \to \tilde{\tau}\nu_\tau\tilde{\tau}\tau \to \tau\tilde{\chi}_1^0\nu_\tau\tau\tilde{\chi}_1^0$. The strategy employed, including the optimization of signal region selections, is developed within the context of the R-parity conserving Minimal Supersymmetric Standard Model (MSSM) with bino (Z-like) $\tilde{\chi}_1^0$ and wino (W-like) $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$, a stau mass defined as $0.5(m_{\tilde{\chi}_1^{\pm}} + m_{\tilde{\chi}_1^0})$, and $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV. The MSSM is an extension of the $SU(3) \times SU(2) \times U(1)$ gauge theory. Within this theory is introduced one extra fermionic doublet Higgs boson. As a consequence of R-parity conservation, superpartners can only be produced in pairs; thus, the MSSM gives a stable lightest supersymmetric particle (LSP), the $\tilde{\chi}_1^0$.

The choice of model parameters above is motivated by examining the phase space where $\tilde{\tau}\tilde{\chi}_1^0$ coannihilation can give rise to the correct DM relic density and studying the areas of SUSY phase space where the current ATLAS and CMS searches have limited sensitivity. This analysis strategy also requires a single soft hadronic τ (τ_h) for two reasons. First, the branching fraction for τs decaying hadronically is higher than for those decaying leptonically. Second, although the above production processes for $\tilde{\tau}$ result in final states with multiple τ leptons, the compressed mass spectra regions of interest result in very soft τ_h candidates, making it difficult to reconstruct and identify multiple τ_h leptons, thus further motivating the single soft-τ_h final state.

Due to the presence of the $\tilde{\chi}_1^0$ LSP from the $\tilde{\chi}_1^{\pm}$ and $\tilde{\chi}_2^0$ decays, which escapes undetected and thus contributes to missing transverse energy (E_T^{miss}), the $\tilde{\chi}_1^{\pm}/\tilde{\chi}_2^0$ masses cannot be fully reconstructed. To successfully distinguish between backgrounds, the visible τ decay products and the E_T^{miss} are used to reconstruct the partial transverse mass of the system:

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

where ϕ is the azimuthal angle between the $\vec{E}_T^{miss}$ and the $\vec{p}_T(\tau_h)$. The z-component of the E_T^{miss} vector is considered zero. The m_T of the system is expected to be large for signal. The general overall strategy of the analysis is as follows. Upon selecting one high quality soft τ_h candidate,

one additional high-p_T jet, and large E_T^{miss}, we utilize the data distribution of m_T to fit for a potential signal that would appear as an excess over the SM expectation in the high m_T region of the distribution. To quantify the significance of any possible excess or set upper limits on the production rate, we perform a binned likelihood fit of the m_T. A representative signal feynman diagram for indirect production of the $\tilde{\tau}$ by decays of heavier SUSY particles is given in Figure 8.

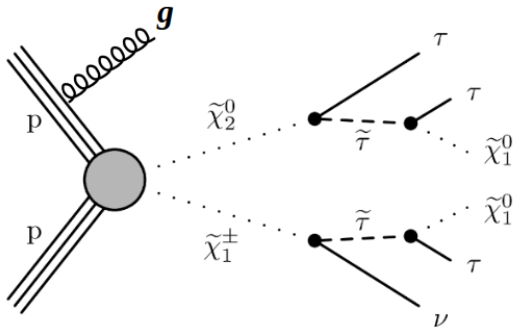

Figure 8: ISR topology for this study.

The signal region selections are more explicitly described in Table 4.

Basic Selection and Event Cleaning
— PV, remove cosmics, instrumental backgrounds "MET filters"
≥ 1 jet with $p_T > 30$ GeV, $\|\eta\| < 2.4$, and "Loose" ("Tight") ID for 2016 (2017)
— $p_T^{\text{lead}}(j) = p_T^{\text{ISR}}(j) > 100$ GeV
$E_T^{miss} > 230$ GeV (PFMet with HF and type-1 corrections)
exactly 1 τ_h with $20 < p_T(\tau_h) < 40$ GeV and $\Delta R(\tau_h, j) > 0.3$
old decay mode finding with 1 prong + "Tight" isolation
QCD rejection cut: $\|\Delta\phi(j_{\text{lead}}, \vec{E}_T^{miss})\| > 0.7$
Veto other leptons ($p_T(e,\mu) > 10$ GeV) and b-jets ($p_T > 30$ GeV)

Table 4: Signal Region Selections

The details of the selections will be described in more detail in the sections to follow. As seen in Table 1, exactly one hadronically-decaying tau lepton (τ_h) is required in our signal event selection targeting compressed mass spectra (where $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) < 50$ GeV). Figure 9 shows the multiplicity ($N(\tau_h)$) of reconstructed τ_h objects for four representative sets of SUSY masses ($m(\tilde{\chi}_1^{\pm}, m(\tilde{\tau}), and m(\tilde{\chi}_1^0)$) for $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV. In these distributions, the signal events include $\tilde{\tau}$ production from $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$, $\tilde{\chi}_1^{\pm}\tilde{\chi}_1^{\pm}$, $\tilde{\chi}_1^{\pm}\tilde{\chi}_1^{\mp}$, and $\tilde{\chi}_2^0\tilde{\chi}_2^0$ decays, where $\tilde{\chi}_1^{\pm}\tilde{\chi}_2^0$ is the dominant mechanism.

They show that the fraction of signal events with more than one τ_h is small, as expected, justifying our multiplicity selection.

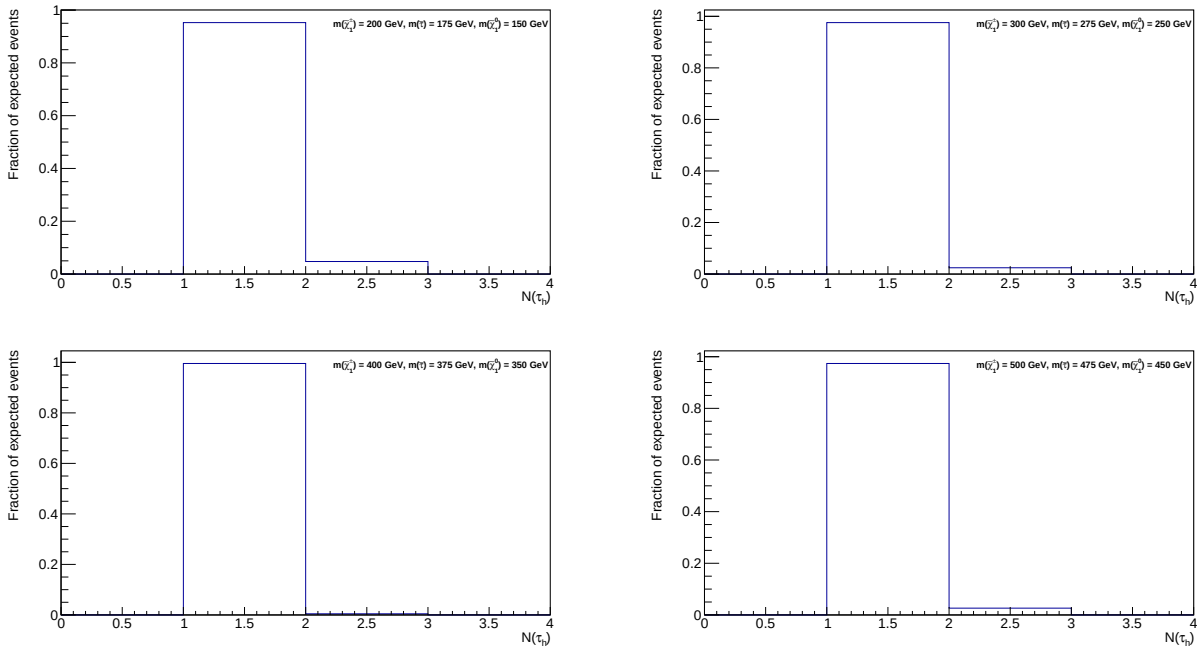

Figure 9: $N(\tau_h)$ for four representative sets of SUSY masses ($m(\tilde{\chi}_1^{\pm})$, $m(\tilde{\tau})$, and $m(\tilde{\chi}_1^0)$) for $\Delta m(\tilde{\chi}_1^{\pm}, \tilde{\chi}_1^0) = 50$ GeV.

The general methodology used for the estimation of background contributions in the signal region (SR) is based on both simulation and data. Since mismodeling of the MC background rates with a genuine τ_h candidate in the SR is expected to come from the combination of ISR and E_T^{miss} selections, the estimate of their background contribution to the search region is determined by obtaining background enriched control regions (CR), which contain negligible signal contamination (CR defined orthogonal to the SR). This is important to (1) validate the correct modeling of the τ_h selections and (2) to measure the data-to-MC correction factors for the ISR and E_T^{miss} selection efficiencies. In all cases, the level of contamination from other backgrounds in the control samples is subtracted off and is also used to derive systematic uncertainties on the background predictions. For backgrounds where a jet is misidentified as a τ_h, the background contribution to the search region is determined using data-driven techniques. The BG estimation studies of this analysis and their purposes are described in what follows. Primary backgrounds to the SR are Z+jets, W+jets, $t\bar{t}$, and QCD multijet.

A key aspect of this analysis is the understanding of the modeling of ISR in MC. In order to examine the ISR selection efficiency first, a Z+jets sample enriched with muons (μs) is obtained. Due to lepton universality, a control sample where μs are produced in decays of the Z can be used

to closely emulate the ISR jet activity in Z+jets events enriched with τs (as in the SR). The μs are far cleaner to identify than the τ_h candidates at CMS. This type of event is also advantageous in that it includes no real E_T^{miss}. In terms of efficiencies, the simplicity of this event type allows for the focus on the ISR selection efficiency.

Necessary event corrections for ISR derived from a Z+jets control sample must be validated in an orthogonal control sample, which for this analysis is a high-purity sample of W+jets events enriched with μs. To contrast with the Z process, exactly one μ is present. This W+jets CR contains real E_T^{miss} from the decay of the W boson to a neutrino. For a signal-like E_T^{miss} selection, ISR corrections are validated. Next, the ISR corrections are consistently applied, and the E_T^{miss} selection is toggled such that this region gives a handle on the level of E_T^{miss} mismodeling as well. This region gives a high degree of confidence that (1) the modeling of the ISR jet activity in MC is understood, (2) the ISR corrections are validated, and (3) the E_T^{miss} in MC is well modeled after appropriate corrections are applied. This study also serves as a validation that the W+jets background contribution to the SR can be taken from simulation.

Next, since the desired signal event topology includes a τ_h, a key focus of the analysis comes from the understanding of the τ_h identification efficiency in data and validating its modeling in simulation. In order to focus on the τ_h identification efficiency, a Z+jets control sample enriched with τs is studied. The efficiency for the requirement of two high quality τ_h candidates is expected to be well modeled by simulation (especially since all proper recommended corrections are applied). With established understanding of the ISR selection efficiency, this region gives a high degree of confidence that the τ_h identification efficiency is understood. As a result of this study (described in further detail to follow), we conclude the Z+jets background can be taken from simulation with proper corrections.

The estimate of the $t\bar{t}$ contribution to the SR is performed in a semi-data-driven way by obtaining four different control samples. The four CRs differ in included multiplicity of jets identified as b quarks (exactly 1 or exactly 2) and in the identification requirements for τ_h (the number of signal tracks and tightness of the isolation). The data-to-MC scale factor in the fourth $t\bar{t}$ CR with 2 b-jets and VTight identification on the τ_h, described in the object reconstruction and identification section to follow, is used to correct the $t\bar{t}$ MC prediction.

Estimate of the QCD multijet contribution to the SR is performed using a fully-data-driven method. A normal fake-factor method is utilized. The shape for QCD in the SR is extracted from a region similar to the SR, except the τ_h candidates in this region pass "Loose" and fail "Tight" isolation. The correct normalization is obtained by reweighting this shape with the $p_T(\tau_h)$-dependent tight-to-loose ratios for the τ_h candidates. These tight-to-loose ratios are obtained from a region of $W(\rightarrow \mu\nu + \tau_h)$, where the τ_h is a jet misidentified as a τ_h by construction.

The background predictions in the SR for smaller backgrounds of single top and diboson are determined entirely from simulation.

Chapter VI

Trigger Efficiency Studies

Trigger Efficiency Measurements

The selection of trigger is a crucial aspect of any analysis. The trigger is necessary to reduce the number of events examined to a manageable rate and to select events of interest with minimal bias in the final analysis. Since the trigger considerations are predominantly based on hardware, difficulty exists in simulating the effects of those considerations on simulation. When the trigger emulation is not included in the MC samples, we take the trigger efficiency from data. When the trigger emulation is properly included in the MC samples, we have to compare the final trigger efficiencies in MC and data. If the two are in disagreement, corrections must be applied to the MC. For the case of muons, this comes via a tag and probe method.

Tag and Probe Method

The tag and probe method makes use of dilepton decays of the Z boson, where the leptons are a pair of electrons (es) or a pair of μs. A clean sample of Z boson decays is achievable since the resolution on the mass of the Z is well-defined. With two leptons present in the event, one can be used to reduce other SM background processes while the other is used to measure trigger efficiency. This methodology is not advantageous for τs since the rate at which a jet fakes a τ lepton is much higher than the rate for faking an e or μ. We describe the typical methodology for determining trigger efficiency in events enriched with τ leptons in a subsequent section.

In what follows, we utilize μs in our examples without loss of generality. First, we obtain a sample of dimuon events. The first μ is required to pass the same selection criteria as the μ in the SR for the analysis. This μ is known as the "tag." The second μ, which is utilized to measure the trigger efficiency, is then the "probe." A high purity sample of dimuon events is obtained by requiring the dimuon mass to fall within three standard deviations of the accepted value of the Z boson mass. The number of events with at least one good tag μ is n_t, and the number of events in which the probe μ passes the given criteria is n_p. Technically, the probe μ could also be the tag μ, so, to account for this, the efficiency (ϵ) is given as follows:

$$\epsilon = \frac{2n_p}{n_p + n_t}. \tag{VI.1}$$

The tag and probe method yields measured efficiencies consistent with true MC efficiencies.

Muon Efficiencies and Scale Factors

The muon object performance group (POG) closely studies trigger and identification efficiencies [39]. The MC samples do emulate the single-μ triggers well. A scale factor is derived to be applied to the MC rates and efficiencies such that they agree with those obtained from data. Systematic uncertainties are also assigned based on the efficiency values. The scale factor derived is the ratio of the trigger efficiency in data to that in MC. We follow POG recommended criteria for the μ selections. The overall trigger efficiencies and corresponding scale factors as measured by the CMS μ POG [39] are outlined in Table 5. The uncertainties reported are statistical.

Scale factor	$\lvert\eta\rvert$ region	Data efficiency [%]	Scale factor
L1 w.r.t. offline	$0.0 < \lvert\eta\rvert < 0.9$	96.86 ± 0.02	0.9914 ± 0.0005
	$0.9 < \lvert\eta\rvert < 2.4$	94.38 ± 0.02	0.9947 ± 0.0005
HLT w.r.t. L1	$0.0 < \lvert\eta\rvert < 0.9$	99.67 ± 0.02	0.9967 ± 0.0005
	$0.9 < \lvert\eta\rvert < 2.4$	99.46 ± 0.02	0.9957 ± 0.0005
Online w.r.t. HLT	$0.0 < \lvert\eta\rvert < 0.9$	97.95 ± 0.02	0.9906 ± 0.0005
	$0.9 < \lvert\eta\rvert < 2.4$	98.28 ± 0.02	0.9931 ± 0.0005

Table 5: Contributions to the isolated single-μ trigger efficiency in 2015 data, integrated over $p_T >$ 22 GeV.

Tau Efficiencies and Scale Factors

As mentioned previously, a clean sample enriched with τ leptons cannot be obtained from tag and probe methods. One reason this is the case is because the jet to τ lepton fake rate is much larger then for jet to μ. In addition, since τ leptons lose energy to their neutrino decay products, resolution on the mass of the Z boson from a ditau decay is poor. The measurement of τ tagging efficiencies and fake rates must be done in such a way that they remain unbiased.

In order to accomplish this, we utilize a sample of $Z(\rightarrow \tau\tau \rightarrow \mu\tau_h)$ events to get a clean sample enriched with τ leptons. Creating a region of this type is advantageous since μs have the lowest jet misidentification rate of the leptons. Thus, requiring a single μ with certain "Tight" identification criteria (to be described later) reduces the contributions from other SM backgrounds greatly.

The main cause for inefficiency in the τ lepton isolation is due to the presence of other particles that fall into the isolation region. In the region enriched with ditau events for this analysis, the τ_hs are required to have only one charged hadron in the isolation region. Thus, the CR used to study the τ lepton triggers is not sensitive to leakage effects for three pronged τ leptons. In addition, the ECAL recovers losses in efficiency due to photon conversions to electron-positron pairs that are present in the isolation region. Therefore, based on the selections for τ leptons in this analysis, the inefficiencies for τ leptons are expected to be similar to those for μs.

The measurement of the τ_h identification efficiencies and scale factors is accomplished by the

τ lepton POG [38]. Table 6 gives the data-to-MC scale factors for different working points of the MVA-based isolation discriminant for τ leptons, using highly boosted Z/γ^* events decaying to τ lepton pairs.

Working point	Scale factor
Very loose	0.97 ± 0.09
Loose	0.99 ± 0.09
Medium	0.98 ± 0.09
Tight	0.96 ± 0.08
Very tight	0.95 ± 0.09
Very-very tight	0.90 ± 0.08

Table 6: Data-to-simulation scale factors for different working points of the MVA-based isolation discriminant, using highly boosted Z/γ^* events decaying to τ lepton pairs [38].

Trigger Efficiency for this Analysis

The selection of the trigger is motivated by the final state of the expected signal: exactly one soft τ_h, a high-p_T jet from ISR, and large E_T^{miss}. Current τ_h triggers available at CMS are inefficient in the low-p_T range of interest for this analysis. So, instead, data events in the signal region are required to fire the HLT_PFMETNoMu120_PFMHTNoMu120_IDTight trigger. This high-level trigger is defined by the values at which it achieves maximum efficiency for the E_T^{miss} and the H_T values for "Tight" lepton identification requirements. The trigger efficiency is measured in a single-μ control sample enriched with W+jets events ($> 95\%$ purity) selected with an isolated-muon trigger, which allows us to study the behavior of the trigger as a function of E_T^{miss} in an orthogonal sample with respect to selected events in the SR. Table 7 outlines the event selection criteria used to obtain the high purity W+jets CR. In addition to the single-μ selection criteria (trigger, "Tight" μ identification, $p_T > 30$ GeV, and $|\eta| < 2.1$), the ISR jet selections used in our final search region are also imposed in order to study events with a signal-like topology.

In order to understand the trigger turn-on curve, the E_T^{miss} threshold is set to 50 GeV. The denominator used in the efficiency calculation is defined as the number of events that pass the selection criteria previously described. The numerator is a subset of the denominator events that additionally fire the HLT_PFMETNoMu120_PFMHTNoMu120_IDTight trigger. The efficiency of the single-μ trigger used in the denominator is effectively factored out since the $p_T(\mu)$ selection threshold is 30 GeV, which sits firmly on the single-μ trigger efficiency plateau.

There exists a correlation between the selection for the p_T of the jet from ISR and the selection made for the E_T^{miss} threshold. Thus, the trigger efficiency is measured as a function of both kinematic quantities. We obtain an efficiency of over 90% for $E_T^{miss} > 200$ GeV and $p_T^{\text{lead}}(j) > 100$ GeV. This result is shown in Figure 10, measured using data.

Selections for the Trigger Study
Trigger1: HLT_IsoMu24_v
N(μ) ≥ 1
$p_T(\mu) > 30$ GeV
$\lvert\eta(\mu)\rvert < 2.1$
ID(μ): "Tight"
$N(j) \geq 1$
$p_T^{\text{lead}}(j) > 50$ GeV
$\lvert\eta(j_{\text{lead}})\rvert < 3$
$N(\text{b}-\text{jets}) == 0$
$E_T^{miss} > 50$ GeV
overlaps removal$\sim \Delta R(\tau_h, j) > 0.3$
Trigger2 (numerator): HLT_PFMETNoMu120_PFMHTNoMu120_IDTight

Table 7: Event selection criteria for trigger efficiency studies.

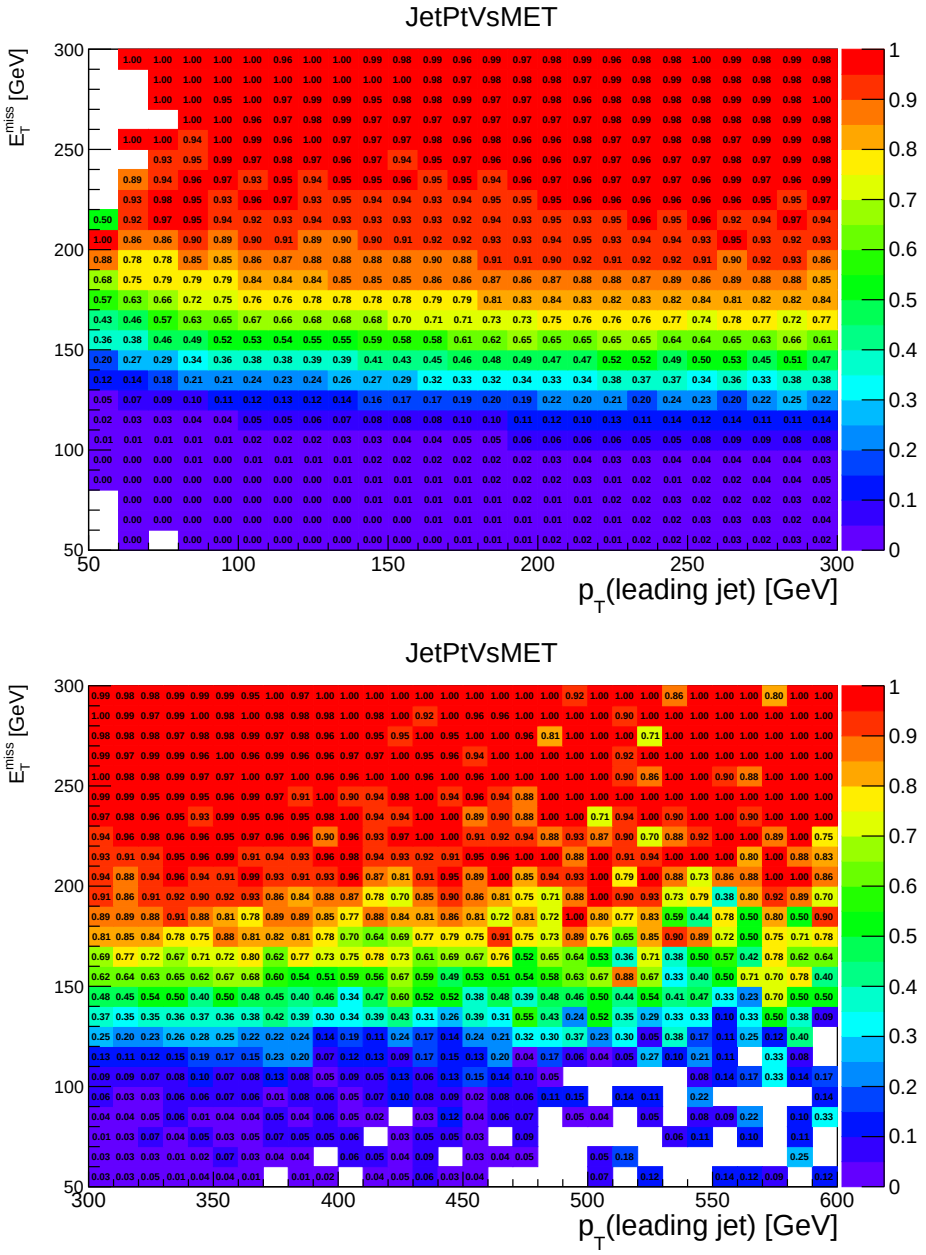

Figure 10: Trigger correlation between ISR jet and the E_T^{miss}: 50 to 300 GeV (top) and 300 to 600 GeV (bottom).

Chapter VII

Optimization Studies

We work to optimize the signal region selections for best discovery potential. The optimization of the event selection criteria is performed using the $\frac{S}{\sqrt{S+B}}$ figure of merit, where S is the signal yield, B is the total background yield, and $\sqrt{S+B}$ is the statistical uncertainty on the total signal plus background yield. To begin the process of optimization, we establish the base selections in Table 8.

Optimization Study Base Selections
$N(\tau_h) \geq 1$
$p_T(\tau_h) > 20$ GeV
$\lvert\eta(\tau_h)\rvert < 2.1$
$N(j) \geq 1$
$p_T^{\text{lead}}(j) > 100$ GeV
$\lvert\eta(j_{\text{lead}})\rvert < 2.4$
$N(\text{b}-\text{jets}) == 0$
$E_T^{miss} > 230$ GeV
overlaps removal$\sim \Delta R(\tau_h, j) > 0.3$

Table 8: Base event selections for optimization studies

To establish the selection for $p_T(\tau_h)$, we take two approaches. First, we study the idea of utilizing a minimal $p_T(\tau_h)$ threshold only. We sample 10 GeV increments from 20 to 60 GeV. The results of this study appear in Figure 11 for three different $m(\tilde{\chi}_1^0)$. Next, we look at the case in which we establish a minimal and a maximal threshold for $p_T(\tau_h)$. We fix the lower $p_T(\tau_h)$ threshold at 20 GeV to accommodate considerations for τ_h reconstruction in the detector. We sample 10 GeV increments from 30 to 70 GeV for the upper p_T threshold. The results of this study appear in Figure 12 for the three $m(\tilde{\chi}_1^0)$. With an established minimal threshold at 20 GeV, we achieve a factor of 2-3 increase in significance in optimizing the maximal $p_T(\tau_h)$ threshold at 40 GeV. Based on these results, the $p_T(\tau_h)$ is selected such that $20 < p_T(\tau_h) < 40$ GeV. We note that, described in a subsequent section, the analysis makes use of the transverse mass ($m(\tau_h, E_T^{miss})$) between the τ_h and the E_T^{miss} as the key observable on which to perform a binned-likelihood fit to search for signal. Hence, since the m_T depends on the $p_T(\tau_h)$ and the E_T^{miss}, these variables are indirectly optimized already when performing the fit to the $m_T(\tau_h, E_T^{miss})$ distribution shape.

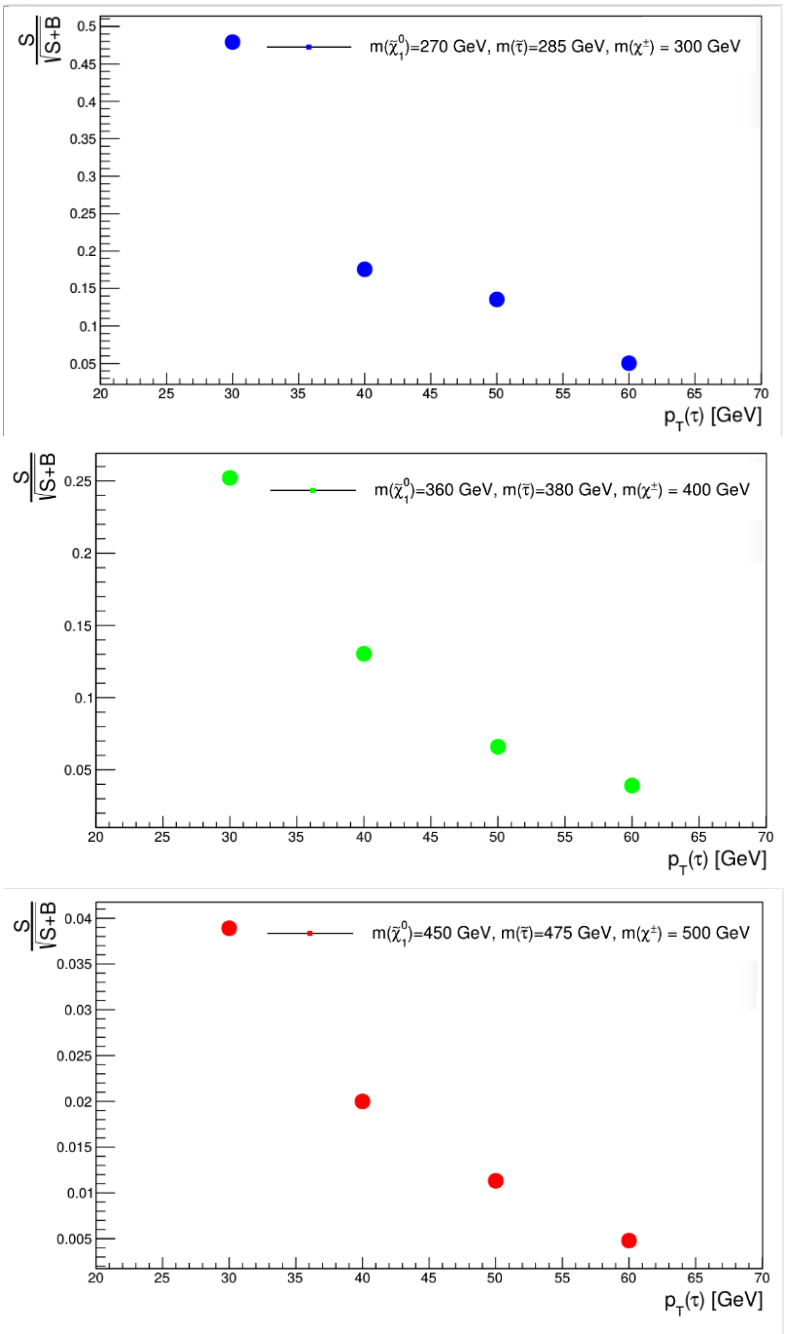

Figure 11: Optimization of the $p_T(\tau_h) > X$ criterion, using signal points with LSP masses of 270 GeV, 360 GeV, and 450 GeV.

The next selection that we optimize is for the p_T of the high-momentum jet from ISR ($p_T^{\text{lead}}(j)$). We study the idea again of utilizing a minimal p_T threshold. We sample 10 GeV increments from 100 to 150 GeV for the signal sample with $m(\tilde{\chi}_1^0) = 270$ GeV and from 110 to 160 GeV for the signal sample with $m(\tilde{\chi}_1^0)$. The results of this study can be seen in Figure 13. We note that there is

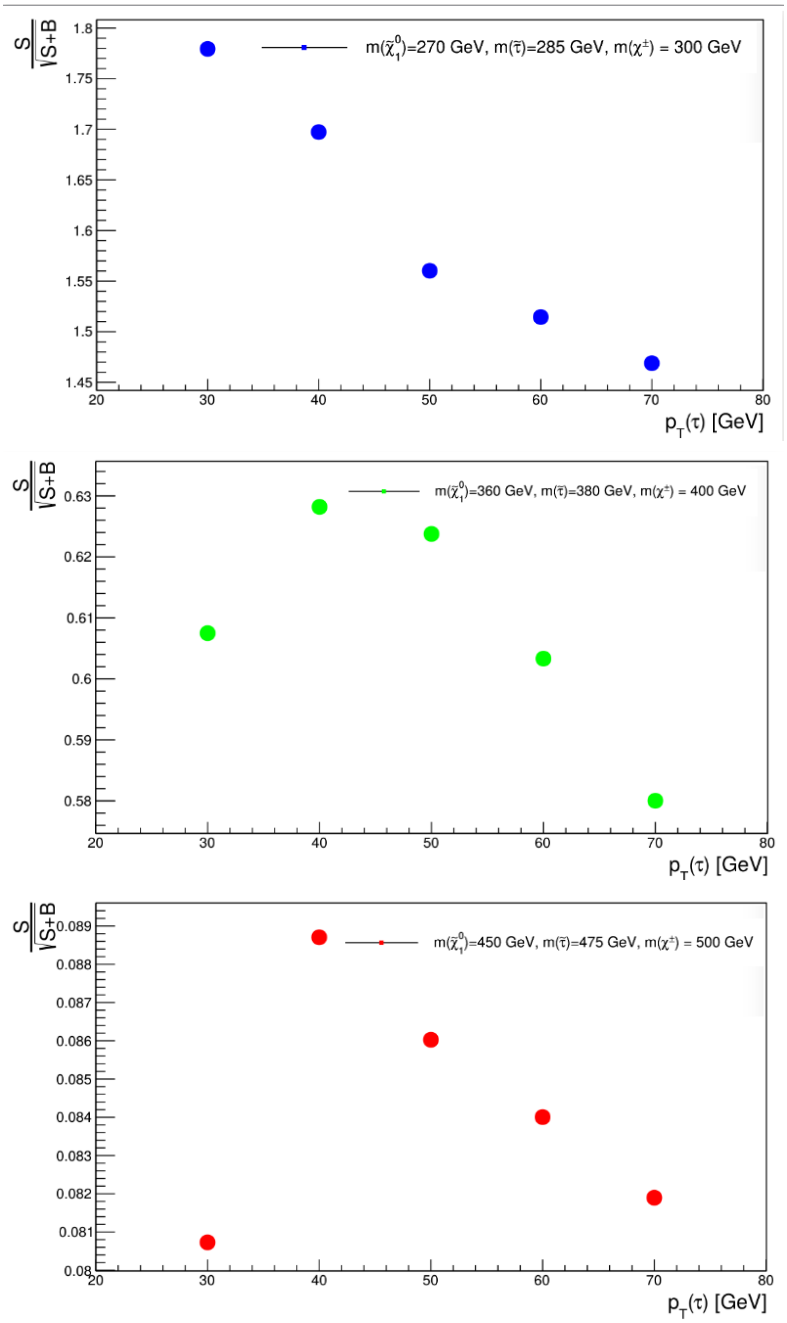

Figure 12: Optimization of the $20 < p_T(\tau_h) < X$ criterion, using signal points with LSP masses of 270 GeV, 360 GeV, and 450 GeV.

minimal gain in significance after the 100 GeV selection. However, we select 100 GeV to minimize overall statistical uncertainty associated with this selection criteria.

