

## ABSTRACT

First search for electroweak supersymmetry in final states with hadronic decays of WW, WZ, or WH boson pair and missing transverse momentum

Ankush Reddy Kanuganti, Ph.D.

Advisor: Kenichi Hatakeyama, Ph.D.

This dissertation presents a new search for particle physics beyond the standard model in the form of electroweak production of supersymmetric particles, specifically charginos and neutralinos. The novel search is performed using Run 2 data from the CMS experiment at the LHC with a total integrated luminosity of  $137 \text{ fb}^{-1}$ . The targeted final states include a pair of bosons (WW, WZ, or WH) with large missing transverse momentum arising from the lightest neutralinos ( $\tilde{\chi}_1^0$ ). Deep neural network large-radius jet taggers are used to identify hadronic decays of the bosons. The standard model backgrounds are estimated using data-driven techniques. No significant excess of events is observed relative to expectations from the standard model. Results are presented in terms of constraints on the cross sections for production of mass-degenerate wino-like superpartners of SU(2) gauge bosons,  $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$ , and higgsino-like supersymmetric particles,  $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0/\tilde{\chi}_3^0$ , in both cases with the bino-like superpartner of U(1) gauge boson as the lightest supersymmetric particle. In the wino-bino scenario, wino masses are excluded up to 960 GeV. In the higgsino-bino scenario, the search excludes higgsino masses from 300 to 650 GeV. These exclusions are most stringent to

date set by the CMS experiment. Sensitivity projections with future high-luminosity LHC data are also presented.

First search for electroweak supersymmetry in final states with hadronic decays of  
WW, WZ, or WH boson pair and missing transverse momentum

by

Ankush Reddy Kanuganti, B.S., M.Sc., M.A.

A Dissertation

Approved by the Department of Physics

---

Lorin S. Matthews, Ph.D., Chairperson

Submitted to the Graduate Faculty of  
Baylor University in Partial Fulfillment of the  
Requirements for the Degree  
of  
Doctor of Philosophy

Approved by the Dissertation Committee

---

Kenichi Hatakeyama, Ph.D., Chairperson

---

Jay R. Dittmann, Ph.D.

---

Andrew Brinkerhoff, Ph.D.

---

Anzhong Wang, Ph.D.

---

David Ryden, Ph.D.

Accepted by the Graduate School  
August 2023

---

J. Larry Lyon, Ph.D., Dean

Copyright © 2023 by Ankush Reddy Kanuganti

All rights reserved



## TABLE OF CONTENTS

LIST OF ACRONYMS .....	viii
LIST OF FIGURES .....	xi
LIST OF TABLES .....	xvi
ACKNOWLEDGMENTS .....	xviii
DEDICATION .....	xix
CHAPTER ONE	
Introduction .....	1
CHAPTER TWO	
Theoretical background .....	6
2.1 The standard model of particle physics .....	6
2.1.1 The standard model Lagrangian .....	10
2.1.2 Quantum chromodynamics .....	12
2.1.3 Quantum electrodynamics .....	13
2.1.4 Weak interactions .....	14
2.1.5 Electroweak interactions .....	16
2.1.6 Higgs mechanism .....	17
2.1.7 Need for beyond standard model physics .....	19
2.2 Supersymmetry .....	20
2.2.1 Fine-tuning problem .....	21
2.2.2 Minimal supersymmetric standard model .....	23
2.2.3 $R$ -parity conserving SUSY .....	24
2.2.4 Electroweak SUSY searches .....	25
CHAPTER THREE	
Experimental apparatus .....	30
3.1 The Large Hadron Collider .....	30
3.1.1 Tuning the LHC beam .....	32
3.2 The Compact Muon Solenoid experiment .....	35
3.2.1 Detector geometry and variables .....	36
3.2.2 Tracker .....	38
3.2.3 Superconducting magnet .....	40
3.2.4 Electromagnetic calorimeter .....	40
3.2.5 Hadronic calorimeter .....	43
3.2.6 Muon system .....	46
3.2.7 Trigger system and data storage .....	49

## CHAPTER FOUR

Event simulation and reconstruction .....	52
4.1 Event simulation .....	53
4.1.1 Event generation .....	53
4.1.2 Detector simulation .....	55
4.2 Event reconstruction .....	56
4.2.1 Tracks .....	56
4.2.2 Calorimeter clusters .....	58
4.2.3 Particle flow .....	58
4.2.4 Physics objects .....	62
4.2.5 MET event filters and further event cleaning .....	75

## CHAPTER FIVE

Search for electroweak supersymmetry in final states with hadronic decays of a WW, WZ, or WH boson pair and large missing transverse momentum ....	78
5.1 Introduction .....	78
5.2 Data and simulation .....	80
5.2.1 Data and triggers .....	80
5.2.2 Signal and background simulation .....	84
5.3 Event selection .....	86
5.3.1 Baseline selection .....	89
5.3.2 b-veto signal region .....	91
5.3.3 b-veto control regions .....	94
5.3.4 b-tag signal regions .....	98
5.3.5 b-tag control regions .....	101
5.4 Background estimation .....	109
5.4.1 Background estimation for the b-veto SR .....	109
5.4.2 Background estimation for the b-tag SRs .....	115
5.5 Systematic uncertainties .....	119
5.5.1 Systematic uncertainties for the b-veto SR .....	119
5.5.2 Systematic uncertainties for the b-tag SRs .....	124
5.5.3 Systematic uncertainties on signal models .....	127
5.6 Validation .....	137
5.6.1 Statistical modeling of backgrounds in the b-veto SR .....	137
5.6.2 Validation regions for the the b-veto SR .....	137
5.6.3 Statistical modeling of backgrounds in the b-tag SRs .....	143
5.6.4 Validation regions for the b-tag SRs .....	144
5.7 Results and interpretations .....	148
5.7.1 Results .....	148
5.7.2 Interpretations .....	149

## CHAPTER SIX

Summary .....	160
---------------	-----

## APPENDIX

Projected future sensitivities for HL-LHC .....	164
A.1 Inputs to HL-LHC projection .....	164

A.2 HL-LHC projection results .....	167
BIBLIOGRAPHY .....	171

## LIST OF ACRONYMS

ALICE	A Large Ion Collider Experiment
APD	Avalanche Photodiode
ASIC	Application Specific Integrated Circuit
BPIX	Barrel Pixel
BSM	Beyond the Standard Model
CERN	European Organization for Nuclear Research
CKM	Cabibbo-Kobayashi-Maskawa
CL	Confidence Level
CMS	Compact Muon Solenoid
CR	Control Region
CSC	Cathode Strip Chamber
CTF	Combinatorial Track Finder
DNN	Dense Neural Network
DT	Drift Tube
ECAL	Electromagnetic Calorimeter
FPIX	Forward Pixel
FSR	Final-State Radiation
GSF	Gaussian Sum Filter
HB	Hadron Calorimeter Barrel
HCAL	Hadron Calorimeter
HE	Hadron Calorimeter Endcap
HF	Hadron Calorimeter Forward
HLT	High Level Trigger
HO	Hadron Calorimeter Outer
HPD	Hybrid Photodiode

ISR Initial-State Radiation  
JEC Jet Energy Correction  
JES Jet Energy Scale  
L1 Level 1  
LHCb Large Hadron Collider beauty  
LHC Large Hadron Collider  
LINAC Linear Accelerator  
LL Lost Lepton  
LSP Lightest Supersymmetric Particle  
MC Monte Carlo  
MET Missing Transverse Energy  
MSSM Minimal Supersymmetric Standard Model  
NLL Next-to-Leading Logarithmic  
NLO Next-to-Leading Order  
NLSP Next-to-Lightest Supersymmetric Particle  
PDF Probability Distribution Function  
PF Particle Flow  
PMT Photomultiplier Tube  
POG Physics Object Group  
PU Pileup  
PV Primary Vertex  
QCD Quantum Chromodynamics  
QED Quantum Electrodynamics  
RPC Resistive Plate Chamber  
SF Scale Factor  
SiPM Silicon Photomultiplier  
SM Standard Model  
SPS Super Proton Synchrotron

SR Search Region  
SUSY Supersymmetry  
SV Secondary Vertex  
TEC Tracker Endcap  
TF Transfer Factor  
TIB Tracker Inner Barrel  
TID Tracker Inner Disks  
TOB Tracker Outer Barrel  
VEV Vacuum Expectation Value  
VPT Vacuum Phototriode  
VR Validation Region  
WP Working Point

## LIST OF FIGURES

Figure 1.1.	Particle discovery timeline.....	2
Figure 1.2.	Summary of standard model fundamental particles.....	4
Figure 2.1.	Feynman diagram showing the electron-positron scattering .....	13
Figure 2.2.	Weak interaction diagrams not involving the Higgs field. ....	15
Figure 2.3.	Illustration of Higgs potential when $\mu^2 < 0$ .....	18
Figure 2.4.	Summary of the SM particles and SUSY particles .....	21
Figure 2.5.	Divergent corrections to the Higgs boson mass .....	22
Figure 2.6.	Gluino pair production diagrams and mass exclusion limits.....	26
Figure 2.7.	Cross section of SUSY particles w.r.t. particle mass .....	27
Figure 2.8.	Feynman diagram showing production of electroweakinos .....	28
Figure 2.9.	Electroweak SUSY searches summary .....	29
Figure 3.1.	The LHC ring overview and the accelerator chain complex.....	31
Figure 3.2.	Cross-section of the LHC dipole magnet.....	33
Figure 3.3.	Total delivered and recorded integrated luminosities for the CMS experiment .....	34
Figure 3.4.	Distribution of the mean number of interactions per proton bunch crossing .....	35
Figure 3.5.	A sectional view of the CMS detector .....	37
Figure 3.6.	A sketch of the CMS tracker system.....	39
Figure 3.7.	Images of the CMS ECAL subdetector.....	41
Figure 3.8.	A sketch showing the coverage of CMS ECAL subsystem.....	42
Figure 3.9.	Photographs of HCAL subdetectors .....	45

Figure 3.10. HCAL depth segmentation before and after the Phase 1 upgrade ....	47
Figure 3.11. A sketch of the muon subsystem.....	48
Figure 3.12. Overview of the L1 trigger system.....	50
Figure 4.1. A sketch showing the interaction of various particles with a slice of CMS detector .....	59
Figure 4.2. Energy fraction in jets as a function of $p_T$ in data and simulation....	67
Figure 4.3. Steps involved in jet energy correction .....	68
Figure 4.4. Jet response with energy corrections.....	69
Figure 4.5. Diagram showing the common principle of identification of b jets initiated by b hadron decays .....	71
Figure 4.6. Efficiency and misidentification rate for Higgs boson identification...	74
Figure 5.1. Simplified SUSY models showing the production of $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ .....	79
Figure 5.2. The MET trigger efficiencies as functions of $p_T^{\text{miss}}$ .....	83
Figure 5.3. Key event-level variable distributions after the baseline selections....	92
Figure 5.4. Distributions of AK8 jet multiplicity, $p_T$ , and mass after the base- line selections.....	93
Figure 5.5. Distributions of the leading AK8 jet's DNN scores after the base- line selections.....	94
Figure 5.6. Signal and background composition in the b-veto search region.....	95
Figure 5.7. Distributions of $p_T^{\text{miss}}$ in b-veto control regions .....	97
Figure 5.8. Background fraction in b-veto regions.....	98
Figure 5.9. Di-lepton mass and leading AK8 distributions in the b-veto dilep- ton control region .....	99
Figure 5.10. Distribution of $p_T^{\text{miss}}$ in all the b-tag search regions .....	101
Figure 5.11. $p_T^{\text{miss}}$ distribution in antitag b-tag control regions .....	103
Figure 5.12. Lepton kinematics in the single lepton control region .....	104



Figure 5.13. Leading AK8 jet kinematics in the single lepton control region .....	105
Figure 5.14. Kinematic distributions in single lepton events .....	106
Figure 5.15. $p_T^{\text{miss}}$ distributions in each of the b-tag regions .....	107
Figure 5.16. Dilepton invariant mass in the b-tag dilepton control region .....	108
Figure 5.17. Distributions of Deep-W and Deep-WMD discriminators measured in the dilepton control regions .....	113
Figure 5.18. Schematic diagram showing the background estimation for a b-tag search region .....	116
Figure 5.19. Transfer factors from $1\ell$ to $0\ell$ b-tag regions.....	117
Figure 5.20. Pass-fail ratios for each of the b-tag signal region .....	117
Figure 5.21. Systematic uncertainty on the 01-res background due to Deep-W and Deep-WMD tagger scale factors .....	120
Figure 5.22. Systematic uncertainty on the 01-res background due to uncertainties in the $t\bar{t}$ , single top production cross sections, and diboson production cross section.....	121
Figure 5.23. Systematic uncertainty on the 01-res background due to JEC .....	122
Figure 5.24. Systematic uncertainty on the 01-res background due to QCD scale uncertainties .....	123
Figure 5.25. Systematic uncertainty on the 01-res background due to PDF uncertainties .....	123
Figure 5.26. Comparison of rare background yields with nominal Deep-W tagger SFs and up variation of the scale factor .....	125
Figure 5.27. Effect of JEC and JER uncertainties on pass-fail ratios .....	128
Figure 5.28. Effect of JEC and JER uncertainties on $0/1\ell$ TFs .....	128
Figure 5.29. Effects of QCD renormalization and factorization scale variations on $\mathcal{R}_{p/f}$ ratios and $0\ell/1\ell$ transfer factors .....	129
Figure 5.30. QCD scale systematic uncertainty on the signal .....	131
Figure 5.31. ISR systematic uncertainty on the signal .....	132

Figure 5.32. Fastsim $p_T^{\text{miss}}$ systematic uncertainty on the signal .....	134
Figure 5.33. b jet response systematic uncertainty on the signal.....	135
Figure 5.34. Nuisance parameters from the fit to the b-veto signal region.....	138
Figure 5.35. Schematic showing validation steps involved in the b-veto analysis..	139
Figure 5.36. Data yields and background predictions in b-veto “1M” valida- tion regions .....	140
Figure 5.37. Data yields and background predictions in “1T1M2J” validation regions .....	141
Figure 5.38. Deep-W and Deep-WMD tagger non-closure corrections .....	142
Figure 5.39. Data yields and background predictions in b-veto “1T2M2J” val- idation regions .....	142
Figure 5.40. Nuisance parameters and their uncertainties for fits to the H tag validation region.....	144
Figure 5.41. Predicted yields in the b-tag “1J” W tag validation region.....	145
Figure 5.42. Nuisance parameters and their uncertainties for fits to the W tag validation region.....	146
Figure 5.43. Data yields and background predictions in the b-tag W and H-tag validation regions.....	147
Figure 5.44. Data yields and the post-fit predictions in the b-veto SR .....	149
Figure 5.45. Observed and post-fit SM background predictions in b-tag signal regions .....	150
Figure 5.46. Distribution of the PDF and $p$ -value calculation .....	154
Figure 5.47. Exclusion limits on TChiWW, TChiWZ, and TChiWH signal models	156
Figure 5.48. Exclusion limits on mass-degenerate wino-like NLSP production ....	157
Figure 5.49. Exclusion limits on mass-degenerate higgsino-like NLSP production	158
Figure 5.50. Projected 95% CL HL-LHC limits on the higgsino production .....	159
Figure A.1. Comparison between privately-computed and centrally-provided Resummino cross-sections for wino-like production .....	166

Figure A.2. Comparison between privately-computed and centrally-provided Resummino cross-sections for higgsino-like production .....	166
Figure A.3. HL-LHC projected mass exclusions and 95% confidence level up- per limits on cross sections for mass-degenerate wino-like $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ and $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production .....	167
Figure A.4. HL-LHC projected mass exclusions and 95% confidence level up- per limits on cross sections for mass-degenerate wino-like $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ or $\tilde{\chi}_1^\pm \tilde{\chi}_2^\mp$ production .....	168
Figure A.5. Expected significance for mass-degenerate wino-like $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ and $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with HL-LHC .....	168
Figure A.6. Projected 95% CL exclusion for mass-degenerate wino-like $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ and $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ production with HL-LHC.....	169
Figure A.7. Ratio of 95% CL upper limit cross sections for TChiWH, TChiWZ, and their combinations with TChiWW.....	170

## LIST OF TABLES

Table 2.1.	SM quarks with their attributes.....	7
Table 2.2.	SM leptons with their attributes .....	8
Table 2.3.	SM gauge bosons and the Higgs boson with their attributes .....	8
Table 2.4.	The particles in MSSM .....	24
Table 4.1.	Radius of the cone with respect to $p_T$ of the lepton .....	64
Table 5.1.	MET primary dataset integrated luminosities .....	81
Table 5.2.	List of SUSY signal samples used in this analysis .....	85
Table 5.3.	List of standard model MC samples used in this analysis for dominant backgrounds.....	87
Table 5.4.	List of standard model MC samples used in this analysis for small backgrounds .....	88
Table 5.5.	Baseline event selections .....	91
Table 5.6.	Summary of tagging requirements for the b-veto SR and CRs .....	96
Table 5.7.	Summary of tagging requirements for the b-tag SRs and CRs .....	103
Table 5.8.	Summary of W-tagging data/MC scale factor measurements.....	112
Table 5.9.	Summary of V-tagging data/MC scale factor measurements .....	114
Table 5.10.	W-tag and bb-vs-light pass-fail ratio SFs derived from dilepton b-tag regions.....	118
Table 5.11.	Summary of systematic uncertainties in the b-veto analysis .....	125
Table 5.12.	Summary of systematic uncertainties in the b-tag analysis .....	130
Table 5.13.	Summary of systematic uncertainties on signal yields .....	136
Table 5.14.	b-veto validation regions .....	139

Table 5.15. Selections in validation pseudo-SR and CRs for b-tagged events .....	145
Table 5.16. Post-fit background predictions vs observation in the b-veto signal region.....	148
Table 5.17. Postfit background predictions vs observation in each of the b-tag SRs.....	151

## ACKNOWLEDGMENTS

It has been a great journey for me while performing this work. I am thankful for Baylor University and the Department of Physics for offering this program in experimental high energy physics. Several individuals helped me succeed during my several years as a graduate student. First, I would like to thank my advisor with heartfelt gratitude, Dr. Kenichi Hatakeyama, for his meticulous guidance and help from the beginning of my graduate school. He helped me in giving valuable advice and finding research opportunities in the field of particle flow development in the CMS Collaboration. My desire to work on the CMS hardware came true with the help from Dr. Jay Dittmann, who provided excellent guidance on the backend electronics used in HCAL subdetector at CMS.

I would like to thank my analysis group members who provided positive feedback from the beginning of this analysis. My special thanks to Dr. Vinay Hegde, for guiding me through various challenges and learned a lot of things. I appreciate the suggestions given by our group members Dr. Andrew Whitbeck, Dr. Keith Ulmer, Dr. John Wilson, and Noah Zipper, and it was a great time working with these amazing people. At Baylor, I would like to thank several people who have helped me and crossed paths with me including Dr. Andrew Brinkerhoff, Dr. Bryan Caraway with whom I spent most of my time as a graduate student, Dr. Caleb Smith, Brooks McMaster and other graduate students in the Baylor HEP group. I am grateful to all the people mentioned above for my success at graduate school.