We use a binned fit of the $m_T(p_T(\tau_h), E_T^{miss})$ distribution to extract the signal and set limits. We note here as well that we select τ_h candidates that have exactly 1 charged track in order to minimize the contamination from jets being mis-identified as τ_h candidates in QCD multijet

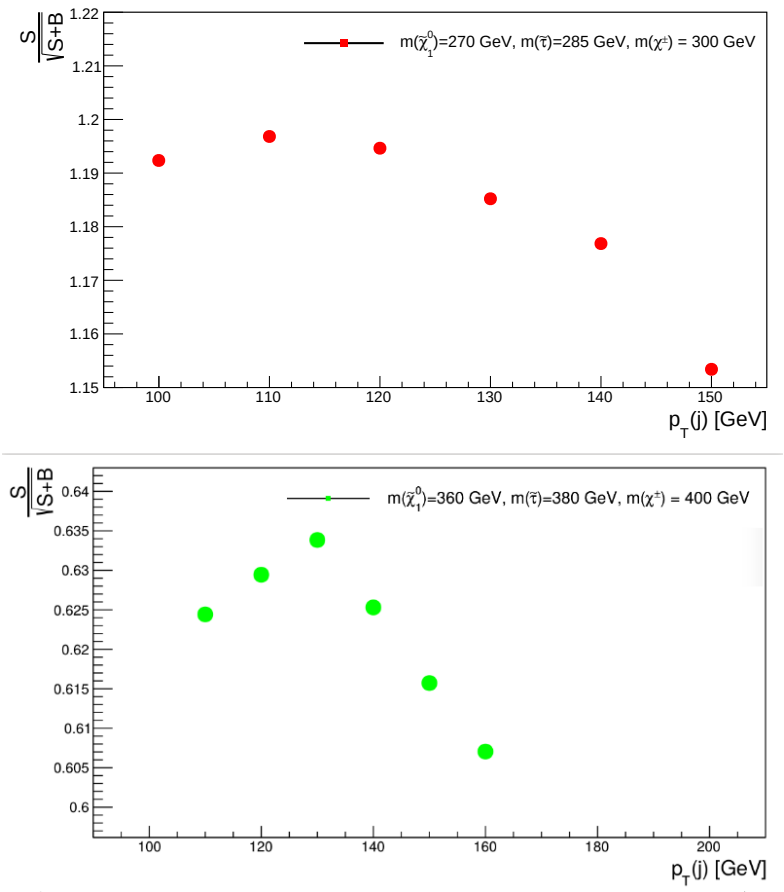

Figure 13: Optimization of the $p_T^{\text{lead}}(j)$ criterion, using signal points with LSP masses of 270 GeV and 360 GeV.

processes. Technically speaking, we select exactly 1 τ_h candidate passing the POG recommended criteria (that is, 1 or 3 charged tracks) and then additionally requiring that this τ_h has exactly 1 charged track. This in effect means that we reject events that have a secondary τ_h with 3 charged tracks. The primary SM background processes contributing to this analysis, as will be seen explicitly in the sections to follow, are from W+jets and QCD multijet processes. The event selection criteria for the SR are listed in Table 9.

Optimized Signal Selections
$N(\tau_h) = 1$
$20 < p_T(\tau_h) < 40$ GeV
$\lvert\eta(\tau_h)\rvert < 2.1$
τ_h ID oldDMF + "Tight" Iso + "againstElectronMVALooseMVA6" + "againstMuonTight3"
N(signal tracks) $= 1$
$N(j) \geq 1$
$p_T(j) \geq 30$ GeV
$\lvert\eta(j)\rvert < 2.4$
$p_T^{\text{lead}}(j) > 100$ GeV
$\lvert\eta(j_{\text{lead}})\rvert < 2.4$
$N(\text{b} - \text{jets}) == 0$
b − jet ID: pfCombinedInclusiveSecondaryVertexV2BJetTags, "Medium"
$E_T^{miss} > 230$ GeV
QCD rejection: $\lvert\Delta\phi(j_{\text{lead}}, \vec{E}_T^{miss})\rvert > 0.7$
veto other leptons $p_T(\ell) > 10$ GeV for $\ell = e, \mu$
overlaps removal$\sim \Delta R(\tau_h, j) > 0.3$

Table 9: Optimized event selection criteria

Chapter VIII

Particle Identification and Event Reconstruction

Jet Reconstruction

The Particle Flow (PF) technique is used to improve the jet p_T and angular resolution in this analysis [1, 2]. The PF technique combines information from different subdetectors to produce a mutually exclusive collection of particles (muons, electrons, photons, charged hadrons, and neutral hadrons) that are used as input for jet clustering algorithms. Jets are clustered using the anti-kT algorithm, with a distance parameter of $\Delta R = 0.4$ [15]. We use these AK4 jets from the collection "slimmedJets," which are made from ak4PFJetsCHS.

The anti-kT algorithm is based on the calculation of a pair of distances:

$$d_{i,j} = \min\left(\frac{1}{k_{t,i}^2}, \frac{1}{k_{t,j}^2}\right)\frac{\Delta R_{i,j}^2}{R^2} \quad \text{(VIII.1)}$$

$$d_{i,B} = \frac{1}{k_{t,i}^2}, \quad \text{(VIII.2)}$$

where $k_{t,i}$ is the transverse momentum of cluster i, R is the jet reach parameter, and $\Delta R_{i,j}$ is the distance in $\eta - \phi$ space between clusters i and j:

$$\Delta R_{i,j} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}. \quad \text{(VIII.3)}$$

For the first iteration of the anti-kT algorithm, $d_{i,j}$ is calculated for each pair ij, while $d_{i,B}$ is calculated for each cluster. In subsequent iterations, cluster pairs with the smallest $d_{i,j}$ are combined to form a "protojet," and the clusters i and j are removed from the cluster for the next iteration. The PF anti-kT algorithm gives a four-momentum obtained from adding the four-momenta of all the associated PF candidates.

The reconstructed jets require energy corrections that are obtained using simulation. The simulated events are generated with PYTHIA, processed through a detector simulation based on GEANT4, and confirmed with in situ measurements of the p_T balance. The overall jet-energy corrections depend on the η and p_T values of the jets. The jet-energy corrections are applied by L1 FastJet, L2 Relative, and L3 absolute corrections. In order to remove the extra energy in jets from underlying events (UE) and pileup, the L1 FastJet corrections use the event-by-event comparison of UE/pileup. The L2 and L3 corrections use jet balancing and γ+jet events to provide a better energy response as a function of the η and p_T for jets. For data, additional residual corrections are applied. Reconstruction efficiency for jets is known to be >99%.

Jet Identification

Jets are required to have $p_T > 30$ GeV and $|\eta| < 2.4$ in this analysis. The selections imposed on jets that are required to pass "Loose" identification are given in Table 10. The recommended "Loose" jet identification is applied to jets in the 2016 iteration of this analysis. The selections imposed on jets that are required to pass "Tight" identification are given in Table 11. The recommended "Tight" jet identification is applied to jets in the 2017 iteration of this analysis.

Table 10: Loose Jet-ID Selections.

Selection	Cut
Neutral Hadron Fraction	< 0.99
Neutral EM Fraction	< 0.99
Number of Constituents	> 1
And for $\eta < 2.4$, $\eta > -2.4$ in addition apply	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

Table 11: Tight Jet-ID Selections.

Selection	Cut
Neutral Hadron Fraction	< 0.90
Neutral EM Fraction	< 0.90
Number of Constituents	> 1
And for $\eta < 2.4$, $\eta > -2.4$ in addition apply	
Charged Hadron Fraction	> 0
Charged Multiplicity	> 0
Charged EM Fraction	< 0.99

For the purposes of the $t\bar{t}$-enriched control samples in this analysis, we use b-tagged jets. This is to make an estimate of the $t\bar{t}$ contribution to the SR. The combined secondary vertex v2 algorithm is used to identify a jet as originating from hadronization of a b-quark [4]. The pertinent discriminator name is "pfCombinedInclusiveSecondaryVertexV2BJetTags." The algorithm combines reconstructed secondary vertex and track-based lifetime information to build a MVA-based discriminator to distinguish between jets from b-quarks and those from the charm or light quarks and gluons. The minimum thresholds on these discriminators define "Loose," "Medium," and "Tight" operating points with a misidentification probability of about 10%, 1%, and 0.1%, respectively, for an average jet p_T of about 80 GeV. The Medium operating point with an efficiency of about 60% is used in this analysis. The methodology of "Event reweighting using scale factors only" is described here: `https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods#1c_Event_reweighting_using_scale`.

Electron Reconstruction

Electrons are reconstructed at CMS using information from both the tracker and the ECAL. When an electron passes through the silicon tracker material, it loses energy due to Bremsstrahlung radiation. The energy of the radiated photons is spread over the ECAL crystals along the electron trajectory (mostly the ϕ-direction). There exist two algorithms for clustering: "Hybrid" for the barrel and "Island" for the endcaps. These are used for the measurement of electron and photon energies [20].

Electron tracks are reconstructed by matching trajectories in the silicon strip tracker to seed hits in the pixel detector. A pixel seed is composed of two pixel hits compatible with the beam spot. A Gaussian Sum Filter (GSF) is used for the reconstruction of trajectories in the silicon strips. The track that best matches an energy supercluster in the ECAL is chosen to be the reconstructed track. This helps to minimize the number of choice trajectories possible due to different combinations of hits.

For effective electron reconstruction, good geometrical matching and good agreement between the momentum of the track and the energy supercluster in the ECAL are necessary. There are a handful of pertinent quantities to consider for geometrical matching. The first is $\Delta\eta_{\text{in}} = \eta_{\text{sc}} - \eta^{\text{Track}}_{\text{vertex}}$, where η_{sc} corresponds to the supercluster η and $\eta^{\text{Track}}_{\text{vertex}}$ corresponds to the position of the track at the interaction vertex extrapolated, as a perfect helix, to the ECAL detector. Next, we consider $\Delta\phi_{\text{in}} = \phi_{\text{sc}} - \phi^{\text{Track}}_{\text{vertex}}$, where ϕ_{sc} corresponds to the supercluster ϕ and $\phi^{\text{Track}}_{\text{vertex}}$ corresponds to the position of the track at the interaction vertex extrapolated, as a perfect helix, to the ECAL detector. The energy-momentum matching is measured by the ratio between the corrected energy in the ECAL supercluster and the momentum of the track taken in the inner layers of the tracker.

Electron Identification

We must guard against electrons that are misidentified from photon conversions. These electrons are removed by requiring that the track associated with the electron has hits in the inner layers of the pixel detector. This is because an electron arising from photon conversion has hits that instead arise much later in subsequent detector components. In all channels, we utilize electron identification and isolation requirements that follow the POG recommended criteria. For identification, we use the "Medium" identification working point for the cut-based identification. The electron identification efficiencies and scale factors are taken from the appropriate Twiki page found here: `https://twiki.cern.ch/twiki/bin/viewauth/CMS/CutBasedElectronIdentificationRun2`. The main electron identification selections are summarized in Table 12.

Table 12: Electron ID Selections

Cut	Barrel	EndCap
H/E	< 0.253	< 0.0878
$\sigma_{i\eta i\eta}$	< 0.00998	< 0.0298
$\lvert\Delta\eta_{\text{in}}\rvert$	< 0.00311	< 0.00609
$\lvert\Delta\phi_{\text{in}}\rvert$	< 0.103	< 0.045
$1/E - 1/p$	< 0.134	< 0.13
Missing inner hits	< 2	< 2
Pass conversion veto	yes	yes
Rel Combined PF Iso (EA corr)	< 0.0695	< 0.0821

Muon Reconstruction

Muon reconstruction begins with information gathered from the muon subdetectors. A standalone muon is reconstructed from hits in the individual drift tube (DT) and cathode strip chambers (CSC). These hits, or "seeds," provide an initial estimate of the muon momentum. Within the muon system, the seeds help to fit for the standalone muon trajectory. We have the ability to reject hits from scenarios like showering or pair production using a selection on the χ^2 for the fit. Connecting the innermost muon station to the outer tracker, we extrapolate for the standalone muon trajectory. A global muon is present when the standalone muon track can be matched to tracks in the silicon tracker. The main source of background with which to contend in the process of reconstructing global muons comes from processes with charged hadrons in the output. The muon POG at CMS works diligently to expand upon muon reconstruction [18]. The reconstruction efficiency is approximately 99% for $p_T > 10$ GeV. Selecting muons with $p_T < 10$ GeV is likely to produce discrepancies between data and simulation.

Muon Identification

In the process of identifying muons, we must guard against fakes from charged pions. Charged hadrons that penetrate the hadronic calorimeter and leave hits in the muon system will deposit most of their energy in the calorimeters. Thus, quality calorimeter compatibility algorithms can help to discriminate against these types of events. However, calorimeter compatibility is not exploited in this analysis due to the uncertainty in how well these algorithms perform in the presence of high concurrent event multiplicity (pileup). The presence of punch-throughs is from pions from the fragmentation of quarks and gluons. These punch-throughs can often be discriminated against by making an isolation - requiring minimal calorimeter deposits and tracks within a region around the muon - requirement. This is similarly the case for non-prompt muons from heavy flavor decays and decays in flight. Further information regarding muon identification is available in multiple sources [18, 19].

Isolated muons are required to have minimal energy from PF neutral and charged candidates in a cone of $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ around the lepton trajectory. We use the PF-based combined relative isolation with $\delta\beta$ corrections around the lepton trajectory. It is customary at the CMS experiment to use isolation weighted by the lepton (in this case muon) p_T. This helps in the discrimination against low-p_T jets. A full list of muon identification criteria is given in Table 13.

Table 13: μ Identification Criteria

Cut
recoMu.isGlobalMuon()
muon::isPFMuon()
recoMu.globalTrack()- > normalizedChi2() < 10
recoMu.globalTrack()- > hitPattern().numberOfValidMuonHits() > 0
recoMu.numberOfMatchedStations() > 1
fabs(recoMu.muonBestTrack()- > dxy(vertex-> position())) < 0.2
fabs(recoMu.muonBestTrack()- > dz(vertex- > position())) < 0.5
recoMu.innerTrack()- > hitPattern().numberOfValidPixelHits() > 0
recoMu.innerTrack()-> hitPattern().trackerLayersWithMeasurement() > 5

We use the "Tight" identification working point and relative isolation < 0.25 in this analysis. The identification and isolation used follow the POG recommended criteria. The muon trigger/identification efficiencies and scale factors used to correct the MC expectations in these analyses have been taken from the Muon POG Run2 recommendations.

Tau Reconstruction

There are several different algorithms that may be used for tau lepton reconstruction at CMS. This analysis employs the Hadrons Plus Strips (HPS) algorithm, which aids in combatting issues arising from generic quark and gluon QCD jets misidentified as tau leptons. We use HPS to identify hadronically decaying tau leptons based on Particle Flow (PF) objects. The PF jets are used as inputs to an algorithm that uses ECAL strips to reconstruct neutral pions. The strips are then combined with charged hadrons within the PF jets. We expand on this idea in what follows.

The jets from tau leptons tend to be better collimated with lower included particle multiplicity than those that arise from QCD events. Thus, PF builds a tau lepton from jets by defining a narrow cone around the central axis of the jet to define the tau lepton constituents. A larger region can then be designed to discriminate against the fake τ_hs from hadronic jets. Two goals to keep in mind are maximizing the probability of correctly identifying a τ_h and minimizing the probability of incorrectly identifying a hadronic jet from QCD as a τ_h. These goals are contingent on establishing the appropriate size of the isolation cone.

It has been established that the angle between the decay products in the detector is seen to

decrease with increasing momentum and energy of the tau lepton. Therefore, constructing a cone that depends on transverse energy (E_T) is necessary to properly isolate the tau lepton. Unlike tau leptons, the distance between the constituent particles is not correlated with the energy of a QCD jet. Given these considerations, the τ_h candidates are built from PF jets by defining a narrow region in $\eta - \phi$ space with respect to the highest p_T track around the central jet axis [35]. A cone size dependent on E_T has the advantages as described above for both the low and high extrema of the $p_T(\tau_h)$ distribution.

The sketch in Figure 14 depicts the tau lepton isolation cone definitions. Figure 15 gives the τ_h E_T resolution for PF based reconstruction, calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections. Particle flow based reconstruction yields significantly better resolution than calorimeter based reconstruction in general. The PF tau lepton resolution is approximately 5%.

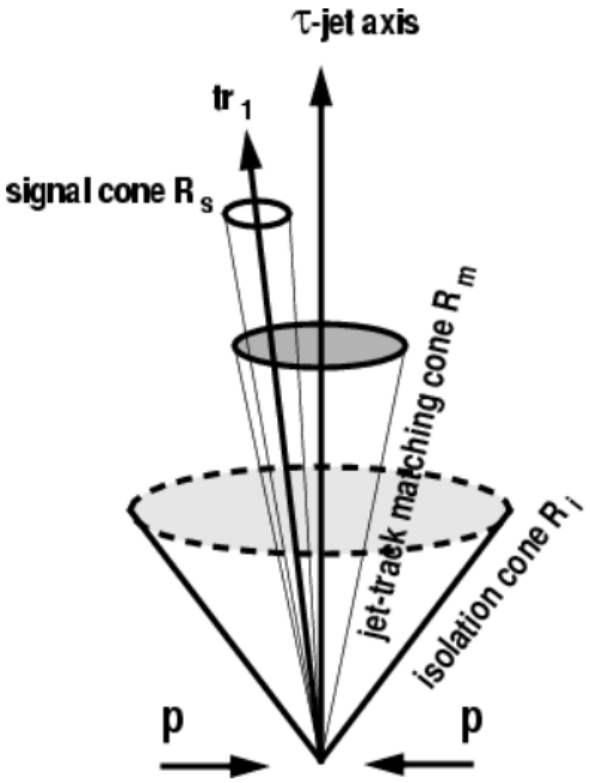

Figure 14: Sketch depicting tau lepton isolation cones.

The main tau lepton decay modes are listed in Table 14. The single hadron plus zero strips decay mode attempts to reconstruct $\tau \to \nu\pi^{\pm}$ decays or $\tau \to \nu\pi^{\pm}\pi^0$ decays where the neutral pion has very low energy. The single hadron plus one or two electromagnetic strips mode attempts to reconstruct tau lepton decays that produce neutral pions where the resulting neutral pion decays produce collinear photons. Listed next, the single hadron plus two strips mode attempts to reconstruct

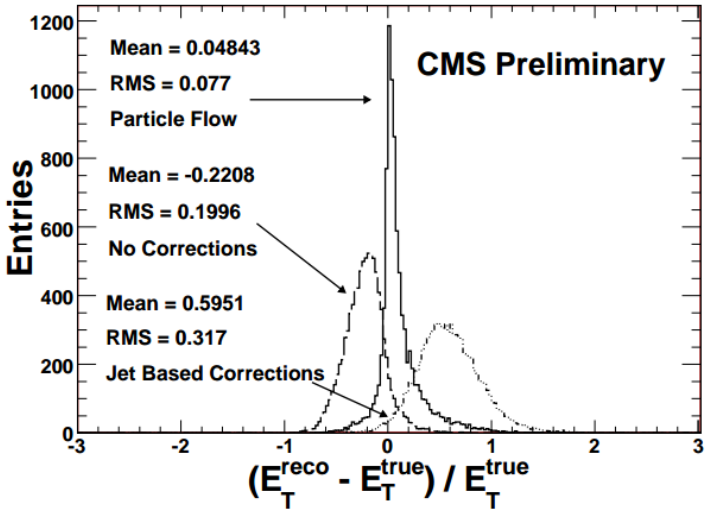

Figure 15: Relative τ_h E_T resolution for PF based reconstruction (solid line), calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections.

tau leptons that decay via e.g. $\tau \to \nu\pi^{\pm}\pi^0$ where the neutral pion decays to well separated photons resulting in two electromagnetic strips. Finally, the three hadrons decay mode: attempts to reconstruct tau lepton decays that occur via $\rho(770)$ resonance.

Table 14: Reconstructed Tau Decay Modes

HPS Tau Decay Modes
Single Charged Hadron + Zero Strip
Single Charged Hadron + One Strip
Single Charged Hadron + Two Strips
Two Charged Hadrons
Three Hadrons

Tau Identification

Since the analysis focuses on compressed spectra with exactly one soft τ_h, we use the multivariate (MVA) tau lepton ID discriminator trained with the "old" decay mode finding. This comes from the recommendations from the tau lepton POG [35]. This discriminator combines the tau lepton isolation and lifetime information and proves to give the best possible discrimination between real τ_h candidates and hadronic jets. We choose the "Tight" working point by POG recommendation as a balance of real τ_h efficiency ($\sim 50\%$) and hadron jet rejection rate (>99% for $p_T(j) > 20$ GeV).

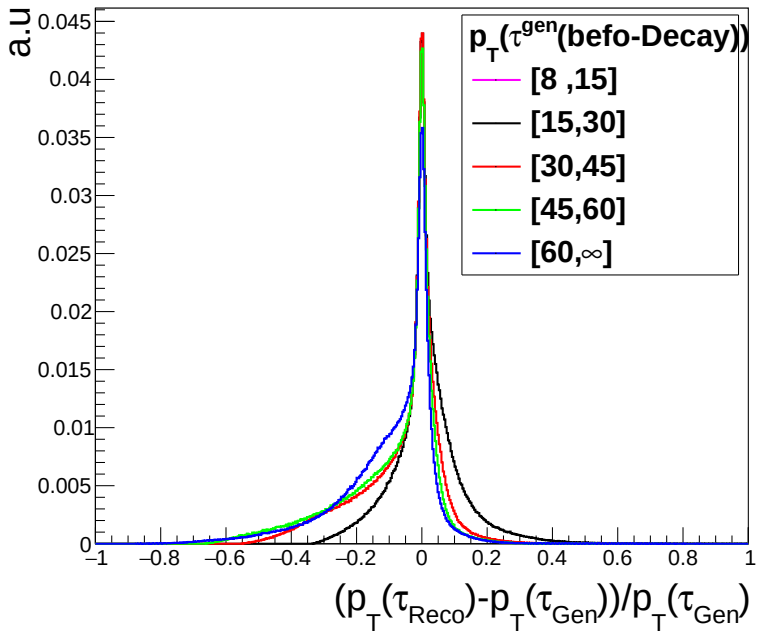

Figure 16: Relative p_T resolution of reconstructed τ_h candidates.

In order to discriminate against muons, HPS tau leptons are required to pass the lepton rejection discriminator which requires the lead track of the tau not be associated with a global muon signature. In order to discriminate against electrons, HPS taus are required to pass an MVA discriminator which uses the amount of HCAL energy associated to the tau lepton with respect to the measured momentum of the track ($\frac{H}{p}$). In addition, the MVA discriminator considers the amount of electromagnetic energy in a narrow strip around the leading track with respect to the total electromagnetic energy of the tau lepton. Finally, HPS taus must not reside in the ECAL cracks, and we guard against this with geometric considerations. The exact discriminator names and working points are listed and described in each their respective sections.

We describe next the tau lepton energy scale and resolution. Since the resolution and scale of m_T reconstruction depends on the effectiveness of the τ_h reconstruction, we summarize studies on the τ_h response and resolution. We define the response as the relative difference between the p_T of a reconstructed τ_h (that has passed all τ_h ID discriminators) and the p_T of a generated τ_h that has been matched $\Delta R < 0.2$ to the reconstructed tau lepton. We see from Figure 15 that the response distribution contains a narrow Gaussian-like component in the bulk, in addition to a relatively long tail (in comparison to electrons and muons). While the tails become less substantial at high p_T, the Gaussian-like bulk of the response distributions broadens at high p_T. The section on systematics later shows how the uncertainty on τ_h energy scale and resolution impacts the predictions for background and signal in our main search regions.

E_T^{miss} Reconstruction

Should there exist a large transverse momentum imbalance at CMS, this could be evidence for new physics like SUSY due to the presence of weakly interacting particles like the LSP, which escape from the detector without producing any direct response. The magnitude of the negative vector sum of the transverse momentum of visible objects is known as missing transverse energy. Missing transverse energy is also one of the most important observables when it comes to discriminating the signal events from background events which do not contain neutrinos (which contribute to real E_T^{miss}). As will be shown, when attempting to reconstruct a mass resonance, any sign of new physics will usually reside in the tails of the mass distribution. Since neutrinos are among the decay products of tau leptons, a good understanding of the tails of the E_T^{miss} distribution is crucial.

This analysis makes use of the E_T^{miss} calculation using reconstructed particle momenta obtained by combining information from all of the CMS subdetectors. This is known as PFMet [3]. When the proton beams approach each other head-on, the initial momentum for the collision system is $\vec{p} = 0$. Thus, the $\vec{E}_T^{miss}$ in the system may be calculated from the following:

$$\begin{aligned} \vec{p} &= 0 \\ &= \vec{E}_T^{miss} + \sum_{i=1}^{N} \vec{p}_{T,i} \\ \vec{E}_T^{miss} &= -\sum_{i=1}^{N} \vec{p}_{T,i}, \end{aligned} \tag{VIII.4}$$

where i, which runs from 1 to N, numbers the visible particle present in the event. Stated otherwise, for our choice of E_T^{miss} reconstruction, this summation would cover all PF candidates in the event. Figure 17 shows the E_T^{miss} resolution for PF-based reconstruction, calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections obtained using a ditau sample of Z+jets MC. Particle flow based E_T^{miss} reconstruction provides significantly better resolution than calorimeter based reconstruction. The resolution for PF-based reconstruction is approximately 5 GeV.

The E_T^{miss} is particularly sensitive to detector effects. These effects might include pileup (PU), underlying-events (UE), or cracks, for instance. This presents a challenge in isolating and understanding systematic effects from subdetectors. Sources of data/MC discrepancy are due to anomalous behavior from hybrid photodiodes (HPDs), problems with readout boxes (RBX), and large energy readouts from electromagnetic crystals (ECAL "spikes"). Large energy deposits occur in the HPDs when the HPD is misaligned with the magnetic field. ECAL spikes are characterized by very large energy readouts in a single crystal; however, these are removed from the E_T^{miss} calculation using a selection on a quantity that compares the energy in an individual cell to all immediately surrounding ECAL cells. RBX noise is a consequence of anomalous behavior in all four HPDs within a RBX. The noise rate is diminished carefully using noise rejection and cleaning.

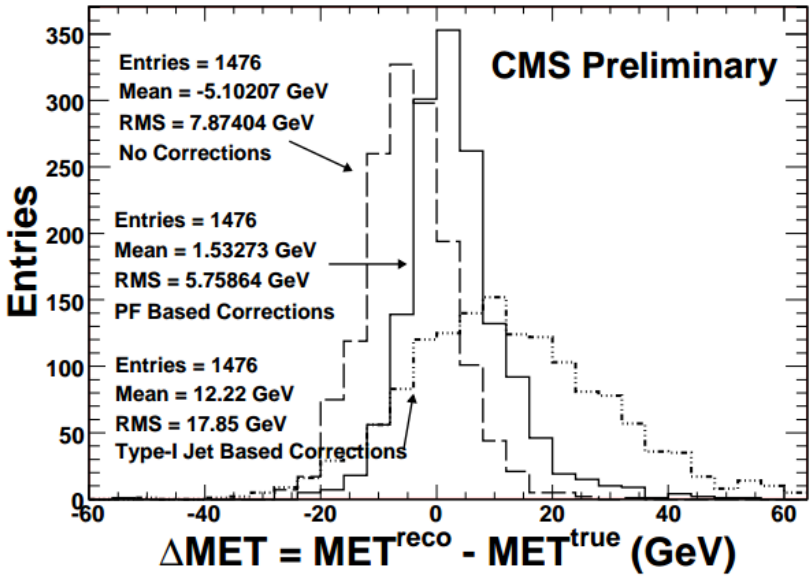

Figure 17: E_T^{miss} resolution for PF-based reconstruction, calorimeter only based reconstruction, and calorimeter based reconstruction with jet based energy scale corrections.

At large values of E_T^{miss}, as are applicable to the SR of this analysis, the noise rate does not present a problem. Even so, the noise rejection variables are included to maintain consistency.

The E_T^{miss} is corrected to account for pileup. The corrections are obtained from a sample of γ+jets events. There is no real E_T^{miss} in this event type. Contributions to E_T^{miss} will be from mismeasurement of the jet energy/momentum, mismeasurement of the photon energy/momentum, and pileup. The projection of the E_T^{miss} onto the direction perpendicular to the γ+jet pair quantifies the E_T^{miss} resolution due to pileup. To accurately apply pileup corrections to the E_T^{miss} calculation, the probability for an event to contain a certain number of primary vertices is extracted from each control region. Once pileup corrections are applied, data and MC distributions exhibit agreement.

Since there are many factors which may result in artificial (fake) E_T^{miss}, comprehensive studies of E_T^{miss} are performed [3]. For the SR of this analysis, we select events with $E_T^{miss} > 230$ GeV (optimized). We make use of FastSim E_T^{miss} in this analysis [8]. The standard and recommended "MET Filters" are utilized in this search `https://twiki.cern.ch/twiki/bin/viewauth/CMS/MissingETOptionalFiltersRun2`.

Chapter IX

Data and Monte Carlo Samples

The 13 TeV data collected by the CMS detector in years 2016 and 2017 is used in this analysis. The MC samples for 2016 are given in Table 15. The data samples for the 2016 iteration of the analysis are given in Table 16. We provide the same information for the 2017 MC samples in Table 17 and for 2017 data in Table 18. The total integrated luminosity of the collision data samples is 35.87 fb^{-1}.

Table 15: MC Samples (2016)

Process	cross-section [pb]	Official CMS Datasets (MINIAODSIM)
$Z \to ll$ HT binned LO	175.3×1.17781 (NNLO)	/DYJetsToLL_M-50_HT-70to100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	147.4 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	147.4 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	40.99 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	40.99 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	5.678 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	5.678 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	1.363 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-600to800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v2/MINIAODSIM
	0.6759 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-800to1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	0.116 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-1200to2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	0.002592 × 1.17781 (NNLO)	/DYJetsToLL_M-50_HT-2500toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
W+jets HT binned LO	1319×1.21378 (NNLO)	/WJetsToLNu_HT-70To100_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	1345×1.21378 (NNLO)	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	1345×1.21378 (NNLO)	/WJetsToLNu_HT-100To200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	359.7×1.21378 (NNLO)	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	359.7×1.21378 (NNLO)	/WJetsToLNu_HT-200To400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	48.91×1.21378 (NNLO)	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	48.91×1.21378 (NNLO)	/WJetsToLNu_HT-400To600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	12.05×1.21378 (NNLO)	/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	12.05×1.21378 (NNLO)	/WJetsToLNu_HT-600To800_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	5.501×1.21378 (NNLO)	/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	5.501×1.21378 (NNLO)	/WJetsToLNu_HT-800To1200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	1.329×1.21378 (NNLO)	/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	1.329×1.21378 (NNLO)	/WJetsToLNu_HT-1200To2500_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
	0.03216×1.21378 (NNLO)	/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	0.03216×1.21378 (NNLO)	/WJetsToLNu_HT-2500ToInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM
Top quark	831.76 (NNLO)	/TT_TuneCUETP8M2T4_13TeV-powheg-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
	35.6	/ST_tW_antitop_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1
	35.6	/ST_tW_top_5f_inclusiveDecays_13TeV-powheg-pythia8_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1
	136.02	/ST_t-channel_top_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1
	80.95	/ST_t-channel_antitop_4f_inclusiveDecays_13TeV-powhegV2-madspin-pythia8_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1
WW	63.21	/WW_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
WZ	22.82	/WZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM
ZZ	10.32	/ZZ_TuneCUETP8M1_13TeV-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM

Table 16: Data Samples (2016)

Run Name	File Location
Run 2016B Met Run2016B-03Feb2017	/Met/Run2016B-03Feb2017-v3/MINIAOD
Run 2016C Met Run2016C-03Feb2017	/Met/Run2016C-03Feb2017-v1/MINIAOD
Run 2016D Met Run2016D-03Feb2017	/Met/Run2016D-03Feb2017-v1/MINIAOD
Run 2016E Met Run2016E-03Feb2017	/Met/Run2016E-03Feb2017-v1/MINIAOD
Run 2016F Met Run2016F-03Feb2017	/Met/Run2016F-03Feb2017-v1/MINIAOD
Run 2016G Met Run2016G-03Feb2017	/Met/Run2016G-03Feb2017-v1/MINIAOD
Run 2016Hv2 Met Run2016H-03Feb2017	/Met/Run2016H-03Feb2017 ver2-v1/MINIAOD
Run 2016Hv3 Met Run2016H-03Feb2017	/Met/Run2016H-03Feb2017 ver3-v1/MINIAOD

Table 17: MC Samples (2017)

Process	cross-section [pb]	Official CMS Datasets (NANOAODSIM)
$Z \to ll$ HT binned LO	224.20	/DYJetsToLL_M-4to50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	37.20	/DYJetsToLL_M-4to50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	3.5810	/DYJetsToLL_M-4to50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	1.1240	/DYJetsToLL_M-4to50_HT-600toInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	213.4	/DYJetsToLL_M-50_HT-100to200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	65.42	/DYJetsToLL_M-50_HT-200to400_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	7.31	/DYJetsToLL_M-50_HT-400to600_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	1.49	/DYJetsToLL_M-50_HT-600to800_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	0.661	/DYJetsToLL_M-50_HT-800to1200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	0.119	/DYJetsToLL_M-50_HT-1200to2500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	0.00280	/DYJetsToLL_M-50_HT-2500toInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
W-jets HT binned LO	1695.0	/WJetsToLNu_HT-100To200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	532.4	/WJetsToLNu_HT-200To400_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	61.6	/WJetsToLNu_HT-400To600_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	12.4	/WJetsToLNu_HT-600To800_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	5.77	/WJetsToLNu_HT-800To1200_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	1.02	/WJetsToLNu_HT-1200To2500_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	0.0248	/WJetsToLNu_HT-2500ToInf_TuneCP5_13TeV-madgraphMLM-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
Top quark	88.29	/TTTo2L2Nu_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	377.96	/TTToHadronic_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
	365.34	/TTToSemiLeptonic_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	136.02	/ST_t-channel_top_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	80.95	/ST_t-channel_antitop_4f_InclusiveDecays_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	38.06	/ST_tW_top_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	38.06	/ST_tW_antitop_5f_inclusiveDecays_TuneCP5_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	3.68	/ST_s-channel_4f_leptonDecays_TuneCP5_PSweights_13TeV-amcatnlo-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
WW	10.480	/WWTo2L2Nu_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6_ext1-v1/NANOAODSIM
	51.723	/WWTo4Q_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6_ext1-v1/NANOAODSIM
	43.530	/WWToLNuQQ_NNPDF31_TuneCP5_PSweights_13TeV-powheg-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6_ext1-v1/NANOAODSIM
WZ	10.730	/WZTo1L1Nu2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	3.0540	/WZTo1L3Nu_13TeV_amcatnloFXFX_madspin_pythia8_v2/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	5.6060	/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	4.4300	/WZTo3LNu_TuneCP5_13TeV-amcatnloFXFX-pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_new_pmx_102X_mc2017_realistic_v6-v1/NANOAODSIM
ZZ	0.5644	/ZZTo2L2Nu_13TeV_powheg_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	3.2220	/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	4.0330	/ZZTo2Q2Nu_TuneCP5_13TeV_amcatnloFXFX_madspin_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM
	1.2040	/ZZTo4L_13TeV_powheg_pythia8/RunIIFall17NanoAODv4-PU2017_12Apr2018_Nano14Dec2018_102X_mc2017_realistic_v6-v1/NANOAODSIM

Table 18: Data Samples (2017)

Run Name	File Location
Run 2017B Met Run2017B-14Dec2018	/MET/Run2017B-Nano14Dec2018-v1/NANOAOD
Run 2017C Met Run2017C-14Dec2018	/MET/Run2017C-Nano14Dec2018-v1/NANOAOD
Run 2017D Met Run2017D-14Dec2018	/MET/Run2017D-Nano14Dec2018-v1/NANOAOD
Run 2017E Met Run2017E-14Dec2018	/MET/Run2017E-Nano14Dec2018-v1/NANOAOD
Run 2017F Met Run2017F-14Dec2018	/MET/Run2017F-Nano14Dec2018-v1/NANOAOD

The leading order generators, PYTHIA8 and MADGRAPH, were mainly used for signal and background MC production. Similar to the MC background samples, the signal samples were produced using full detector simulation. The predicted background yields in simulation are determined using next-to-leading-order (NLO) or next-to-next-to-leading-order (NNLO) cross sections, while the signal yields and distributions in all plots shown for this analysis were normalized using the leading-order (LO) cross sections. These were confirmed with the Resummino program. The reason for using LO cross sections for the signal yields is because there is currently no functionality in MADGRAPH to calculate the NLO cross sections. This is true across SUSY analyses.

The LO $t\bar{t}$, W+jets, and Z+jets MADGRAPH samples were used in this analysis, and the predicted yields in simulation, as mentioned, were determined using NLO or NNLO cross sections. The MC simulated samples have a pileup distribution that does not match that of data. Therefore, the MC is reweighted to fit the pileup distribution observed in data. The reweighting of MC events is performed by determining the probabilities to obtain n interactions in data ($P_{\text{data}}(n)$) and MC ($P_{MC}(n)$) and using the event weights:

$$w_{PU}(n) = \frac{P_{\text{data}}(n)}{P_{\text{MC}}(n)} \tag{IX.1}$$

to reweigh MC events based on the number of interactions. The recommended minibias cross section of 69.2 mb is used to determine these weights.

The following corrections have been applied to the MC predictions: (i) trigger efficiency weights to match the efficiency in data (this includes not only the trigger efficiency weights for the E_T^{miss} trigger used to define the SR but also the single-muon trigger efficiency scale factors when we look at the muon control samples); (ii) b-tagging SFs as recommended by the POG; (iii) τ_h ID efficiency scale factor as recommended by the τ lepton POG; (iv) pileup weights to match the number of interactions per bunch crossing in data (using the recommended cross section); (v) ISR jet/boost corrections (see subsequent sections for full explanation) to correct the modeling of the Z/W boost (and thus the correlated ISR kinematics); (vi) top-quark p_T weights to correct the mismodeled top-quark boost in the $t\bar{t}$ MC; and (vii) muon identification efficiency scale factors.

Chapter X

Background Estimation: Boost and Recoil Studies

The Importance of $Z(\rightarrow \mu\mu)$+ISR

Several important factors are to be considered when determining the expected background and signal yields for the desired topology in this analysis. The SR includes a single hadronic decay of a τ lepton, large E_T^{miss}, and a high p_T jet from ISR. The expected Z+jets yield in the SR, which is primarily $Z(\rightarrow \tau\tau \rightarrow \tau_h\tau_h)$ with one lost τ_h combined with an ISR jet, can be parameterized as follows:

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

where σ_Z is the best known cross section of the process, L_{int} is the integrated luminosity, ϵ_{τ_h} is the selection efficiency on τ_h, (1-ϵ_{τ_h}) is the efficiency associated with the lost second τ lepton leg, $\epsilon_{E_T^{miss}}$ is the efficiency of the E_T^{miss} selection, ϵ_{ISR} is the efficiency of the ISR jet requirement, and ϵ_{other} is the product of any remaining efficiencies to be considered (e.g. b-jet veto cut efficiency, efficiency of the $\Delta\phi$ QCD rejection selection, and efficiencies of vetoes of additional leptons).

A key aspect of this analysis is the understanding of the modeling of ISR in MC. In order to focus-in on the ISR efficiency, a sample enriched with $Z(\rightarrow \mu\mu)$+ISR jet events is studied. There are several advantages to studying a sample of this type initially. First, due to lepton universality, a control sample where μs are produced in decays of the Z can be used to closely emulate the ISR jet activity in $Z(\rightarrow \tau_h\tau_h)$ events. This is highly advantageous since the identification of μs is far cleaner than that of τ_hs. A second important advantage to note is that this event includes no real E_T^{miss}. The number of events in the $Z(\rightarrow \mu\mu)$ control sample is to be parameterized as follows:

$$N^{CR}_{Z\rightarrow\mu\mu} = \sigma_Z \cdot L_{\text{int}} \cdot \epsilon_\mu^2 \cdot \epsilon_{\text{ISR}} \cdot \epsilon_{\text{other}}, \tag{X.2}$$

where σ_Z is again the cross section of this particular process, L_{int} is the integrated luminosity, ϵ_μ is the muon selection efficiency, ϵ_{ISR} is the efficiency of the ISR jet criteria, and ϵ_{other} is the product of any remaining efficiencies to be considered.

It is evident from these two parameterizations that the ISR efficiency is more easily obtained through examination of the $Z(\rightarrow \mu\mu)$ CR as opposed to the $Z(\rightarrow \tau_h\tau_h)$ since the E_T^{miss} and the τ_h selection efficiencies are factored out for $Z(\rightarrow \mu\mu)$ and because of the improved ability to detect μs as opposed to τ_hs.

The event selection criteria for the for the $Z(\rightarrow \mu\mu)$+ISR CR, which closely pertain to the SR,

are given in Table 19. These event selection criteria pertain to the $Z(\rightarrow \mu\mu)$+ISR CR in both the 2016 and 2017 iterations of the analysis.

Table 19: $Z(\rightarrow \mu\mu)$+ISR Event Selections

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
$p_T(j_{\text{lead}})$	≥ 100 GeV
Trigger	HLT_IsoMu24_ (HLT_IsoMu27 OR HLT_IsoMu29) [2016] ([2017])

The $\mu\mu$ pair has an invariant mass compatible with the Z-mass hypothesis ($80 < m(\mu, \mu) < 100$ GeV). A selection is made on jet pseudorapidity ($\eta(j)$) in order to avoid potential jet resolution issues in the forward regions of the detector. To guard against jets faking leptons, jets must be separated from identified μs by $\Delta R > 0.3$, and "Loose" identification is applied to jets for the 2016 analysis. For 2017, we use the recommended "Tight" jet identification. In addition, the requirement of at least one highly energetic ISR jet, specifically with transverse momentum $p_T(j_{\text{lead}}) > 100$ GeV, boosts the $Z(\rightarrow \mu\mu)$ system in a similar fashion to the boosted Z in the SR.

Included in Figure 18 are plots for this CR with selections as listed in Table 19 and 2016 data and MC. To obtain these distributions, events firing the HLT_IsoMu24_ trigger are selected, and single-muon datasets are used. The requirement of a high-p_T ISR jet dictates that the Z boson is not produced at rest; instead, the Z gains some transverse momentum that is subsequently transferred to the dimuon pair. Hence, the boost of the Z boson (Z-Boost) is calculated by vectorially summing the transverse momenta of the two muons. The left plot of Figure 18 is for the Z-Boost distribution in the CR, and the right plot is the H_T distribution, defined as the scalar sum of the p_T of all jets with $p_T > 30$ GeV and $|\eta(j)| < 2.4$.

There exists substantial disagreement between data and the default MC for the boost of the Z boson. The two plots in Figure 18 indicate a potential convolution of effects, namely ISR/boost mismodeling and incorrect jet resolution in MC. There is a potential correlation between the two effects, but the H_T plot shown at the right in Figure 18 indicates that separately examining the ISR modeling and examining the jet resolution have the potential to be fruitful avenues of exploration. We first calculate a necessary set of weights based on the ISR mismodeling (from Figure 18 left) and then examine the resultant effects on the jet resolution (to be described in what follows).

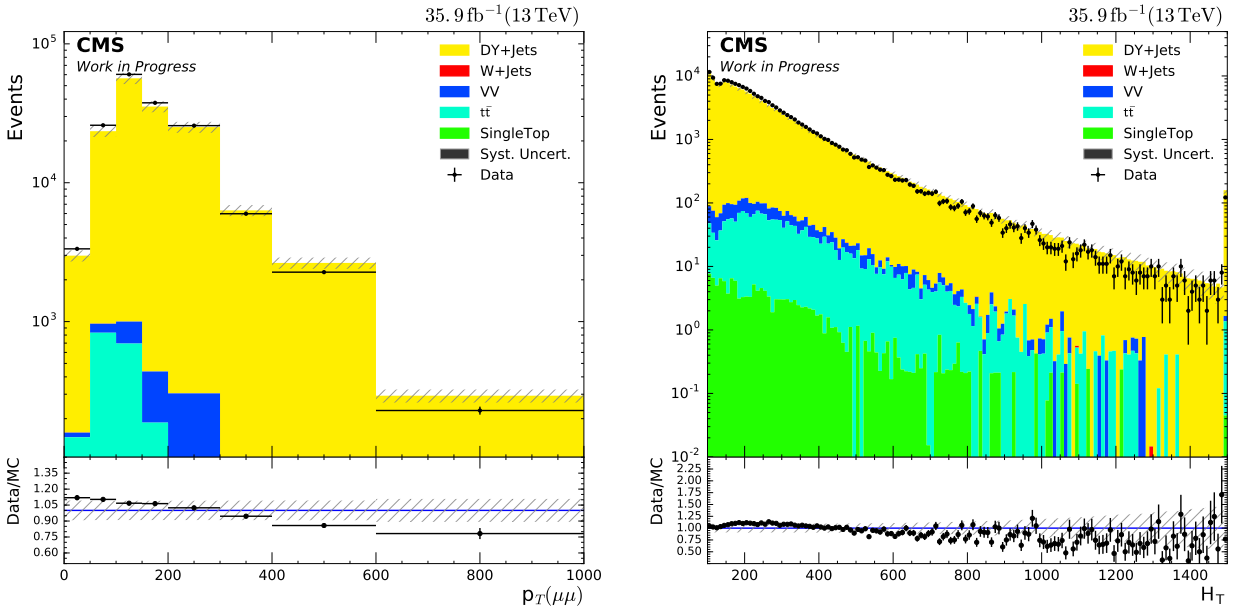

Figure 18: Z-Boost [GeV] (left) and H_T [GeV] (right) distributions for $Z(\rightarrow \mu\mu)$+ISR (2016).

Shifting Focus to Jet Resolution and ISR Modeling for $Z(\rightarrow \mu\mu)$+ISR

To delve deeper into the ISR kinematics and the jet resolution, the jet recoil is studied. The recoil, hereafter defined by $\vec{u}$ is defined as follows:

$$\vec{u} = -(\vec{E}_T^{miss} + \vec{Z}_T), \tag{X.3}$$

where $\vec{E}_T^{miss}$ is the vectorial missing transverse energy for the event, and $\vec{Z}_T$ is the vectorial transverse momentum of the boosted Z boson. Defining the jet recoil in this way takes into consideration the full energy spectrum for the event since the missing momentum is included in the previous equation. Note that $\vec{u}$, as defined, has no component in the z-direction. The component of the recoil parallel to the boost of the Z boson (r_T) is considered by projecting $\vec{u}$ along the direction of $\vec{Z}_T$. This parallel component of the recoil is shown in Figure 19.

Figure 19 is binned in the same fashion as the Z-Boost in Figure 18. Note that r_T is binned along the negative x-axis, however, since the jet recoil has a component parallel but in the opposite direction to the Z-Boost. Evident by this plot is disagreement between data and MC. Since the end goal will be to develop weights/corrections that can be applied to other DY-like processes (e.g. W+jets, signal, etc.), we expand the study in Figure 19 and calculate $\frac{|r_T|}{|Z_T|}$ and plot the mean (i.e. profile) as a function of $|\vec{Z}_T|$. The profile is shown in Figure 20. In this Figure 20, the black trend corresponds to data, and the red corresponds to MC.

At lower values of $|\vec{Z}_T|$ (near the selection of 100 GeV on $p_T(j_{\text{lead}})$), the parallel component of

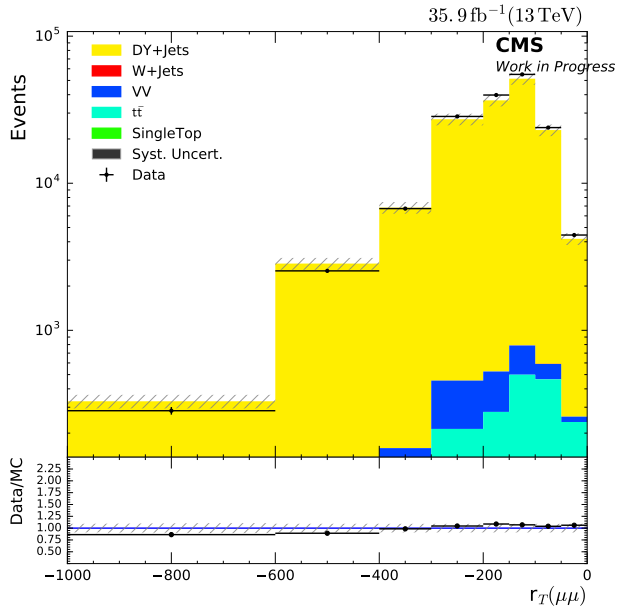

Figure 19: Jet recoil parallel to Z-Boost [GeV] (2016).

the recoil is not modeled to balance the boost of the Z boson. This is clear in the fraction's deviation from 1, which is depicted with the horizontal blue line in Figure 20. At low boost of the Z boson, there must be additional hadronic activity in the direction of the Z to conserve momentum with the ISR jet. This is not modeled well, even at NLO, with the MC. Across the spectrum of values for $|\vec{Z}_T|$, there is disagreement in the profile between data and MC for $\frac{|r_T|}{|\vec{Z}_T|}$. In order to address the mismodeling of the boson boost overall in MC, the boost of the Z is corrected in MC, and the resultant effects on the recoil momentum are examined. By first correcting for the mismodeling of

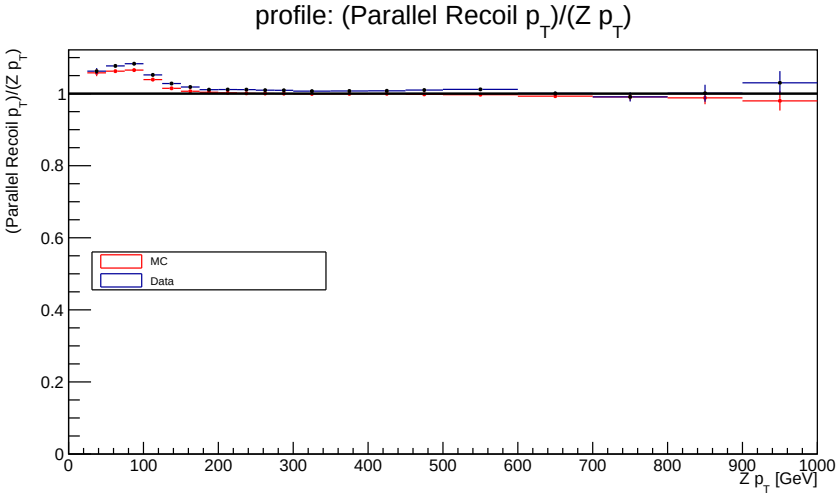

Figure 20: Profile of $\frac{|r_T|}{|\vec{Z}_T|}$ (2016).

the Z-Boost (i.e. the ISR kinematics), we can then study the effects on jet resolution without bias. In order to determine the corrections for $|\vec{Z}_T|$, the applicable weight for each $|\vec{Z}_T|$ bin is extracted from the Data/MC ratio plot in Figure 21 (left). These weights are subsequently applied to each event satisfying the selection criteria. The corrected $|\vec{Z}_T|$ distribution shown in Figure 21 (right) has perfect agreement between data and MC by construction.

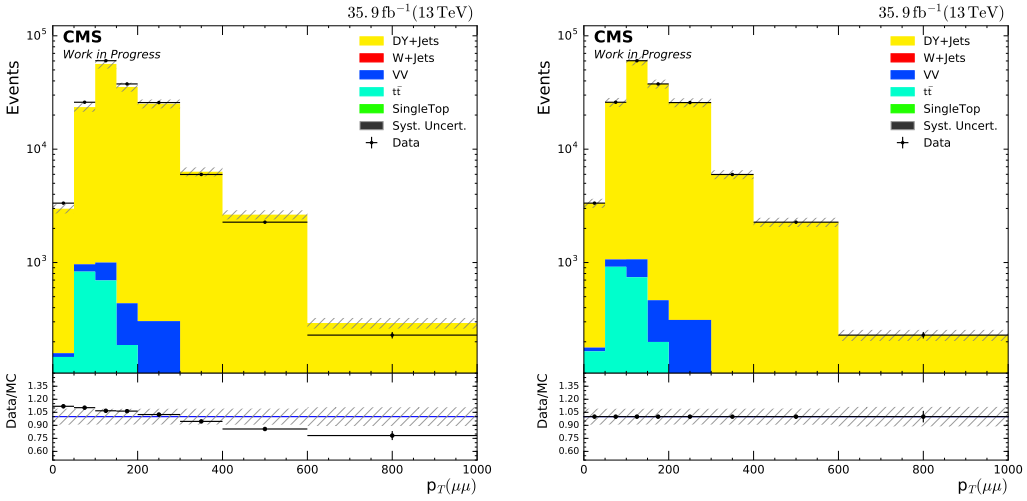

Figure 21: Z-Boost [GeV] distribution without boost weights (left) and with weights (right) for events in the $Z(\rightarrow \mu\mu)$ + ISR jet control region (2016).

Table 20 lists explicitly the weights that apply to each $|\vec{Z}_T|$ range. Associated uncertainties are statistical. Near the selection for $p_T(j_{\text{lead}})$ (100 GeV), the correction falls between six and eleven percent for 2016, which is significant and also consistent with weights otherwise derived by the SUSY group. For 2017, we calculate the ISR weights from Figure 22, and these weights are listed explicitly in Table 20 as they apply to each $|\vec{Z}_{\text{T}}|$ range. Near the selection for $p_T(j_{\text{lead}})$ (100 GeV), the correction falls between seven and twelve percent for 2017, which is significant and also again consistent to the weights otherwise derived by the SUSY group. Within statistical uncertainties, the Z-Boost weights derived for the 2016 and 2017 iterations of the analysis are in agreement.

Following the application of the Z-Boost weights, the effect on the parallel component of the recoil is examined. Figure 23 (left) is the parallel recoil without boost weights applied, and the right plot in Figure 23 is the parallel recoil with weights applied for 2016 data and MC. There is marked improvement in the agreement between data and MC with the application of the weights. In addition, the Data/MC ratio plot in Figure 23 (right) shows agreement within accepted ranges for jet resolution at CMS (∼5%). Thus, we conclude that no further corrections for jet energy resolution are necessary.

We draw a similar conclusion utilizing 2017 data and MC for the parallel component of the recoil. Figure 24 is the parallel recoil with weights applied using 2017 data and MC. There is again

Table 20: Event Weights by Z-Boost (2016 & 2017)

Z-Boost Bin	Weight (2016)	Weight (2017)
1: 0-50 GeV	1.12 ± 0.03	1.09 ± 0.02
2: 50-100 GeV	1.10 ± 0.01	1.11 ± 0.01
3: 100-150 GeV	1.07 ± 0.01	1.08 ± 0.01
4: 150-200 GeV	1.06 ± 0.01	1.07 ± 0.01
5: 200-300 GeV	1.02 ± 0.01	1.02 ± 0.01
6: 300-400 GeV	0.95 ± 0.02	0.94 ± 0.01
7: 400-600 GeV	0.86 ± 0.03	0.84 ± 0.02
8: 600+ GeV	0.78 ± 0.11	0.79 ± 0.05

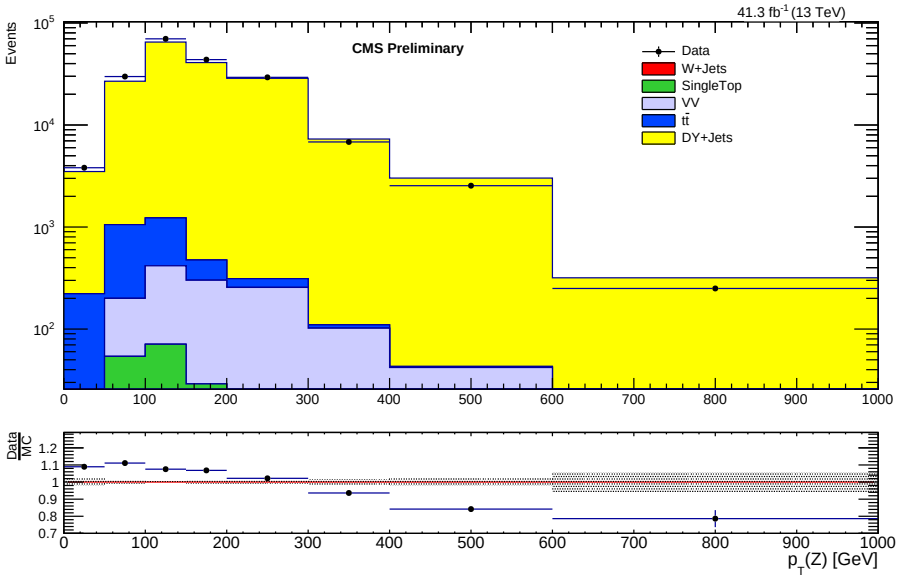

Figure 22: Z-Boost [GeV] distribution without boost weights for events in the $Z(\rightarrow \mu\mu)$ + ISR jet control region (2017).

quality agreement between data and MC with the application of the weights. In addition, the Data/MC ratio plot in Figure 24 again shows agreement within accepted ranges for jet resolution at CMS (∼5%). Thus, we conclude once more that no further corrections for jet energy resolution are necessary.

In the following subsection, the agreement between data and MC distributions of other relevant variables is examined.

Conclusions from Boost and Recoil Studies on $Z(\rightarrow \mu\mu)$+ISR

Figure 25 includes $Z(\rightarrow \mu\mu)$+ISR CR plots using 2016 data and MC for μ momentum (i.e. two entries per event), momentum of the first leading jet, and H_T before and after application of the

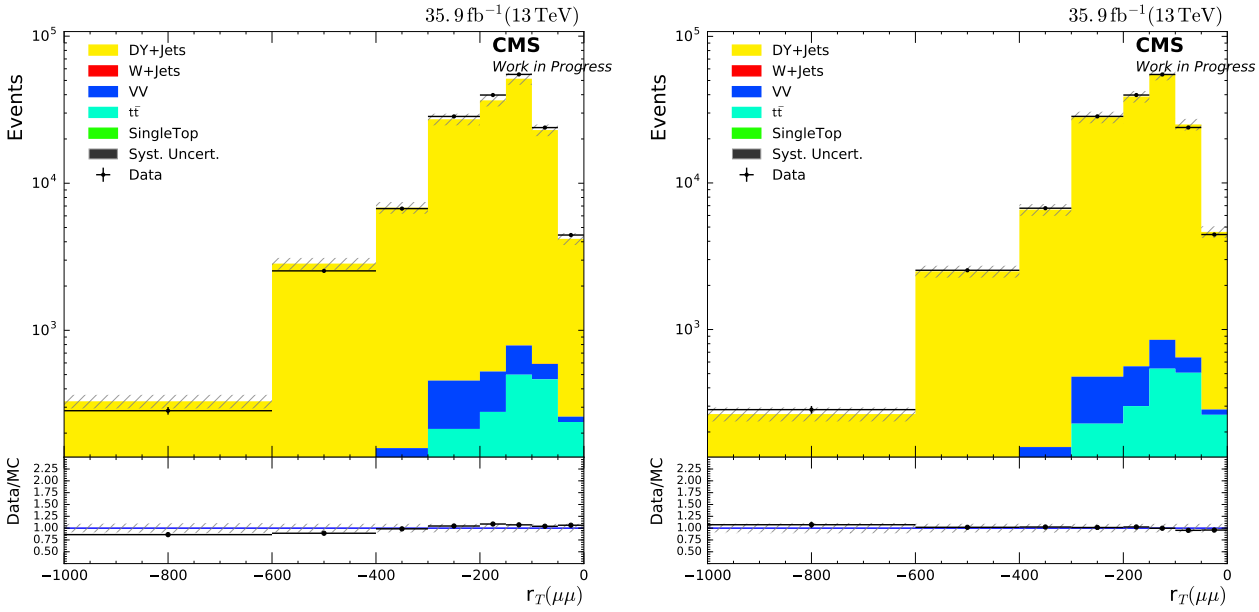

Figure 23: Parallel Recoil r_T [GeV] without boost weights (left) and with weights (right) (2016).

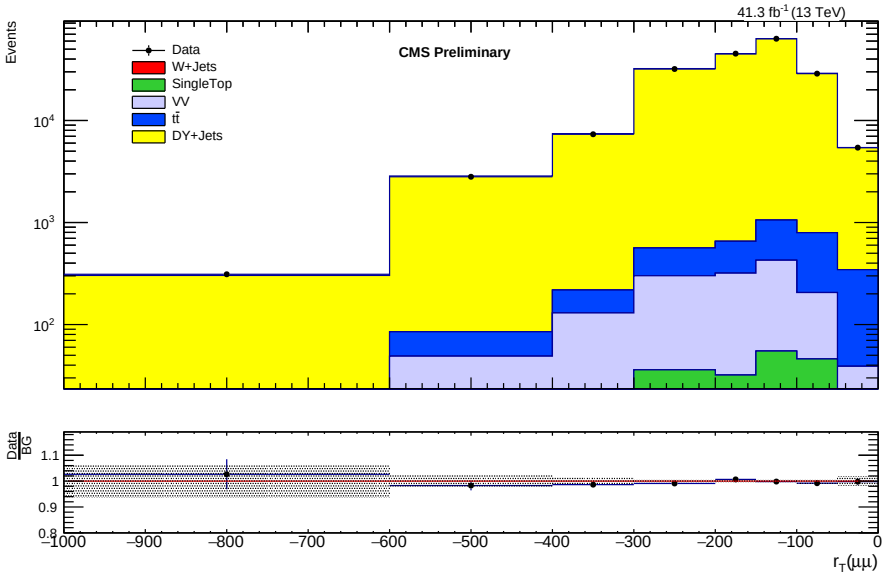

Figure 24: Parallel Recoil r_T [GeV] with weights (2017).

ISR weights. For the $p_T(\mu)$, there is marked improvement in agreement between data and MC for both low- and high-momentum μs after application of the ISR weights. The same ideas hold for the momentum of the first leading jet and the H_T distributions.

These plots in Figure 25 are a few examples to exemplify the benefit of applying the ISR weights to correct ISR mismodeling. In general, the data/MC comparison improves by $\sim 10 - 20\%$ across the analysis depending on the range and distribution.