*This work is dedicated to my parents,  
for their support during all times.*

## CHAPTER ONE

### Introduction

The current understanding of the fundamental nature of matter and the forces that govern subatomic particles began in the early 19th century. A hot, dense condition known as a singularity is where the universe first appeared 13.8 billion years ago, according to the well-known dominant hypothesis known as the “Big Bang”. From this singularity, the universe rapidly expanded and cooled, leading to the formation of subatomic particles such as protons and neutrons. Development of cathode ray tubes and cloud chambers instilled curiosity in scientists since they could observe the behavior of particles and their interactions. The discovery of new particles such as positron and muon in the 1930s brought more excitement to the field as these particles were not predicted by the existing theories at that time.

Particle physics emerged as a field of study in the early 20th century, when scientists first began to probe the subatomic structure of matter. Later, developments of high-energy particle accelerators, such as the cyclotron and the synchrotron, allowed for the study of particles at higher energies and much smaller distance scales. In the early 1970s, the discovery of the weak neutral current and the charm quark made a breakthrough in the field, leading to the development of the electroweak theory and the standard model (SM) of particle physics. Today, particle physics continues to be an active and rapidly-evolving field, with active research in areas pertaining to studies of dark matter, elusive neutrinos, and searches for new particles beyond the standard



model (BSM). The Large Hadron Collider (LHC) is the world's largest particle accelerator which started its first operation called Run 1 in 2009, and Run 2 operations were carried out during 2015–2018. In order to unlock the mysteries of nature, and interesting physics which may exist at higher energies, the LHC started Run 3 operations in 2022 with an unprecedented proton-proton collision center-of-mass energy of 13.6 TeV and will collect more new collision data than existing data.

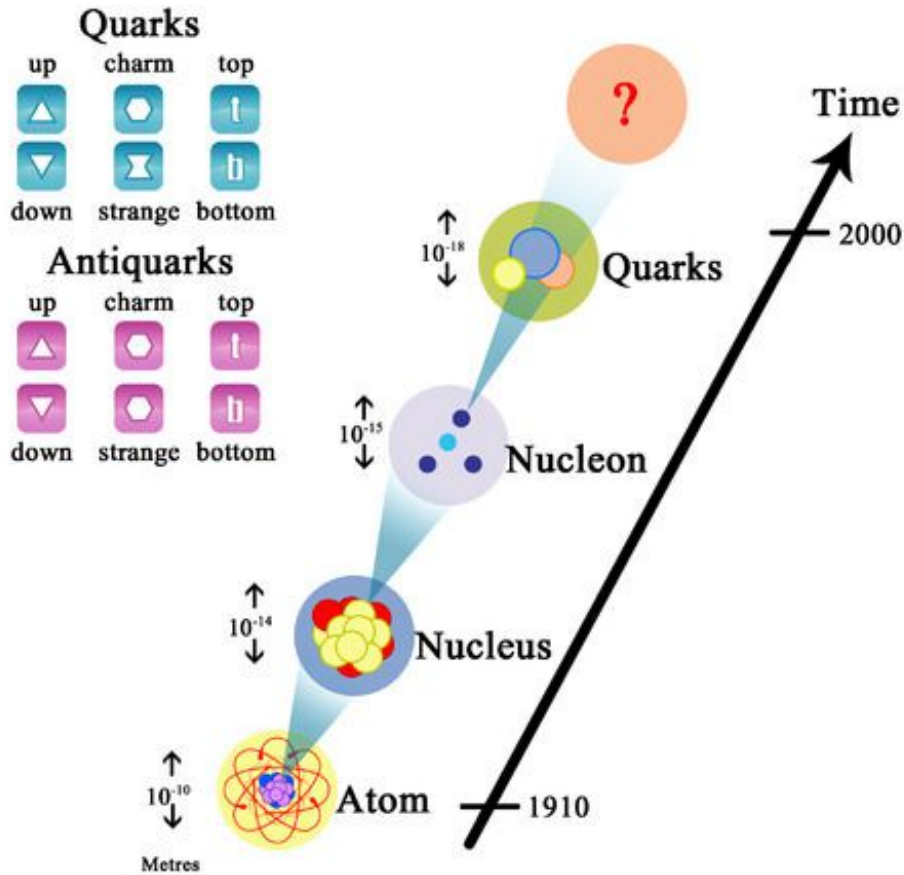


Figure 1.1: Timeline showing the discovery of particles from an atom to quarks. Figure source [1].

Figure 1.1 shows the microscopic view of matter that we are able to probe until today. In a nutshell, the nucleons are composed of quarks. Matter is made up of particles such as electrons and quarks, which are held together by the strong force and electromagnetic force. As of today, there are four fundamental forces which are listed in order of strength:

- Strong force: The gluon is the fundamental particle that binds quarks and acts at short distances.
- Electromagnetic force: Electrically charged particles experience this force through a photon.
- Weak force: W and Z bosons are the weak force carriers that govern particle decays at short distances.
- Gravitational force: Any particle with mass experiences this force through a hypothesized force carrier particle (graviton), which has not been discovered yet.

The observed fundamental particles can be described by the standard model of particle physics that includes six quarks (in three generations), six leptons (in three generations), four gauge bosons, and the scalar Higgs boson along with antimatter particles as shown in Fig. 1.2.

However, there are good reasons to question the completeness of the standard model. For example, it fails to explain the gravitational force that we experience in our quotidian lives. From cosmological observations, dark matter accounts for 85% of the total matter in the universe [3]; however, again the SM fails to provide a dark matter candidate. From the theoretical point of view, the incompleteness of the SM may be postulated by the so-called fine-tuning problem. We discovered the Higgs boson

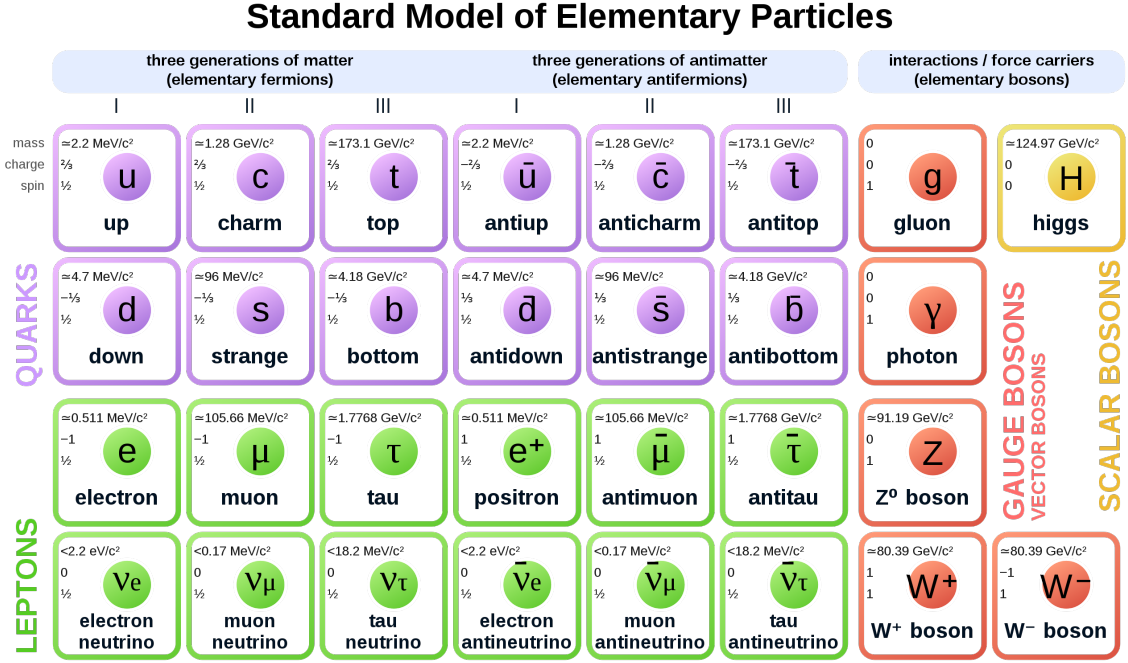


Figure 1.2: Summary of SM fundamental particles. Figure source [2].

at much lower mass than the Plank scale, which is many orders of magnitude larger than 125 GeV, and only fine-tuning the bare mass of the Higgs boson can get us the mass value we observe experimentally. The fact that there is no physical explanation on the fine-tuning parameters in the SM with large corrections may indicate that the SM is an incomplete theory.

In order to answer some of these questions, many extensions to SM have been proposed, one of them being supersymmetry (SUSY). SUSY introduces space-time symmetries, giving each SM particle a “superpartner” that differs in spin by half a unit. Since we have failed to observe supersymmetric particles in a similar mass range to their counterpart SM particles, it led to an idea that supersymmetry is broken and the superpartners of SM must be at heavier mass scales. There is a broad range of SUSY models. Among them, a model called the Minimal Supersymmetric Standard

Model (MSSM) [4] contains the smallest number of new particle states and new interactions consistent with phenomenology. SUSY provides a dark matter candidate: the lightest supersymmetric particle (LSP). The LSP is stable under conservation of  $R$ -parity that does not allow supersymmetric particles to decay entirely into SM particles. The LSP is a weakly interacting matter particle which does not leave any signatures in the CMS detector. One of the several ways to search for SUSY signatures involves hunting for the LSP, which escapes the detector and can be traced by missing transverse momentum in an event. This dissertation also extensively explores the use of deep machine learning based reconstruction of SM particles produced in SUSY particle decays in order to enhance sensitivities to SUSY scenarios not explored previously.

In Chapter Two, a theoretical background is laid out to elaborate the concepts of the SM through the Lagrangian formulation along with its limitations. Section 2.2 describes SUSY in detail. The LHC and the CMS detector, which provided data used in this analysis, are discussed in Chapter Three. Once the collision data are obtained, the event reconstruction is performed, which is elaborated in Chapter Four, together with event simulation used for comparison with the collision data. Chapter Five presents a motivation for specific SUSY models, and describes the complete methodology used to perform the search along with the results and their interpretation. Chapter Six provides summary of the search and prospects for future SUSY analyses.

## CHAPTER TWO

### Theoretical background

A theoretical overview of particle physics with an emphasis on the standard model (SM) and beyond is presented in this chapter. The SM is a successful theoretical model that describes the fundamental particles in the universe and their interactions. The SM is covered in detail in Section 2.1, although it falls short of explaining several of the known phenomena, such as gravity. Section 2.1.7 lists further limitations on the SM. A few solutions to the limitations of the SM are brought by an extension to SM called Supersymmetry (SUSY). In Section 2.2, I will try to motivate why SUSY searches are crucial, and in particular searches for SUSY in the electroweak sector. The theory discussed in this chapter is derived from these sources [5, 6].

#### *2.1 The standard model of particle physics*

The SM of particle physics lists all the known fundamental particles including mediator particles of three (out of four) fundamental forces. Figure 1.2 shows the SM particles along with their anti-particles (with opposite electric charge). The fundamental particles can be classified based on their spin into fermions ( $\frac{1}{2}$  integer spin) and bosons (whole integer spin), where spin is the intrinsic angular momentum of a particle. Fermions are divided into two categories, quarks and the leptons (classified into three generations), as shown in Tables 2.1 and 2.2. Bosons are classified into vector or gauge bosons and a scalar boson (the Higgs boson) which are the force mediators for the matter particles as shown in Table 2.3. This section describes

the interactions between the fermions and bosons along with the fundamental forces explained by the standard model.

Section 2.1.1 provides specifics on the SM particles and interactions between them. The SM requires a local gauge symmetry of  $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ , where the unitary group  $U(1)_Y$  describes electrodynamics as detailed in Section 2.1.3. The  $SU(2)_L$  group describes the weak interactions between the fermions and gauge bosons (discussed in Section 2.1.4). The strong interactions that hold the partons in the nucleus together is described by the  $SU(3)_C$  group according to quantum chromodynamics (discussed in Section 2.1.2). The electroweak theory which combines electromagnetic and weak forces together is generated by the groups  $SU(2)_L \otimes U(1)_Y$  as discussed in Section 2.1.5.

Table 2.1: SM quarks with their attributes

Generation	Particle	Charge	Mass (MeV)	Interactions
1	$u$	$\pm\frac{2}{3}$	$\approx 2.2$	Strong, EM, Weak
1	$d$	$\mp\frac{1}{3}$	$\approx 4.7$	Strong, EM, Weak
2	$c$	$\pm\frac{2}{3}$	$1.275 \times 10^3$	Strong, EM, Weak
2	$s$	$\mp\frac{1}{3}$	$\approx 95$	Strong, EM, Weak
3	$t$	$\pm\frac{2}{3}$	$173.1 \times 10^3$	Strong, EM, Weak
3	$b$	$\mp\frac{1}{3}$	$\approx 4.18 \times 10^3$	Strong, EM, Weak

From the famous Dirac equation for free particles, fermions are the spin  $\frac{1}{2}$  fields that are represented as solutions to the Dirac equation.

$$i\gamma^\mu\partial_\mu\psi - m\psi = 0. \quad (2.1)$$

Table 2.2: SM leptons with their attributes

Generation	Particle	Charge	Mass (MeV)	Interactions
1	$\nu_e$	0	$< 2.2 \times 10^{-6}$	Weak
1	$e$	$\pm 1$	0.511	EM, Weak
2	$\nu_\mu$	0	$< 0.17$	Weak
2	$\mu$	$\pm 1$	105.7	EM, Weak
3	$\nu_\tau$	0	$< 15.5$	Weak
3	$\tau$	$\pm 1$	1777	EM, Weak

Table 2.3: SM gauge bosons and the Higgs boson with their attributes

Particle	Charge	Mass (GeV)	Spin	Force mediator	Mixing fields
$g$	0	0	1	Strong	$g$
$\gamma$	0	0	1	EM	$W_3, B$
W	$\pm 1$	80.4	1	Weak	$W_1, W_2$
Z	0	91.2	1	Weak	$W_3, B$
H	0	125.2	0	-	$H_u, H_d$

The quark fields are defined as a triplet of Dirac spinors, as shown by:

$$u = \begin{pmatrix} u_{\text{red}} \\ u_{\text{green}} \\ u_{\text{blue}} \end{pmatrix}, \quad (2.2)$$

where each of the color component is represented by the four-component Dirac spinors,

as shown by:

$$q = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}, \quad (2.3)$$

where  $q$  is any quark color flavor and each component is a complex function. The

same field is applicable to leptons except that there is no color. The SM adopts chiral

theory, the fields have left-handed and right-handed components that are linked by the projection operators ( $P_L$  and  $P_R$ ).

$$P_L\psi_L = \frac{1-\gamma^5}{2}\psi_L, \quad P_L\bar{\psi}_R = \bar{\psi}_R, \quad P_L\psi_R = P_L\bar{\psi}_L = 0, \quad (2.4)$$

$$P_R\psi_R = \frac{1+\gamma^5}{2}\psi_R, \quad P_R\bar{\psi}_L = \bar{\psi}_L, \quad P_R\psi_L = P_R\bar{\psi}_R = 0. \quad (2.5)$$

The charge conjugate of a right-handed field is left-handed, where  $\gamma^5$  is the fifth gamma matrix ( $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$ ). The left-handed quarks and leptons are arranged in three generations by doublets as shown by:

$$Q^\alpha = \begin{pmatrix} u^\alpha \\ d^\alpha \end{pmatrix}, \begin{pmatrix} c^\alpha \\ s^\alpha \end{pmatrix}, \begin{pmatrix} t^\alpha \\ b^\alpha \end{pmatrix}, \quad (2.6)$$

$$L = \begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \quad (2.7)$$

where  $\alpha \in \{\text{red, green, blue}\}$  is the color index for quarks. The right-handed singlets are considered for the quarks and three generations of leptons, other than the neutrinos. The weak force allows the interactions from the left-handed neutrinos only and the SM does not include the right-handed neutrinos.

The SM includes gauge bosons (three weak bosons, one photon, and eight gluons) which form twelve gauge fields.  $G_\mu^{1\dots 8}$  are the eight gluon fields that carry the color charge and are bi-colored. The electroweak boson fields include  $W_\mu^{1,2,3}$  and  $B_\mu^0$ , and these fields mix to create physically-observable states such as the massive neutral Z boson and charged W boson, and the massless neutral photon. The gauge bosons that interact with the Higgs boson are given mass by the complex scalar Higgs field,  $(\phi)$ , while the fermion masses are obtained through Yukawa coupling interactions.



Each force has mediator bosons, as discussed in Chapter One. The strong force is mediated by gluons, while the weak force is mediated by W and Z bosons, the electromagnetic force by photons, and the gravitational force presumably by gravitons (not explained by the SM and yet undiscovered). With a coupling constant of  $\alpha_s \approx 1$ , the strong force dominates all other forces by at least 137 times. It operates at a short range of around  $10^{-15}$  m (about the size of a nucleon), where the effective range of a force is the smallest distance between particles that could have a high chance of interaction. The electromagnetic force, which has an unlimited range and a coupling constant of  $\alpha = \frac{1}{137}$  at low energy scales, is regarded as the second most powerful force. In the 1 GeV energy range, the weak force has a coupling constant of about  $10^{-7}$  and operates between  $10^{-17}$  to  $10^{-16}$  m. The coupling constant for gravity, which has the least value of coupling constant equal to  $10^{-39}$ , is the lowest of all the forces with an unlimited range.

### 2.1.1 The standard model Lagrangian

The SM Lagrangian describes all the interactions between the matter particles and the force carriers. The gauge interactions with the fermions can be described using the covariant derivative  $D_\mu$ . When the covariant derivative acts on the fermion fields, the expression is given by:

$$D_\mu = \partial_\mu - \frac{1}{2}ig_s\Gamma_s\lambda_\alpha G_\mu^\alpha - ig|I_3|\sigma_j W_\mu^j - \frac{1}{2}ig'YB_\mu, \quad (2.8)$$

$$D_\mu\psi_L = (\partial_\mu + gA_\mu)\psi_L,$$

where  $g_s$  is the coupling strength for SU(3),  $g$  is the coupling strength for SU(2),  $\Gamma_s$  is 1 for quarks and 0 for other fields,  $\lambda_\alpha$  for  $\alpha \in \{1, 2, 3, 4, 5, 6, 7, 8\}$  are the eight generators for SU(3),  $G_\mu^\alpha$  are the eight SU(3) gauge fields,  $I_3$  is the third component

of weak isospin,  $\sigma_j$  for  $j \in \{1, 2, 3\}$  are the three generators for SU(2),  $W_\mu^j$  are the three SU(2) gauge fields,  $g'$  is the coupling strength for U(1),  $Y$  is hypercharge,  $A_\mu$  is the spin-1 massless photon field, and  $B_\mu$  is the U(1) gauge field. The field tensor can be constructed using the above covariant derivative,

$$F_{\nu\mu} = -\frac{i}{g}[D_\mu, D_\nu]. \quad (2.9)$$

The Lagrangian for all the interactions can be summarized using Eq. (2.10):

$$\begin{aligned} \mathcal{L}_{\text{SM}} = & -\frac{1}{4}F_{\mu\nu}^a F_a^{\mu\nu} + \bar{\Psi}_L i\gamma^\mu D_\mu \Psi_L + \bar{\psi}_R i\gamma^\mu D_\mu \psi_R \\ & + (D^\mu \phi)^\dagger (D_\mu \phi) + \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \\ & - y_{ij} \bar{\phi} \bar{Q}_i u_j - y_{ij} \phi \bar{Q}_i d_j - y_{ij} \phi \bar{L}_i e_j + hc, \end{aligned} \quad (2.10)$$

where  $\phi$  represents the scalar Higgs field and the middle three terms correspond to the Higgs mechanism, while the first three terms show the interactions between the fermions and the bosons. The last line shows the Yukawa couplings and  $hc$  stands for Hermitian conjugate. Each of the term is explained in the following subsections.

The covariant derivatives acting on the fields can be summarized as:

$$\begin{aligned} D_\mu \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} &= \left[ \partial_\mu - \frac{ig_1}{2} B_\mu + \frac{ig_2}{2} W_\mu \right] \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \\ D_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} &= \left[ \partial_\mu + \frac{ig_1}{6} B_\mu + \frac{ig_2}{2} W_\mu + ig G_\mu \right] \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \end{aligned} \quad (2.11)$$

where

$$\begin{aligned} D_\mu \nu_R &= \partial_\mu \nu_R, \\ D_\mu e_R &= [\partial_\mu - ig_1 B_\mu] e_R, \\ D_\mu u_R &= \left[ \partial_\mu + \frac{2ig_1}{3} B_\mu + ig G_\mu \right] u_R, \end{aligned} \quad (2.12)$$

$$D_\mu d_R = \left[ \partial_\mu - \frac{ig_1}{3} B_\mu + ig G_\mu \right] d_R,$$

$$D_\mu \phi = \left[ \partial_\mu + \frac{ig_1}{2} B_\mu + \frac{ig_2}{2} W_\mu \right] \phi.$$

### 2.1.2 Quantum chromodynamics

The strong interactions between the quarks and the gluons is described by the theory of quantum chromodynamics (QCD). The gluon (spin-1 massless vector boson) is the mediator of this strong force. QCD is the field theory based on the  $SU(3)_C$  gauge group,  $C$  represents color (red, green, and blue), and the gluons can self-interact according to  $SU(3)_C$ . The  $3 \times 3$  color combinations form an octet:

$$|r\bar{g}\rangle, |r\bar{b}\rangle, |g\bar{b}\rangle, |g\bar{r}\rangle, |b\bar{r}\rangle, |b\bar{g}\rangle,$$

$$\frac{1}{\sqrt{2}}(|r\bar{r}\rangle - |g\bar{g}\rangle), \frac{1}{\sqrt{6}}(|r\bar{r}\rangle + |g\bar{g}\rangle - 2|b\bar{b}\rangle),$$
(2.13)

and a singlet:

$$\frac{1}{\sqrt{3}}(|r\bar{r}\rangle + |g\bar{g}\rangle + |b\bar{b}\rangle).$$
(2.14)

Isolated quarks or gluons are not observed in nature due to a requirement called *color confinement* in QCD. Therefore, all the baryons (three-quark combinations) and mesons (quark-antiquark pair combinations) are colorless. After a collision event, the free quarks and gluons produced new hadrons from a process called hadronization. The strength of the strong force becomes smaller at high energies and increases with distance which is described by asymptotic freedom behavior [7].

In order to maintain the  $SU(3)$  symmetry, the following transformations are applied on the quark and gluon fields:

$$q \xrightarrow{SU(3)} e^{ig_s \xi_a \lambda_a / 2} q,$$

$$G_\mu^c \xrightarrow{SU(3)} G_\mu^c - \frac{1}{g_s} \partial_\mu \xi_c - f_{abc} \xi_a G_\mu^b,$$
(2.15)

where  $\xi$  is a 8-dimensional arbitrary vector and  $f_{abc}$  are the structure constants of SU(3). The gluon field tensor and the QCD contribution to SM Lagrangian is shown by:

$$\begin{aligned} G_{\mu\nu}^a &= \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c, \\ \mathcal{L}_{\text{QCD}} &= \bar{q} (i\gamma^\mu D_\mu - m) q - \frac{1}{4} G_{\mu\nu}^a G_a^{\mu\nu}. \end{aligned} \tag{2.16}$$

### 2.1.3 Quantum electrodynamics

The photon, a massless vector boson, serves as the electromagnetic force's mediator in the SM. The interactions between photons and matter particles are theoretically described by quantum electrodynamics (QED). Charge conservation is maintained through the local gauge symmetry of QED, U(1), which is invariant under rotations. Despite having no electrical charge, photons interact with all charged particles. Fig. 2.1 illustrates an instance of electron-positron scattering resulting from the exchange of a photon. Now, following the U(1) symmetry, the Lagrangian of a

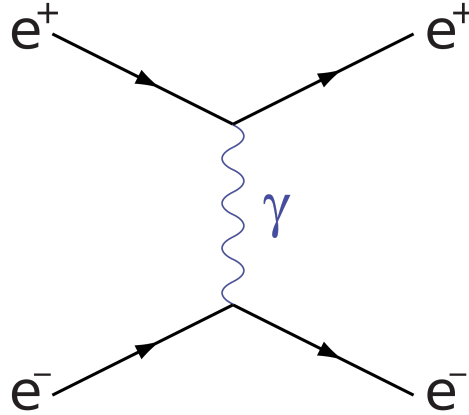


Figure 2.1: Feynman diagram showing the electron-positron scattering. Figure source [8].

free-particle Dirac fermion is invariant under rotations. In order to maintain the invariance, we must introduce a spin-1 vector field  $A_\mu$  (massless photon). The fermion field, covariant derivative and  $A_\mu$  are defined as:

$$\begin{aligned}\phi(x) &\xrightarrow{\text{U}(1)} e^{iq\alpha}\phi(x), \\ D_\mu &= \partial_\mu - ieA_\mu, \\ A_\mu &\xrightarrow{\text{U}(1)} A_\mu + \frac{1}{e}\partial_\mu\alpha.\end{aligned}\tag{2.17}$$

Then using the electromagnetic field tensor  $F_{\mu\nu}$ , we can define the QED contribution to the SM Lagrangian to be:

$$\begin{aligned}F_{\mu\nu} &\equiv \partial_\mu A_\nu - \partial_\nu A_\mu, \\ \mathcal{L}_{\text{QED}} &= \bar{\psi}(i\gamma^\mu D_\mu - m) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}.\end{aligned}\tag{2.18}$$

#### 2.1.4 Weak interactions

The weak forces are mediated by the  $W^\pm$  bosons and the Z boson. The local gauge symmetry that models weak interactions between the fermions and the bosons is the SU(2) group, which is the rotational symmetry of a sphere. Parity is not conserved in weak interactions since the weak interactions depend on chirality [9], but the weak isospin ( $T_3$ ) is conserved. Weak isospin for left-handed fermions is  $\pm\frac{1}{2}$ , based on the charge of the fermion, while the right-handed fermions have weak isospin of 0. Due to the large mass of the W boson, particle decays mediated through weak interactions have longer lifetime. In the case of a neutral pion, the lifetime is about  $10^{-16}$  seconds because the decay rate is proportional to the term  $\frac{g_W^2}{m_W^2 c^2}$  in the limit when the condition  $Q \ll m_W c$  is satisfied, where  $Q$  is the momentum transferred by the W boson.

The weak interactions can involve three types of interactions as shown in Fig. 2.2 which can be summarized as:

- Neutral current interaction (shown in top left diagram of Fig. 2.2) where a quark/lepton can emit/absorb a Z boson.
- Charged current interaction where the quark flavor can change by absorbing or emitting a W boson. For example, the beta decay process where  $n \rightarrow p + W^-$ .
- A charged lepton can also have charged current interaction by absorbing/emitting a W boson.

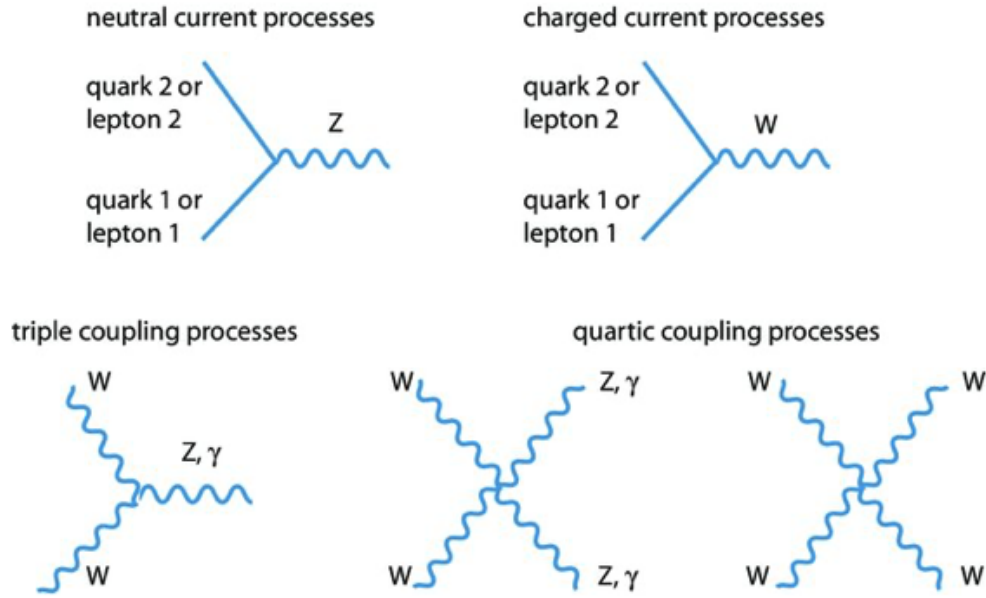


Figure 2.2: Weak interaction diagrams not involving the Higgs field. Figure source [10]

The weak force contribution to the SM Lagrangian is given by:

$$\mathcal{L}_{\text{weak}} = -\frac{g_W}{\sqrt{2}} \left[ \bar{\psi}_p \gamma^\lambda \left( 1 - \frac{g_A}{g_v} \gamma_5 \right) \psi_n \right] [\bar{\psi}_e \gamma_\lambda (1 - \gamma_5) \psi_\nu] + hc \quad (2.19)$$

where  $g_W$  is the weak coupling constant,  $g_v$  and  $g_A$  represent the vector and axial vector coupling constants. For quarks,  $g_v = G_F V_{ud}$ , where  $G_F$  is the Fermi coupling constant and  $V_{ud}$  is an element from the CKM mixing matrix [11].  $\lambda$  is defined as the ratio of  $g_A$  to  $g_v$  and still is one of the tunable parameters in the standard model.

### 2.1.5 Electroweak interactions

The electromagnetic, and the weak force (discovered in the 1930s), are combined in the electroweak theory. Electroweak theory is described by the gauge group  $SU(2) \times U(1)_Y$ , where  $Y$  is the hypercharge that was added to make the generator of  $U(1)_{EM}$  commute with  $SU(2)$ . The bosons that mediate weak interactions have mass; however, this theory calls for non-physical massless bosons of weak isospin ( $W^1, W^2, W^3$ ) and one of weak hypercharge ( $B^0$ ). Spontaneous symmetry breaking (SSB), which is covered in section below on how gauge bosons acquire their masses, provides the solution to this issue.

Weak isospin ( $T$ ) and weak hypercharge ( $Y$ ) are generated using the  $SU(2)$  and  $U(1)$ , respectively. These quantities determine the charge ( $q$ ) of a particle:

$$q = T_3 + \frac{1}{2}Y. \quad (2.20)$$

The electroweak Lagrangian is defined as:

$$\begin{aligned} \mathcal{L}_{EWK} &= \bar{\Psi}_L i \gamma^\mu D_\mu \Psi_L + \bar{\psi}_R i \gamma^\mu D'_\mu \psi_R - \frac{1}{4} W_a^{\mu\nu} W_{\mu\nu}^a - \frac{1}{4} B^{\mu\nu} B_{\mu\nu}, \text{ where} \\ D_\mu &= \partial_\mu - i \frac{g'}{2} Y B_\mu - i \frac{g_W}{2} \sigma^a W_\mu^a, \\ B_\mu &\xrightarrow{U(1)} B_\mu - i g' \partial_\mu \alpha, \\ W_\mu^a &\xrightarrow{SU(2)} W_\mu^a - g_W \partial_\mu \xi^a - g_W \epsilon_{abc} \xi^b W_\mu^c, \end{aligned} \quad (2.21)$$

where  $\Psi_L$  ( $\psi_R$ ) represent the left-handed (right-handed) fermions. The new spin-1 vector fields  $B_\mu$  and  $W_\mu^a$  remain invariant after the U(1) and SU(2) transformations, respectively.

### 2.1.6 Higgs mechanism

The experiments at CERN in the early 1980s [12] proved that the W and Z bosons are massive. All spin-1 bosons must be massless in order for the local gauge invariance to hold, which is true for photons and gluons but not for the W and Z bosons. To solve this problem, a spontaneous symmetry breaking mechanism was proposed by Englert, Brout, and Higgs [13, 14] in which the gauge bosons acquire their masses by interacting with a new scalar field, called the Higgs field  $\phi$ :

$$\phi = \begin{pmatrix} \phi_+ \\ \phi_0 \end{pmatrix}. \quad (2.22)$$

The Lagrangian is given by:

$$\begin{aligned} \mathcal{L}_\phi &= (D^\mu \phi)^\dagger (D_\mu \phi) - V(\phi^\dagger \phi), \\ V(\phi) &= \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2, \end{aligned} \quad (2.23)$$

where  $D^\mu$  is the covariant derivative defined in Eq. (2.21) and  $V(\phi)$  defines the potential term. The constant  $\lambda$  has to be positive to prevent unstable potential and the value of  $\mu^2$  can either be positive or negative. If  $\mu^2$  is positive, then the potential is positive with a vacuum expectation value (VEV) of zero. If  $\mu^2$  is negative, then minimum potential becomes  $\sqrt{-\frac{\mu^2}{2\lambda}}$ , as shown in Fig. 2.3. At the minima, the ground state of Higgs field is non-zero and symmetric. This symmetry is broken when the system moves to the available minima and has non-zero VEV which is referred to as SSB.



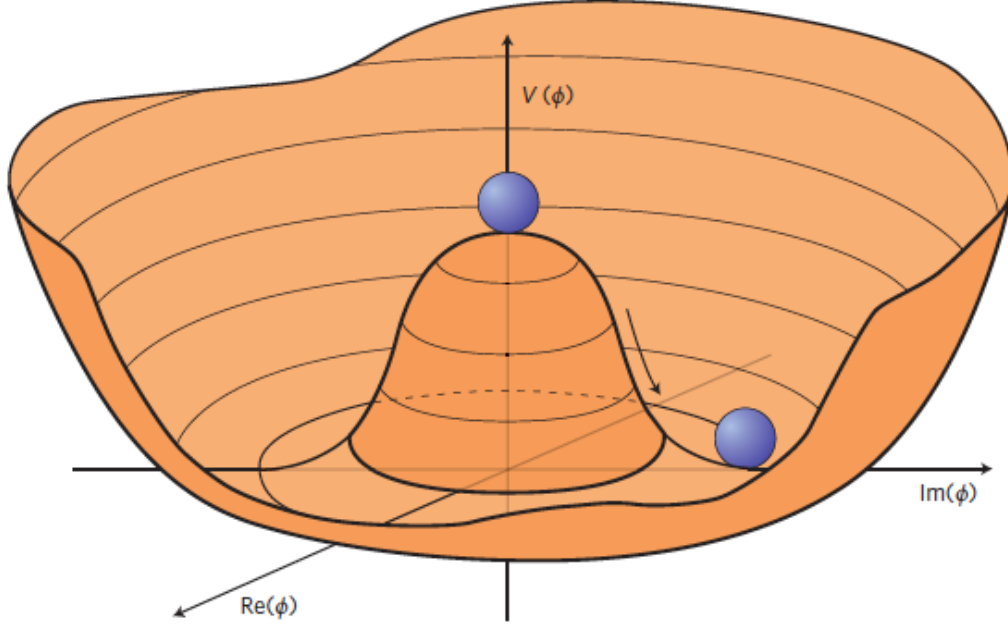


Figure 2.3: Illustration of Higgs potential when  $\mu^2 < 0$ . Figure source [15].

Then the Higgs field is re-defined and the ground state is defined by:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \quad (2.24)$$

where  $v$  is the VEV and  $H$  defines the Higgs field. After the SSB, the electroweak fields mix to give gauge boson fields that are massive:

$$\begin{aligned} W_\mu^\pm &= \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2), \\ Z_\mu &= \cos \theta_w W_\mu^3 - \sin \theta_w B_\mu, \\ A_\mu &= \sin \theta_w W_\mu^3 + \cos \theta_w B_\mu, \end{aligned} \quad (2.25)$$

where  $A_\mu$  corresponds to the photon field,  $\theta_w$  is the Weinberg mixing angle defined as  $\theta_w = \tan^{-1} \left( \frac{g'}{g_W} \right)$ . The masses of the gauge bosons and Higgs boson are then given by:

$$m_W = \frac{v}{2}g_W,$$

$$\begin{aligned}
m_Z &= \frac{m_W}{\cos \theta_w}, \\
m_A &= 0, \\
m_H &= \sqrt{2\lambda}v.
\end{aligned} \tag{2.26}$$

### 2.1.7 Need for beyond standard model physics

Even though the SM includes the Higgs boson and can explain many of the experimental observations, this theory is unable to account for all of the data. Each of the following constraints is addressed by one or more extensions to the SM Lagrangian.

- *Gravity:* One of the fundamental forces we encounter on a daily basis that the SM can not account for is gravity. The force mediator of gravity, called graviton, has been proposed as one extension to the SM.
- *Matter-Antimatter Asymmetry:* The SM predicts that the big bang should have produced roughly equal quantities of matter and anti-matter. Our universe is currently only dominated by matter. The puzzle of matter-antimatter asymmetry is: what happened to antimatter?
- *Dark Matter:* Furthermore, our universe is made of 95% dark matter and dark energy [16], and the SM, once again, lacks a dark matter particle or particles.
- *Neutrino Masses:* The SM proposes the neutrinos masses to be zero as there are no right-handed neutrinos in the SM, but the neutrino oscillations [17,18] prove that the neutrinos have nonzero masses.

- *Unification of the Forces:* The fundamental forces are predicted to unify around the planck energy scale, yet within the SM, these forces do not unify at a single point. The unification of these forces is predicted by an extension to the SM outlined in Section 2.2.2.
- *The Hierarchy Problem:* With quantum corrections, the Higgs boson mass should be many orders of magnitude higher than the current measured mass of 125 GeV. The observed mass at this energy scale can be explained by substantial adjustments or fine tuning (described in Section 2.2.1), which is unnatural and is known as the hierarchy problem.