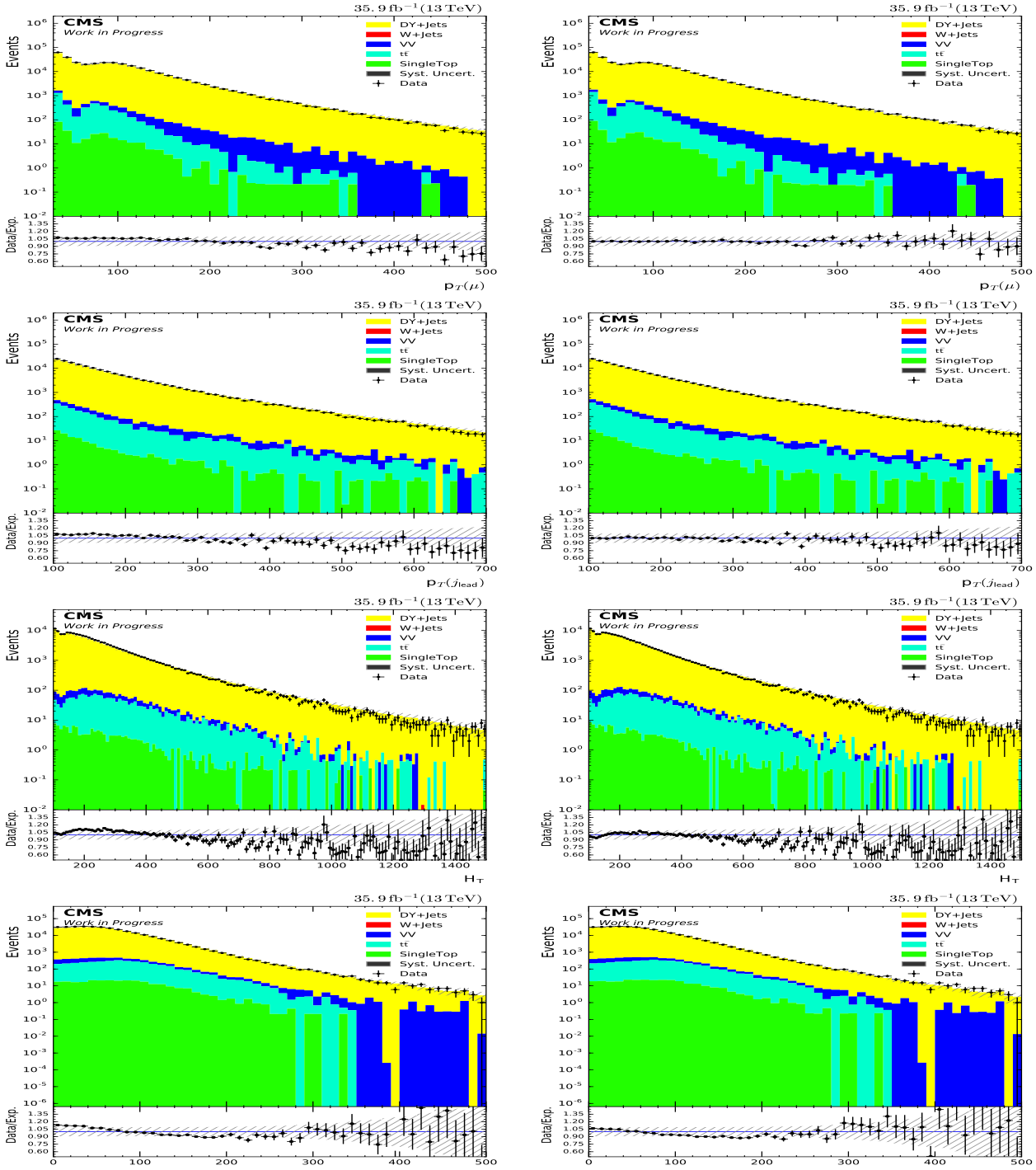

Figure 25: Top Row: Muon p_T without boost weights (left) and with weights (right); Second Row: First leading jet p_T without boost weights (left) and with weights (right); Third Row: H_T without boost weights (left) and with weights (right); Bottom Row: $m_T(\tau_h, E_T^{miss})$ without boost weights (left) and with weights (right) (2016).

Validating ISR Weights on $W(\rightarrow \mu\nu)$+ISR

Since the ISR weights are derived from a Z+jets control sample, it is important to validate those weights in an orthogonal control sample. In this section, we describe how those weights are validated on a W+jets process: $W(\rightarrow \mu\nu)$+ISR. The event selection criteria defining this W control region are given in Table 21.

Table 21: $W(\rightarrow \mu\nu)$+ISR Event Selections

Selection	Cut
$N(\mu)$	1
$p_T(\mu)$	> 30 GeV
$\lvert\eta(\mu)\rvert$	< 2.1
E_T^{miss}	> 230 GeV
$m_T(\mu, E_T^{miss})$	[60, 100] GeV
$\lvert\eta(j)\rvert$	< 2.4
$N(j)$	≥ 1
$\mathrm{p}_T^{\mathrm{lead}}(j)$	≥ 100 GeV
QCD rejection	$\lvert\Delta\phi(j_{\mathrm{lead}}, E_T^{miss})\rvert > 0.7$
Trigger	HLT_IsoMu24_ (HLT_IsoMu27 OR HLT_IsoMu29) [2016] ([2017])

These event selection criteria pertain to this region in both the 2016 and 2017 iterations of the analysis. To contrast with the Z process, now exactly one μ is required. Unlike the $Z(\rightarrow \mu\mu)$+ISR CR where there is no real E_T^{miss}, this $W(\rightarrow \mu\nu)$+ISR CR contains real E_T^{miss} from the decay of the W boson to a neutrino. Therefore, a selection on $E_T^{miss} > 230$ GeV (similar to the SR) is applied. A selection is made on the transverse mass of the μ and E_T^{miss} ($60 < m_T(\mu, E_T^{miss}) < 100$ GeV) in order to pick out events near the Jacobian $m(W)$ peak. A selection is made on jet η in order to avoid potential jet resolution issues in the forward regions of the detector. To guard against jets faking leptons, jet candidates must not overlap the identified μ ($\Delta R(\mu, j) > 0.3$), and "Loose" ("Tight") jet identification is applied for 2016 (2017). In addition, we impose the requirement of at least one highly energetic ISR jet, specifically with transverse momentum greater than 100 GeV and separated from the direction of E_T^{miss} ($\Delta\phi(j_{\mathrm{lead}}, E_T^{miss}) > 0.7$), to produce a boosted event topology. Generally speaking, the W+ISR jet CR is defined with a similar criteria to the SR, except with events firing the single-muon trigger, a requirement of one well-defined μ (instead of a soft τ_h), and an m_T selection to choose on-mass-shell W bosons.

The requirement of the ISR jet dictates that the W boson is not produced at rest in $W(\rightarrow \mu\nu)$+ISR. Instead, the W gains some transverse momentum that it subsequently transfers to the μ and the neutrino. The ISR weights are applied based on the momentum of the W boson, which is accessed at the generator level particle collection. Included in Figure 26 are the $p_T(\mu)$ and the E_T^{miss} both without and with ISR weights applied for 2016 data and MC in this region. These two

kinematic quantities in particular illustrate the validation of the ISR weights extracted from the Z CR. In particular, Figure 26 shows an improvement of ∼7% in the global data-to-MC scale factor, in addition to a significant improvement in the modeling of the relevant shapes (i.e. data-to-MC ratio is flat). Note that the plots of Figure 26 are constructed with a variable binning procedure such that the relative MC statistical uncertainty in a given bin is less than 15%. While the variable binning procedure is used to obtain smooth distributions, the chosen binning is also finer than the one used in the SR.

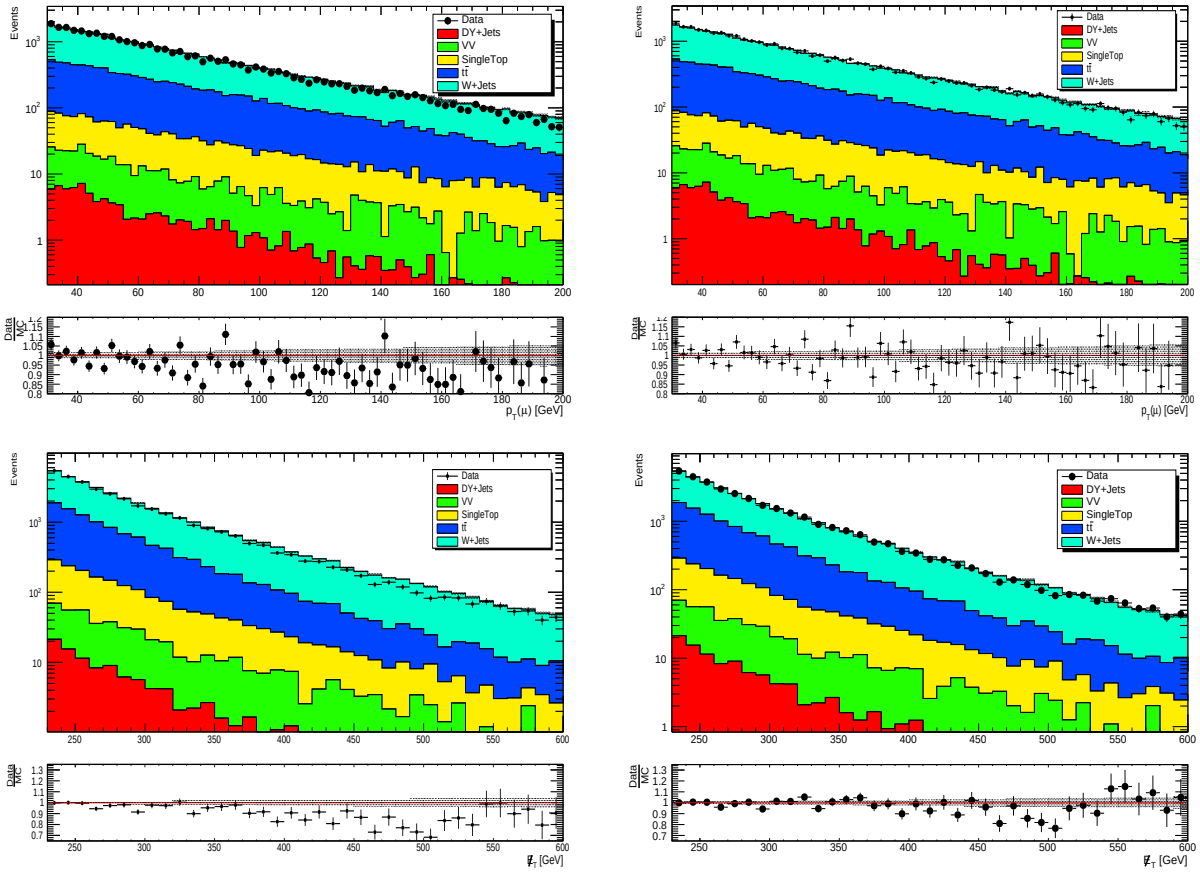

Figure 26: Top Row: Muon p_T without boost weights (left) and with weights (right); Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) (2016).

Included in Figure 27 is the E_T^{miss} distribution with ISR weights applied for 2017 data and MC. This figure illustrates the validation of the ISR weights extracted from the Z control region by the improvement in agreement between data and MC. A similar set of plots using 2016 data and MC are shown in Figure 28 with instead a lower E_T^{miss} selection of 50 GeV and a lepton veto on electrons and τ_hs. The motivation for lowering the E_T^{miss} threshold is to simultaneously validate the modeling of the E_T^{miss} with the other selections required. The uncertainty band in the ratio plots of Figure 28 (right) includes the statistical uncertainty and also the effect due to the ISR weight uncertainty. These plots again show good agreement between data and MC after the application

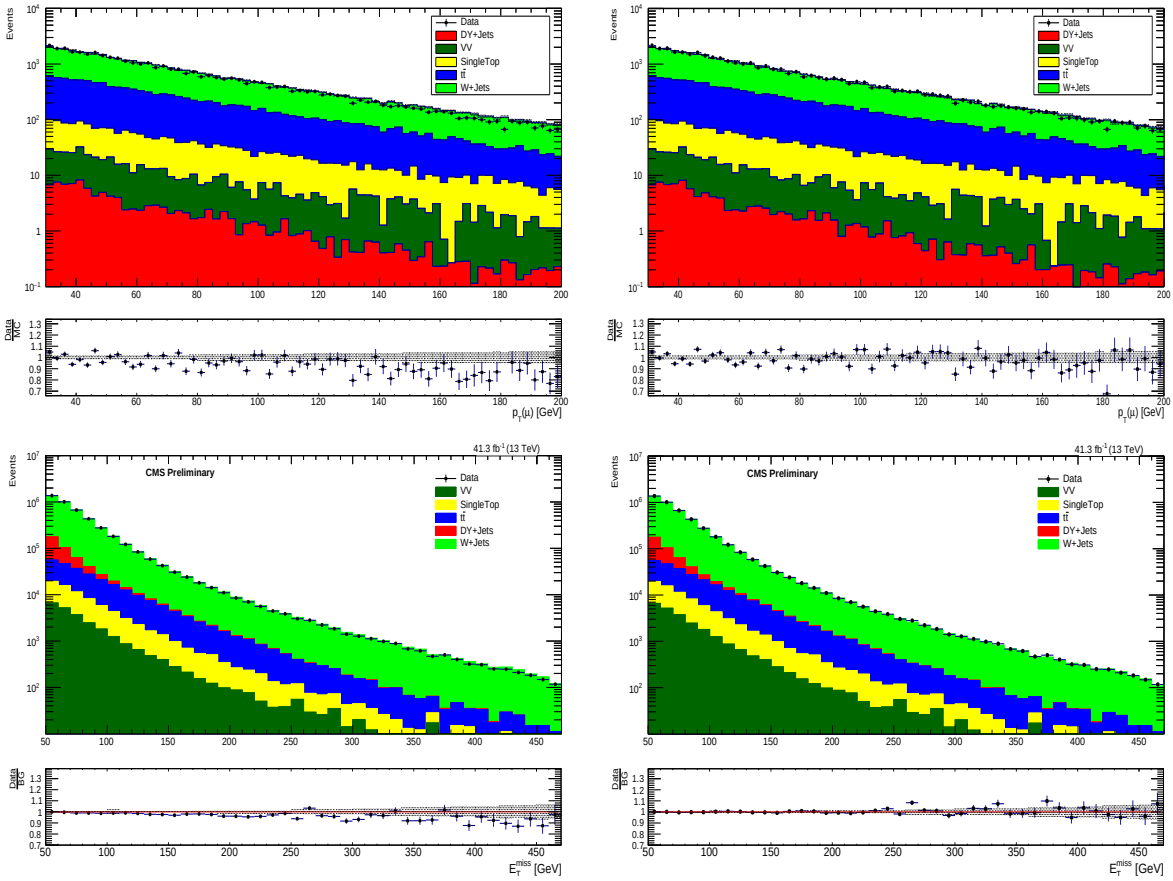

Figure 27: Top Row: Muon p_T without boost weights (left) and with weights (right) using 2017 data ($E_T^{miss} > 230$ GeV cut). Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) using 2017 data ($E_T^{miss} > 50$ GeV).

of the ISR weights to the W+jets MC, especially in the tails of the distributions.

The combination of the plots in Figure 26, Figure 27, and Figure 28 gives us a high degree of confidence that (1) the modeling of the ISR jet activity in MC is understood, (2) the ISR weights are correct, and (3) the E_T^{miss} in MC is well modeled after appropriate corrections are applied. Table 22 lists the predicted background yields in simulation (after corrections discussed) as well as the observed yield in data for 2016. These yields are derived from the region resulting in the plots of Figure 27 with the higher selection on the E_T^{miss} threshold. Since the scale factor is consistent with unity, the MC prediction for the W+jets yield in the SR is taken directly from simulation for this analysis. Systematic uncertainty on the m_T shape is described in the following sections. We also note that the signal contamination is negligible for all signal samples used in this analysis (with chargino masses ranging from 100 to 500 GeV).

Figure 29 shows the results for $m_T(\mu, E_T^{miss})$ in this CR using 2017 data and MC, where quality agreement is observed between data and MC for this study.

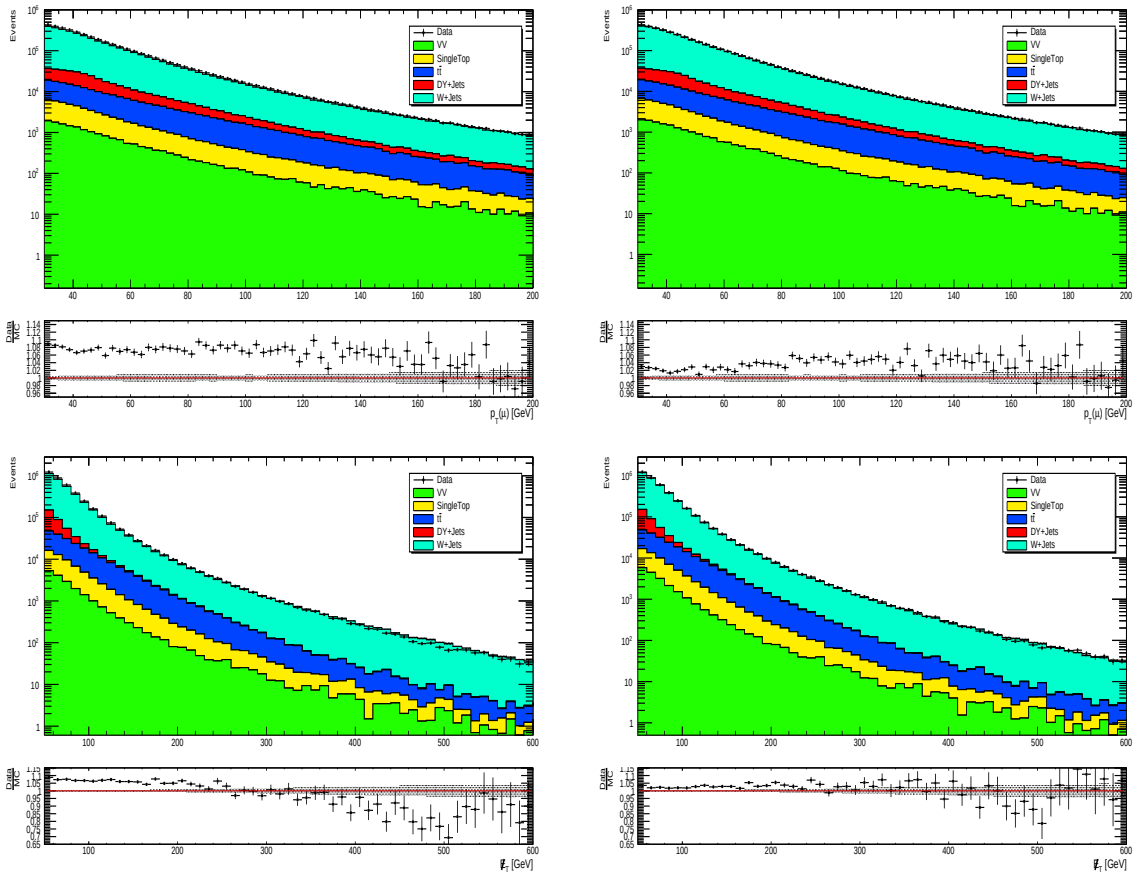

Figure 28: Top Row: Muon p_T without boost weights (left) and with weights (right); Bottom Row: E_T^{miss} without boost weights (left) and with weights (right) (2016).

Process	Yield
Data	34344
W + jets	23702.10 ± 84.32
DY + jets	107.11 ± 3.19
Diboson	383.78 ± 11.53
$t\bar{t}$	9062.12 ± 59.48
Single Top	1452.41 ± 15.39
SF_{BG}	0.98 ± 0.01

Table 22: Background and data yields in the W+jets control regions (2016)

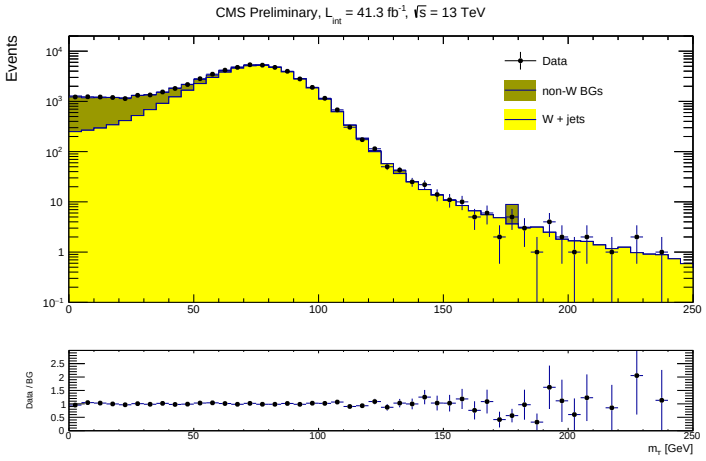

Figure 29: Results for $W(\rightarrow \mu\nu)$+ISR control region using 2017 data and MC.

Chapter XI

Background Estimation: τ_h Identification Studies

The Importance of $Z(\rightarrow \tau\tau)$+ISR

As discussed previously, the expected Z+jets yield in the SR, which is primarily $Z(\rightarrow \tau\tau \rightarrow \tau_h\tau_h)$ with one lost τ_h combined with an ISR jet, can be parameterized as follows:

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

where σ_Z is the best known cross section of the process, L_{int} is the integrated luminosity, ϵ_{τ_h} is the selection efficiency on τ_h, (1-ϵ_{τ_h}) is the efficiency associated with the lost second τ lepton leg, $\epsilon_{E_T^{miss}}$ is the efficiency of the E_T^{miss} selection, ϵ_{ISR} is the efficiency of the ISR jet requirement, and ϵ_{other} is the product of any remaining efficiencies to be considered (e.g. b-jet veto cut efficiency, efficiency of the $\Delta\phi$ QCD rejection selection, and efficiencies of vetoes of additional leptons).

Since the desired signal event topology includes a single soft τ_h, a key component of this analysis comes from understanding the τ_h identification efficiency in data and validating its modeling in simulation. In order to focus-in on the τ_h identification efficiency, a control sample enriched with $Z(\rightarrow \tau\tau \rightarrow \tau_h\tau_h)$+ISR events is studied. The efficiency for the requirement of two high quality τ_h candidates is expected to be well modeled by simulation, especially since all proper POG recommended corrections are applied. Another advantage to note for this event is that this sample is collected with a double-τ_h trigger with higher $p_T(\tau_h)$ thresholds than those defining the SR. Therefore, in addition to the requirement of two τ_h candidates (as opposed to one), the kinematic criteria provides a control sample with negligible signal contamination. Other criteria (described in what follows) are used to ensure minimal signal contamination over a broad range of signal models (including large mass gaps). The number of $Z(\rightarrow \tau\tau)$ events in the CR is to be parameterized as follows:

$$N^{CR}_{Z\rightarrow\tau\tau} = \sigma_Z \cdot L_{\text{int}} \cdot \epsilon^2_{\tau_h} \cdot \epsilon_{\text{ISR}} \cdot \epsilon_{\text{other}}, \tag{XI.2}$$

where σ_Z is again the cross section of this particular process, L_{int} is the integrated luminosity, ϵ_{τ_h} is the selection efficiency on one τ_h candidate, ϵ_{ISR} is the efficiency of the ISR cut, and ϵ_{other} is the product of any remaining efficiencies to be considered. Since ISR modeling, described previously, is well-understood and proper corrections to the MC have been measured and validated with muon control samples, those ISR weights are used to correct the Z+jets prediction from MC in this di-τ_h control sample. No cut on E_T^{miss} is applied in order to factorize the effect from τ_h identification

efficiency as much as possible.

The event selection criteria for the $Z(\rightarrow \tau\tau)$+ISR CR, which closely pertain to the SR, are given in Table 23. These event selection criteria pertain to this region in both the 2016 and 2017 iterations of the analysis. In order to simultaneously select events from on-mass-shell Z bosons and also to minimize the contribution from the QCD multijet background, the $\tau_h\tau_h$ pair is required to have an invariant mass greater than 50 GeV but less than 100 GeV. A selection is made on jet pseudorapidity ($\eta(j)$) in order to avoid potential jet resolution issues in the forward regions of the detector. To guard against jets faking leptons, jet candidates must not overlap the identified τ_h ($\Delta R(\tau_h, j) > 0.3$), and "Loose" identification is applied to jets in the 2016 iteration of the analysis. The equivalent of "Loose" identification is not available for jets in the 2017 iteration. For 2017, the recommended "Tight" jet identification is utilized. In addition, we impose the requirement of at least one highly energetic ISR jet, specifically with transverse momentum ($p_T(j_{\text{lead}})$) greater than 100 GeV, to produce a boosted event topology.

For purity in this region, the τ_h candidates are required to pass anti-electron and anti-muon criteria (described in the object identification section) to minimize events where electrons and muons can fake τ_h. The name of the anti-electron discriminator is againstElectronMVALooseMVA6. The name of the anti-muon discriminator is againstMuonTight3. The τ_h candidates are also required to pass the byTightIsolationMVArun2v1DBnewDMwLT isolation discriminator. For 2017 data and MC, the equivalent of this is the "Tight" isolation, given by the discriminator [8]. Events firing the HLT_DoubleMediumIsoPFTau35_Trk1_e trigger (double-τ_h trigger) are selected, and "Tau" primary datasets are used. The choice of trigger and its corresponding efficiency motivate the lower limit of 60 GeV on $p_T(\tau_h)$.

Table 23: $Z(\rightarrow \tau\tau)$+ISR Event Selections.

Selection	Cut
$N(\tau_h)$	≥ 2
$p_T(\tau_h)$	> 60 GeV
$\lvert\eta(\tau_h)\rvert$	< 2.1
$Q_{\tau_{h,1}} * Q_{\tau_{h,2}}$	< 0 [OS]
$m_{\tau_h\tau_h}$	[50, 100] GeV
$\lvert\eta(j)\rvert$	< 2.4
$N(j)$	≥ 1
$p_T(j_{\text{lead}})$	≥ 100 GeV
$\text{N(b} - \text{jets)}$	$= 0$
Trigger	HLT_DoubleMediumIsoPFTau35_Trk1_eta2p1

Estimating the QCD Contribution to the $Z(\rightarrow \tau\tau)$+ISR Control Region

Of interest is the background contribution from QCD multijet events to this CR. The typical probability of misidentifying a QCD jet as a τ_h is at least an order of magnitude higher than that for a QCD jet to be misidentified as a light lepton. Thus, a pair of jets from $q\bar{q} \rightarrow g \rightarrow q'\bar{q}'$ can be misidentified as two τ_hs. In terms of identification, the jets are "charge blind." This means that in some scenarios (about 45% of the time), the QCD multijet event will have two jets identified as a pair of same-sign τ_hs, and in others (about 55% of the time), the pair will be identified as opposite-sign. A similar possible complication arises for a QCD multijet event that produces a pair of bottom quarks ($b\bar{b}$). Sometimes, those bottom quarks decay into a tau lepton, a neutrino, and a charm quark. Hence, this event could be misidentified as a $Z(\rightarrow \tau\tau)$+ISR event. Even so, these $b\bar{b}$ events contain tau leptons that are not isolated, making them less likely to present significant challenge to the event topology. Again, on average, a misidentified QCD multijet event is identified as a pair of opposite-sign leptons 55% of the time and as a pair of same-sign leptons 45% of the time.

Estimation of QCD background contribution is not taken from MC due to poor statistics. Instead, a data-driven method is used to estimate the background contribution from QCD multijet events. To create a control sample with a high purity of QCD multijet events, the nominal opposite-sign di-τ_h requirement is switched to same sign, and all other selections from the $Z(\rightarrow \tau\tau)$+ISR CR remain the same. Then, these selections are used to analyze again all "Tau" data samples and all background MC samples. Following this run, the three non-QCD contributing backgrounds of Z+jets, W+jets, and $t\bar{t}$ are subtracted from data to extract the QCD contribution in the same-sign CR. The QCD contribution in the same-sign region is scaled by an OS-to-SS transfer factor of 1.22$\pm$0.12 (1.20$\pm$0.15) for 2016 (2017) to account for how often QCD events give opposite-sign versus same-sign on average. This transfer factor is measured from an orthogonal control sample where the invariant mass of the di-τ_h system is inverted (i.e. required to be greater than 100 GeV). The OS-to-SS transfer factor is extracted from a di-τ_h control sample selected with a di-τ_h trigger and requiring both τ_h candidates to fail "Tight" but pass "Loose" isolation. The control sample is close to >95% purity of QCD. The purity in the OS and SS control samples varies between 96-99%, and thus the subtraction of non-QCD background has negligible effect.

The OS mass sideband used to derive the OS-to-SS ratio for QCD multijet events is given in Figure 30 (left), and the SS mass sideband used to derive the OS-to-SS ratio for QCD multijet events is given in Figure 30 (right). These two plots are utilized to make a data-driven estimation of the QCD in each region (data-nonQCD bg) and then to determine subsequently the OS/SS ratio for QCD multijet production.

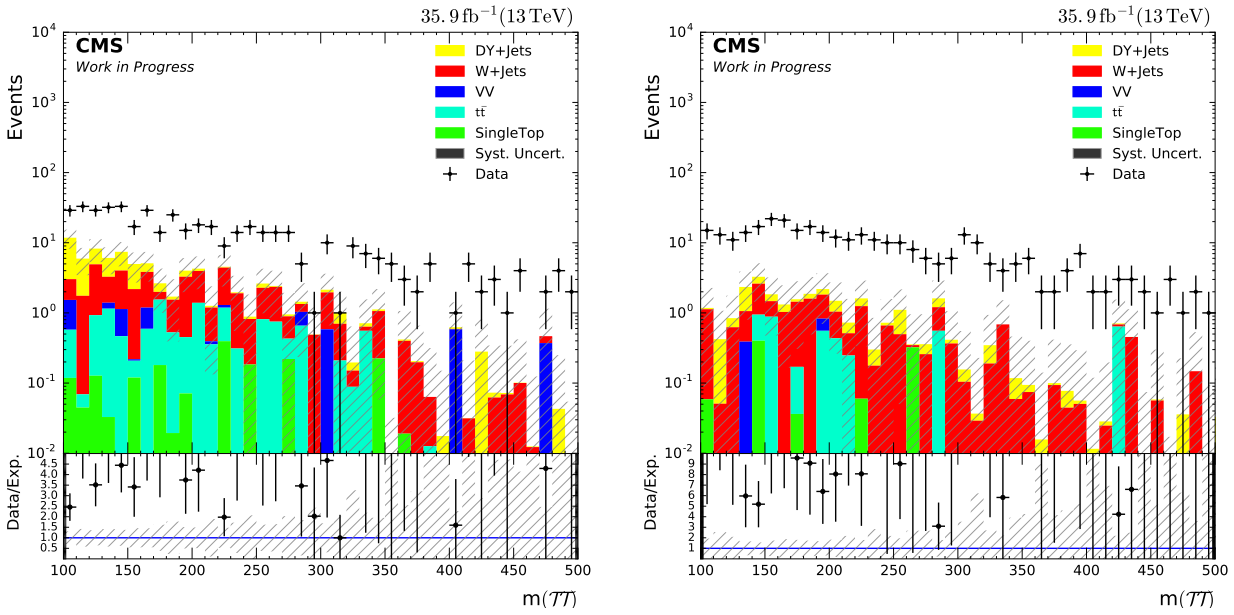

Figure 30: Mass sideband used to find the QCD OS-to-SS ratio for OS (SS) [left] ([right]).

Conclusions to Draw from $Z(\rightarrow \tau\tau)$+ISR

Included in Figure 31 are plots for the $Z(\rightarrow \tau\tau)$+ISR CR with selections as listed in Table 23 using 2016 data and MC. In addition to the data-driven QCD estimation as described in the subsection above, these plots include the boson boost weights as previously described. Beyond that, the contribution from Z+jets is scaled by 0.95^2 (0.83^2) to account for the POG recommended 5% (17%) correction to the τ_h identification efficiency for 2016 (2017) data and MC. There is significant QCD contribution to note in the tail of the di-τ_h reconstructable mass plot and at low E_T^{miss}. The plots indicate good agreement between data and MC over the full p_T spectrum. This set of plots establishes that the τ_h identification efficiency, being the focus of this CR, is well understood.

Table 24 lists the predicted background yields in simulation (after the corrections discussed) as well as the observed yield in data for 2016.

The measured data-to-MC SF for Z+jets is 0.92±0.05; therefore, the Z+jets prediction in the SR is estimated by correcting the MC prediction with this scale factor. Since the m_T shape in this control region (Figure 32) is well-modeled by MC, the Z+jets shape in the SR is taken directly from simulation for 2016 with systematic uncertainty described in subsequent sections.

Table 25 lists the predicted background yields in simulation (after the corrections discussed) as well as the observed yield in data for 2017. The measured data-to-MC ratio for Z+jets is 0.95±0.04; therefore, the Z+jets prediction in the SR is estimated by correcting the MC prediction with this scale factor. Since the m_T shape in this CR (Figure 33) is well-modeled by MC, the Z+jets shape in the SR is taken directly from simulation for 2017 with systematic uncertainty

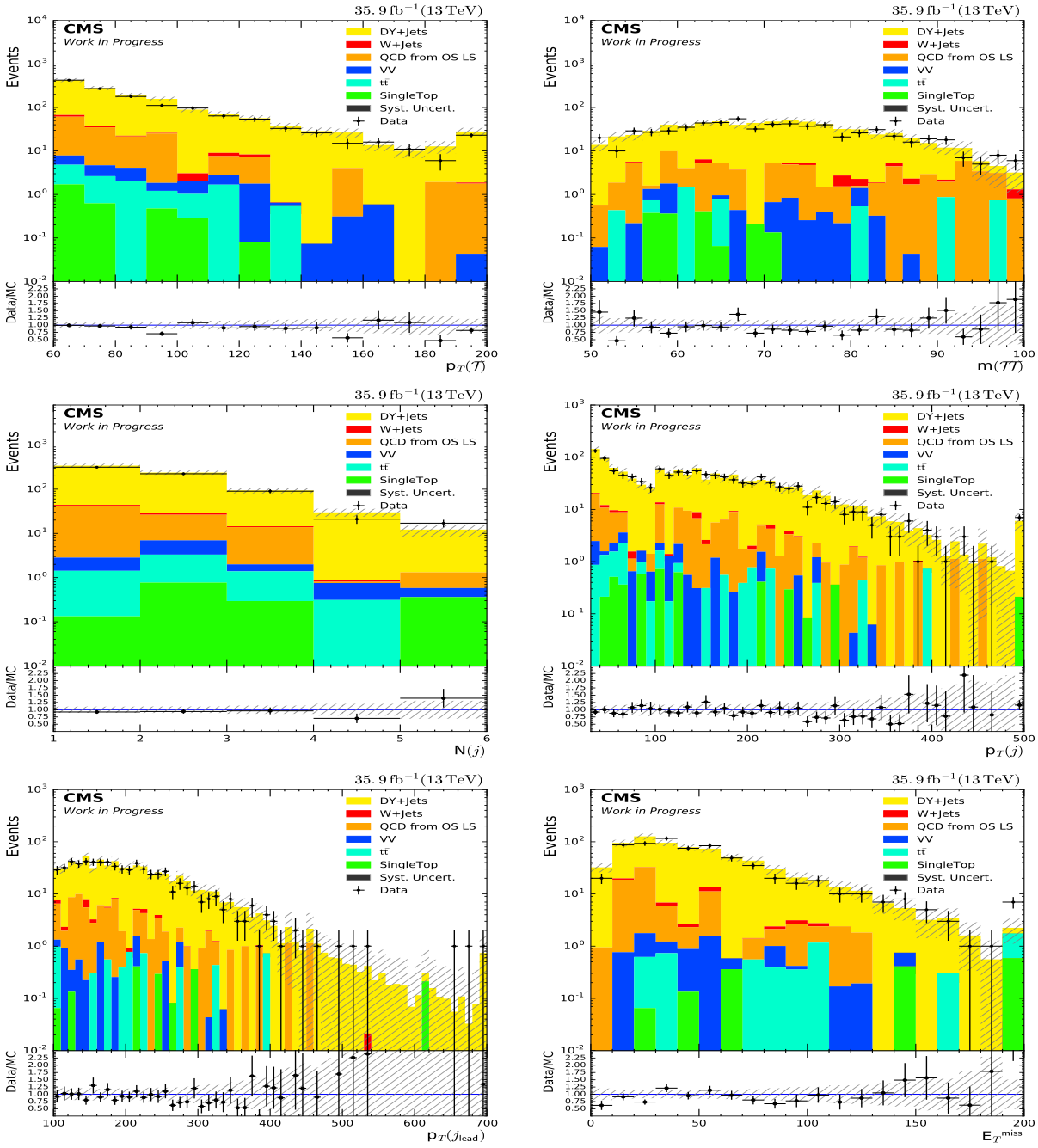

Figure 31: Top Row: $p_T(\tau_h)$ and $m(\tau_h, \tau_h)$; Second Row: $N(j)$ and $p_T(j)$; Third Row: $p_T^{\text{lead}}(j)$ and E_T^{miss} (2016).

described in subsequent sections.

Process	Yield
Data	665
W + jets	6.86 ± 1.55
DY + jets	623.93 ± 12.57
Diboson	6.32 ± 1.33
$t\bar{t}$	5.24 ± 1.42
QCD	69.92 ± 10.96
Single Top	1.56 ± 0.53
SF_{BG}	0.92 ± 0.05

Table 24: Background and data yields in the $Z(\rightarrow \tau\tau)$+ISR control region (2016).

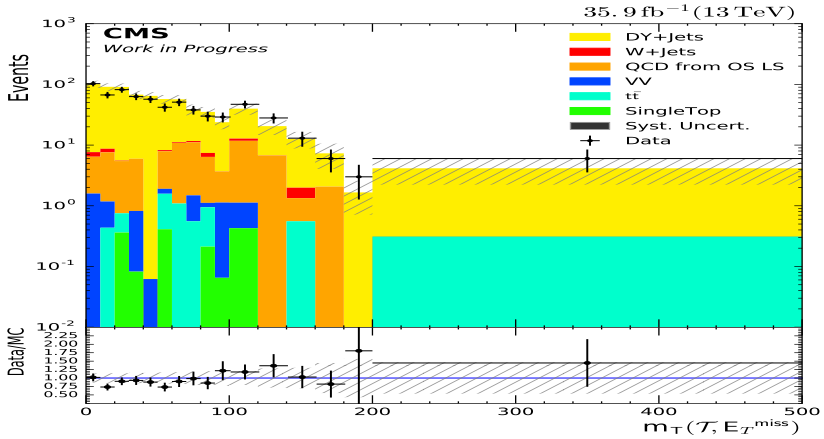

Figure 32: $m_T(\tau_h, E_T^{miss})$ in the $Z(\rightarrow \tau\tau)$+ISR control region (2016)

Process	Yield
Data	801
W + jets	8.8 ± 2.0
DY + jets	728.1 ± 14.7
Diboson	6.8 ± 1.5
$t\bar{t}$	9.0 ± 2.4
QCD	84.0 ± 13.2
Single Top	3.2 ± 1.1
ΣBG	839.9 ± 20.1
SF_{BG}	0.95 ± 0.04

Table 25: Background and data yields in the $Z(\rightarrow \tau\tau)$+ISR control region (2017).

Emulating Hadronic Tau Kinematics Using Muons

In order to further validate the m_T shape and expected yields for W+jets ($W(\rightarrow \tau_h\nu_\tau)$) can be taken from MC with the corrections as described previously, we also perform an emulation of the

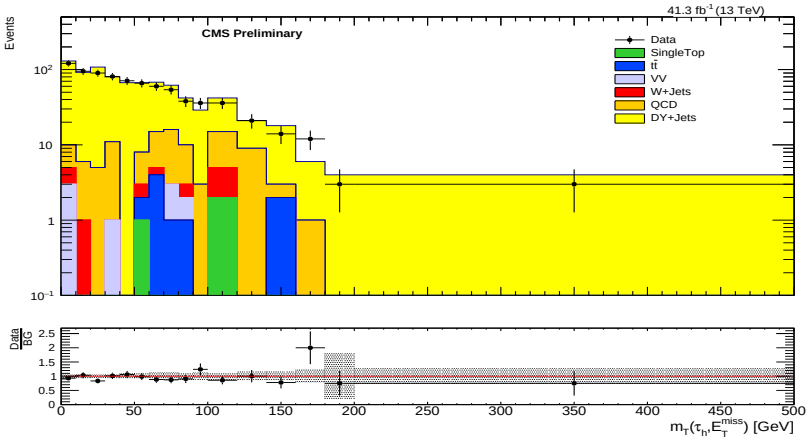

Figure 33: $m_T(\tau_h, E_T^{miss})$ in the $Z(\rightarrow \tau\tau)$+ISR control region (2017).

τ_h kinematics in W+jets events, using a data CR with muons. The method can be described in the following steps:

- Determine the fraction of τ lepton momentum taken by its visible decay products (i.e. how much does the hadronically decaying τ take). This fraction is defined as $R(p_T^{\tau_{gen}}) = p_T^{\tau_h,RECO}/p_T^{\tau_{gen}}$, where $p_T^{\tau_{gen}}$ is the generator level τ momentum before it decays, and $p_T^{\tau_h,RECO}$ is the reconstructed transverse momentum of the visible decay products (i.e. the reconstructed τ_h). The ratio $R(p_T^{\tau_{gen}})$ ("response template") is studied using W+jets MC. The minimum p_T of the reconstructed τ_h candidates saved in our files for analysis (ntpules) is 15 GeV. The templates are derived by first finding a true generator level visible hadronic τ_h (with a given gen. visible p_T) and requiring that it be matched to a reco. level τ_h using a matching requirement of $\Delta R < 0.4$. Once we have a matching pair (gen-reco match), then we produce response histograms in bins of generator level τ_h p_T.

- Calculate matching and reconstruction (ID) efficiencies (i.e. how often does a "true" gen-level τ_h actually get reconstructed and identified by our algorithms).

- Emulate a reconstructed τ_h by using muons (i.e. treat a muon as a gen. τ before decay in the above equations).

- Recalculate the E_T^{miss} vectorially by adding back in the 4-momentum vector of the emulated τ_h since the p_T contribution is missed in the muon sample.

Lepton universality and the small momentum scale uncertainty for muons (i.e. $\sigma(p_T(\mu)) < 2\%$) is small compared to other E_T^{miss} and τ_h uncertainties) allows us to describe $p_T^{\tau_{gen}} \sim p_T^{\mu}$. The p_T^{μ} must be weighted by the matching and reconstruction efficiencies of the τ_h, as well as unfolding the efficiency associated with the muon. Hence, the emulated momentum is as follows:

$$p_T(\tau_h)_{emulated} = p_T(\mu) \cdot R(p_T^{\tau_{gen}}) \cdot \epsilon_{Matching} \cdot \epsilon_{ID}/\epsilon_{\mu}. \quad \text{(XI.3)}$$

Figure 34, shows the $R(p_T^{\tau_{gen}}) = p_T^{\tau,RECO}/p_T^{\tau_{gen}}$ fraction, for different $p_T^{\tau_{gen}}$ ranges.

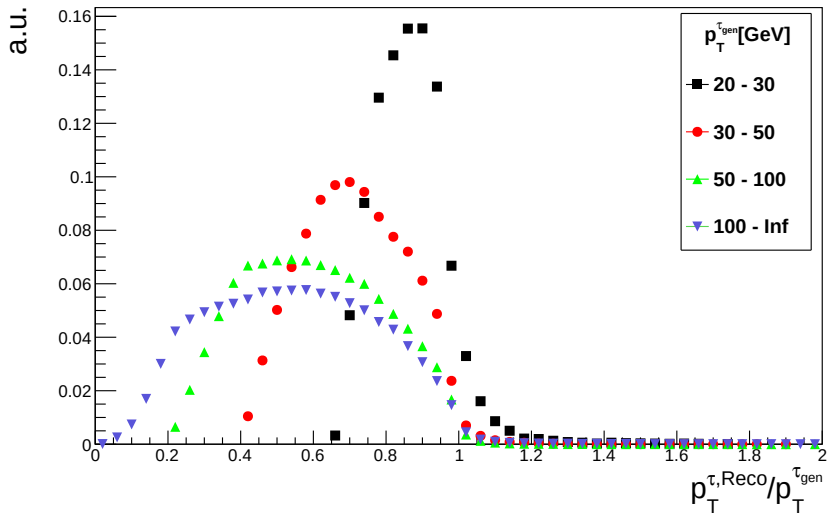

Figure 34: Fraction of momentum carried away by τ_h in τ lepton hadronic decays.

The matching and reconstruction efficiencies are calculated by using the following equations:

$$\epsilon_{Matching} = \frac{N(p_T(\tau_h^{RECO}(matched)))}{N(p_T(\tau_h^{gen}))} \tag{XI.4}$$

$$\epsilon_{ID} = \frac{N(p_T(\tau_h^{RECO}(matched + ID)))}{N(p_T(\tau_h^{RECO}(matched)))}. \tag{XI.5}$$

Since the $W(\rightarrow \tau\nu)$ samples are binned in H_T, the matching and reconstruction efficiencies are studied for the different H_T ranges.

Figure 35 shows the emulated E_T^{miss} and $m_T(\tau_h, E_T^{miss})$ distributions, using muon events to emulate the τ_h kinematics as described above. The distribution labeled W+jets in the legend is the emulated m_T using a single-muon dataset and treating the muon as a τ_h. The distribution labeled Data in the legend of Figure 35 is constructed from events in the Met primary data satisfying all the SR criteria because with $40 < p_T(\tau_h) < 60$ GeV in order to provide a validation sample orthogonal to the SR. Note there is overall good agreement between real $W(\rightarrow \tau\nu_\tau \rightarrow \tau_h\nu_\tau)$ data and the emulated distribution using single-muon data events with MC-derived response τ_h templates. This good agreement gives us further confidence to use the MC (with previously described corrections) to estimate the W+jets contribution in the SR. We note the last bin in Figure 35 represents the overflow bin and thus includes events out to infinity.

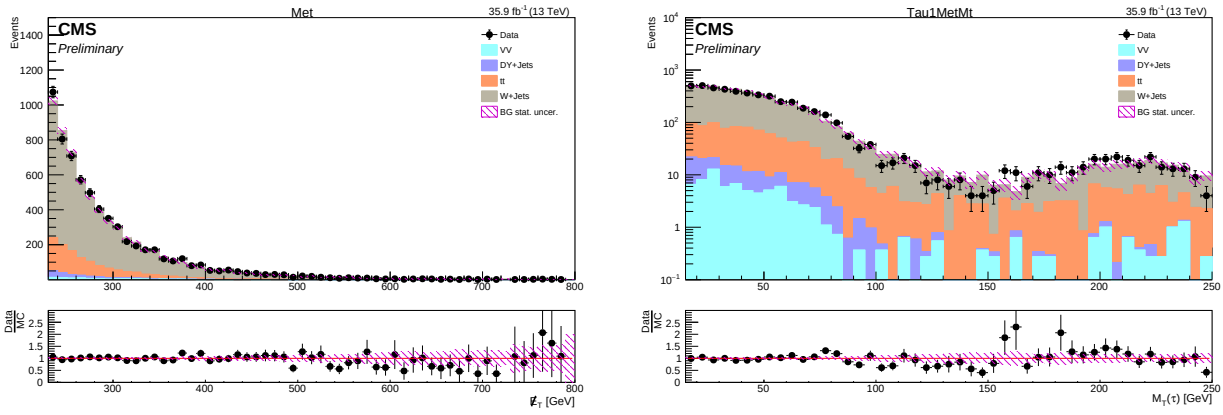

Figure 35: Emulated E_T^{miss} and $m_T(\tau_h, E_T^{miss})$ distributions, using muon events to emulate the τ_h kinematics (2016)

Chapter XII

Background Estimation: $t\bar{t}$ Contribution

The QCD production of $t\bar{t}$ events represents only a small fraction of the total background in the SR: ∼10-15% of the total background at $m_T < 200$ GeV and ∼25% at $m_T > 200$ GeV. According to simulation, the predicted $t\bar{t}$ signal rate for 2016 data and MC is 1353.78±25.83 for $m_T < 100$ GeV and 18.57±3.10 for $m_T > 200$ GeV. The estimate of the $t\bar{t}$ contribution to the SR is performed in a semi-data-driven way by obtaining four different control samples. The four CRs differ in the b-jet multiplicity and in the identification requirements for τ_h. These differences are summarized in Table 26 and apply to the 2016 and 2017 iterations of the analysis.

Region	$N(\mathrm{b-jet})$	τ ID isolation	N(signaltracks)
Signal	$== 0$	Tight	1
CR 1	$== 1$	Tight	1or2or3
CR 2	$== 2$	Tight	1or2or3
CR 3	$== 1$	VTight	1
CR 4	$== 2$	VTight	1

Table 26: Event selection differences between signal region and $t\bar{t}$ control regions.

Introducing one b-jet into the event selection requirements enriches the high-purity contribution from $t\bar{t}$ due to the decay of the top quark. Varying the b-jet multiplicity from 1 to 2 reduces the total contribution from QCD multijet backgrounds while minimally affecting the global $t\bar{t}$ data-to-MC ratio. In changing from a "Tight" to a "Very Tight" identification requirement for τ_h, the number of events with potential fake-τ_hs is substantially reduced. Changing from 1, 2, or 3 signal tracks to requiring 1 signal track similarly reduces the contribution from QCD multijet backgrounds, thus again reducing the number of events with potential fake-τ_hs. The τ_h, E_T^{miss}, and jet-specific event selection criteria, which closely pertain to the SR, are given for CR 1 explicitly in Table 26. These event selection criteria pertain to this region in both the 2016 and 2017 iterations of the analysis.

The event selection requirements for τ_h are highly pertinent to the SR. Similar to the SR, a single τ_h is required, and this τ_h is soft with a transverse momentum between 20 and 40 GeV. A selection is made on jet pseudorapidity ($\eta(j)$), requiring $|\eta(j)| < 2.4$, where jet resolution is best understood, and "Loose" identification is applied to jets for the 2016 iteration of the analysis. The equivalent of the "Loose" identification is not available for 2017. For 2017, the recommended "Tight" jet identification is utilized. The highly energetic ISR jet with momentum ($p_T(j_{\mathrm{lead}})$) greater than 100 GeV boosts the system similar to the SR. Events firing the HLT_PFMETNoMu120_PFMHTNoMu120_IDTight trigger are selected, and E_T^{miss} datasets are

Selection	Cut
$N(\tau_h)$	1
$p_T(\tau_h)$	[20, 40] GeV
$\lvert\eta(j)\rvert$	< 2.5
$N(j)$	≥ 1
$p_T^{\text{lead}}(j)$	≥ 100 GeV
E_T^{miss}	≥ 230 GeV
$N(\text{b}-\text{jet})$	1
τ_h ID	"Tight"
N(signaltracks)	1or2or3hps
QCD rejection	$\lvert\Delta\phi(j_{\text{lead}}, E_T^{miss})\rvert > 0.7$
Trigger	HLT_PFMETNoMu120_PFMHTNoMu120_IDTight

Table 27: $t\bar{t}$ Control Region Event Selections

used. The trigger efficiency is applied as a weight to the MC as outlined in the Triggers section.

In order to produce a high-purity $t\bar{t}$ control sample with little signal contamination, the first CR requires a single b-jet (see Table 26). This CR provides good purity, but there is non-negligible contamination from QCD fake-τ_h backgrounds in the high-m_T regions (see Figure 36 for 2016 data and MC).

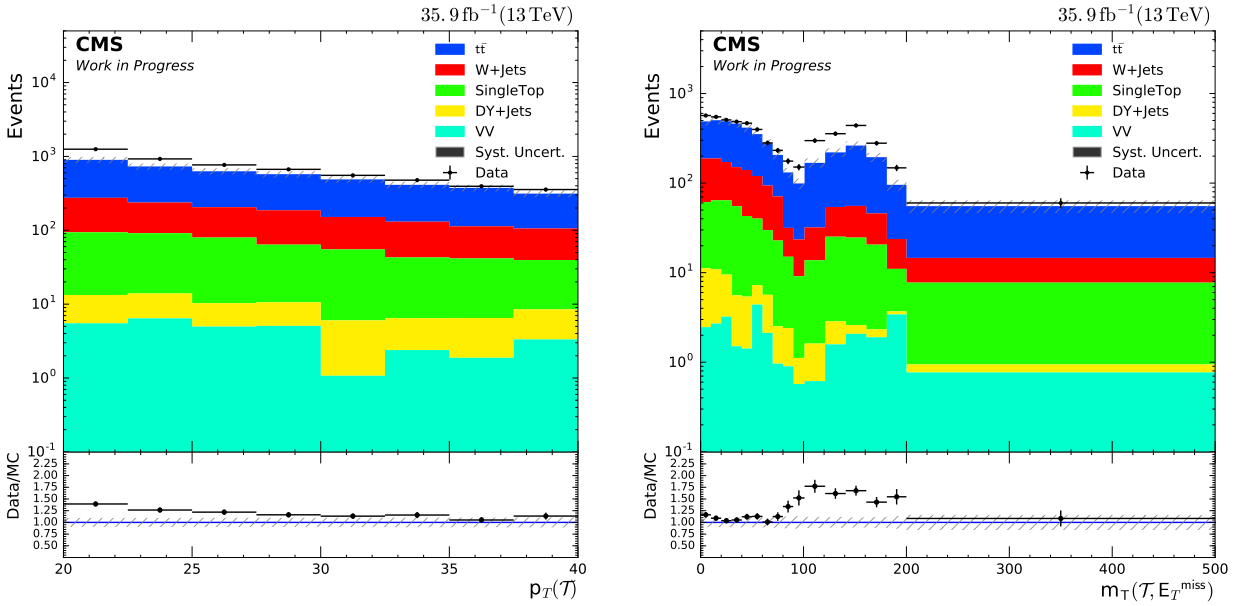

Figure 36: $p_T(\tau_h)$ (left) and $m_T(\tau_h, E_T^{miss})$ (right) for events with exactly 1 b-jet and "Tight" ID for τ_h (2016)

To reduce the QCD multijet contamination at high-m_T, the second CR for $t\bar{t}$ is developed, which requires exactly 2 b-jets. This reduces the overall contribution from QCD multijet by roughly a factor of the b-jet fake rate (f_{b-jet}) while minimally affecting the global $t\bar{t}$ data-to-MC scale factor (dominated by lower m_T). In Figure 37, the plots for $p_T(\tau_h)$, jet multiplicity, and E_T^{miss} are included

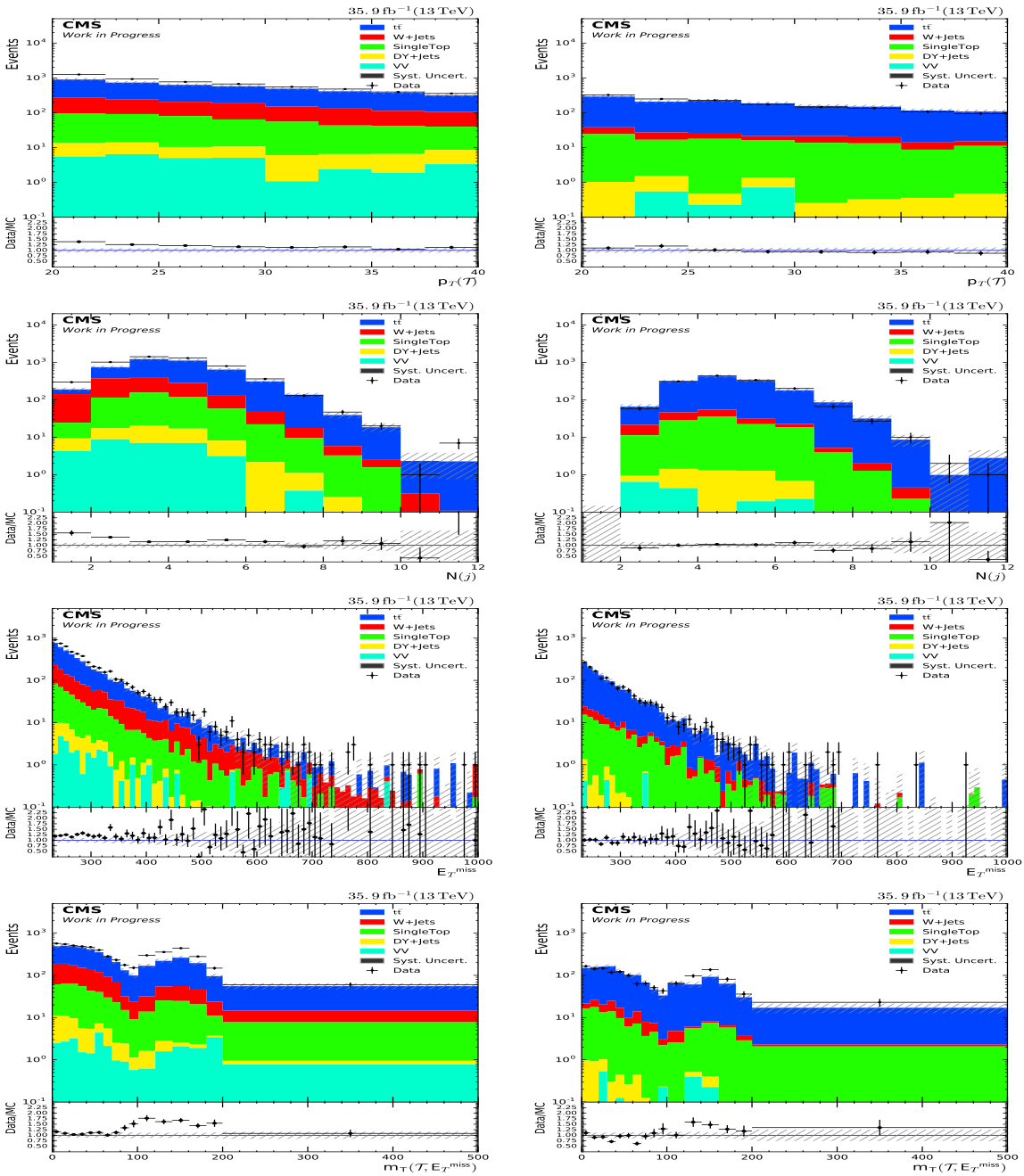

Figure 37: Top Row: $p_T(\tau_h)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Middle Row: $N(j)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Next Row: $m_T(\tau_h, E_T^{miss})$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Bottom Row: E_T^{miss} with exactly 1 b-jet (left) and for exactly 2 b-jets (right) for events with "Tight" τ_h ID (2016).

for these two CRs and 2016 data and MC. The $p_T(\tau_h)$ shows the most disagreement at lower p_T for CRs 1 and 2. A disagreement is present in the 1 and 2 jet bins of the jet multiplicity distribution (as was the case in CR 1) that improves in control region 2. The presence of fake-τ_hs is suggested by these two observations and the improvement in the data-to-MC ratio for the E_T^{miss} in moving

from CR 1 to 2. In other words, the level of agreement between data and MC has improved in CR 2 due to the lower QCD multijet contamination in CR 2.

The third and fourth CRs are utilized to reduce the presence of fake-τ_hs further. In the third and fourth CRs, similar selections to the first two CRs are used; however, instead of a "Tight" isolation requirement on the τ_h identification isolation, a "Very Tight" (VTight) τ_h identification isolation is applied. Changing from a "Tight" to a "Very Tight" τ_h identification isolation reduces the jet-to-τ_h fake rate by $\sim 50\%$. Additionally, instead of requiring τ_h with 1, 2, or 3 signal tracks, regions 3 and 4 are required to have a τ_h with 1 signal track. This should reduce the fake rate since fake-τ_hs commonly result in 2 tracks. The results of these two changes using 2016 data and MC appear in Figure 38 for $p_T(\tau_h)$, jet multiplicity, and m_T.

Tightening the τ_h selections improves the agreement between data and MC in $t\bar{t}$-enriched CRs (Figures 37 and 38). More specifically, discrepancy between data and MC at large values of m_T, low values of $p_T(\tau_h)$, low jet multiplicity, and low E_T^{miss} has gone away, again suggesting that the discrepancies in the first two CRs were from the presence of fake-τ_hs from QCD multijet backgrounds. The results of these four CRs indicate that the contribution from $t\bar{t}$ in the SR is well-understood.

Table 28 lists the predicted background yields in simulation (after the corrections discussed) as well as the observed yields in data for 2016 data and MC. The data-to-MC scale factor in the fourth $t\bar{t}$ CR has a measured value of 0.94 ± 0.05. Therefore, the $t\bar{t}$ prediction in the SR is estimated by correcting the MC prediction with this scale factor. Since the m_T shape in these CRs is well-modeled by MC, the $t\bar{t}$ m_T shape in the SR is taken from simulation. Systematic uncertainty on the m_T shape is described in the following sections.

Process	Yield
Data	783
W + jets	38.7 ± 3.3
DY + jets	2.7 ± 0.5
Diboson	0.6 ± 0.5
$t\bar{t}$	705.5 ± 18.9
Single Top	80.5 ± 4.4
SF_{BG}	0.94 ± 0.05

Table 28: Background and data yields in the $t\bar{t}$ control regions (2016).

Table 29 lists the predicted background yields in simulation (after the corrections discussed) as well as the observed yield in data for 2017 data and MC. The data-to-MC scale factor in the fourth $t\bar{t}$ CR has a measured value of 0.95 ± 0.04. Therefore, the $t\bar{t}$ prediction in the SR is estimated by correcting the MC prediction with this scale factor. Since the m_T shape in these CRs is well-modeled by MC, the $t\bar{t}$ shape in the SR is taken directly from simulation for the 2017 iteration as well. Systematic uncertainties are taken into account once again. Figure 39 shows the m_T

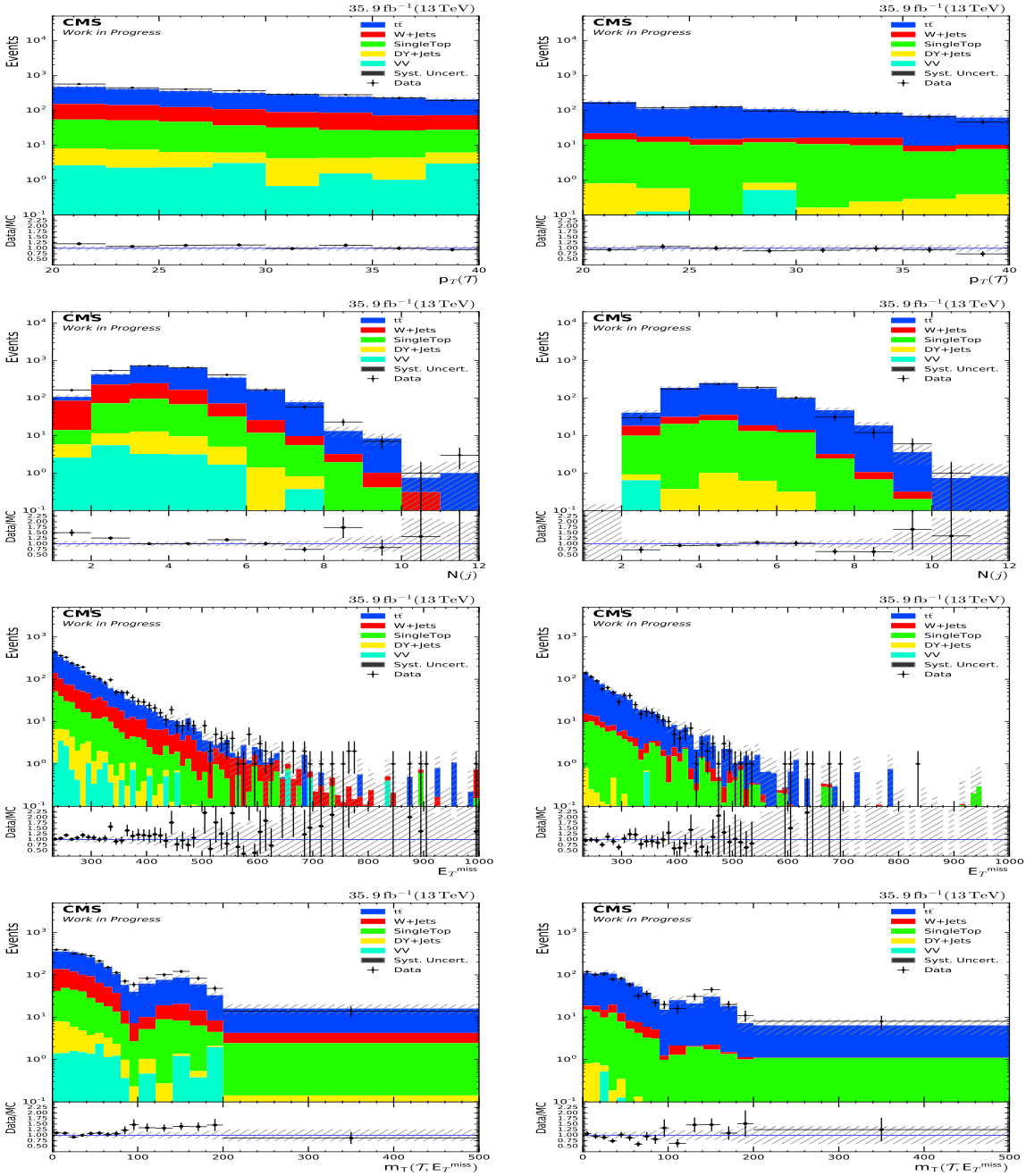

Figure 38: Top Row: $p_T(\tau_h)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Second Row: $N(j)$ with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Third Row: E_T^{miss} with exactly 1 b-jet (left) and for exactly 2 b-jets (right); Bottom Row: m_T with exactly 1 b-jet (left) and for exactly 2 b-jets (right) for events with "VTight" ID for τ_h (2016).

distribution in $t\bar{t}$ CR 4 using 2017 data, where quality agreement is observed between data and MC.