## 2.2 Supersymmetry

SUSY is one of the theoretical extensions to the SM. This dissertation is concerned with the direct search for specific SUSY particles. This section seeks to provide a quick overview of SUSY, with a more extensive explanation in [19]. SUSY suggests that, for every SM particle, there exists a super companion with a spin difference of  $\frac{1}{2}$ . In other terms, SUSY is a space-time symmetry that transforms the boson and fermion fields using an operator.

$$\begin{aligned} Q|\text{Boson}\rangle &= |\text{Fermion}\rangle, \\ Q|\text{Fermion}\rangle &= |\text{Boson}\rangle. \end{aligned} \tag{2.27}$$

Figure 2.4 shows the SUSY particles in comparison with the SM particles. The present energy limits at the LHC range up to TeV scale for particle production, and since we have not found any signs of SUSY, this indicates that SUSY is a “broken” symmetry. If SUSY were a unbroken symmetry, SUSY particles would have the same masses as their counterpart SM particles. Broken symmetry proposes the masses

of most SUSY particles to be greater their SM counterparts with naturalness [20] conditions.

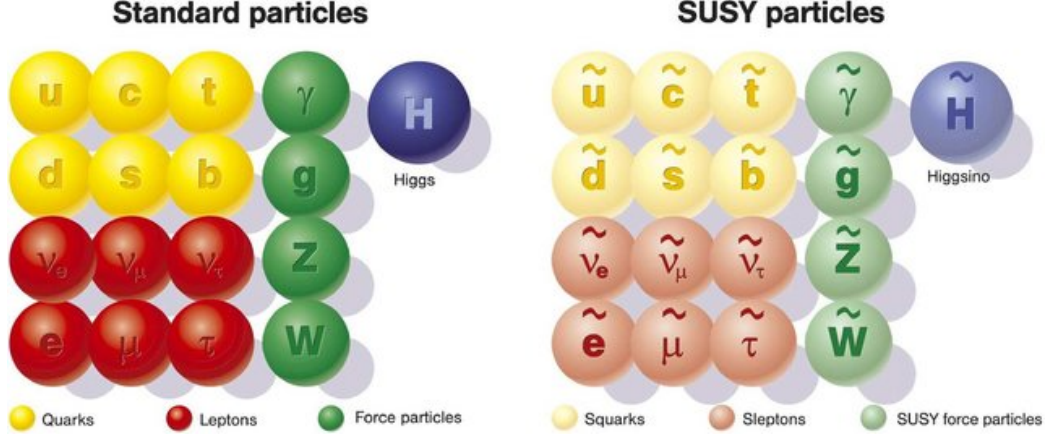


Figure 2.4: Summary of the SM particles and SUSY particles.

### 2.2.1 Fine-tuning problem

The concept of fine-tuning stems from iterative modifications in model parameters to align with data, and no well-known process explains why specific parameters should have certain values. One example is the calculation of the Higgs mass. The mass of the Higgs boson depends on the loop correction terms, with the top quark loop corrections contributing the most. The coupling of the Higgs to the top quark can be described by the Lagrangian as follows:

$$\mathcal{L} = -\frac{y_t}{\sqrt{2}}\phi\bar{t}_L t_R + h.c., \quad (2.28)$$

where  $y_t$  is the Yukawa coupling with the top quark. At low energy scale where the masses of the bosons are roughly on the same magnitude and gravity can be ignored, the couplings are valid. But at the highest scale where QFT holds, the Plack scale, top quark loop corrections are applied to the Higgs mass as illustrated in Fig. 2.5 and

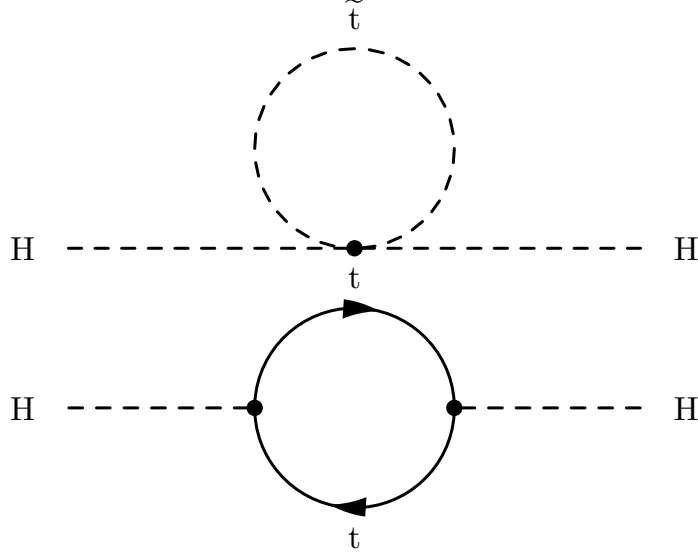


Figure 2.5: (Top) Corrections from the scalar partner of the top quark. (Bottom) Divergent loop corrections to the Higgs boson mass.

as the correction term in Eq. (2.29):

$$\Delta m_H^2|_{\text{top}} = -\frac{3|y_t|^2}{8\pi^2} \left[ \Lambda^2 - 3m_t^2 \ln \left( \frac{\Lambda^2 + m_t^2}{m_t^2} \right) \right], \quad (2.29)$$

where  $\Lambda$  is the Planck scale value. This contribution shows only the asymptotic terms as  $\Lambda \rightarrow \infty$ . These corrections are several orders of magnitude larger than the Higgs mass, so the bare mass of Higgs boson must be fine-tuned in order to derive the observed mass of Higgs boson. The Higgs bare mass and the observed Higgs mass is (with large corrections) are approximated by:

$$m_{Bare}^2 \approx 125^2 + \frac{3m_{top}^2}{4\pi^2 v^2} \Lambda^2 + \mathcal{O}(\Lambda) \approx (\mathcal{O}(10^{19}) \text{ GeV})^2, \quad (2.30)$$

$$m_H^2 = m_{Bare}^2 - \frac{3m_{top}^2}{4\pi^2 v^2} \Lambda^2 + \mathcal{O}(\Lambda) \approx (125 \text{ GeV})^2, \quad (2.31)$$

where  $m_H$  denotes the observed Higgs boson mass. This excessive fine-tuning can be avoided by introducing a scalar supersymmetric top quark (also referred as top squark) ( $\tilde{t}$ ) with identical Yukawa coupling as top quark which can cancel the loop

corrections. The introduction of this scalar top quark is one of the building blocks for SUSY.

### 2.2.2 Minimal supersymmetric standard model

The concept of SUSY is rather complicated to test for at the colliders; therefore a formalism is developed that includes SUSY taking into account minimum amount of fields that is based on the SM gauge symmetry  $(\text{SU}(3)_C \otimes \text{SU}(2)_L \otimes \text{U}(1)_Y)$  known as minimal supersymmetric standard model (MSSM) [19]. Similar to the SM particles, the quarks and leptons have their superpartners that consist two-component Weyl fermion and a complex scalar field. These particles are denoted by having an “s” prepended to the fermion’s name, e.g. the superpartner of bottom quark is called “sbottom”. The electroweak bosons have their spin- $\frac{1}{2}$  superpartners ( $\widetilde{W}^\pm, \widetilde{W}^0, \widetilde{B}^0$ ) called as winos and binos. The list of the MSSM fields is shown in Table 2.4.

Similar to the SM electroweak symmetry breaking, various gauge fields mix to form MSSM particles [19]. In this dissertation, we are searching for the direct production of charginos and neutralinos. The superpartner of the Higgs boson is called a higgsino. Charginos ( $\tilde{\chi}_1^\pm, \tilde{\chi}_2^\pm$ ) are the mass eigenstates formed by mixing the gauge eigenstates of charged winos and charged higgsinos. Neutralinos ( $\tilde{\chi}_1^0, \tilde{\chi}_2^0, \tilde{\chi}_3^0, \tilde{\chi}_4^0$ ) are formed by mixing the gauge eigenstates of neutral winos, binos and two higgsinos. A single higgsino would lead to gauge anomaly since addition of Higgs superpartner is weak isospin doublet with hypercharge ( $Y = \pm\frac{1}{2}$ ). Therefore, another Higgs doublet is introduced with spin-0 having the opposite hypercharge to the SM Higgs doublet.

Table 2.4: Gauge and mass eigenstates in MSSM [19]

Names	Spin	Gauge eigenstates	Mass eigenstates
squarks	0	$\tilde{u}_L \tilde{u}_R \tilde{d}_L \tilde{d}_R$	same
squarks	0	$\tilde{c}_L \tilde{c}_R \tilde{s}_L \tilde{s}_R$	same
squarks	0	$\tilde{t}_L \tilde{t}_R \tilde{b}_L \tilde{b}_R$	$\tilde{t}_1 \tilde{t}_2 \tilde{b}_1 \tilde{b}_2$
sleptons	0	$\tilde{e}_L \tilde{e}_R \tilde{\nu}_e$	same
sleptons	0	$\tilde{\mu}_L \tilde{\mu}_R \tilde{\nu}_\mu$	same
sleptons	0	$\tilde{\tau}_L \tilde{\tau}_R \tilde{\nu}_\tau$	$\tilde{\tau}_1 \tilde{\tau}_2 \tilde{\nu}_\tau$
neutralinos	1/2	$\tilde{B}^0 \tilde{W}^0 \tilde{H}_u^0 \tilde{H}_d^0$	$\tilde{\chi}_1^0 \tilde{\chi}_2^0 \tilde{\chi}_3^0 \tilde{\chi}_4^0$
charginos	1/2	$\tilde{W}^\pm \tilde{H}_u^\pm \tilde{H}_d^\pm$	$\tilde{\chi}_1^\pm \tilde{\chi}_2^\pm$
gluino	1/2	$\tilde{g}$	same
Higgs bosons	0	$\tilde{H}_u^0 \tilde{H}_d^0 \tilde{H}_u^\pm \tilde{H}_d^\pm$	$h^0 H^0 A^0 H^\pm$
gravitino	3/2	$\tilde{G}$	same

(2.32)

### 2.2.3 *R*-parity conserving SUSY

In the MSSM, baryon number and lepton number are no longer conserved by renormalizable couplings in the theory. Multiplicative quantum number is introduced for every particle based on its baryon and lepton numbers which is called as *R*-parity and given by:

$$P_R = (-1)^{3B-L+2s} \quad (2.33)$$

where  $s$  is the spin,  $B$  corresponds to baryon number, and  $L$  corresponds to lepton number. The *R*-parity for all the SM particles corresponds to  $P_R = 1$ , while for the SUSY particles corresponds to  $P_R = -1$ . In this dissertation, we consider that *R*-parity is conserved, which implies:

- At any vertex this number must be conserved, and consequently supersymmetric particles are produced in pairs.
- A supersymmetric particle can never decay only into SM particles.

- The lightest supersymmetric particle (LSP), which in this dissertation is the lightest neutralino ( $\tilde{\chi}_1^0$ ), must be weakly interacting, stable, and escapes our detector. This LSP could be a candidate for dark matter.

The signal models we chose to investigate have two LSPs in the final state that elude the CMS detector. This indicates that the momentum in a collision event is unbalanced. When collision events are sorted into a dataset, there is a requirement on the missing transverse momentum in the events, as explained in Section 3.2.7, which narrows our search for these types of events.

#### 2.2.4 Electroweak SUSY searches

Supersymmetric particle searches have not yet been successful; however, many SUSY analyses are becoming more sensitive to a wide variety of SUSY signal models. Some SUSY searches emphasize gluino pair production, which has the largest cross section for SUSY particle formation as shown in Fig. 2.7. If the LSP is stable in these searches,  $R$ -parity is conserved, resulting in missing transverse momentum and the gluinos lead to numerous quark pairs, which ultimately result in a large number of jets in the detector. The decay of the gluino can produce a quark-antiquark pair. Figure 2.6 displays some of the mass exclusions from recent searches for gluinos decaying into top-antitop quark pair (along with other light quark pair) by the CMS experiment. As shown in the bottom plot of Fig. 2.6, the strongest exclusion limits now imposed for the gluino pair production are up to 2300 GeV.

This dissertation is driven by the desire to look for electroweak production of charginos and neutralinos. Searches for direct production of charginos and neutralinos are challenging due to its low cross section compared to strong production



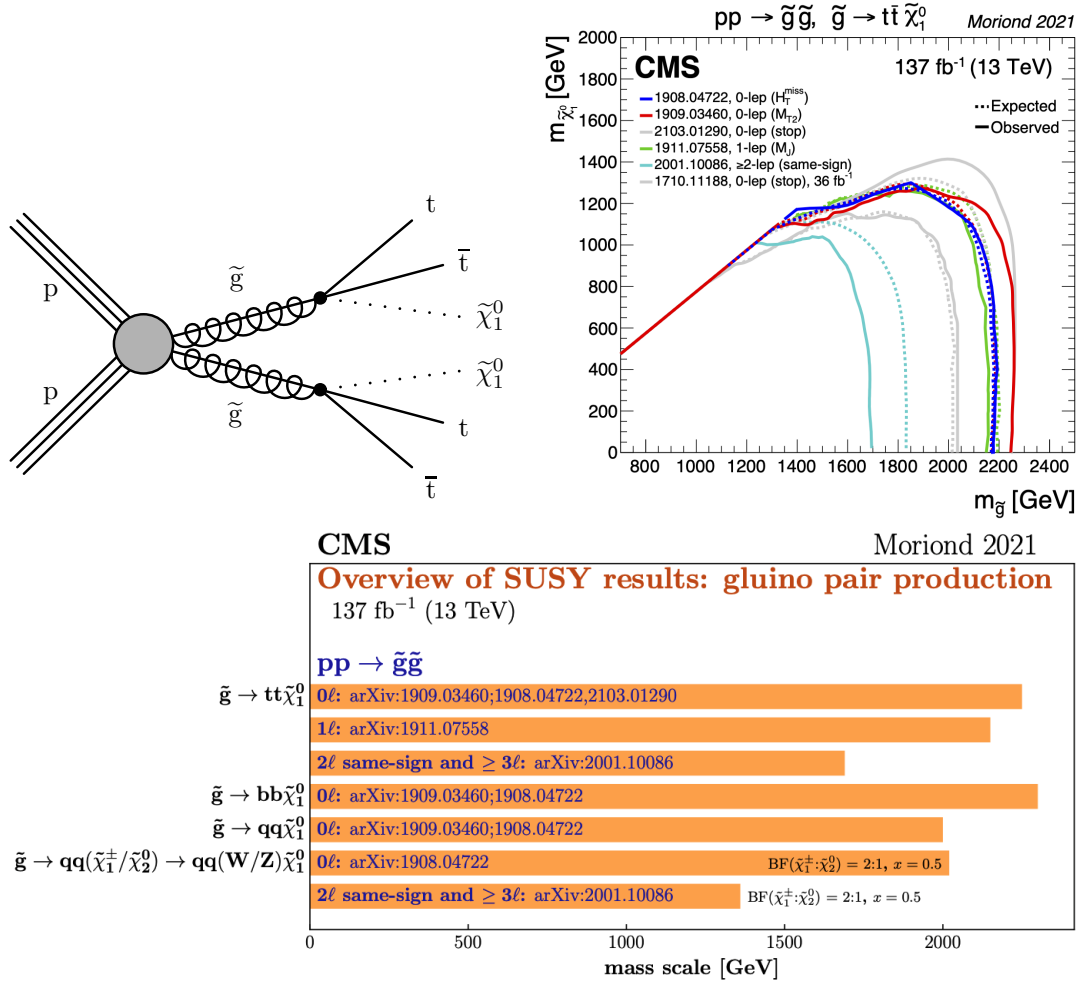


Figure 2.6: (Top left) Diagram showing the decays of gluinos into top-antitop quark pairs along with an LSP. (Top right) Mass exclusion limits placed at 95% CL obtained in the context of simplified model shown above with each line corresponds to the limits set by the analysis performed indicated by their arXiv numbers. (Bottom) Overview of the mass reach obtained by gluino pair production through various decay models of gluino using Run 2 CMS data. Figure source [21].

of gluinos and squarks, as show in Fig. 2.7. In the search performed in Chapter Five, the electroweak SUSY particles decay into SM gauge boson or Higgs boson along with LSP. Figure 2.8 displays a few electroweak SUSY models.

The excellent resolution of the CMS tracker, ECAL, and the muon detector, as mentioned in Chapter Three, results in clean signatures for searches involving leptonic final states. The hadronic searches aid in extending search sensitivities to higher

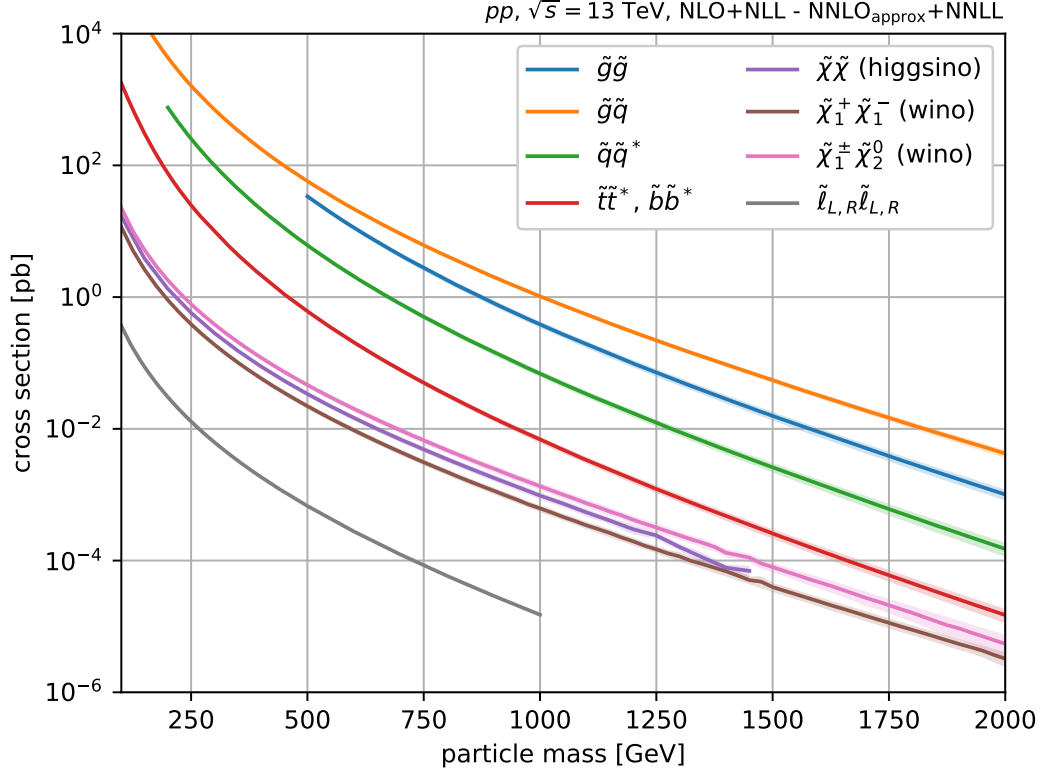


Figure 2.7: Cross sections for SUSY particles shown as a function of particle mass. Figure source [22].

chargino/neutralino masses, as shown in Fig. 2.9, which is represented by the curves with the ‘ $0\ell'$ ’ label. Searches in leptonic final states are effective for electroweakino models at the low chargino/neutralino mass range and at high LSP masses.

The mass exclusion limits set by the search described in Chapter Five for the mass of  $\tilde{\chi}_2^0$  extends up to 970 GeV, and the mass of  $\tilde{\chi}_1^0$  up to 250 GeV. More information on these findings is covered in the dissertation’s analysis section and also in this reference [23]. Other electroweak analyses that used Run 2 CMS data to search for the electroweak SUSY models are listed in these references [23–27].

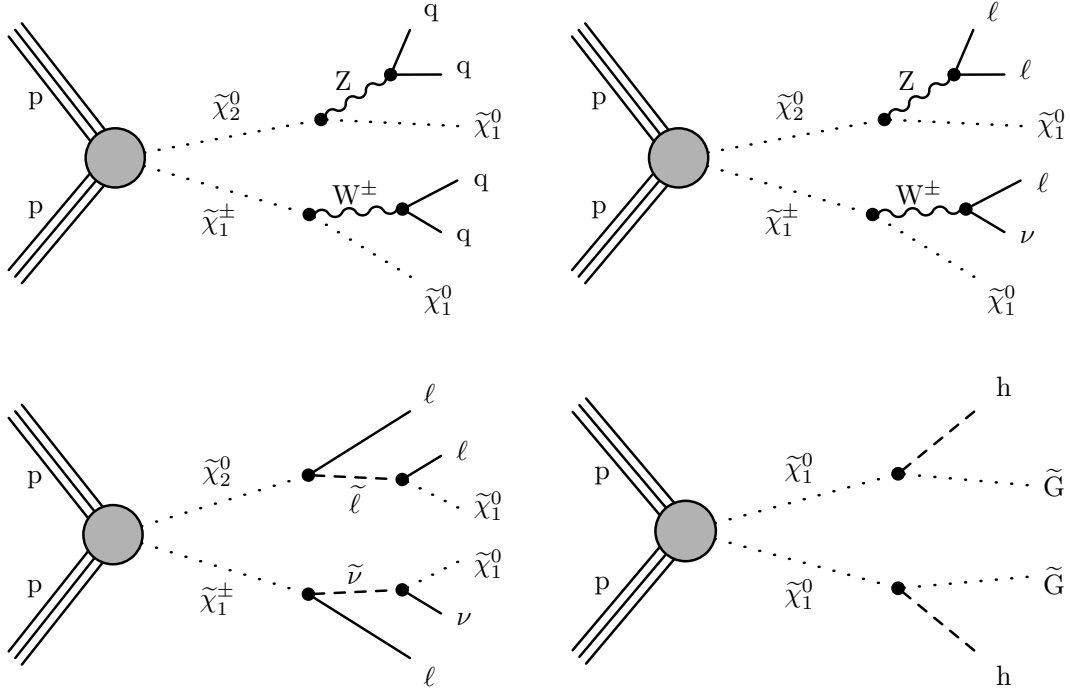


Figure 2.8: Diagrams for several scenarios of direct electroweak production of charginos and/or neutralinos. (Top) Diagrams of direct production of charginos and neutralinos where they decay to SM bosons and  $\tilde{\chi}_1^0$ , and SM boson decaying hadronically (left) or leptonically (right). (Bottom) Diagrams of direct production of charginos and neutralinos where they finally decay to SM leptons and  $\tilde{\chi}_1^0$  (left), or Higgs boson and  $\tilde{\chi}_1^0$  (right).

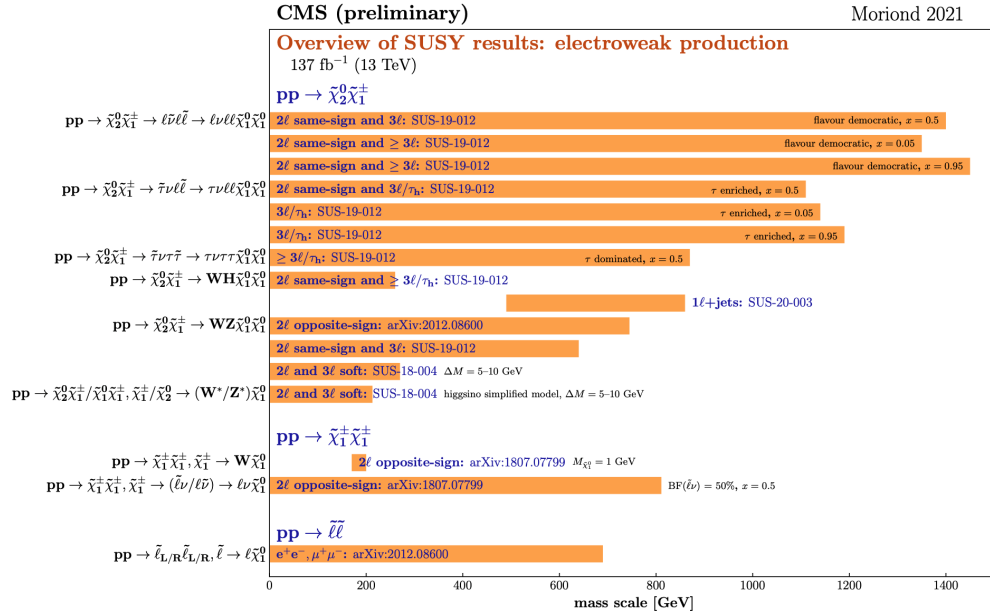
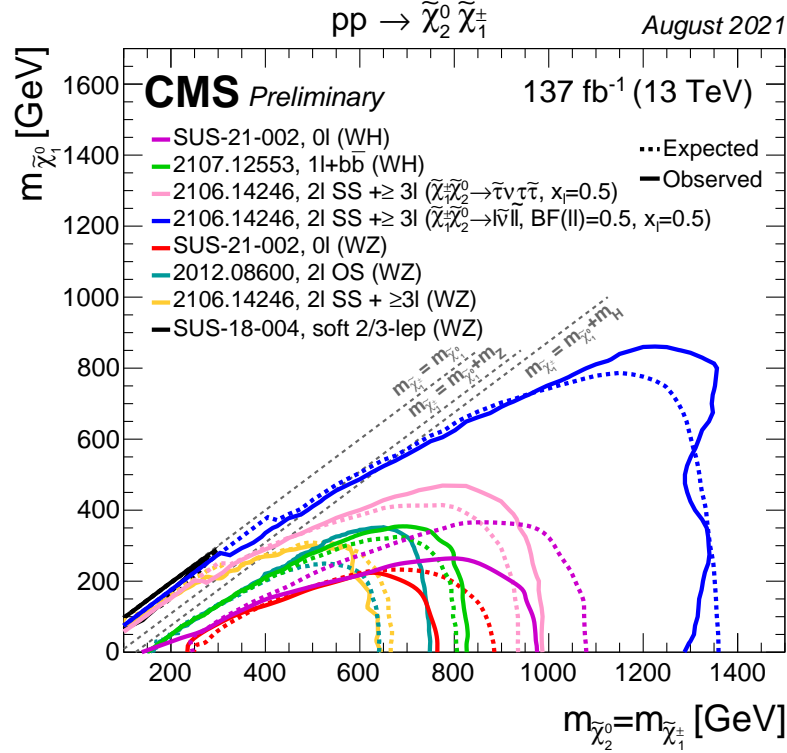


Figure 2.9: (Top) The latest exclusion limits by various electroweak SUSY analyses identified by their arXiv numbers (SUS-21-002 corresponds to this analysis). (Bottom) The mass reach by various electroweak production. Figure source [21].

## CHAPTER THREE

### Experimental apparatus

One of the most effective ways to explore the fundamental particles and their interactions is colliding protons at high energies, which enables the identification of fundamental particles. While fixed target experiments involve colliding proton beams with a stationary target, the energy produced in these collisions is lower than when two beams collide head-on. Currently, the Large Hadron Collider (LHC) is the most powerful and largest particle collider equipped with multiple particle detectors, providing valuable insights into the properties of these particles and their behavior during collisions. One of them is the Compact Muon Solenoid (CMS) detector. This dissertation is based on the data collected by the CMS detector during the years 2016–2018.

#### *3.1 The Large Hadron Collider*

The LHC, which is situated at the European Organization for Nuclear Research (CERN) around Geneva, Switzerland, is the largest circular particle accelerator ever constructed. It is approximately 27 kilometers in circumference and located about 100 meters beneath the ground. The LHC has been designed to accelerate protons to energies of up to 7 TeV, while maintaining a peak luminosity of  $10^{34} \text{ cm}^{-2}\text{s}^{-1}$ . The proton beams are brought into collision at four different interaction points, where four major experiments are located to observe and record the decay products of the protons. The accelerator complex along with the major experiments is shown in Fig. 3.1.

## The CERN accelerator complex *Complexe des accélérateurs du CERN*

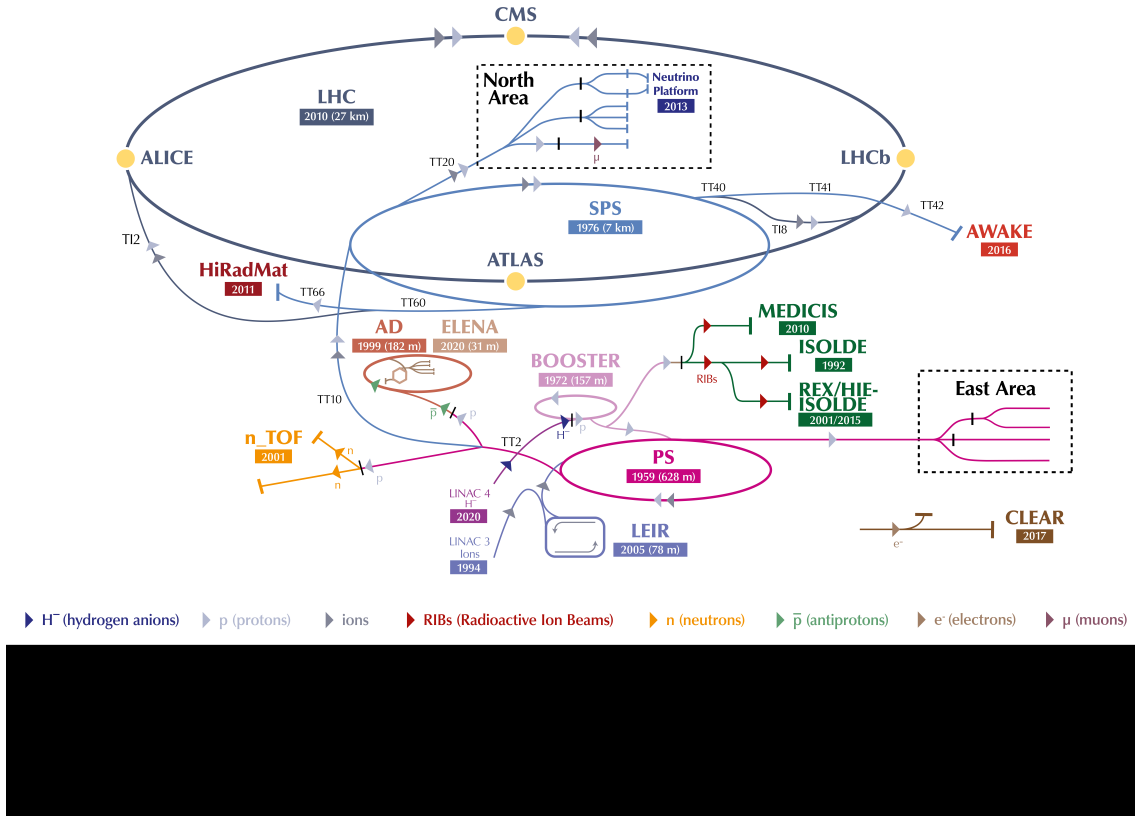


Figure 3.1: The LHC ring overview and the accelerator chain complex. Figure source [28].

Hydrogen gas serves as the starting point for the entire chain of processes. When a duoplasmatron is filled with hydrogen gas, the electrons from a heated cathode are used to split  $H_2$  molecules into  $H^+$  ions and protons. The proton energy is then raised to 50 MeV by accelerating these protons using the LINAC 2 linear accelerator. The next stage is to use a tiny circular accelerator called a proton synchrotron booster to raise the energy even higher, up to 1.4 GeV. Protons from the booster are further accelerated by the proton synchrotron to an energy of 25 GeV before being gathered together and prepared for injection into the super proton synchrotron (SPS). Protons within the SPS have energies of 450 GeV and are spaced 25

nanoseconds apart. Two proton beams are injected into the LHC ring and made to circulate in opposite directions. While the LHC has the capability to accelerate heavy ions such as Pb and Ar, for the purpose of this dissertation, only proton collisions are of relevance. The LHC is equipped with four detectors, namely ALICE, ATLAS, CMS, and LHCb, situated at four specific locations along the LHC ring where the beams intersect (known as bunch crossing). A precise set of beam parameters must be carefully adjusted to ensure the successful collision of the proton beams. The details of these beam parameters are discussed in the subsequent subsection.

### 3.1.1 *Tuning the LHC beam*

Once the beam is injected from the SPS into the large and circular LHC ring, it is crucial for the beam to maintain its circular trajectory. This is achieved with the help of 1232 superconducting dipole magnets, which are made of niobium-titanium (Nb-Ti) coils. These magnets are capable of generating a magnetic field of 8.4 T by cooling them to a temperature of 1.9 K, as depicted in Fig. 3.2.

Using 16 radio frequency (RF) coils that run at a frequency of 400 MHz, the proton's energy is raised to 6.5 TeV. The beam reaches its maximum energy after traveling through the cavities  $10^7$  times for about 20 minutes. After this acceleration step, the beam undergoes squeezing at the detector entrances using 392 quadrupole magnets, which have four symmetrically arranged magnetic poles to focus the beam sharply (nearly  $16\text{ }\mu\text{m}$ ) at the center of the detector. One of the key operational parameters of the LHC is the instantaneous luminosity ( $\mathcal{L}$ ). For proton colliders, the instantaneous luminosity can be expressed by:

$$\mathcal{L} = \frac{N^2 n_b f}{4\pi \rho_x \rho_y}, \quad (3.1)$$

## LHC DIPOLE : STANDARD CROSS-SECTION

CERN AC/DI/MM - HE107 - 30 04 1999

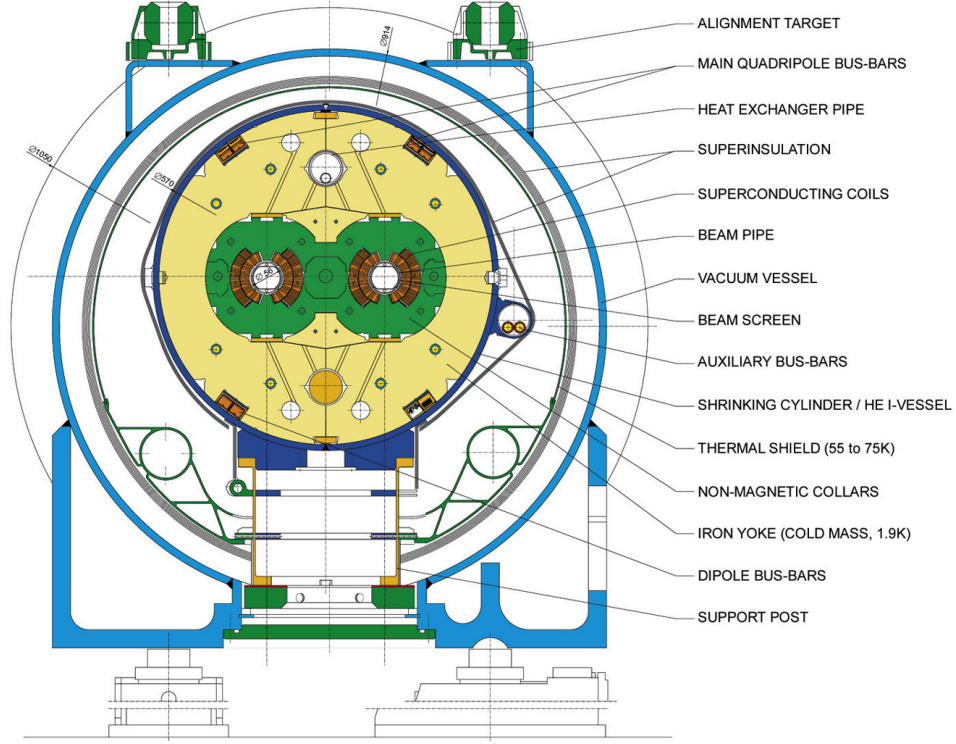


Figure 3.2: Cross-section of the LHC dipole magnet. Figure source [29].

where  $n_b$  are the total number of proton bunches,  $N$  is the number of protons per bunch, and  $f$  is the frequency of revolution. The quantities  $\rho_x$  and  $\rho_y$  refer to the transverse spreads of the beam in  $x$  and  $y$  directions at the interaction point. The number of events  $N_{\text{events}}$  expected for a process with cross section of  $\sigma$  is given by:

$$N_{\text{events}} = \sigma \int \mathcal{L} dt. \quad (3.2)$$

The integrated luminosity  $\int \mathcal{L} dt$  is used to describe the amount of collected data during a data-taking run.

The total integrated luminosity is displayed in Fig. 3.3, where the delivered integrated luminosity measures the volume of data sent from the LHC to the CMS. Occasionally, the CMS detector must stop for technical reasons, making it impossible



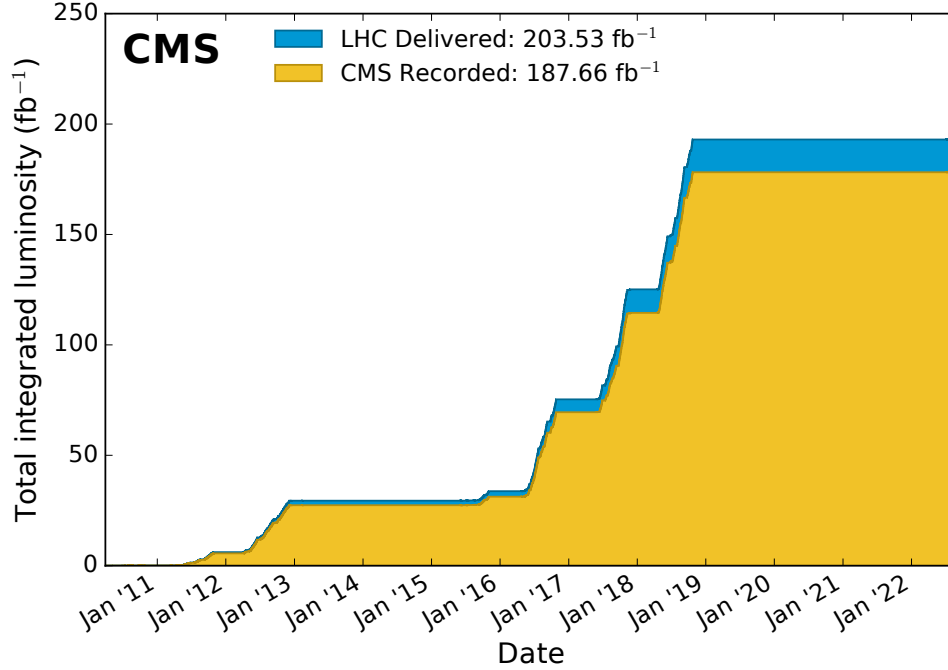


Figure 3.3: Total delivered and recorded integrated luminosities for CMS from the beginning of the LHC until the beginning of 2022. This dissertation is based on the data collected during the 2016–2018 period. Figure source [30].

to capture all collisions. The amount of proton-proton collision data that the CMS detector has recorded is shown as the recorded integrated luminosity.