Process	Yield
Data	915
W + jets	41.0 ± 4.5
DY + jets	1.9 ± 0.4
Diboson	0.7 ± 0.7
$t\bar{t}$	805.8 ± 21.7
Single Top	112.1 ± 5.0
ΣBG	961.5 ± 22.7
SF_{BG}	0.95 ± 0.04

Table 29: Background and data yields in the $t\bar{t}$ control regions (2017).

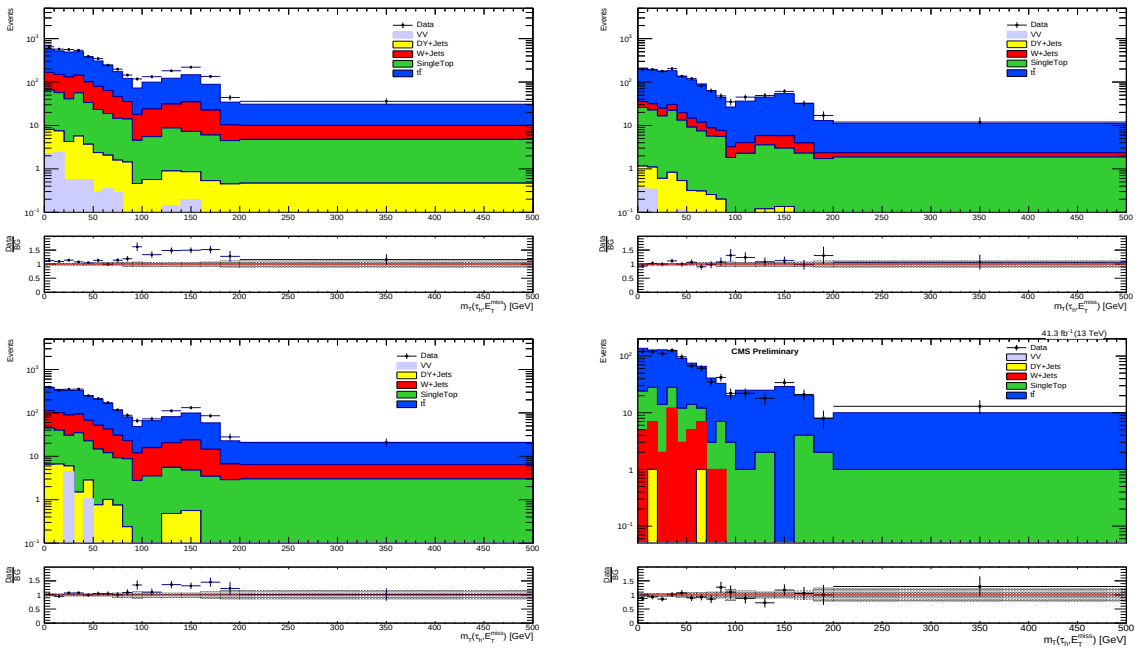

Figure 39: Results for $t\bar{t}$+ISR control regions using 2017 data. The top left plot is for CR1, top right for CR2, bottom left for CR3, and bottom right for CR4.

Chapter XIII

Background Estimation: QCD Contribution

To determine the contribution from QCD multijet events in the SR, a fully-data-driven estimation is used. Our data-driven estimation utilizes the classic ABCD method (sketched in Figure 40), where four high-purity QCD multijet CRs are defined. The four control regions differ in their requirements for two key variables: τ_h isolation and $\Delta\phi(j_{\text{lead}}, E_T^{miss})$. The $\Delta\phi(j_{\text{lead}}, E_T^{miss})$ is the difference in the azimuthal angles for the leading jet and the missing transverse energy. Region A (signal) of Figure 40 will have events with $|\Delta\phi(j_{\text{lead}}, E_T^{miss})| > 0.7$ and a τ_h candidate passing the "Tight" isolation requirement. The CR B will only differ from the SR in that the τ_h candidate included will pass the "Loose" but fail the "Tight" τ_h isolation requirement. To create regions CR C and CR D, the $\Delta\phi(j_{\text{lead}}, E_T^{miss})$ is inverted ($|\Delta\phi(j_{\text{lead}}, E_T^{miss})| < 0.7$) to create two more QCD-enriched regions. The τ_h candidates in CR C pass the "Loose" but fail the "Tight" isolation requirement, whereas the τ_h candidates in CR D are required to pass "Tight." This method will allow for the determination of the correct normalization and m_T shape for QCD events in the SR using a transfer factor. This transfer factor is defined as the number of events with a τ_h candidate passing "Tight" divided by the number of events with a τ_h candidate passing "Loose" but failing "Tight" (a.k.a. "tight-to-loose ratio"). This methodology, including how we obtain the tight-to-loose ratio, is described in further detail in what follows.

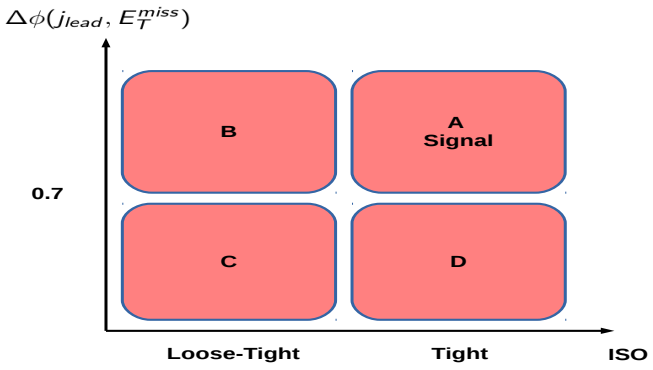

Figure 40: Sketch for ABCD methodology for the QCD background estimation.

Ideally the transfer factor (or "tight-to-loose" ratio) can be determined as the ratio of events in CR D to CR C. This factor, once obtained, will allow for proper normalization of the number of QCD events in CR A, which will be obtained from the number of QCD events in CR B scaled by the tight-to-loose ratio. However, while the QCD multijet background includes τ_h candidates that

come from fakes, CR C and CR D contain non-negligible contributions from non-QCD backgrounds with genuine τ_h candidates. For this reason, we instead measure the tight-to-loose ratio using two other CRs enriched with Z/W+jets events: $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ and $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$. These regions are quite clean due to the muons in the final state. Since a jet has a much lower probability to get misidentified as a prompt muon than a genuine τ_h, the τ_h candidates in these CRs are jets being misidentified as genuine τ_h candidates. The topologies of the two CRs are depicted in the Feynman diagrams in Figure 41.

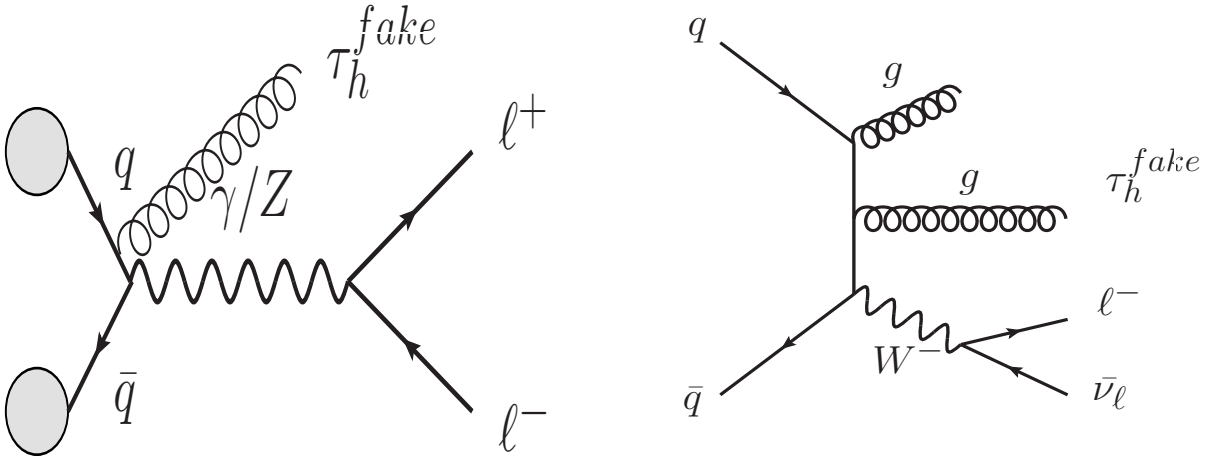

Figure 41: Feynman diagram of $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ and $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$.

Table 30 summarizes the event selection criteria used in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ region. Events in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ region are required to have a well-identified muon pair with opposite electric charge and a reconstructed dimuon mass falling in a mass window around the Z-peak. In addition, a τ_h candidate with the same kinematic and identification criteria as that for the SR is requested. Since the two muons are decay products of the Z boson, the additional selected τ_h candidate results from a hadronic jet misidentified as a τ_h. The second CR enriched with $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ events differs only in that a single muon is required, and a single τ_h (a fake) is again selected. Note there is a E_T^{miss} requirement in both CRs. The E_T^{miss} in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ region results mainly from the mismeasurement of the associated ISR jet's p_T, while in the $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CR, E_T^{miss} comes from the corresponding muon-neutrino and also from mismeasured jets. In order to measure the tight-to-loose ratio with selection criteria as similar as possible, and to have good purity, the same E_T^{miss} selection of > 30 GeV is used in both CRs. Nevertheless, it has been checked that the ratios do not depend on the E_T^{miss} threshold.

The tight-to-loose ratio ($Ratio^{Tight}_{Loose-nonTight}(data - driven)$) for both CRs ($Z/W$) is given by Equation XIII.1, where $N^{Tight}_{Z/W+Jets}(data - driven)$ is the number of events (data - non$Z$MC for Z region and data - non$W$MC for W region) passing the "Tight" τ_h isolation requirement and $N^{Loose-nonTight}_{Z/W+Jets}(data - driven)$ is the number of events (data - non$Z$MC for Z region and data - non$W$MC for W region) passing the "Loose" but failing the "Tight" τ_h isolation requirement. This

Central Selections	$Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$	$W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$
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	-
$N(e)$&$N(b)$	-	$== 0$
$p_T(e)$	-	10 GeV
$p_T(b)$	-	30 GeV
E_T^{miss}	> 30 GeV	> 30 GeV
$m(\mu, \mu)$	> 70 & < 110 GeV	-
$m_T(\mu, E_T^{miss})$	-	> 50 & < 120 GeV

Table 30: Summary of the cuts used for the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ and $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CRs.

methodology for calculating the tight-to-loose ratio applies to the 2016 and 2017 iterations of the analysis.

$$Ratio^{Tight}_{Loose-nonTight}(data - driven) = \frac{N^{Tight}_{Z/W+Jets}(data - driven)}{N^{Loose-nonTight}_{Z/W+Jets}(data - driven)}. \quad \text{(XIII.1)}$$

Figure 42 shows the $p_T(\tau_h)$ distribution for events passing the "Tight" isolation requirement (left) and passing the "Loose" but failing the "Tight" isolation requirement (right) in the Z region. The top row of Figure 42 includes results from 2016 data and MC, and the bottom row corresponds to the same results using 2017 data and MC.

Figure 43 shows the ratio of the left column ("Tight") to the right column ("Loose," fail "Tight") (after subtracting "other" backgrounds from data) in each bin of the $p_T(\tau_h)$ distribution for the 2016 (left) and 2017 (right) iterations of the analysis. Thus, Figure 43 shows the tight-to-loose ratio versus $p_T(\tau_h)$ measured from the Z region. The values of the tight-to-loose ratio and their errors based only on statistics are presented in Table 31 for 2016 data and MC. Results for 2017 data and MC are presented in Table 32. Reported uncertainties correspond with only the statistical contribution from data and MC. Within statistical uncertainties, the ratios obtained for the Z region in the 2016 and 2017 iterations of the analysis are in agreement. We fit the $p_T(\tau_h)$-dependent tight-to-loose ratios each time with a polynomial of degree-1, as shown in Figure 43. We utilize these fits to describe the tight-to-loose ratio for fake-τ_h candidates in terms of the $p_T(\tau_h)$ dependence henceforth. The motivation for using the fit instead of the bin-by-bin values is to reduce the systematic uncertainty on our background prediction for the QCD multijet background in the SR.

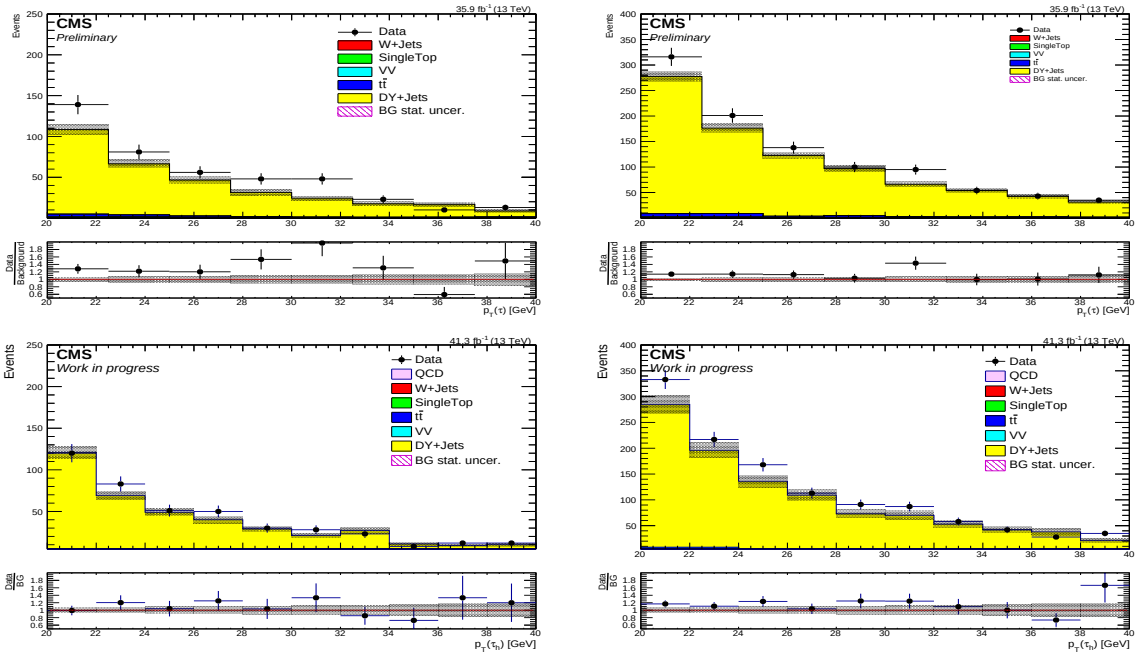

Figure 42: Top (Bottom) row: $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ CR w/ (1) "Tight" isolation and (2) passing "Loose" but failing "Tight" for 2016 (2017).

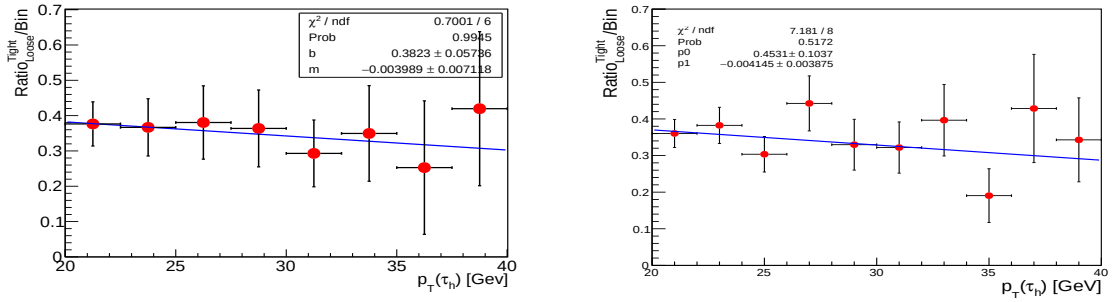

Figure 43: $p_T(\tau_h)$ dependence of the tight-to-loose ratio in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ region for 2016 (2017), left (right).

The selection criteria used to construct the $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CR is outlined in Table 30. The tight-to-loose ratio for the τ_h is also calculated using Equation XIII.1. Figure 44 shows the $p_T(\tau_h)$ distribution for events passing the "Tight" isolation requirement (left) and passing the "Loose" but failing the "Tight" isolation requirement (right) in the W region. The top row of Figure 44 includes results from 2016 data and MC, and the bottom row corresponds to the same results using 2017 data and MC. Figure 45 shows the ratio of the left column ("Tight") to the right column ("Loose," fail "Tight") (after subtracting "other" backgrounds from data) in each bin of the $p_T(\tau_h)$ distribution for the 2016 (left) and 2017 (right) iterations of the analysis.

$p_T(\tau_h)$ [GeV]	$Ratio^{Tight}_{Loose}$ (2016)
$[20, 22.5]$	0.37 ± 0.06
$[22.5, 25]$	0.36 ± 0.08
$[25, 27.5]$	0.38 ± 0.10
$[27.5, 30]$	0.36 ± 0.10
$[30, 32.5]$	0.29 ± 0.09
$[32.5, 35]$	0.34 ± 0.13
$[35, 37.5]$	0.25 ± 0.19
$[37.5, 40]$	0.42 ± 0.22

Table 31: Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the Z region (2016).

$p_T(\tau_h)$ [GeV]	$Ratio^{Tight}_{Loose}$ (2017)
$[20, 22]$	0.36 ± 0.04
$[22, 24]$	0.38 ± 0.05
$[24, 26]$	0.30 ± 0.05
$[26, 28]$	0.44 ± 0.08
$[28, 30]$	0.33 ± 0.07
$[30, 32]$	0.32 ± 0.07
$[32, 34]$	0.40 ± 0.10
$[34, 36]$	0.19 ± 0.07
$[36, 38]$	0.43 ± 0.15
$[38, 40]$	0.34 ± 0.11

Table 32: Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the Z region (2017).

Thus, Figure 45 shows the tight-to-loose ratio versus $p_T(\tau_h)$, measured from the W region. The values of the tight-to-loose ratio and their errors based on only statistics are presented in Table 33 for 2016 data and MC. Results for 2017 data and MC are presented in Table 34. The reported uncertainties correspond with only the statistical contribution from data and MC.

Again, we see here that the ratios obtained for the W region in the 2016 and 2017 iterations of the analysis are in close agreement. We fit the $p_T(\tau_h)$-dependent tight-to-loose ratios each time with a polynomial of degree-1, as shown in Figure 45. These results are consistent with the values obtained in the $Z(\rightarrow \mu\mu) + \tau_h^{\text{fake}}$ CR. Thus, we will utilize the results from the W region since it gives smaller uncertainties for the p_T-dependent tight-to-loose ratio for the fake τ_h candidates. The difference with respect to the measured tight-to-loose ratios obtained from the Z region is used to assign systematic uncertainty.

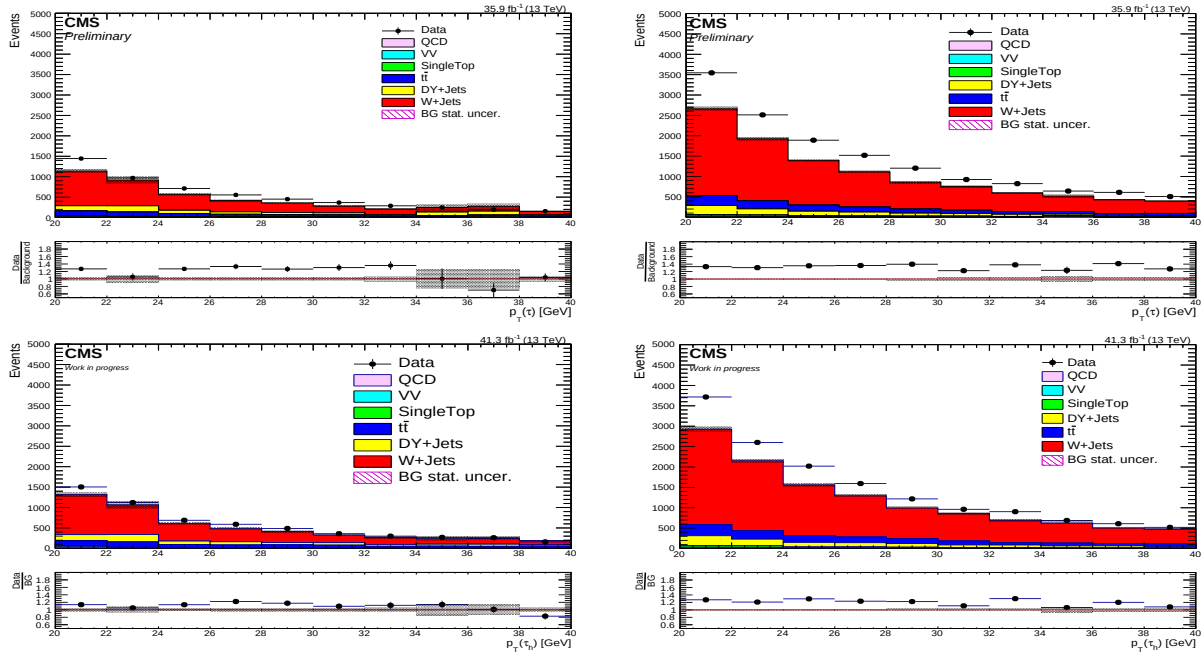

Figure 44: Top (Bottom) row: $W(\rightarrow \mu\nu) + \tau_h^{\text{fake}}$ CR w/ (1) "Tight" isolation and (2) passing "Loose" but failing "Tight" for 2016 (2017).

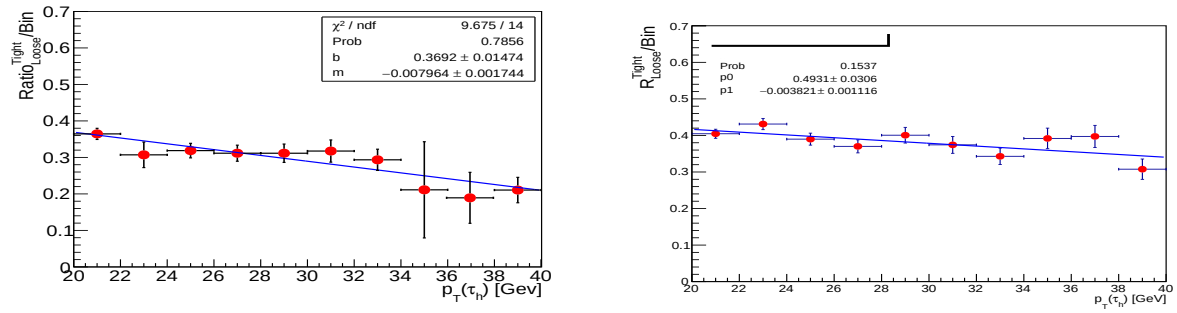

Figure 45: $p_T(\tau_h)$ dependence of the tight-to-loose ratio in the $W(\rightarrow \mu\nu) + \tau_h^{fake}$ region for 2016 (2017), left (right)

As mentioned previously, we take the shape for QCD multijet events from CR B and re-weight it using the $p_T(\tau_h)$-dependent tight-to-loose ratios in order to get the correct shape in the SR. To obtain the proper normalization for QCD events in the SR, we normalize the reweighted m_T shape to the predicted yield obtained using the tight-to-loose method described above in the $W(\rightarrow \mu\nu)+\tau_h^{\text{fake}}$ region. Since the $p_T(\tau_h)$ is correlated with the $m_T(\tau_h, E_T^{miss})$, we first create a 2D histogram of m_T vs. $p_T(\tau_h)$ and then apply the p_T-dependent tight-to-loose ratios as a weight to the 2D histogram. This is a suitable strategy to get the proper QCD shape and normalization in the SR. To show our confidence in this data-driven methodology, we perform a closure test in data by utilizing CR C

$p_T(\tau_h)$ [GeV]	$Ratio_{Loose}^{Tight}$ (2016)
$[20, 22]$	0.36 ± 0.01
$[22, 24]$	0.30 ± 0.03
$[24, 26]$	0.31 ± 0.01
$[26, 28]$	0.31 ± 0.02
$[28, 30]$	0.31 ± 0.02
$[30, 32]$	0.31 ± 0.03
$[32, 34]$	0.29 ± 0.02
$[34, 36]$	0.21 ± 0.13
$[36, 38]$	0.20 ± 0.06
$[38, 40]$	0.21 ± 0.03

Table 33: Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W region (2016).

$p_T(\tau_h)$ [GeV]	$Ratio_{Loose}^{Tight}$ (2017)
$[20, 22]$	0.40 ± 0.01
$[22, 24]$	0.43 ± 0.02
$[24, 26]$	0.39 ± 0.02
$[26, 28]$	0.37 ± 0.02
$[28, 30]$	0.40 ± 0.02
$[30, 32]$	0.37 ± 0.02
$[32, 34]$	0.34 ± 0.02
$[34, 36]$	0.39 ± 0.03
$[36, 38]$	0.40 ± 0.03
$[38, 40]$	0.31 ± 0.03

Table 34: Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W region (2017).

and CR D. For this purpose, we predict the QCD m_T distribution in CR D by taking the 2D data distribution of m_T vs. $p_T(\tau_h^{fake})$ from CR C, after subtracting non-QCD contributions derived from MC, and reweighting that distribution by the p_T-dependent tight-to-loose ratios derived from the W CR.

Figure 46 (left) represents the derived yields of QCD multijet events in CR C, where the histogram has been rebinned with the same binning as the SR for m_T. Figure 46 (right) is similar to Figure 46 (left), except the bin-by-bin yields have been reweighted with the $p_T(\tau_h)$-dependent tight-to-loose ratios obtained in the W CR. If we examine a particular m_T bin of the 2D histogram in Figure 46 (right), we see that the τ_h can take on various p_T values. To determine the 1D m_T distribution for QCD in CR D, we take the projection of the 2D histogram in Figure 46 (right) onto the y-axis.

Figure 47 gives the prediction for QCD multijet events in CR D using the tight-to-loose ratios obtained from the Z CR (left) and from the W region (right) (for comparison and validation).

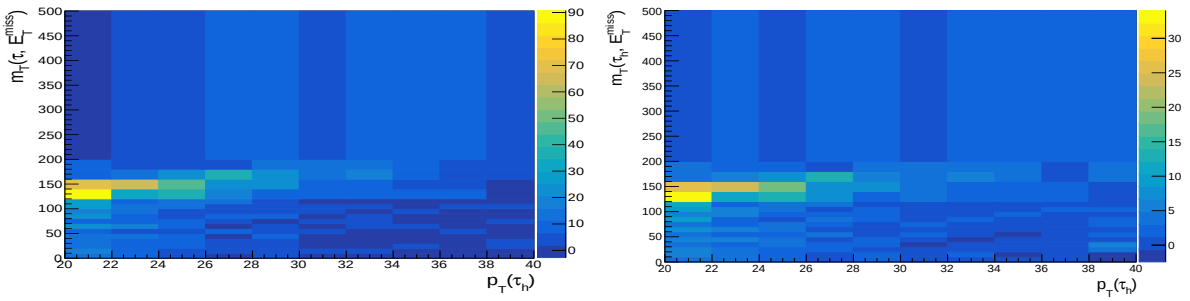

Figure 46: 2D histogram representing the predicted yield of QCD multijet events in CR C (left) and CR D (right), as a function of m_T and $p_T(\tau_h)$.

We perform this check with both 2016 and 2017 data and MC. Thus, in Figure 47, plots in the first row correspond to the 2016 iteration, and plots in the second row correspond to the 2017 iteration. Note there is very good agreement between the data and the expected background. The good agreement, in both shape and normalization, gives confidence in the method to estimate QCD multijet events. The relative difference between the data and background prediction in this CR will be used to assign systematics on the shape and the overall normalization.

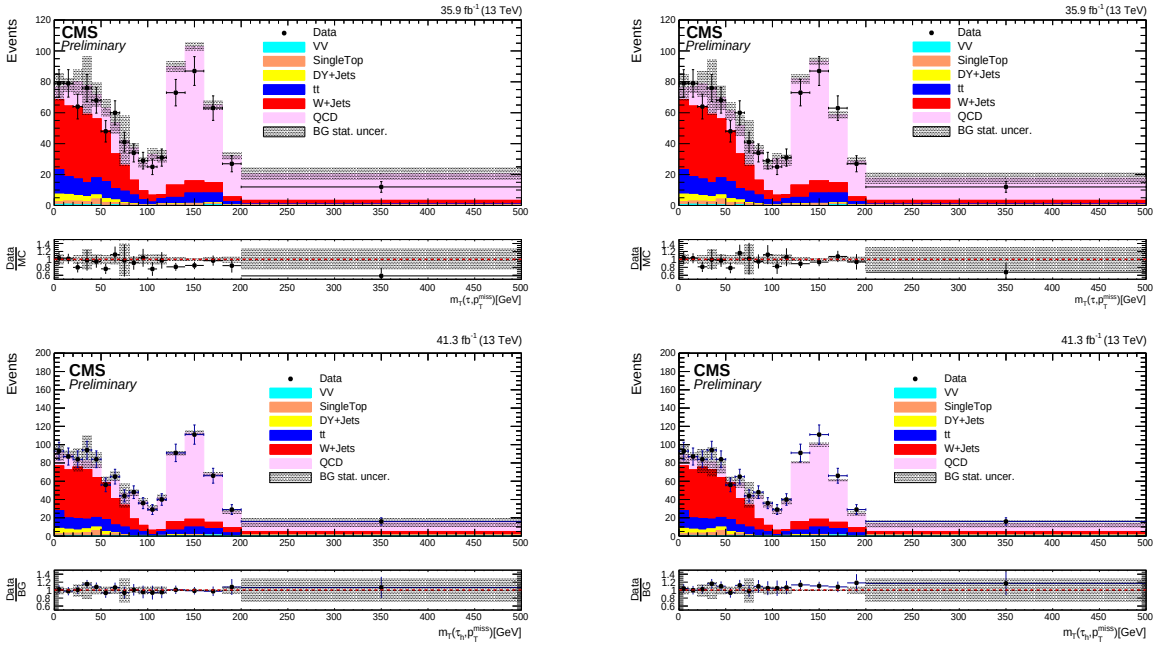

Figure 47: Final yield of QCD estimate in CR D. Response 2D histogram from CR C is re-weighted for Z CR (left) and W CR (right) for 2016 (top row) and 2017 (bottom row).

As a final closure test for the use of the $p_T(\tau_h)$-dependent tight-to-loose ratios as a methodology to estimate the QCD multijet background, we provide a similar closure test as described above, except with only MC. We use the 2D histogram of m_T vs. $p_T(\tau_h)$ obtained from the QCD MC in

CR C and predict the yield of the QCD background in CR D by reweighting this histogram by the tight-to-loose ratios (which are also derived from MC). Figure 48 shows a result that is consistent within the statistical uncertainty for 2016 (left) and 2017 (right) data and MC, and it represents the agreement between data and MC using only MC information. With both the data and MC closure tests showing good agreement, we establish that the fully-data-driven estimation is a good methodology for estimating the QCD multijet contribution to the SR.

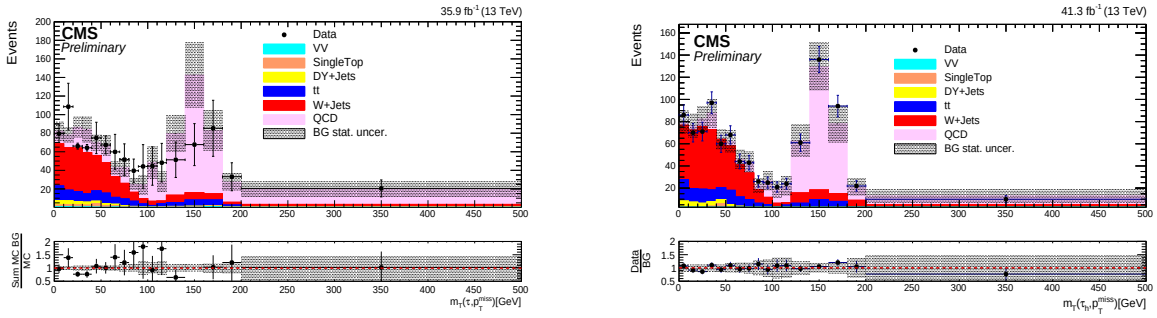

Figure 48: CR D yield from QCD prediction using the shape extracted from MC in CR C in 2016 (2017), left (right).