Multiple proton-proton collisions can occur during a proton-proton bunch crossing. These additional collisions other than the one of our main interest are referred to as in-time pileup (PU). Out-of-time pileup refers to the additional proton-proton collisions which occur just before or after the proton bunch crossing of our main interest. By changing the bunch spacing, the number of protons in a bunch, and the strength with which the proton bunches are squeezed before the collision, the mean number of pileup interactions can be changed. Pileup may appear to be detrimental because it might conceal interesting physics from the primary proton-proton

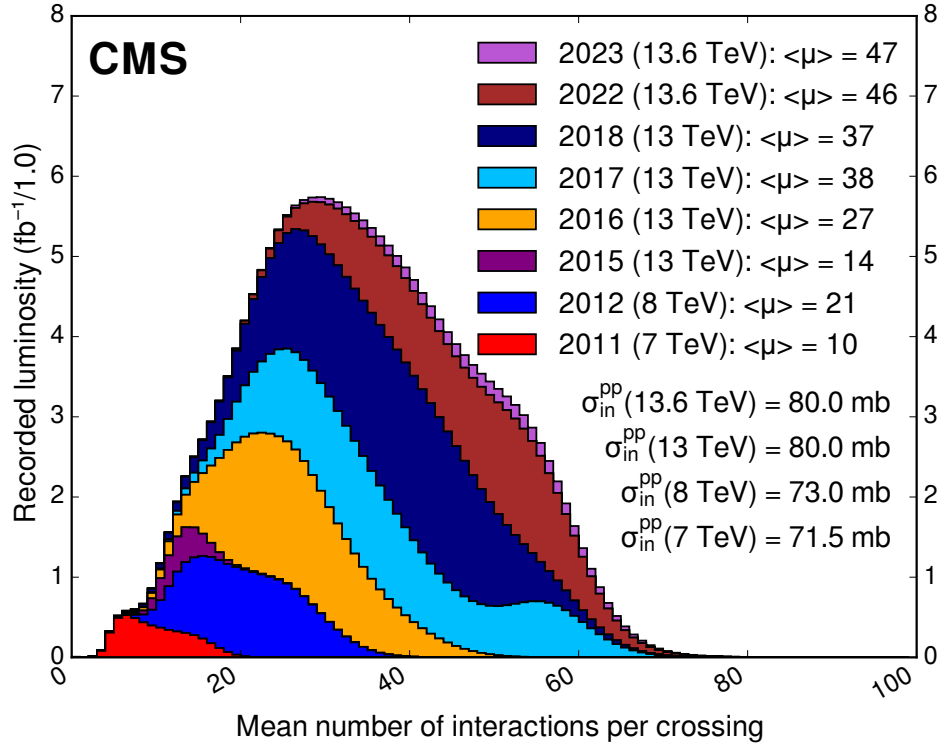


Figure 3.4: Distribution of the mean number of interactions per proton bunch crossing for various data-taking years. The average number of interactions per year is also shown. Figure source [30].

collision, but it actually has the advantage of increasing the likelihood of developing interesting and rare processes. During the initial data-taking runs at the LHC, the mean number of pileup interactions was relatively small. However; it has increased over time, as well as the collision energy, as shown in Fig. 3.4. With improved detector capabilities and data analysis software, physicists managed to handle data with increasing levels of pileup interactions.

### 3.2 The Compact Muon Solenoid experiment

The CMS detector is one of the general-purpose detectors situated at the LHC. It is considered “compact” as both the tracker and the calorimeters are designed to fit seamlessly within the solenoid magnet, and it is both massive and dense. Despite

weighing approximately 14,000 tonnes, the detector measures 28.7 meters in length and 15 meters in diameter. It is located near Cessy, France, at interaction point 5 (P5), approximately 100 meters below the ground level. The solenoid magnet generates a magnetic field of 3.8 Tesla (T) which is essential to bend the track of the charged particles coming out of the collision and is used to precisely measure its momenta.

Figure 3.5 shows a sectional view of the cylindrical-shaped CMS detector, with subsystems interleaved between each other. The closest subdetector to the beam pipe is the silicon tracker, which is enclosed by the electromagnetic calorimeter (ECAL) and then the hadronic calorimeter (HCAL). The muon system consists of three subsystems and are located outside the magnet. Aluminum drift tubes (DT) are arranged in the barrel ( $|\eta| < 1.3$ ) region. Cathode strip chamber(CSC) in the endcaps ( $0.9 < |\eta| < 2.4$ ) provide good position and momentum resolution for muons. Resistive plate chambers (RPC) are spread throughout the barrel and endcap regions to provide excellent timing resolution. The CMS detector subsystems are described in detail in the following sections.

### *3.2.1 Detector geometry and variables*

The coordinate system is defined with its origin at the nominal collision point, and this coordinate system is used to reference multiple portions of the detector. The  $x$  axis is oriented toward the center of the LHC ring, the  $y$  axis is oriented vertically upward, and the  $z$  axis is oriented anticlockwise along the beam direction when viewed from above. Cylindrical coordinates are often used with the azimuthal angle  $\phi$ , defined in the  $x$ - $y$  plane. The polar angle  $\theta$ , with respect to the  $z$  axis is

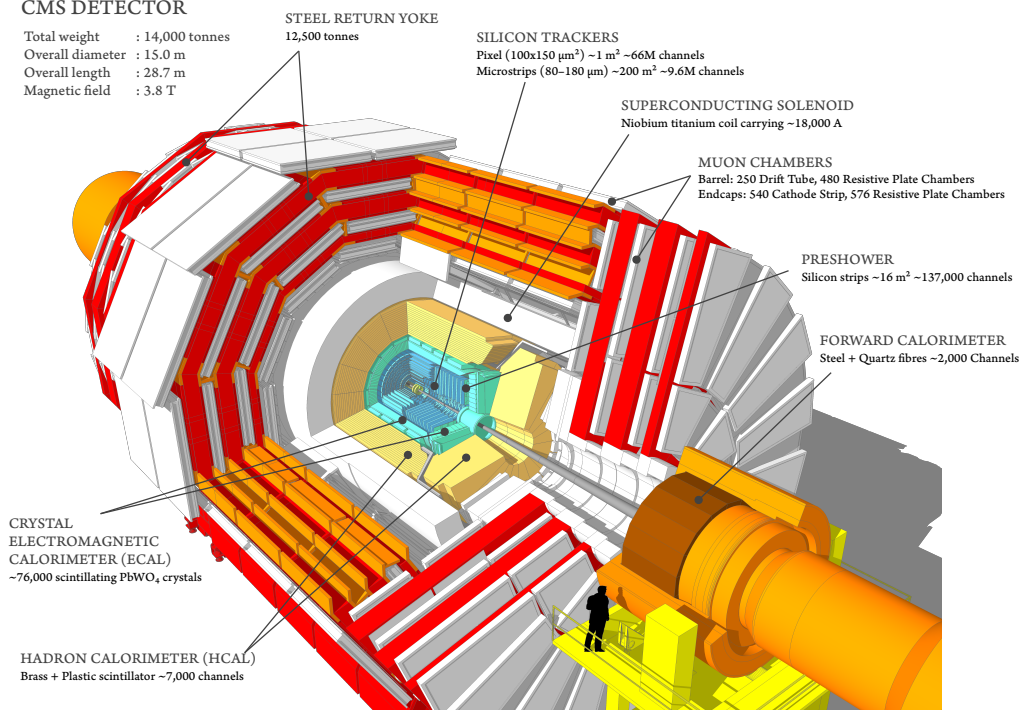


Figure 3.5: A sectional view of the CMS detector with various subsystems. To illustrate the scale, a person is shown in the image. Figure source [31].

often expressed in terms of pseudorapidity ( $\eta$ ).

$$\eta = -\ln \left[ \tan \left( \frac{\theta}{2} \right) \right]. \quad (3.3)$$

Pseudorapidity is an estimate for rapidity variable ( $y$ ) in the limit  $E \approx |\mathbf{p}|$ , which is defined as:

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right), \quad (3.4)$$

where  $E$  is the energy of the particle and  $p_z$  is the momentum of the particle along the  $z$  axis. The quantities  $E$  and  $p_z$  are difficult to measure precisely when the mass of the particles is unknown. Difference in rapidity is Lorentz invariant and the  $z$ -component of a particle's momentum cannot be measured accurately. In the case,  $E \gg m$ , or transverse momentum in  $x$ - $y$  plane ( $p_T$ )  $\gg m$ , rapidity can be approximated by pseudorapidity, which is often used at collider experiments. Another variable defined

in terms of  $\eta$  is the angular separation ( $\Delta R$ ) which is a parameter used to quantify the distance between particles.

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}. \quad (3.5)$$

The initial momentum of the colliding partons in the transverse ( $x$ - $y$ ) plane is almost zero because the beam is injected through the  $z$  axis. Momenta of the collision products are often measured in the transverse plane, since  $p_z$  of the colliding partons is unknown.

### 3.2.2 Tracker

Following the proton-proton collisions, the collision products pass through the beam pipe and interact with the inner tracker at first as it is closest to the interaction point. The tracker is used to determine the momentum of charged particles with  $p_T > 1$  GeV. Hits in numerous layers of the tracker material define the trajectory of the charged particles. It also helps in locating the primary pp collision vertex as well as the secondary vertex if any particle decay occurs after it has traversed some distance away from the primary vertex.

Silicon is used as detector material since it has high carrier mobility and small band gap which is important to have minimal interactions with the incident particles because of the large flux of incoming particles. Also, it is crucial to keep the bremsstrahlung, nuclear interactions, and particle energy loss to a minimal level since there are other subdetectors upstream which are designed to measure the energy and other properties.

The tracker is composed of multiple concentric layers of Silicon (Si) modules, as illustrated in Fig. 3.6. It comprises two main components, namely the pixel tracker

and strip tracker. The pixel detector consists of three barrel layers (BPix) and two endcap disks (FPix) situated very close to the interaction point. With a pixel size of  $100 \times 150 \mu\text{m}^2$ , there are over 1400 small Si modules. Geometrically, the three barrel layers are positioned at 4.4 cm, 7.3 cm, and 10.2 cm from the beam line. The endcap disks, on the other hand, are located at  $z = \pm 34.5$  cm and  $z = \pm 46.5$  cm from the center of the detector, with inner and outer radii of 6 cm and 15 cm, respectively.

The Si strip tracker surrounds the pixel and consists of over 15,000 Si strip modules consisting of the tracker inner barrel (TIB) with four layers and tracker inner disks (TID), which have two disc modules. Further away is the tracker outer barrel (TOB) with six module layers and tracker endcaps (TEC) that are aligned vertically similar to TID with nine disk modules. The track reconstruction efficiency is about 99% for muons with  $p_T$  in the range 1–100 GeV and lower for the charged pions because of the interactions with the tracker material.

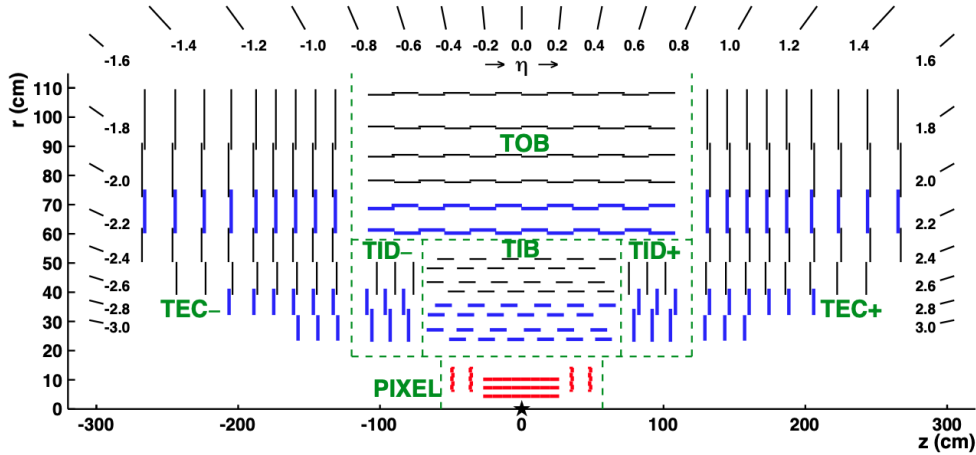


Figure 3.6: A sketch of the tracker system consisting of the pixel detector and outer Si strip tracker showing the upper part only. The tracker extends its coverage up to  $|\eta| = 2.4$ . Figure source [32].

### 3.2.3 Superconducting magnet

The CMS magnet has an inner diameter of 5.9 m and length of 12.5 m which is made up of Nb-Ti coils and can generate a magnetic field of 3.8 T. This strong magnetic field plays crucial role in bending the trajectory of the charged particles (electrons, charged hadrons, and muons) and identifying the charge of the particles. The magnetic flux is made homogeneous with the help of 10 kilo ton iron return yoke. The magnet is operated at 4.6 K with the help of liquid helium and 19.5 kA of current flows through the solenoid.

### 3.2.4 Electromagnetic calorimeter

The ECAL is compact, granular, and radiation tolerant. The main purpose of ECAL is to measure the energy and position of electromagnetically interacting particles such as the electrons, photons, and the electromagnetic particle component of jets. Unlike the tracker, which does not absorb energy of incident particles, ECAL is a calorimeter that absorbs most of the electromagnetically interacting particle's energy producing light signals and thus being destructive.

ECAL is made up of 61,200 lead tungstate ( $\text{PbWO}_4$ ) crystals in the ECAL barrel (EB) region and 7,324 crystals in the ECAL endcap(EE) region. The image of a lead tungstate crystal is shown in top right image of Fig. 3.7. The crystals have a dimension of  $28.6 \times 28.6 \times 220$  mm. The EB has pseudorapidity coverage up to  $|\eta| < 1.479$  and the EE further extends the coverage up to  $|\eta| < 3.0$ . To identify the neutral pions, a preshower detector (ES) is placed in the beginning of EE with  $1.653 < |\eta| < 2.6$  consisting of two lead radiations and Si strips that are orthogonal to each other. The sketch of the ECAL components with the crystals is shown in the

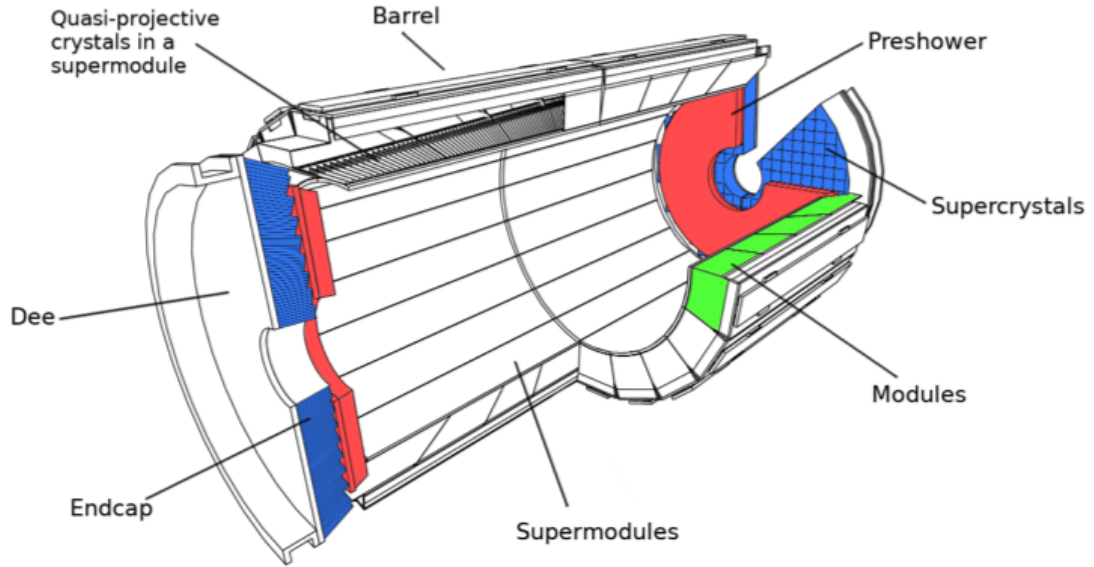
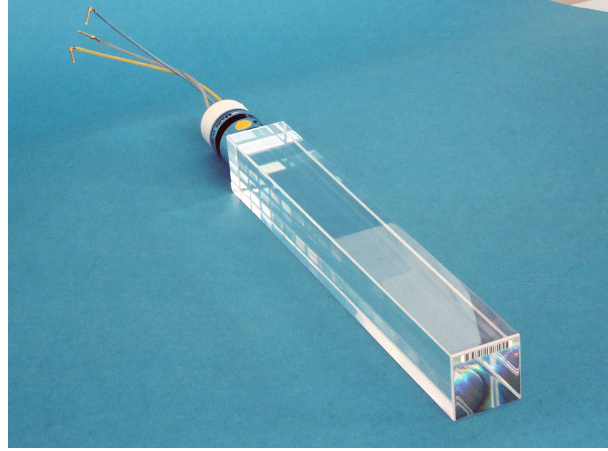
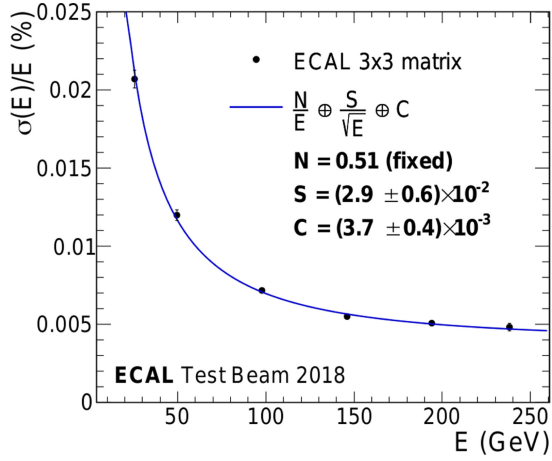


Figure 3.7: (Top left) Energy resolution in a  $3 \times 3$  cm crystal as a function of energy. (Top right) Image of a lead tungstate (PbWO<sub>4</sub>) crystal. (Bottom) Schematic of ECAL showing the barrel, endcaps and the preshower [33,34].



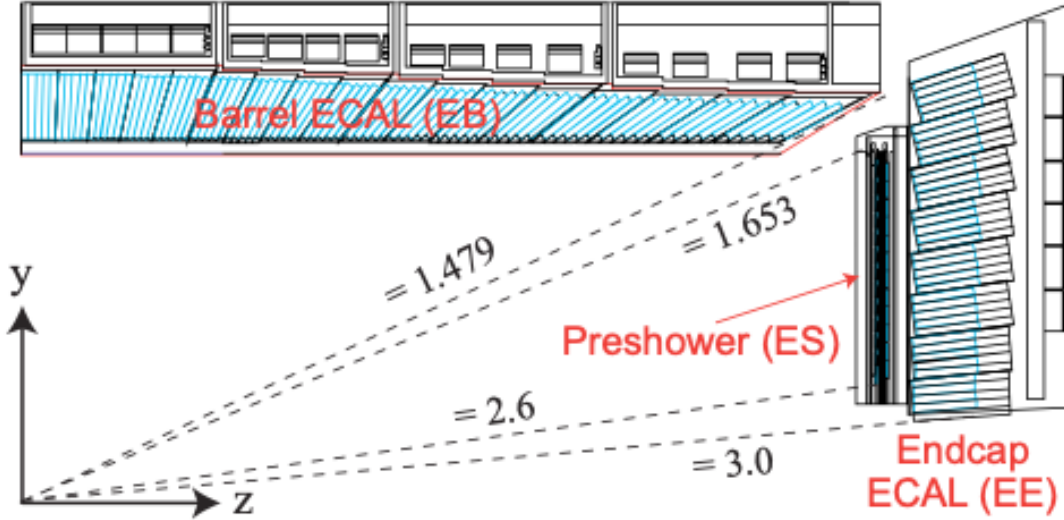


Figure 3.8: A sketch showing the coverage of ECAL subsystem. Figure source [35].

bottom picture of Fig. 3.7 and coverage of the barrel and endcap of ECAL is shown in Fig. 3.8. The granularity of the Si strips helps in identifying the photons from a Higgs decay and a neutral pion decay.

When a particle interacts with ECAL, it causes the  $\text{PbWO}_4$  supercrystals, to emit an electromagnetic shower. After interacting with the ECAL crystals, photons above a threshold energy (usually in the range of a few MeV) produce electron-positron pairs and cause bremsstrahlung for incoming electrons. These processes will continue until the particle's energy drops below the 1 MeV threshold, at which point Compton scattering and the photoelectric effect will take control of the decay chain. A silicon avalanche photodiode (APD) in EB and a vacuum phototriode (VPT) in EE gather the light deposited in the crystal and transform it into electrical signals, which are then amplified for further processing. VPTs are utilized to survive high radiation levels because of the greater flux in the endcaps.

The energy resolution of the ECAL is experimentally measured by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2 \quad (3.6)$$

where  $S$  is the stochastic term,  $N$  is noise term, and  $C$  is a constant. The energy resolution improves for higher energy particles as seen from the above equation and a plot showing the performance in recent test beam experiment is shown in top left plot of Fig. 3.7. The  $S$  term is affected by the shower fluctuations and the low number of events. The  $C$  term is due to the non-uniformity of light collection, energy leak from back of the crystal. The noise from the electronics, pileup contribution, and digitization contribute to the noise term  $N$ . Adding the energy from  $3 \times 3$  ECAL crystals, the typical values for these quantities are  $S = 2.8\%$ ,  $N = 0.12$  GeV,  $C = 0.3\%$ .

### 3.2.5 Hadronic calorimeter

The HCAL detector is located outside, all around the ECAL detector. HCAL is a sampling and heterogeneous calorimeter with alternate layers of absorber made of brass and plastic scintillator. The task of HCAL detector is to detect neutral and charged hadrons. When neutral and charged hadrons interact with the material nucleus in the HCAL detector, hadronic showers are produced. Eventually, the particles lose energy by depositing most of their energy in the absorber material; thereby, by producing a spray of particles. Light that is proportionate to the energy of the shower is produced by the plastic scintillator.

HCAL extends up to  $|\eta| = 5.2$  and is divided into barrel (HB), endcap (HE), forward (HF), and outer calorimeter (HO). The HB and HE form the center part of the detector and are located inside the magnet as shown in Fig. 3.9. HO is located outside

the solenoid magnet, and it is designed to capture the decays of very high energetic particles that pass through HB, but most of the particles get fully absorbed in HB. For the above three subdetectors, the material component remains the same except for the design. When the particles pass through the plastic scintillator, scintillation occurs and blue light is emitted which is captured by the wave length shifting (WLS) fibers that surround the scintillator tiles. The wavelength of the light is shifted (corresponding to the color shift) from blue light to green light to minimize the light loss from the sides of the fiber. The light is carried to hybrid photodiodes (HPDs) which convert the optical signal to digital signal. During the Run 1 and Run 2 LHC operation, the HPDs along with the scintillator tiles had suffered from radiation damage, and only the HPDs were replaced by silicon photomultipliers (SiPMs) that are radiation resistant. The HPDs were replaced by SiPMs in HE before 2018 data taking and in HB following the Run 2 LHC operation. After the end of Run 3 LHC operation, the entire HCAL endcap will be replaced by high granularity calorimeter.

HF is located in the forward region of  $|\eta| > 3.0$ . It is different from the rest of the HCAL subdetectors in the sense that it has around 100 km of quartz fibers as the active sampling medium enclosed within a steel absorber. As highly energetic particles traverse through the quartz fibers, Cherenkov light is produced and carried by the fibers into the photomultiplier tubes (PMTs) which convert the optic signal to a digital signal. Cherenkov light is emitted only when the particle's velocity is greater than the speed of light in the medium which it is traveling. The steel block helps in producing particle showers proportional to the energy of the incident particle and works best in high radiation environment. Half of the quartz fibers extend over the entire depth of HF (165 cm), whereas the remaining half commence at a distance of

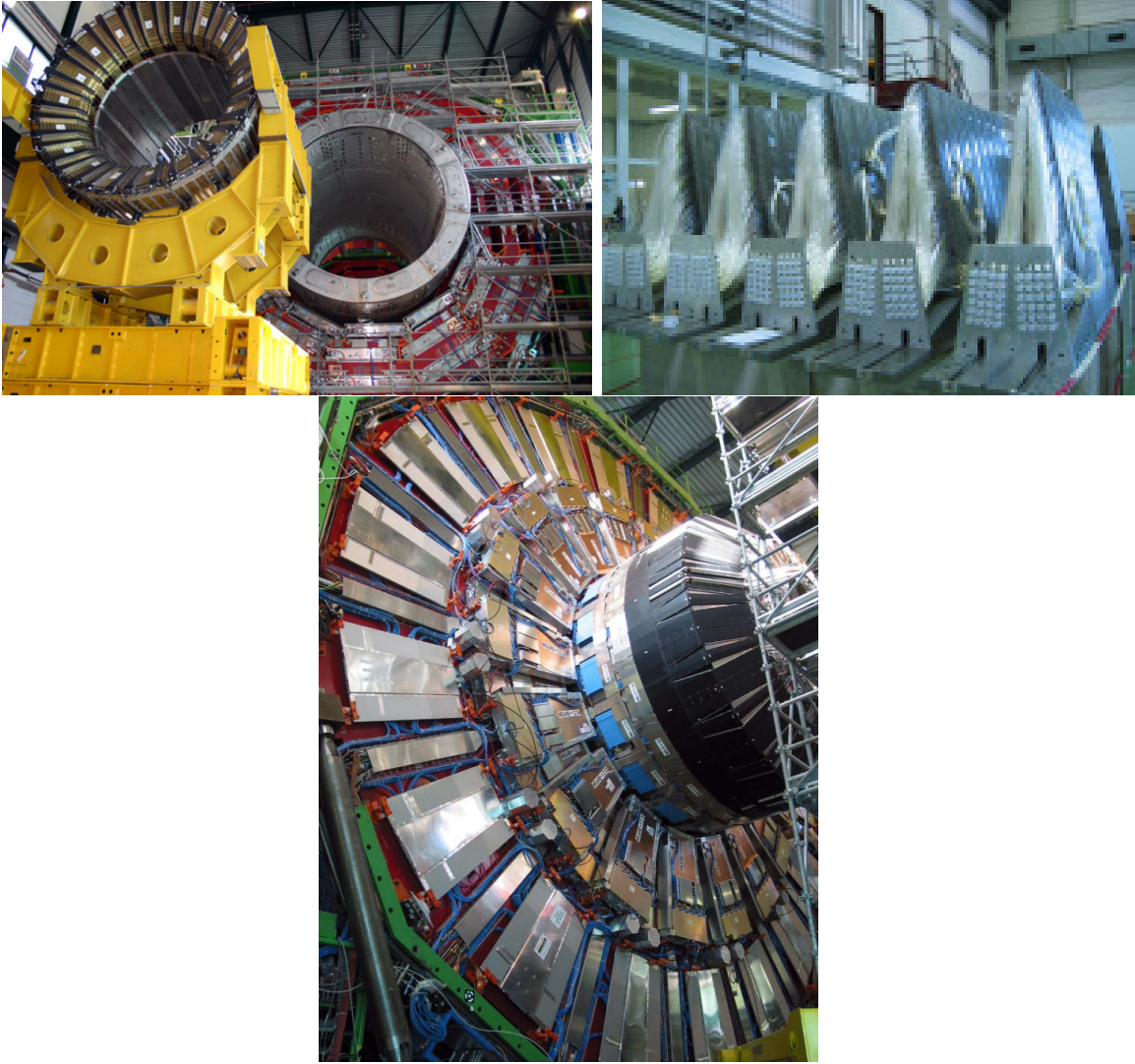


Figure 3.9: Photographs of HB installation into the magnet (top left), HF quartz fibers (top right), and HE installation (bottom). Image sources [36–38].

22 cm from the detector's front face, denoted as long and short fibers, respectively. This setup is designed to capture the behavior of electromagnetic showers generated by electrons and photons, which primarily deposit energy in the initial layers of HF, resulting in signals that predominantly activate long fibers. In contrast, hadronic showers generally occur later and continue deeper, primarily activating short fibers and, to a lesser extent, long fibers.

### 3.2.6 Muon system

By volume the muon system is the highest among the rest of subsystems. As the name suggests, it is used to detect muons that can survive traversing the sections of ECAL and HCAL subdetectors. The muons can be detected through the ionization effect it creates when they pass through the gas filled chambers of the muon system. Also, the direction of the magnetic field in this region is opposite to the direction of the magnetic field inside the magnet. The muon systems are classified into barrel and endcap regions and are composed of drift tubes (DTs), resistive plate chambers (RPCs), and cathode strip chambers (CSCs) as shown in Fig. 3.11.

As the muon passes through the gas chambers, it ionizes the gas in these chambers. The DTs are located in MB regions consist of gas ionization chambers with anode wire which are designed to measure the drift time of free electrons that are created after the muon passes ionizes the gas. Extending up to  $|\eta| < 1.2$ , there are about 250 DTs with each about  $2\text{ cm} \times 2.5\text{ m}$  in size with 1 or 2 RPCs attached to it, running parallel to the beam pipe as shown in Fig 3.11. There are 610 RPCs that are used to identify the position of the muon as it passes through two highly resistive plates charged oppositely with gas filled between them. The free electrons created

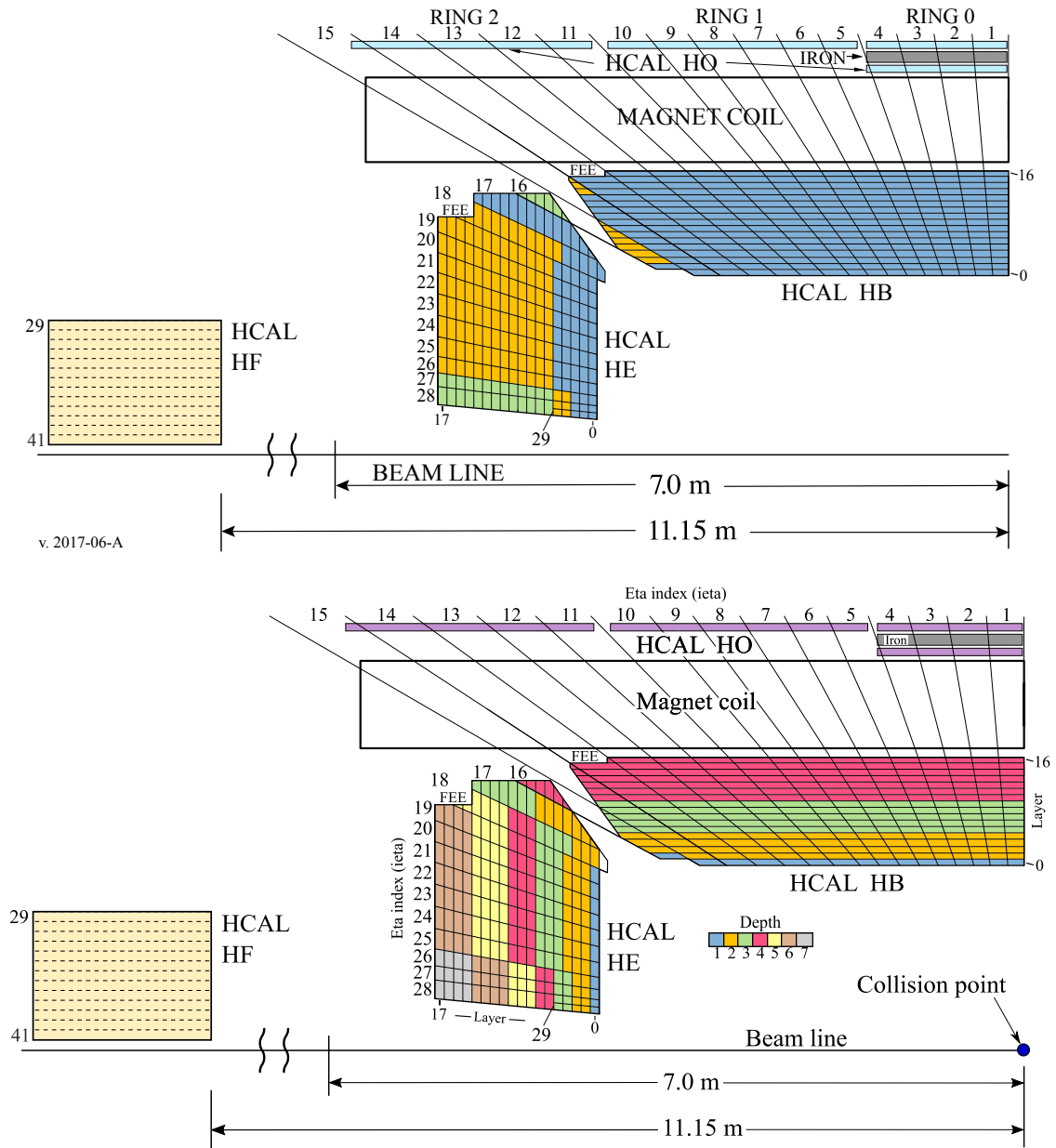


Figure 3.10: HCAL depth segmentation. Segmentation of tiles for the legacy HCAL (top). After the Phase 1 HCAL upgrade, further depths are added (bottom). Image source [39].

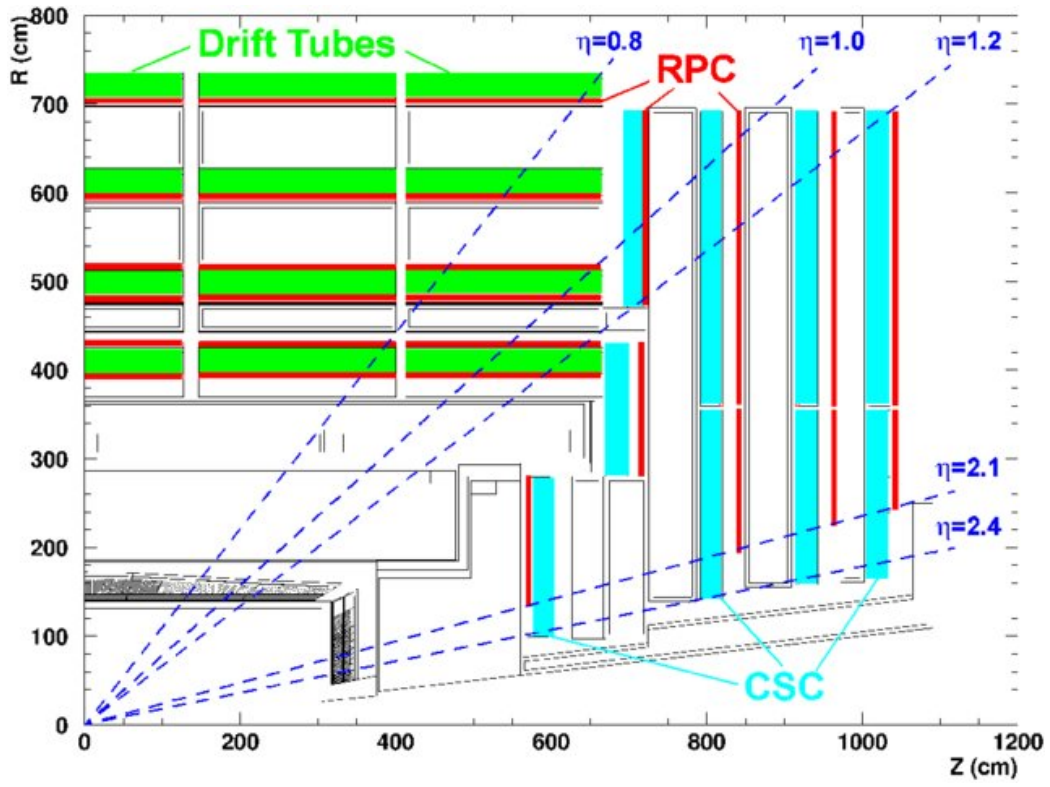


Figure 3.11: A sketch of one quarter of the muon subsystem showing the layout of DT, RPC, and CSC. Image source [40].

cause an avalanche that moves towards the positively charged plate. RPCs provide good timing resolution of about 1 ns which is ideal for triggering events compared to the DTs which take longer than 25 ns. In the ME regions, there are around 468 CSCs that are radiation hard compared to DTs and are used in combination with RPCs. CSCs are almost the same as DTs except that they also have negatively charged wire to collect the positively charged ions after the ionization. For a muon with  $p_T < 100$  GeV, the  $p_T$  resolution is approximately 1% in MB and 3% in ME regions, and when combined with momentum information from the tracker, there is further improvement in the resolution.

### 3.2.7 *Trigger system and data storage*

During the normal operation of the LHC, the proton bunches cross at the center of the detector every 25 ns or at a rate of 40 MHz. It is nearly impossible to capture and store the data for each event. Also, only a small fraction of events would be of interest for offline analyses on new physics searches, precision physics measurements, and calibration studies. Therefore, in order to select meaningful events, the trigger system comes into action. Depending on a few criteria, the trigger flags an event promptly and makes a decision if it wants to store or discard the event. The trigger system composes of two main parts, the level-1 (L1) trigger and the high level trigger (HLT).