To summarize, the 2D histogram of m_T vs. $p_T(\tau_h)$ for QCD will be extracted from CR B (SR-like but inverted τ_h isolation requirement) and will be reweighted using the $p_T(\tau_h)$-dependent tight-to-loose ratios obtained in the W CR. With this full data-driven background estimation methodology, the QCD multijet contribution in the SR would be: $N^{SR}_{Data-Driven} = 1530.4 \pm 18.6$ in 2016 data and MC for Z weights (and $N^{SR}_{Data-Driven} = 1359.9 \pm 13.4$ for W weights), where the corresponding uncertainties are only statistical. The formula to calculate the number of events in the SR is given by:

$$N^{SR}_{Data-Driven} = N^{CRB}_{Data-NonQCD}(p_T(\tau_h)) \times Ratio^{Tight}_{Loose}(p_T(\tau_h)). \quad \text{(XIII.2)}$$

Chapter XIV

Systematics

The following systematics have been considered:

- Parton Distribution Functions (PDF): The systematic effect due to imprecise knowledge of the parton distribution functions (i.e. the fraction of the total momentum of the nucleon taken by each of the partons) is determined by comparing CTEQ6.6L, MSTW08, and NNPDF10 PDF sets [32, 31, 51] with those from the default PDF set. The maximal deviation from the central value is used to assign the overall systematic due to PDFs. We obtain a value of 6.0%.

- Initial State Radiation (ISR): We apply uncertainty to the ISR weights bin-by-bin for the boost of the Z boson. The relative uncertainties are summarized in Table 35. The first step in investigating the uncertainty associated with the ISR weights was to determine the statistical uncertainty bin-by-bin for the boost of the Z-boson. These statistical uncertainties are described previously in the section on Z-Boost. One can see that the uncertainty on the weight (combination of Poisson error in data and the statistical uncertainty on MC) varies with the p_T of the Z-boson and ranges from 1-3% at low Z-p_T to $\sim$ 11% at high Z-p_T. However, these values are only the uncertainty on the weights, and they do not necessarily reflect how varying the weights by these uncertainties affects the signal and background yields and shapes in a given region (CR or SR).

Table 35: Event Weight Uncertainties by Z-Boost

Z-Boost Bin	Weight Uncertainty
1: 0-50 GeV	2.23%
2: 50-100 GeV	1.21%
3: 100-150 GeV	1.09%
4: 150-200 GeV	1.18%
5: 200-300 GeV	1.29%
6: 300-400 GeV	1.95%
7: 400-600 GeV	3.23%
8: 600+ GeV	14.45%

- To determine the effect of these weight uncertainties on the background and signal yields and distributions, we then varied the weights by $\pm 1\sigma$, where σ is the actual uncertainty on the weight, to examine the possible systematic effect on the m_T distribution in the SR for

Z+jets and W+jets. These are the plots in Figure 49, where red is the m_T distribution obtained with the central value of the weights, green is the variation of the weights down by 1σ (i.e. m_T distribution obtained when we use weight = nominalWeight - 1σ), and blue is for the variation of the weights up by 1σ (i.e. m_T distribution obtained when we use weight = nominalWeight + 1σ). From these plots in Figure 49, we derive a shape-based uncertainty for the ISR weights, as is indicated in Table 37. The shape-based uncertainty is a few percent at low values of m_T and as much as 15% at high m_T values. These uncertainties are applied to Z+jets, W+jets, and signal, as stated in Table 37. We obtain normalization uncertainties of $\sim 1\%$.

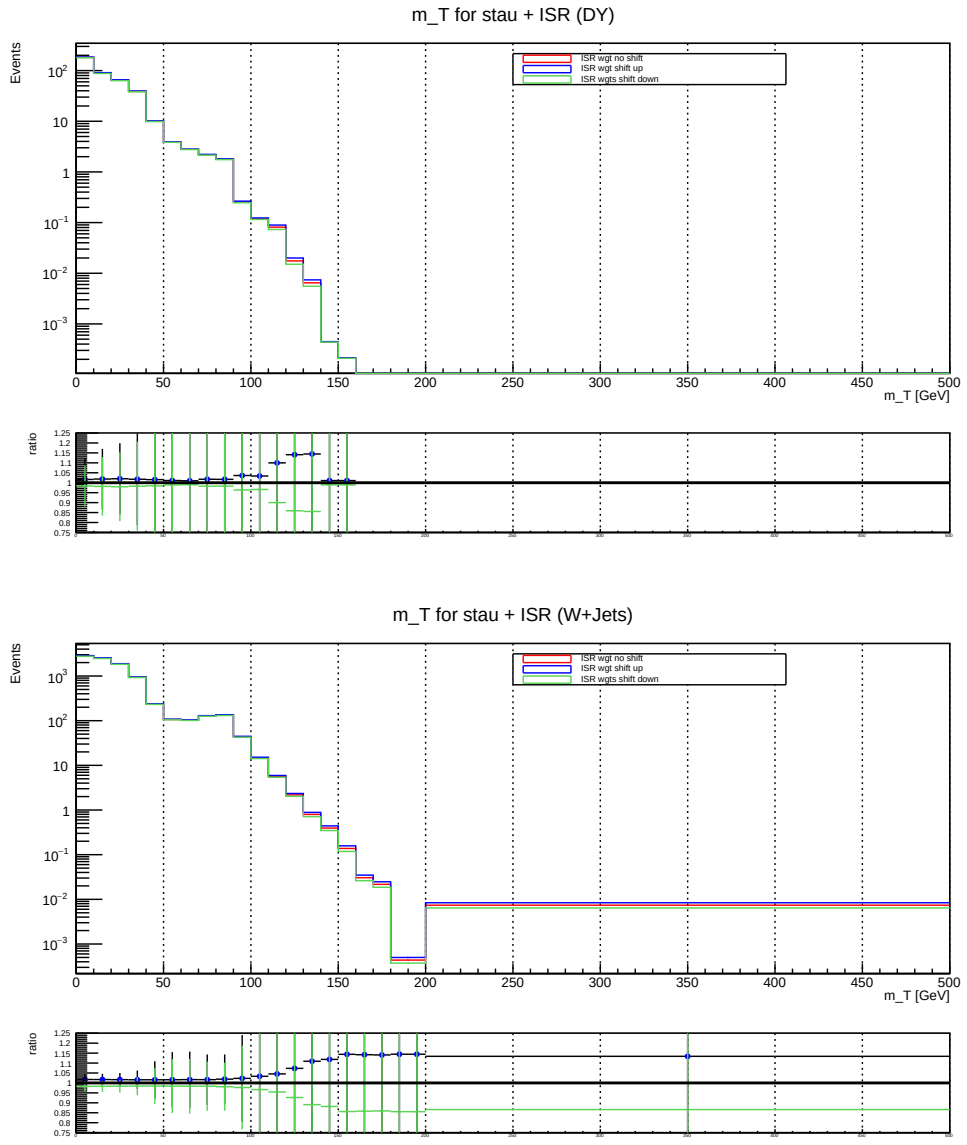

Figure 49: $m_T(\tau_h, E_T^{miss})$ for Z+jets (top) and for W+jets (bottom) in the SR.

- Luminosity: We consider a 2.5% (2.3%) uncertainty on the measured luminosity for 2016 (2017) [21, 22].

- Trigger, Reconstruction, and Selection: An overall uncertainty is applied for the trigger uncertainties. We consider a 3.0% uncertainty due to the trigger efficiency, which is a result of a fit to the trigger efficiency turn-on curve. The 3.0% is the uncertainty on the fit at the trigger efficiency plateau (where the trigger is maximally efficient), which is driven by the statistics in data. Scale factors for τ_h identification are taken from the τ lepton POG and obtained using a fit of data in a $Z(\rightarrow \tau\tau)$ enhanced region and fixing the cross section to that measured using $ee/\mu\mu$. The τ_h identification uncertainty is 6%. The uncertainty for muons and electrons is up to $\sim 2\%$, each depending on η of p_T, and only contributing through the extra-lepton vetoes.

- b-tagging Efficiency: We consider the uncertainty on the mis-tag rate as measured by the b-tagging POG ($\sim 10\%$) [5]. For the case of our signal, the systematic uncertainty on the requirement of 0 jets mis-tagged as b-jets is determined by propagating the 10% uncertainty on the mis-tag rate through the following equation (which represents the signal efficiency for requiring 0 jets mis-tagged as b-jets):

$$\epsilon^{\text{NBtag}<1} = 1 - \sum_{n=1} P(n) \cdot \sum_{m=1}^{n} C(n,m) \cdot f^m \cdot (1-f)^{n-m} \tag{XIV.1}$$

  where $P(n)$ is the probability to obtain n additional jets (non-tau and non-lepton) in the event, $C(n,m)$ the combinatorics of n *choose* m, and f the mis-tag rate. The probability to obtain at least one additional jet in the event is $\sim 10\%$. Therefore, based on the above equation, the mis-tag rate and uncertainty, and the probability to obtain at least one additional jet, we calculate a systematic effect of $\sim 1\%$ on our signal due to the mis-tag rate. The b-tagging/mis-tagging systematics are considered 100% correlated across MC-based backgrounds with similar composition (e.g. W+jets and Z+jets where there are typically no real b-jets), but completely uncorrelated to backgrounds that have different composition (e.g. $t\bar{t}$ vs. Z+jets). We note the b-tagging uncertainties are evaluated on a per-event basis, following the POG recommendation and as referenced in the object reconstruction section of this thesis.

- Electron Energy Scale: We consider the effect on the signal acceptance of a 1% (2.5%) shift on the electron energy scale in the barrel (endcap) region. The systematic effect is $< 1\%$.

- Muon Momentum Scale: We consider the effect on the signal acceptance of a 1% momentum scale uncertainty on the muon momentum. The systematic effect is $< 1\%$.

- Tau Energy Scale: We consider the effect of the 5% tau lepton energy scale uncertainty measured by the tau lepton POG on the signal acceptance. The tau lepton 4-momentum is scaled by a factor of $k = 0.95$ or 1.05 ($p_{smeared} = k \cdot p_{default}$) and variables are recalculated

using $p_{smeared}$. We use the varied m_T templates as "shape variations" due to tau energy scale.

- Jet Energy Scale: We consider the effect of a 2-5% jet energy scale uncertainty on the signal acceptance (depending on the η and p_T of the considered jets as prescribed by the JetMET POG). For example, for a 5% jet energy scale uncertainty, the jet 4-momentum is scaled by a factor of $k = 0.95$ or 1.05 ($p_{smeared} = k \cdot p_{default}$) and variables are recalculated using $p_{smeared}$. We use the varied m_T templates as "shape variations" due to jet energy scale.

- MET: The uncertainty on the E_T^{miss} for our signal process is driven by the jet energy scale (non-tau jets), tau energy scale, light lepton energy/momentum scale, and unclustered energy. The systematic effect from E_T^{miss} due to each of the energy scales is included based on the descriptions above. We find that a 10% uncertainty on the unclustered energy results in at most a 0.5% fluctuation on the signal acceptance.

- FastSim MET Correction (for signal): To account for the worse resolution of FastSim MET, the signal yields are corrected as $S_{\text{nominal}} = (S_{\text{gen-MET}} + S_{\text{reco-MET}})/2$, where $S_{\text{gen-MET}}$ is the yield obtained when we use gen-MET in our cuts, while $S_{\text{reco-MET}}$ is the yield obtained when we use reco-MET in our cuts [8]. The uncertainty on the signal yield is equal to one-half the difference between the two acceptances. This amounts to about a 5.5% systematic uncertainty on the signal significance. This is shown in Figure 50. This is a shape-based systematic uncertainty in terms of m_T. This shows that while the difference in shape is small at low m_T, it can be about 10% at high m_T.

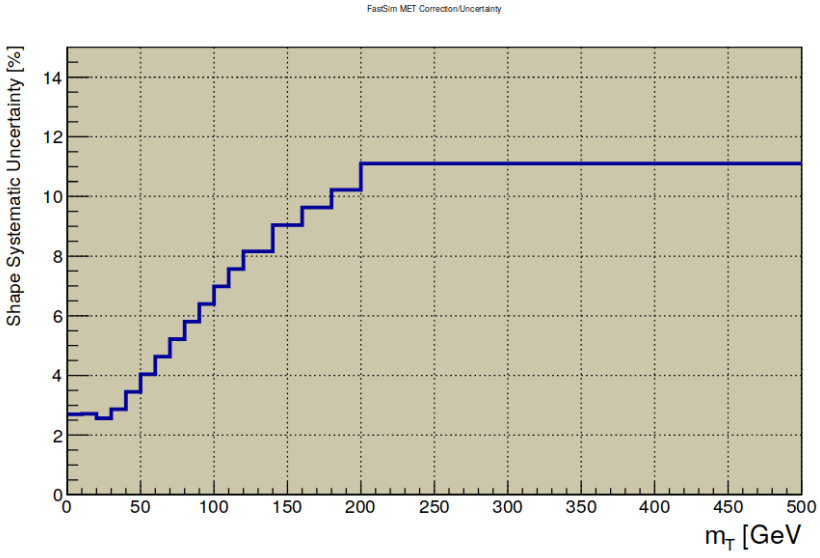

Figure 50: FastSim MET correction for signal

- Additional shape systematics: We also assign systematic uncertainties on the shapes based on the level of agreement between data and MC distributions in the control samples. We take the data/MC ratios and fit them with a polynomial of degree-1. The deviation from the "flat line" fit, as a function of mass or m_T, is treated as a systematic uncertainty on the shape. This results in up to $\sim 20\%$ systematic uncertainty in a given bin. The data-to-MC ratios, as a function of m_T, and the fits are shown in Figure 51.

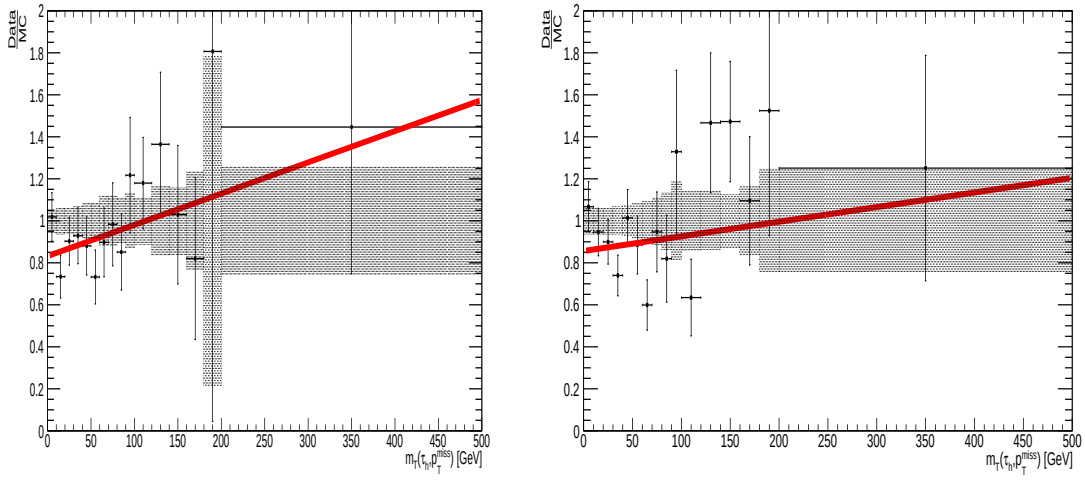

Figure 51: Data/MC ratios, as a function of m_T, in the $Z(\rightarrow \tau\tau)$ and $t\bar{t}$ control regions that are used to assign shape-based systematics.

- Non-QCD Background Closure/Normalization: Additionally, since the background estimation involves correcting the MC predictions using scale factors obtained in data CRs, the deviation in the scale factors from unity is assigned as uncertainty in the background predictions.

- QCD Background Closure/Normalization: As previously described, we use the $p_T(\tau_h)$-dependent tight-to-loose ratios, measured using W+jets events, to estimate the QCD m_T distribution in the SR. This method closes well in both data and MC. However, we also measure the $p_T(\tau_h)$-dependent tight-to-loose ratios using Z+jets events as a way to assign systematic uncertainty on the QCD background prediction. The $p_T(\tau_h)$-dependent tight-to-loose ratios from both CRs are shown in Table 36, along with the relative difference of the ratios between both regions. The third column represents the $p_T(\tau_h)$-dependent systematic uncertainty, which is propagated to the m_T distribution, resulting in bin-by-bin uncertainties (i.e. shape uncertainty).

$p_T(\tau_h)$ [GeV]	W+jets $Ratio^{Tight}_{Loose}$	Z+jets $Ratio^{Tight}_{Loose}$	R. Difference
$[20, 22]$	0.361 ± 0.039	0.378 ± 0.268	4.5%
$[22, 24]$	0.345 ± 0.033	0.371 ± 0.227	7.0%
$[24, 26]$	0.329 ± 0.028	0.363 ± 0.191	9.3%
$[26, 28]$	0.313 ± 0.024	0.355 ± 0.163	11.8%
$[28, 30]$	0.298 ± 0.022	0.347 ± 0.146	14.1%
$[30, 32]$	0.282 ± 0.022	0.339 ± 0.146	16.8%
$[32, 34]$	0.266 ± 0.024	0.332 ± 0.163	19.8%
$[34, 36]$	0.250 ± 0.028	0.324 ± 0.191	22.8%
$[36, 38]$	0.234 ± 0.033	0.316 ± 0.227	25.9%
$[38, 40]$	0.218 ± 0.039	0.308 ± 0.268	29.2%

Table 36: Values of the tight-to-loose ratio for each bin of $p_T(\tau_h)$ in the W+jets and Z+jets regions and their relative differences.

Table 37: Systematics values given in percent- "s" indicates "shape" uncertainties.

Source	W	DY	$t\bar{t}$	VV	QCD	Signal
Lumi	2.5	2.5	2.5	2.5	–	2.5
μ ID	< 1	< 1	< 1	< 1	–	1
e ID	< 1	< 1	< 1	< 1	–	1
τ_h ID	6	8	9	9	–	9
Trigger	3	3	3	3	–	3
b ID	2	2	7	2	–	2
JES	s	s	s	s	–	s
TES	s	s	s	s	–	s
MMS	< 1	< 1	< 1	< 1	–	< 1
EES	< 1	< 1	< 1	< 1	–	< 1
Pileup	5.0	5.0	5.0	5.0	–	5.0
PDF	4.8	4.2	4.2	3.5	–	6.0
bin-by-bin stat.	s	s	s	s	–	s
Closure+Norm.	2	8	6	–	23	–
ISR	s	s	–	–	–	s
Prefiring	–	–	–	–	–	s
$Ratio^{Tight}_{Loose}$	–	–	–	–	s	–
Gen. Scale	1	1	3.5	–	–	2
Fast Sim.	–	–	–	–	–	s

Chapter XV

Results and Conclusions

The m_T distribution in the SR is shown in Figure 52 for the 2016 (left) and 2017 (right) iterations of the analysis. The dominant backgrounds are W+jets, followed by $t\bar{t}$. To illustrate the sensitivity of this search, the results are presented in the context of the R-parity conserving Minimal Supersymmetric Standard Model (MSSM) and considering cases such as those shown in Figure 53 for pure electroweak production of charginos and neutralinos with an ISR jet. In the SUSY nomenclature, $\tilde{\chi}_2^0$ is the second lightest neutralino, and $\tilde{\chi}_1^\pm$ is the lightest chargino. SUSY models with Bino-like (Z-like) $\tilde{\chi}_1^0$ and Wino-like (W-like) $\tilde{\chi}_2^0$ and $\tilde{\chi}_1^\pm$ are considered. Since the latter two gauginos belong to the same gauge group multiplet, we set $m(\tilde{\chi}_2^0) = m(\tilde{\chi}_1^\pm)$ and present results as a function of this common mass and the LSP mass ($m(\tilde{\chi}_1^0)$). The mass $m_{\tilde{\tau}_1}$ of the intermediate stau is parameterized in terms of a variable $x_{\tilde{\tau}_1}$ as

$$m_{\tilde{l}} = m_{\tilde{\chi}_1^0} + x_{\tilde{\tau}_1}(m_{\tilde{\chi}_1^\pm} - m_{\tilde{\chi}_1^0}) \tag{XV.1}$$

where $0 < x_{\tilde{\tau}_1} < 1$. We present results for $\Delta m = m_{\tilde{\chi}_1^\pm} - m_{\tilde{\chi}_1^0} = 50$ GeV and $x_{\tilde{\tau}_1} = 0.5$.

Theoretically, $\tilde{\tau}$s can be produced directly in pairs or through cascade decays of $\tilde{\chi}_1^\pm$, and the next-to-lightest neutralino, $\tilde{\chi}_2^0$, in processes such as $\tilde{\chi}_1^\pm\tilde{\chi}_1^\mp \to \tilde{\tau}\tilde{\tau}\nu_\tau\nu_\tau$ and $\tilde{\chi}_1^\pm\tilde{\chi}_2^0 \to \tilde{\tau}\nu_\tau\tilde{\tau}\tau$. Therefore, the signal samples were produced "inclusively" in stau production to contain: (1) direct production of $\tilde{\tau}$ pairs and an ISR jet, and (2) additional production of of $\tilde{\tau}$ events through cascade decays of $\tilde{\chi}_1^\pm$ or $\tilde{\chi}_2^0$, including $\tilde{\chi}_1^\pm\tilde{\chi}_1^\mp$ and $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$ processes. Therefore, we note that Figure 53 is only one of the representative diagrams contributing to the total signal cross section and interpretation.

The CL$_S$ [23, 29, 36] criterion is used is used (full LHC-style CL$_S$) to calculate upper limits on the cross sections for electroweak pair production of charginos and neutralinos with one associated jet as a function of $m(\tilde{\chi}_1^\pm) = m(\tilde{\chi}_2^0)$. The results are interpreted by assuming $Br(\tilde{\chi}_2^0 \to \tau\tilde{\tau} \to \tau\tau\tilde{\chi}_1^0) = 100\%$, $Br(\tilde{\chi}_1^\pm \to \nu\tilde{\tau} \to \nu\tau\tilde{\chi}_1^0) = 100\%$. These are the assumptions about the branching fractions for our simplified model. Figure 54 shows the signal acceptance as a function of $m(\tilde{\chi}_1^\pm)$ for production of $\tilde{\tau}$s through cascading decays of $\tilde{\chi}_1^\pm$ and $\tilde{\chi}_2^0$. Figure 55 shows the expected and observed limits as well as the theoretical cross section as functions of $m(\tilde{\chi}) = m(\tilde{\chi}_1^\pm) = m(\tilde{\chi}_2^0)$. The bands on the expected limits represent the two standard deviations obtained using a large sample of pseudo-experiments based on the background-only hypothesis for each bin of the transverse mass distribution. The upper limit on $m(\tilde{\chi})$ corresponds to the point where the observed limit crosses the theoretical line. We expect to exclude $\tilde{\chi}_2^0/\tilde{\chi}_1^\pm$ with masses below 290 GeV for $m(\tilde{\chi}_1^\pm) - m(\tilde{\tau}) = 25$ GeV with combined 2016 and 2017 data.

Explicit bin-by-bin yields for each of the plots in Figure 52 are listed in Table 56 for 2016 and Table 57 for 2017. The CL$_S$ criterion [23, 29, 36] is used again (full LHC-style CL$_S$) to calculate

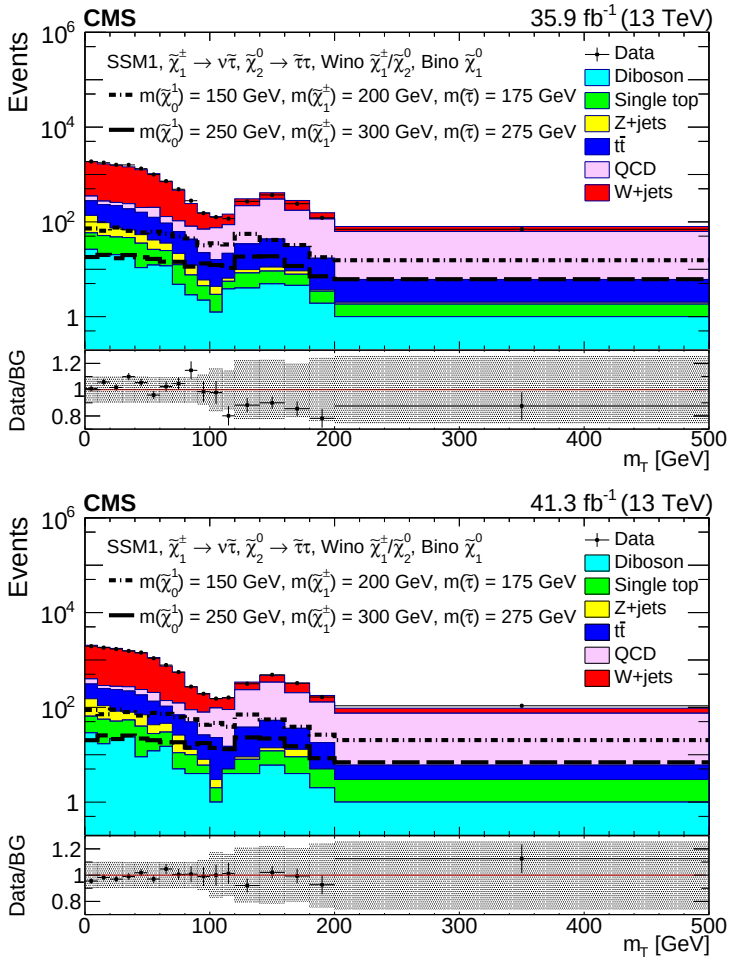

Figure 52: Expected signal and background m_T distributions in the search region with 2016 data (top) & 2017 data (bottom) iterations.

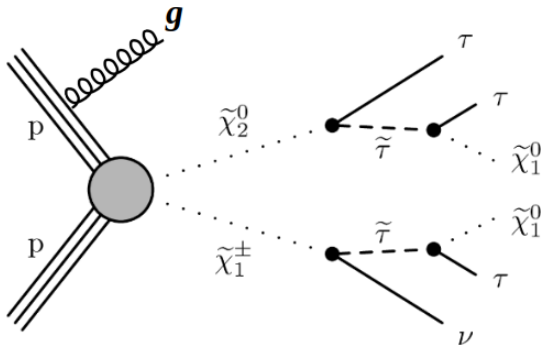

Figure 53: ISR topology for this study.

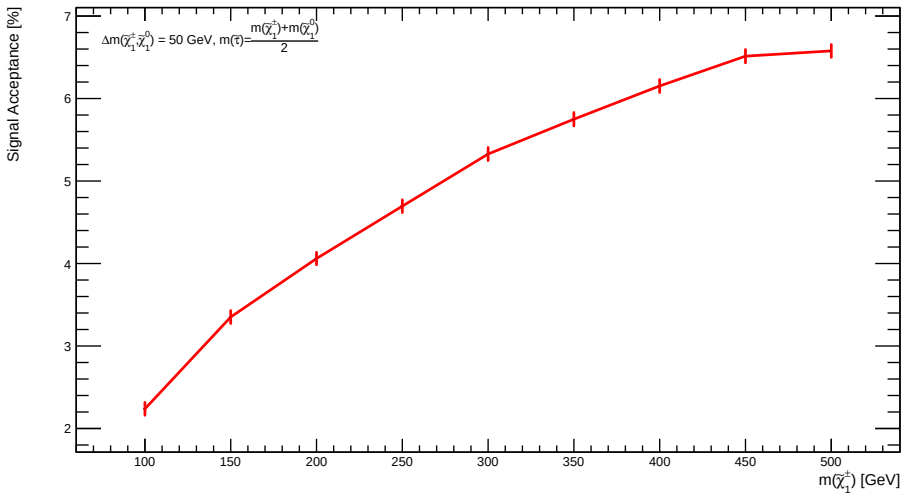

Figure 54: Signal acceptance for cascading production of $\tilde{\tau}$s as a function of $m(\tilde{\chi}_1^{\pm})$, where $\Delta m(\tilde{\chi}_1^{\pm},\tilde{\chi}_1^0) = 50$ GeV and $m(\tilde{\tau}) = \frac{m(\tilde{\chi}_1^{\pm})+m(\tilde{\chi}_1^0)}{2}$.

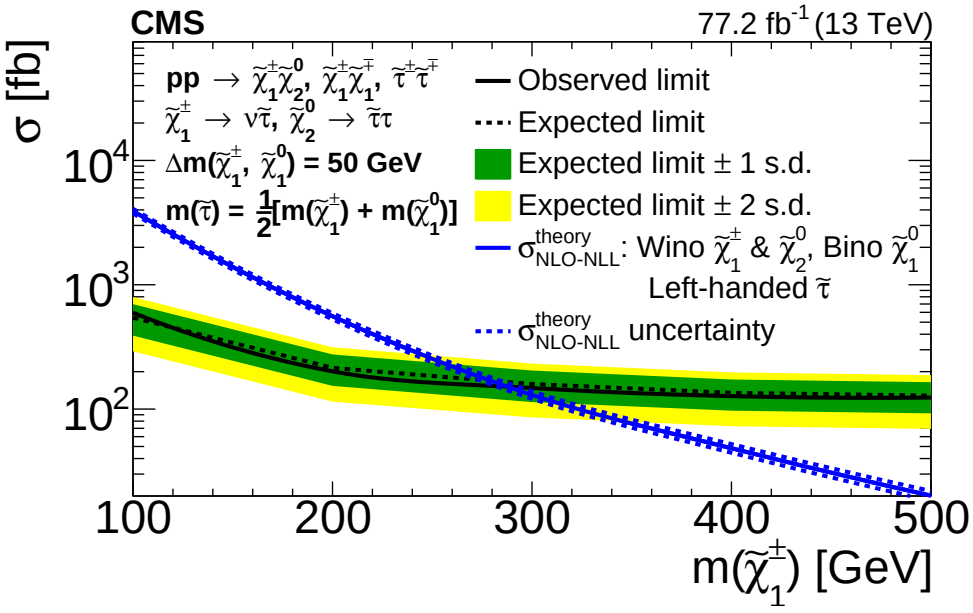

Figure 55: The 95% confidence level upper limits on the production cross sections ($\sigma_{95\%\,CL}$) as a function of $m(\tilde{\chi}_1^{\pm})$. The solid blue line shows the theoretical cross section, and the dashed blue line its uncertainty. The observed limit is shown with the solid black line, while the expected limit is shown with the dashed black line. The green (yellow) band corresponds to the one (two) standard deviation range about the central value of the expected limit.

upper limits on the cross sections for direct $\tilde{\tau}$ production with one associated jet as a function of $m(\tilde{\tau})$. Figure 58 shows the signal acceptance as a function of $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$ for direct $\tilde{\tau}$ production

with various specified values of $m(\tilde{\tau})$.

SUS-19-002 2016 Yields

m_T	W+Jets	QCD	$t\bar{t}$	Drell-Yan	Single-top	Diboson	$m_{\tilde{\chi}_1^\pm} = 200$	$m_{\tilde{\chi}_1^\pm} = 300$	ΣBG	Data
0-10	1510 ± 20	73 ± 5	141 ± 8	79 ± 3	32 ± 3	26 ± 3	73 ± 4	18 ± 1	1870 ± 30	1881
10-20	1400 ± 20	44 ± 4	134 ± 8	45 ± 2	33 ± 3	19 ± 3	65 ± 4	20 ± 1	1680 ± 30	1775
20-30	1310 ± 20	37 ± 4	146 ± 8	26 ± 2	28 ± 2	20 ± 3	76 ± 4	17 ± 1	1570 ± 20	1598
30-40	1200 ± 20	54 ± 5	120 ± 7	21 ± 1	28 ± 2	19 ± 3	65 ± 4	20 ± 1	1450 ± 20	1594
40-50	1060 ± 20	42 ± 4	105 ± 7	18 ± 1	25 ± 2	11 ± 2	60 ± 4	18 ± 1	1260 ± 20	1331
50-60	840 ± 20	78 ± 5	83 ± 6	15 ± 1	14 ± 2	12 ± 2	62 ± 4	17 ± 1	1040 ± 20	999
60-70	580 ± 20	36 ± 4	58 ± 5	11 ± 1	13 ± 2	12 ± 2	56 ± 4	14 ± 1	700 ± 20	724
70-80	360 ± 10	43 ± 4	40 ± 4	9 ± 1	8 ± 1	5 ± 1	50 ± 4	14 ± 1	470 ± 10	488
80-90	163 ± 8	38 ± 4	31 ± 4	4.7 ± 0.8	4.7 ± 0.8	2.9 ± 0.9	45 ± 3	11 ± 1	240 ± 10	280
90-100	85 ± 6	48 ± 4	16 ± 3	1.5 ± 0.3	2.3 ± 0.6	2.2 ± 0.9	32 ± 3	13 ± 1	155 ± 8	153
100-110	54 ± 5	59 ± 5	12 ± 2	1.4 ± 0.5	1.7 ± 0.5	1.2 ± 0.7	35 ± 3	12 ± 1	129 ± 7	126
110-120	59 ± 6	66 ± 5	15 ± 3	0.9 ± 0.3	1.7 ± 0.6	4 ± 1	33 ± 3	11 ± 1	146 ± 8	117
120-140	88 ± 7	185 ± 8	24 ± 3	1.5 ± 0.4	4.3 ± 1.0	4 ± 1	56 ± 4	19 ± 1	300 ± 10	271
140-160	111 ± 7	258 ± 10	32 ± 4	2.0 ± 0.7	4 ± 1	5 ± 1	42 ± 3	19 ± 1	410 ± 10	371
160-180	100 ± 40	145 ± 7	21 ± 3	1.4 ± 0.6	3 ± 1	5 ± 1	32 ± 3	12 ± 1	280 ± 40	242
180-200	41 ± 4	98 ± 6	13 ± 2	0.27 ± 0.07	1.5 ± 0.7	1.9 ± 0.9	18 ± 2	7 ± 1	156 ± 8	122
200+	17 ± 2	57 ± 5	4 ± 2	0.18 ± 0.04	0.8 ± 0.4	1.0 ± 0.5	16 ± 2	6 ± 1	80 ± 5	70

Figure 56: 2016 SR yields for BG and signal, as a function of m_T bin. Two benchmark signal samples are considered: (i) $m(\tilde{\chi}_1^\pm) = 200$ GeV, $m(\tilde{\chi}_1^0) = 150$ GeV; (ii) $m(\tilde{\chi}_1^\pm) = 300$ GeV, $m(\tilde{\chi}_1^0) = 250$ GeV.

SUS-19-002 2017 Yields

m_T	W+Jets	QCD	$t\bar{t}$	Drell-Yan	Single-top	Diboson	$m_{\tilde{\chi}_1^\pm} = 200$	$m_{\tilde{\chi}_1^\pm} = 300$	ΣBG	Data
0-10	1644 ± 24	89 ± 6	158 ± 9	87 ± 3	36 ± 3	29 ± 4	88 ± 5	20 ± 4	2043 ± 27	1952
10-20	1565 ± 26	39 ± 4	141 ± 8	53 ± 3	39 ± 3	17 ± 2	72 ± 5	26 ± 4	1854 ± 28	1821
20-30	1476 ± 25	40 ± 4	154 ± 8	32 ± 2	29 ± 2	22 ± 3	90 ± 5	21 ± 4	1753 ± 27	1700
30-40	1292 ± 22	52 ± 4	141 ± 9	25 ± 2	29 ± 2	25 ± 4	80 ± 5	25 ± 4	1564 ± 25	1548
40-50	1170 ± 22	56 ± 5	112 ± 7	30 ± 2	31 ± 3	9 ± 2	67 ± 4	22 ± 4	1408 ± 24	1434
50-60	915 ± 20	72 ± 5	91 ± 7	15 ± 1	16 ± 2	12 ± 2	76 ± 5	20 ± 4	1121 ± 22	1088
60-70	607 ± 16	34 ± 4	59 ± 5	16 ± 1	15 ± 2	15 ± 3	73 ± 5	18 ± 4	746 ± 18	781
70-80	429 ± 15	54 ± 5	43 ± 5	15 ± 2	6 ± 1	5 ± 2	62 ± 5	18 ± 4	552 ± 17	555
80-90	182 ± 9	40 ± 4	36 ± 4	3 ± 0.5	6 ± 1	4 ± 1	56 ± 4	14 ± 3	271 ± 11	273
90-100	116 ± 8	53 ± 5	18 ± 3	2 ± 0.4	2 ± 1	4 ± 2	43 ± 4	18 ± 3	195 ± 10	193
100-110	56 ± 5	74 ± 5	20 ± 4	1 ± 0.4	1 ± 0.3	1 ± 0.6	47 ± 4	14 ± 3	153 ± 8	153
110-120	69 ± 6	75 ± 5	10 ± 2	0.4 ± 0.1	1 ± 0.4	4 ± 1	40 ± 3	13 ± 3	159 ± 8	161
120-140	107 ± 8	197 ± 9	29 ± 4	1 ± 0.2	4 ± 1	4 ± 1	71 ± 5	23 ± 4	342 ± 13	315
140-160	137 ± 9	286 ± 10	39 ± 5	2 ± 0.7	6 ± 2	6 ± 2	56 ± 4	22 ± 3	476 ± 15	486
160-180	123 ± 10	167 ± 8	23 ± 3	4 ± 2	5 ± 2	4 ± 1	39 ± 4	15 ± 3	326 ± 13	324
180-200	49 ± 5	111 ± 7	13 ± 2	0.3 ± 0.1	3 ± 1	2 ± 1	27 ± 3	8 ± 2	178 ± 9	165
200+	21 ± 3	69 ± 5	3 ± 1	0.3 ± 0.1	2 ± 1	1 ± 0.5	20 ± 3	7 ± 2	96 ± 6	108

Figure 57: 2017 SR yields for BG and signal, as a function of m_T bin. Two benchmark signal samples are considered: (i) $m(\tilde{\chi}_1^\pm) = 200$ GeV, $m(\tilde{\chi}_1^0) = 150$ GeV; (ii) $m(\tilde{\chi}_1^\pm) = 300$ GeV, $m(\tilde{\chi}_1^0) = 250$ GeV.

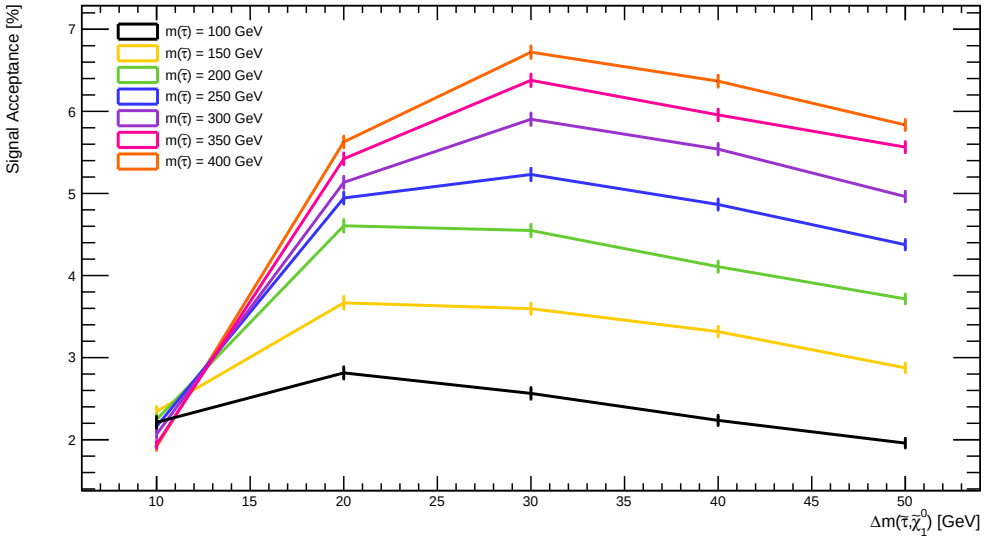

Figure 58: Signal acceptance for direct $\tilde{\tau}$ production of as a function of $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$ and specified values of $m(\tilde{\tau})$.

In Figure 59, we show the observed upper limit on the signal cross-section as a function of $m(\tilde{\tau})$ and $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$ for direct $\tilde{\tau}$ production. The extremely small direct $\tilde{\tau}$ production cross sections make these scenarios very challenging, especially when $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0) < 50$ GeV. This plot of combined 2016 and 2017 data shows that we do not expect to exclude $\tilde{\tau}$ masses at this time with the current luminosity. However, we do note improved sensitivity in this analysis compared to previous non-ISR searches. For example, for a $\tilde{\tau}$ mass of 100 GeV and a $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0) = 30$ GeV, the observed limit is 12 times the theoretical cross section. This is to be compared with the most sensitive non-ISR search from the ATLAS Collaboration [10], which shows $\approx \times 4$ improvement in the cross section upper limit for the scenario with $m(\tilde{\tau}) = 150$ GeV and $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0) = 50$ GeV.

In order to show which m_T bins provide the highest sensitivity for a given signal hypothesis, we have produced Table 60 to show the 95% C.L. upper limits on the signal cross section (for the electroweakino model) for each m_T bin and for two benchmark signal samples. The first is for $m(\tilde{\chi}_1^{\pm}) = 200$ GeV, and the second is for $m(\tilde{\chi}_1^{\pm}) = 300$ GeV.

The main result of this paper is that the stau-LSP coannihilation region with $\Delta m < 50$ GeV, where experimental sensitivity is limited from current searches performed at CMS and the LHC, can be probed using a search strategy of one soft hadronically decaying tau lepton and large missing transverse energy recoiling against a high-p_T jet from initial state radiation. These regions of SUSY also play a decisive role in thermal Bino DM cosmology models which require stau-LSP coannihilation to obtain the correct relic DM density observed today. A major highlight of the proposed search strategy is the ability to select low-p_T hadronic tau decays, facilitated by the use of a E_T^{miss} trigger from the boost effect of the high-p_T ISR jet, in order to maximize signal

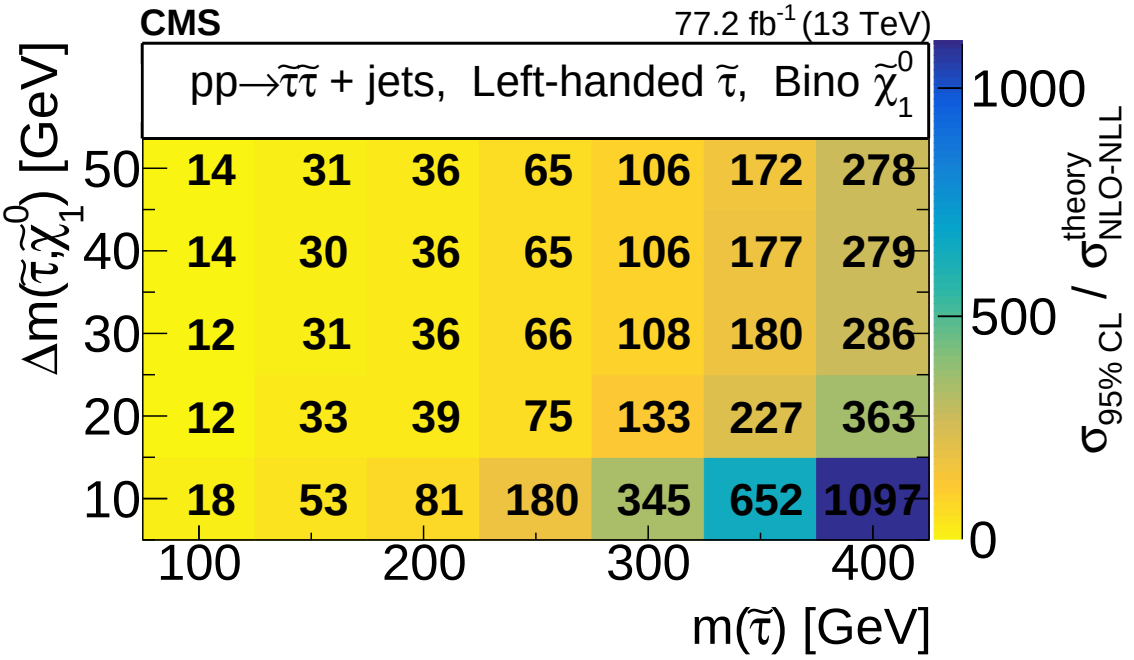

Figure 59: 95% C.L. upper limits on the direct $\tilde{\tau}$ pair production signal cross sections, as a function of $m(\tilde{\tau})$ and $\Delta m(\tilde{\tau}, \tilde{\chi}_1^0)$.

acceptance in these compressed scenarios while simultaneously providing large reduction against SM backgrounds. We find that for $m(\tilde{\chi}_1^{\pm}) - m(\tilde{\tau}) = 25$ GeV, gaugino masses up to 290 GeV can be excluded with approximately 77 fb^{-1} of 13 TeV data from the LHC. We emphasize that the experimental constraints for the SUSY parameter with $m(\tilde{\chi}_1^{\pm}) - m(\tilde{\tau}) = 25$ GeV using non-ISR searches have not exceeded those of the LEP experiments, and thus this new search nicely complements the current analyses performed at CMS.

Information for Reinterpretation of Results

The results of this analysis have the potential to be reinterpreted in the context of other models that incorporate coannihilation. Coannihilation, as discussed, is a potential bridge to the gap between particle physics and cosmology. Currently, there exist a variety of models that have groups of particles with particular masses and couplings that lead to a DM relic density commensurate with that from astronomy. For the analysis described in this thesis, the chosen model is SUSY, but in no way is the physics constrained to just SUSY. For instance, there exist leptoquark-portal DM models where the mass difference between the LQ, DM, and co-annihilation partner is small, thus resulting in a final state with a soft τ lepton and E_T^{miss} [40]. Signal acceptance is mostly entirely dependent on the mass gap and the mass of the produced SUSY particles. For this reason, the results of this analysis can be re-interpreted in other coannihilation models such as this by re-scaling

m_T	σ_{95CL} UL [fb] (2016)	σ_{95CL} UL [fb] (2017)
0-10	3431.4511	3109.0829
10-20	3473.1114	3535.4022
20-30	2828.375	2682.7636
30-40	3036.2772	2714.0421
40-50	2942.7079	2901.1807
50-60	2339.6318	2058.9239
60-70	1850.8886	1497.375
70-80	1356.9545	1273.767
80-90	847.4477	753.8784
90-100	847.4477	761.7313
100-110	712.3512	631.6926
110-120	800.7296	816.3023
120-140	1013.8227	902.0187
140-160	1970.5455	1871.6522
160-180	2001.6909	2110.8329
180-200	1882.1671	1580.5625
200+	1388.233	1471.4205

m_T	σ_{95CL} UL [fb] (2016)	σ_{95CL} UL [fb] (2017)
0-10	3205.875	3159.2124
10-20	2614.5	2303.25
20-30	2925.75	2614.5
30-40	2295.4812	1984.2312
40-50	2272.125	2062.0188
50-60	2046.4812	1797.4812
60-70	1665.2124	1388.9718
70-80	1093.2594	1003.7688
80-90	778.125	694.461
90-100	476.586	431.8656
100-110	476.586	496.0578
110-120	581.664	564.1344
120-140	710.0484	638.0376
140-160	1015.4718	1097.1438
160-180	1287.7782	1260.5376
180-200	1104.9624	1155.5094
200+	813.1344	1007.6532

Figure 60: 95% CL upper limits on the signal cross-section (for the electroweakino model) for each m_T bin and for two benchmark signal samples: (i) $m(\tilde{\chi}_1^{\pm}) = 200$ GeV (top); (ii) $m(\tilde{\chi}_1^{\pm}) = 300$ GeV (bottom).

the signal production cross section. Regardless of the model chosen, it is important to note that

the analysis presented here and possible reinterpretations take important steps toward establishing a DM candidate particle whose properties lead to a DM relic density commensurate with that measured by astronomers.

Chapter XVI

Discussion of Limits

As noted in the main text, the Higgs combine tool was used to obtain 95% C.L. upper limits for each channel and to perform the subsequent combination of the results, using a CL_s method. A shape-based analysis is performed, using the $m_T(\tau_h, E_T^{miss})$ distribution as the fit discriminant to determine the likelihood of observing signal in the presence of the predicted background rate, given the observed yield in data. The tool takes as input data cards with the yields and nuisance parameters in each $m_T(\tau_h, E_T^{miss})$ bin. We have one data card per bin per signal sample. The cards corresponding to each signal sample were then combined using the "CombineCards.py" tool provided by the Higgs Limit Tool, resulting in a single combined data card. The individual limits were obtained by running the combine tool over each combined card separately.

Correlations of systematic effects between signal, data, and backgrounds are taken into consideration. In order to handle correlations, the following approach was used. Each nuisance parameter was defined with an index to identify the type of process (j = Signal = 0, j = W+jets = 1, j = Z+jets = 2, $j = t\bar{t}$ = 3, j = VV = 4, j = QCD multijet = 5, j = Single Top = 6). Since the Limit Tool handles nuisance parameters with the same name as fully correlated, correlations across processes were specified by utilizing the same channel index.

Table 38 shows some of the nuisance parameters taken into consideration. The first part of the table shows the 2.5% systematic uncertainty considered for the luminosity. Note that there are no indices on the name of the nuisance parameter given to the Higgs Limit Tool. The reason for this is that the tool considers nuisance parameters with the same name as 100% correlated. Therefore, since the systematic effect resulting from the measurement of the luminosity is correlated not only for signal and backgrounds for a given bin, but also across bins, the name given to the systematic must be the same. The bottom part of the table ("Not Correlated Across Channels") shows the systematic effect due to the closure on the QCD multijet background estimation. Since this systematic is specific for the QCD background, it must not be correlated with any other process, and thus a unique index labels the background.

We have performed various fit diagnostic tests to ensure the binned likelihood fit is producing reasonable results. Figure 61 shows the pulls and impact plots of the nuisance parameters used for the 2016 limit calculation. The following naming convention is used:

- **lumi:** luminosity uncertainty
- **SCL:** generator scale uncertainty
- **PDF:** uncertainty due to variations in the parton distribution functions

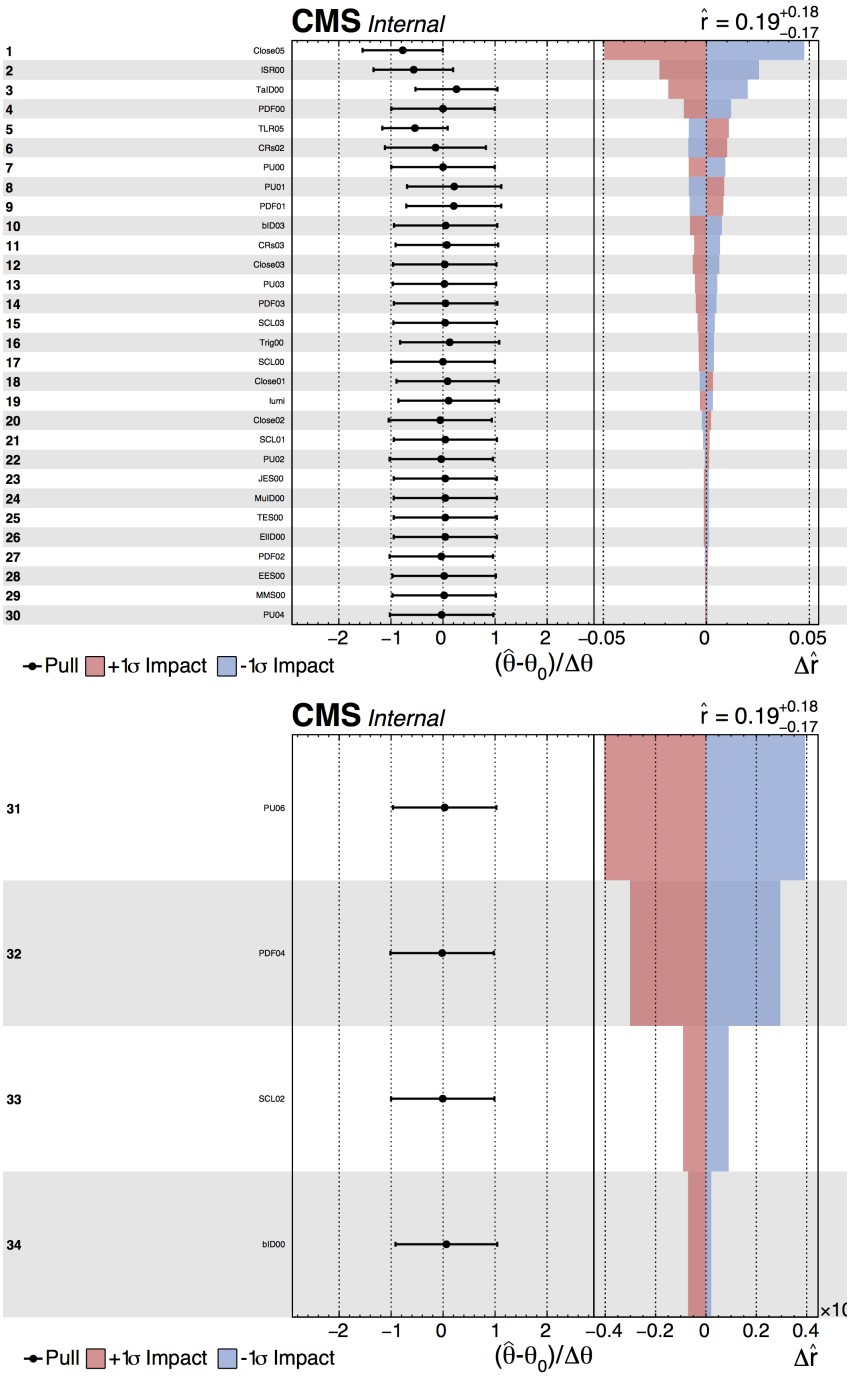

Figure 61: Pulls and impact plots of the nuisance parameters used for 2016 limit calculation.

Table 38: Example of some correlated and uncorrelated nuisance parameters.

Nuisance	Signal	W+Jets	DY+Jets	$t\bar{t}$	VV	QCD	SingleTop
Correlated							
lumi	1.025	1.025	1.025	1.025	1.025		1.025
Not Correlated Across Channels							
Close10							
Close11		1.02					
Close12			1.08				
Close13				1.06			
Close14							
Close15						1.23	
Close16							

- **Trig:** trigger efficiency uncertainty
- **MuID:** muon ID uncertainty
- **ElID:** electron ID uncertainty
- **TaID:** τ_h ID uncertainty
- **bID:** b-tagging uncertainty
- **EES:** uncertainty due to electron energy scale
- **MMS:** uncertainty due to muon momentum scale
- **TES:** uncertainty due to τ_h energy scale
- **JES:** uncertainty due to jet energy scale
- **ISR:** bin-by-bin shape uncertainty due to the ISR weights
- **Close:** uncertainty on the background normalization due to the data-driven estimate
- **TLR:** bin-by-bin shape uncertainty due to the tight-to-loose ratio
- **CRs:** bin-by-bin shape uncertainty due to the level of agreement between data and MC m_T shapes in the CRs

Figure 62 shows the goodness of fit test for the 2016 data. The black histogram represents the probability density function for the test statistic, obtained using 500 toys, assuming the "saturated" model as prescribed by the CMS statistics committee. The observed value of the goodness of fit indicator (i.e. using the data) is 0.01, which is indicated by the blue arrow in the plot. The p-value is the integral above the observed value, which represents the probability of obtaining the observed data distributions assuming a null hypothesis (i.e. assuming no signal exists).

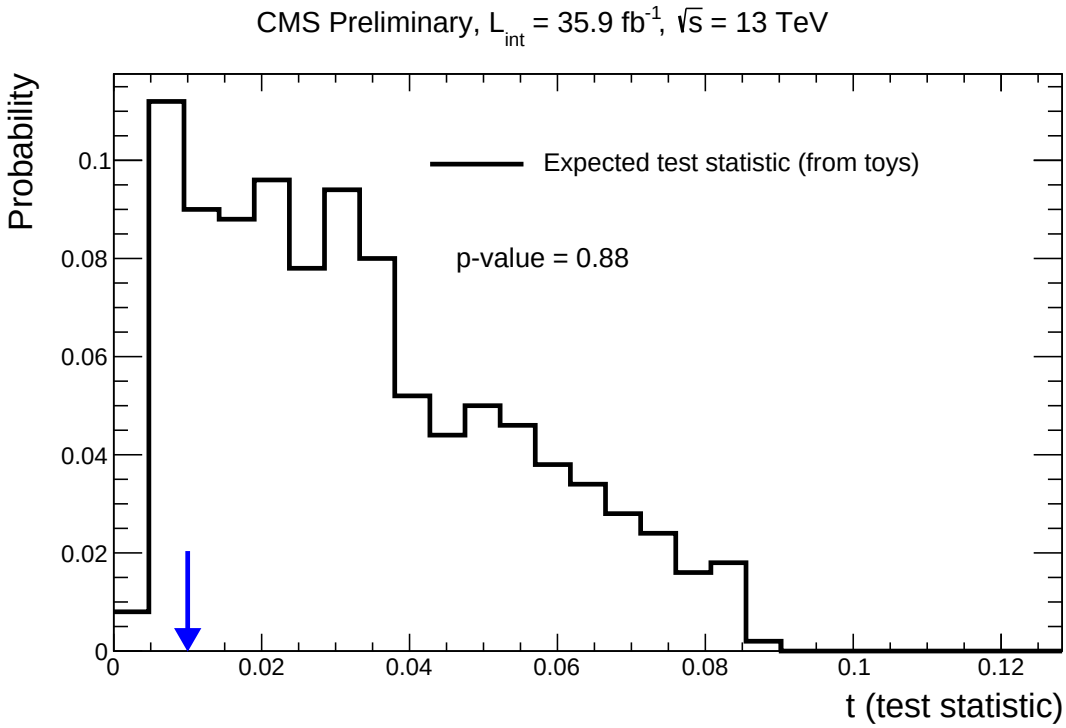

Figure 62: Goodness of fit test for the combination: The black histogram represents the probability density function for the test statistic, obtained using 500 toys, assuming the "saturated" model as prescribed by the CMS statistics committee. The blue arrow represents the observed value of the goodness of fit indicator (i.e. using the data). The p-value is the integral above the observed value, which represents the probability of obtaining the observed data distributions assuming a null hypothesis.

Bibliography

[1] Particle-flow event reconstruction in CMS and performance for jets, taus, and met. *CMS Physics Analysis Summary*, CMS-PAS-PFT-09-001, 2009.

[2] Commissioning of the particle-flow event reconstruction with the first LHC collisions recorded in the CMS detector. *CMS Physics Analysis Summary*, CMS-PAS-PFT-10-002, 2010.

[3] Missing transverse energy performance of the CMS detector. *JINST*, 6:09001, 2011.

[4] Identification of b-quark jets with the CMS experiment. *JINST*, 8:04013, 2013.

[5] Performance of b–jet identification in CMS. *JINST*, 8:P04013, 2013.

[6] Search for pair production of tau sleptons in $\sqrt{s} = 13$ TeV pp collisions in the all-hadronic final state. Technical Report CMS-PAS-SUS-17-003, CERN, Geneva, 2017.

[7] Taking a Closer Look at the LHC. Luminosity. `https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.luminosity`, 2019. Accessed: 2019-12-31.

[8] S. Abdullin, P. Azzi, F. Beaudette, P. Janot, and A. Perrotta. The fast simulation of the CMS detector at LHC. *J. Phys. Conf. Ser.*, 331:032049, 2011.

[9] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.

[10] ATLAS Collaboration. Search for direct stau production in events with two hadronic τ-leptons in $\sqrt{s} = 13$ TeV *pp* collisions with the ATLAS detector. Submitted to *Phys. Rev. D*, 2019.

[11] Julien Baglio. *Phenomenology of the Higgs at the hadron colliders: from the Standard Model to Supersymmetry.* PhD thesis, Orsay, LPT, 2011.

[12] Neta A. Bahcall. Hubble's law and the expanding universe. *Proceedings of the National Academy of Sciences*, 112(11):3173–3175, 2015.

[13] J. D. Barrow. What is the principal evidence for the cosmological principle? *Quarterly Journal of the Royal Astronomical Society*, 30:163–167, Jun 1989.

[14] Alessandro Bartoloni, S Baccaro, L Barone, Francesca Cavallari, Ioan Dafinei, D Re, M Diemoz, E Marco, M Grassi, E Longo, Paolo Meridiani, F Micheli, G Organtini, S Nourbakhsh, Riccardo Paramatti, Fabio Pellegrino, S Rahatlou, Tiziano Rovelli, Michael Sigamani, and L Soffi. The CMS ECAL barrel HV system. *Journal of Instrumentation*, 8:C02039, 02 2013.

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

[16] CERN. CERN's Accelerator Complex. `https://home.cern/science/accelerators/accelerator-complex`, 2019. Accessed: 2019-12-31.

[17] CERN. The Large Hadron Collider. `https://home.cern/science/accelerators/large-hadron-collider`, 2019. Accessed: 2019-12-31.

[18] CMS Collaboration. Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV. 2012. Submitted to *J. Inst.*

[19] CMS Collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. Submitted to *JINST*, 2013.

[20] CMS Collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ tev. *JINST*, 10:P06005, 2015.

[21] CMS Collaboration. CMS luminosity measurements for the 2016 data-taking period. CMS Physics Analysis Summary CMS-PAS-LUM-17-001, 2017.

[22] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} =$ 13 TeV. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[23] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: 10.1140/epjc/s10052-013-2501-z].

[24] S. Dawson. *Introduction to electroweak symmetry breaking*, http://arxiv.org/abs/hep-ph/9901280v1, (1999).

[25] Stefanos Dris, C Foudas, and J. Troska. Performance of the CMS tracker optical links and future upgrade using bandwidth efficient digital modulation. 04 2010.

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

[27] David Griffiths. *Introduction to Elementary Particles; 2nd rev. version.* John Wiley & Sons, 2008.

[28] Mukund Gupta. Calculation of radiation length in materials. Technical Report PH-EP-Tech-Note-2010-013, CERN, Geneva, July 2010.

[29] Thomas Junk. Confidence level computation for combining searches with small statistics. *Nucl. Instrum. Meth. A*, 434:435, 1999.

[30] Michele Livan and Richard Wigmans. Misconceptions about calorimetry. *Instruments*, 1(1):3, May 2017.

[31] A. D. Martin, W. J. Stirling, R. S. Thorne, and G. Watt. Update of parton distributions at NNLO. *Phys. Lett. B*, 652:292, 2007.

[32] Pavel M. Nadolsky, Hung-Liang Lai, Qing-Hong Cao, Joey Huston, Jon Pumplin, Daniel Stump, Wu-Ki Tung, and C.-P. Yuan. Implications of CTEQ global analysis for collider observables. *Phys. Rev. D*, 78:013004, 2008.

[33] University of Zurich: Physik Institut. Simple Example of 3d Axes with Spherical Coordinates. `https://wiki.physik.uzh.ch/cms/latex:example_spherical_coordinates`, 2017. Accessed: 2020-01-01.

[34] University of Zurich: Physik Institut. Standard Model. https://www.physik.uzh.ch/en/researcharea/lhcb/outreach/StandardModel.html, 2018. Accessed: 2019-09-09.

[35] M. Pioppi et al. Tau reconstruction and identification with particle-flow techniques using the CMS detector at LHC. *Nucl. Phys. B, Proc. Suppl. 189 (2009) 311 - 316.*

[36] A. L. Read. Presentation of search results: the CL_s technique. *J. Phys. G*, 28:2693, 2002.

[37] Live Science. What Is the Large Hadron Collider. https://www.livescience.com/64623-large-hadron-collider.html, 2019. Accessed: 2019-12-31.

[38] Albert M Sirunyan et al. Performance of reconstruction and identification of τ leptons decaying to hadrons and ν_τ in pp collisions at $\sqrt{s} = 13$ TeV. *JINST*, 13:P10005, 2018.

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

[40] A.M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, M. Dragicevic, J. Erö, A. Escalante Del Valle, and et al. 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.

[41] D. N. Spergel et al. First year wilkinson microwave anisotropy probe (WMAP) observations: determination of cosmological parameters. *Astrophys. J. Suppl.*, 148:175, 2003.

[42] Lucas Taylor. Hadron Calorimeter. http://cms.web.cern.ch/news/hadron-calorimeter, 2011. Accessed: 2020-01-02.

[43] Lucas Taylor. Muon Detectors. http://cms.web.cern.ch/news/muon-detectors, 2011. Accessed: 2020-01-02.

[44] Lucas Taylor. Silicon Pixels. http://cms.web.cern.ch/news/silicon-pixels, 2011. Accessed: 2020-01-01.

[45] Lucas Taylor. Silicon Strips. http://cms.web.cern.ch/news/silicon-strips, 2011. Accessed: 2020-01-01.

[46] Lucas Taylor. Superconducting Magnet. http://cms.web.cern.ch/news/superconducting-magnet, 2011. Accessed: 2020-01-01.

[47] Lucas Taylor. Tracker Detector. http://cms.web.cern.ch/news/tracker-detector, 2011. Accessed: 2020-01-01.

[48] Lucas Taylor. Triggering and Data Acquisition. http://cms.web.cern.ch/news/triggering-and-data-acquisition, 2011. Accessed: 2020-01-02.

[49] J. Thomason, Roland Garoby, S. Gilardoni, L. Jenner, and J. Pasternak. Proton driver scenarios at CERN and Rutherford Appleton Laboratory. *Physical Review Special Topics - Accelerators and Beams*, 16, 05 2013.

[50] ST Tsou. *Symmetry and symmetry breaking in particle physics*, https://arxiv.org/pdf/hep-th/9803159.pdf, (1998).

[51] Maria Ubiali. NNPDF1.0 parton set for the LHC. *Nucl. Phys. Proc. Suppl.*, 186:62, 2009.