The L1 trigger is a specially created hardware-based system with a decision-making time of  $3.2 \mu\text{s}$  and an output frequency of about 100 kHz. The L1 trigger hardware is located close to the CMS detector in the service cavern to reduce delay. As shown in Fig. 3.12, electrons, muons, and jets are quickly rebuilt using data from



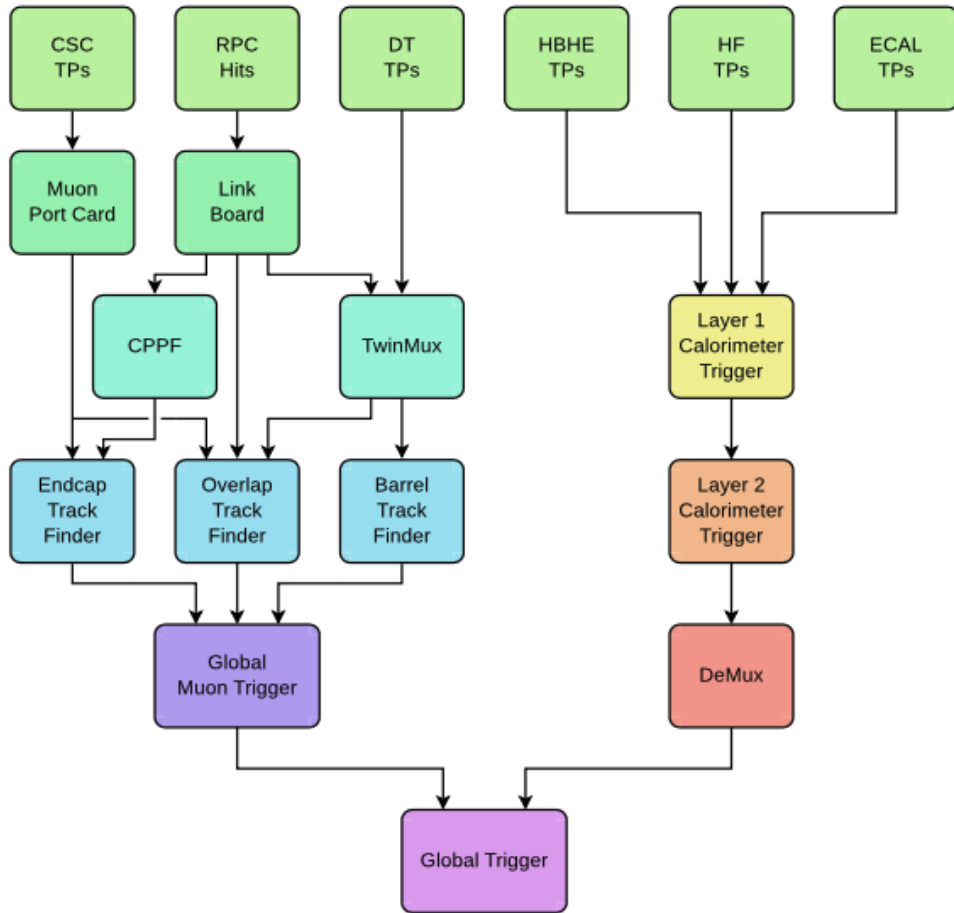


Figure 3.12: Overview of the L1 trigger system. Image source [41].

the ECAL, HCAL, and muon subsystems. The subsequent stage involves passing the events that exceed a predetermined threshold for object  $p_T$  and other kinematic variables to the HLT.

The HLT is in charge of making the final decision regarding whether to store an event. It is a computer farm near the CMS control room that can lower the event rate from around 100 kHz to 1 kHz and make decisions in roughly 160 ms. At this level, the prerequisites for passing an event are known as trigger routes, which are made up of several logic requirements, keeping in mind the requirement of future physics analysis. The full event information is currently available to allow for better decision-making. The trigger paths assist in sorting the data into datasets, which is helpful for physics analyses to select the pertinent datasets, and around 99.9% of the events are eliminated. The Worldwide LHC Computing Grid assists in replicating and transferring the output from the HLT to the grid for storage after it is initially stored at the CERN data center. The information is then utilized to reconstruct an event, which is covered in the following chapter.

## CHAPTER FOUR

### Event simulation and reconstruction

The LHC operations and the numerous CMS subdetectors that are utilized to record the collision data can be inferred from the preceding chapter. In this chapter, we will discuss at how an event is recorded and how reconstruction—the process of examining the data to piece together the 4-vectors of the particles that emerge from the collision—is carried out. The production of different particles for each proton bunch crossing at the detector’s center is referred to as an event. The techniques utilized to identify different physics objects provide the foundation of this event reconstruction. In the CMS experiment, particle flow algorithm is used, which uses information from various subsystems to reconstruct physics objects such as electrons, photons, muons, charged hadrons, and neutral hadrons with their position and momenta.

Simulation of the events is performed with help of special software along with the CMS detector geometry information. The simulation follows a sequential process starting with collisions from the hard scattering process, and the subsequent decay chain of particles from the hard scattering process. The simulation also takes into account interactions of produced particles in proton collisions and the detector materials. The event reconstruction is performed in the same way for both data and simulated events. The reconstructed physics objects are then calibrated so that they closely match the true particle-level quantities. Although event simulation is possible

for a wide variety of physics processes, I will focus on simulation of a few prominent standard model background processes and SUSY signal models relevant to this dissertation.

#### 4.1 *Event simulation*

Testing theoretical hypotheses at the LHC requires precise simulations of proton-proton collisions in the CMS detector. In order to model these collisions, Monte Carlo (MC) [42] simulation techniques are used. The main steps involved in event simulation include the hard scattering of partons, parton shower, hadronization, underlying event, and detector simulation. These are described in the upcoming subsections. The process of event reconstruction is the same for data and simulated events which is discussed in the next section.

##### 4.1.1 *Event generation*

In proton-proton collisions, it is the partons which interact with each other through a process called hard scattering. For simulation of various hard scattering processes in proton-proton collisions, a clear representation of the proton, which is composed of quarks and gluons is necessary. However, the interior structure of a proton made up of partons is complicated, these partons communicate with one another by exchanging the energies among themselves. The initial energies of partons in this energy-sharing mechanism are represented by probabilistic distributions known as the parton distribution functions (PDFs). The PDFs are used to calculate kinematics of the partons, and using the SM and the Feynman rules, we can predict the initial particles produced by the hard scatter. The determination of the proton PDFs is a

huge task by itself. A few groups, such as the NNPDF Collaboration [43, 44], determine the proton PDFs from global analyses of available hard scattering data. Along with the proton PDFs, matrix element generators such as MADGRAPH5\_aMC@NLO (NLO) are used to simulate hard scattering of various background and SUSY particle production processes relevant to the analysis presented in this dissertation.

The hard-scattered partons often radiate gluons according to QCD, and those gluons may further split into either quark-antiquark pairs or gluons. In addition, the hard-scattered charged particles may also radiate photons according to QED. These radiation processes are referred to as initial state radiation (ISR) or final state radiation (FSR) depending on whether they come from the initial partons involved in the hard process or from the outgoing particles.

The hard-scattered partons undergo cascades of radiation, and this cascade process is called the parton shower (PS). The colored quarks and gluons cannot remain free and pull partons out of the vacuum in order to form color-neutral states. Color confinement predicts that partons couple together to produce stable color-neutral hadrons when the perturbative QCD starts to diverge after a certain scale called  $\Lambda_{\text{QCD}}$ . This process of forming color-neutral hadrons is called the hadronization. Non-perturbative techniques are used to simulate the hadronization process.

The hard process does not involve the entire proton in the interaction. The additional hadrons that are not a part of the hard scattering are described by a process called the underlying event (UE) [45]. The non-hard scattering partons may undergo soft scattering or the parton remnants undergo the PS step that lead to the UE. The whole process of PS, UE, and hadronization is handled by software programs, most commonly by PYTHIA [46], in the CMS experiment. At this point,

we have simulated production of stable particles in the event, the next process in the simulation chain involves the interaction of these particles with the detector material as discussed below.

#### 4.1.2 *Detector simulation*

Up to this point, the physical description of the detector has no impact on the event simulation process. In the detector simulation step, the stable particles that are generated through the parton shower and hadronization interact with each of the CMS subdetector materials. This requires accurate modeling of the detector material and its interactions with the stable particles that have lifetimes greater than the time required to traverse the detector material. For this step, a software called GEANT4 [47] is used, which models interactions of particles with the detector active material, various structural supports, absorbers, etc. At this step, the simulation of pileup interactions (described in Chapter Three) is also added to the simulated event.

The search performed in this dissertation considers several signal models, each signal model with a mass scan of the parent and daughter SUSY particles (around 1000 for each signal model). Producing these samples using GEANT4 would be laborious in terms of the computational resources and processing time required. A framework for Fast Simulation of Particle Interactions (fastsim) [48] is being utilized to optimize this process, with event generation rates that are 100 times higher than those of a Full Simulation. For the simulation of particle interactions with the tracker, fastsim uses a streamlined tracker geometry, and fastsim employs parameterized simulation of longitudinal and lateral shower profiles for electromagnetic and hadronic

shower developments. The event reconstruction procedure (discussed in Section 4.2) is the same for both fullsim and fastsim.

## 4.2 Event reconstruction

The event reconstruction procedure attempts to identify particles produced in individual proton-proton collision events. The data from simulations or collisions, such as the track data from the tracker, track and energy deposits in muon chambers, and the energy deposits from the calorimeters are taken into consideration at this stage. The particle flow (PF) algorithm [49], which can reconstruct candidates of electrons, photons, muons, charged hadrons, and neutral hadrons, has been developed by the CMS experiment. The PF algorithm gathers data from all subdetectors and produces a comprehensive description of each event.

Then main ingredients for the particle flow algorithm are charged particle tracks and calorimeter energy deposits. This section describes the procedures to reconstruct charged particle tracks and energy deposits in the calorimeters. It also discusses how various final state PF candidates and physics objects are reconstructed. Section 4.2.3 describes how various PF candidates are reconstructed.

### 4.2.1 Tracks

The reconstruction of charged particles can be challenging. A vast number of proton collisions result in about 20 charged particles [50], and with the presence of pileup interactions, we often have to reconstruct hundreds of charged particle tracks in each event. The *iterative track algorithm* utilizes the Kalman Filter [32] to select a series of hits in different layers that are produced by a charged particle passing through the silicon modules of the tracker element. The algorithm solves the issue of

tracking multiple particles by identifying clear tracks and removing associated hits. The remaining hits are utilized for track identification and fitting through multiple iterations. The process includes finding seeds, finding tracks, fitting tracks, and selecting tracks based on the number of hits, impact parameters, and hits from muon chambers. Later, the remaining hits are used for track identification and fitting through multiple iterations. A track reconstruction iteration includes:

- *Finding seeds:* A hit from a particle that passes some threshold is recorded, then a pair or triplets of these hits constitute a seed. The track seeding algorithm generates a list of seeds that matches a charged particle trajectory.
- *Finding tracks:* Once a list of seeds is generated, the combinatorial track finder (CTF) algorithm [51] extrapolates the trajectory to find matching hits in the inner(outer) pixel layer with the outer(inner) pixel layer as depicted in Fig. 3.6. Working outwards, this algorithm looks for any seed that needs to be matched to the seeds in the inner layers.
- *Fitting tracks:* To smooth the trajectory, fitting is performed and track parameters such as origin and  $p_T$  are obtained. A few tracks appear to be away from the beam spot or primary vertex due to the delay in the decay of collision particles from b hadrons. Additional constraints are placed to correct and fit such type of tracks.
- *Selecting tracks:* Track selection is performed based on the number of hits (including the isolated hits), impact parameters of the track, and the hits from muon chambers.

After tracks are built, the next step is to reconstruct electron-induced track candidates using track information and energy deposits in the ECAL. Due to the tracker material,



approximately 84% of electrons radiate a photon, called bremsstrahlung, as they pass through it, causing a sudden change in track momentum and direction. To address this issue, a technique called Gaussian-sum-Filter (GSF) [52] is used instead of the Kalman Filter to check the track candidates generated by the combinatorial track finder (CTF) algorithm. In the final step, a boosted decision tree (BDT) classifier score is employed to identify electron tracks, which takes information from both the KF and GSF tracks. The process of reconstructing muon tracks is similar to that of electron tracks, but it starts from the outer muon chamber tracks and connects them with the tracks from the inner tracker.

#### 4.2.2 *Calorimeter clusters*

The purpose of the calorimeters (ECAL and HCAL) in CMS is to measure the energy and position of particles. When a particle interacts with the dense material of the calorimeter, it creates a particle shower that can spread over multiple layers of absorber material. In order to isolate the energy deposits from each particle, an algorithmic approach is used to cluster the energies from different absorber layers. The clustering techniques are unique to ECAL and HCAL because the shower profile is different for electrons and hadrons. In ECAL, a seed is planted based on the most energetic hit, and superclusters are formed by combining neighboring small clusters with low energy, taking into account the shower profile.

#### 4.2.3 *Particle flow*

The particle flow algorithm adds significantly to the raw information from the tracks and the calorimeters clusters by linking basic elements from all the subdetectors and forming a global event description of an event. Figure 4.1 gives an example of

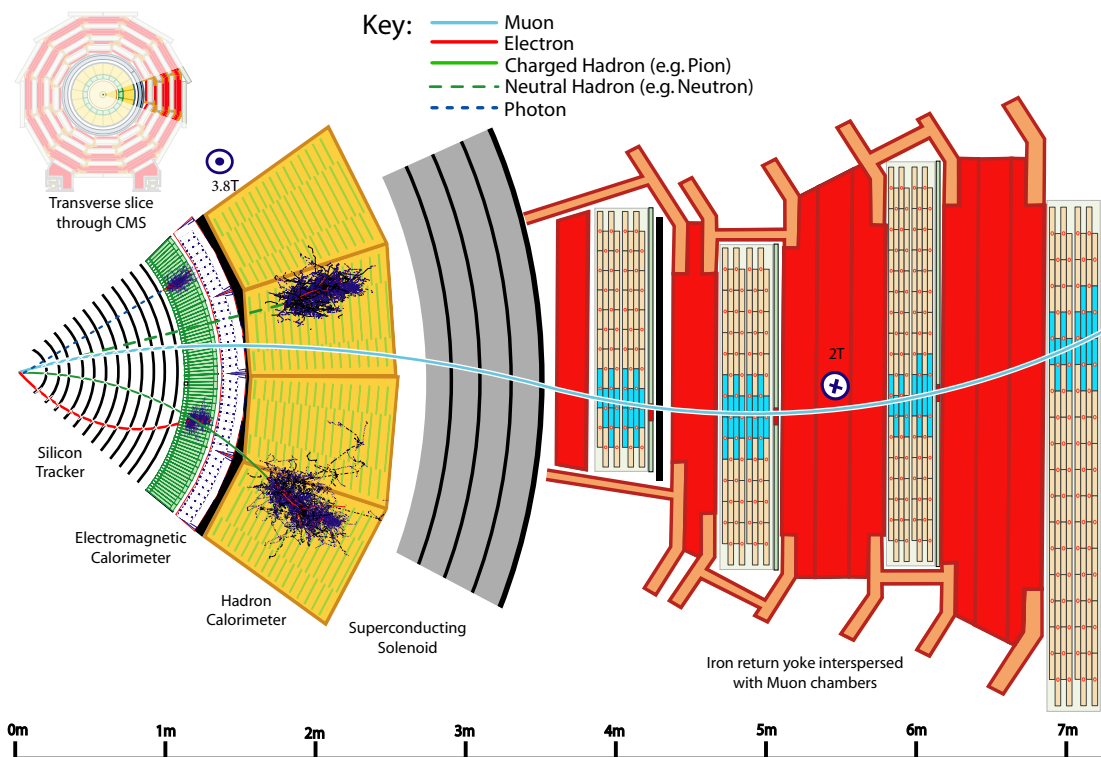


Figure 4.1: A sketch showing the interaction of various particles with a slice of CMS detector. Figure source [49].

particle interactions with the detector elements. From the information gathered from tracks, calorimeter clusters, and hits in muon systems, and with the help of *linking algorithm*, the particle flow algorithm reconstructs final state particle candidates, called PF candidates, and determines their 4-momenta and types. The PF candidate identification allows us to apply specific energy corrections to certain types of particles based on dedicated calibration studies. The PF algorithm provides a very clean information required to carry out physics analyses. The PF algorithm is a complex algorithm. Therefore, it is difficult to describe the whole algorithm chain in detail. Broadly speaking, muon candidates are reconstructed first, then electrons, isolated photons, charged hadrons, non-isolated photons, and neutral hadrons. The main

features of the algorithm and reconstruction of different types of PF candidates are described below:

- (1) Muons generally pass through the whole CMS detector with hits in the inner-tracker, calorimeters, and muon systems. Muon tracks and compatible inner-tracker tracks, taking into account their curvatures in the complex magnetic field in the solenoid magnetic and in the return yoke, are reconstructed as PF muons. The corresponding PF elements (tracks and clusters) associated to muon candidates are removed from further considerations of other types of PF candidate reconstruction.
- (2) Electron-induced tracks are identified as GSF tracks, as discussed in Section 4.2.1. ECAL superclusters are matched to electron tracks according to track trajectories and momenta. PF electrons are reconstructed from these linked ECAL superclusters and GSF tracks. In addition, isolated photon candidates are retained if they are isolated from other tracks and calorimeter clusters, and if the ECAL supercluster energy distribution and the ratio between the HCAL and ECAL energies are compatible with those expected from a photon shower.
- (3) Next, we reconstruct charged hadrons which leave tracks in the tracker and deposit their energies in the HCAL (and some portion in the ECAL). At this step, the remaining tracks should be from charged hadrons since electrons and muons are already identified, and their tracks are removed from consideration. The linking algorithm matches tracks to calorimeter clusters when the track trajectories pass through the cluster areas. Charged hadron candidates are formed from individual tracks. When the calorimeter cluster

energy sum is compatible with the sum of associated track momenta, no neutral particle is formed, and the charged-hadron momenta are redefined based on weighted averages of tracker and calorimeter measurements if the tracking measurement uncertainties are relatively large.

- (4) Non-isolated photons (e.g. from  $\pi^0$  decays) and neutral hadrons do not leave a signature in the tracker because of their neutral charge. Photons deposit most of their energies in the ECAL. The remaining ECAL clusters which are not identified as electrons and do not have any associated HCAL deposits are used to reconstruct non-isolated PF photons. Neutral hadrons deposit energies in HCAL and often also in ECAL. Based on the geometric compatibility, the HCAL and ECAL clusters aid in reconstructing neutral hadrons. The formation of PF non-isolated photons and PF neutral hadrons can also arise from excess energies in ECAL and HCAL clusters compared to the linked track momentum measurements.

The particle flow algorithm creates a list of all the reconstructed particles in an event which is useful but physics analyses often require further refined version of the physics objects. As the detector operating condition evolves, several optimizations are performed for physics objects by physics object groups (POGs) who provide standard object definitions to be used for obtaining optimal performance for a given physics object. The following section describes various physics objects used in this analysis.

#### 4.2.4 Physics objects

The event comprises a list of reconstructed PF particle candidates after particle flow. Despite the value of this information, a complete physics analysis calls for higher-level physics objects that more accurately capture the primary physics production of the hard process. Physics object groups (POGs) function in the CMS Collaboration in order to develop and standardize object definitions. Each POG consists of specialists who, among other things, are in charge of making suggestions for object identification and enhancing object-related algorithms. The physics objects used in this analysis are based on recommendations from various POGs. This section discusses the standard procedure to reconstruct higher-level physics objects.

**4.2.4.1 Interaction vertex** The proton-proton collisions take place around the center of CMS detector and since there are a bunch of collisions taking place in each proton bunch crossing, there can be several interaction vertices. This reconstruction aims to locate the positions of interaction vertices with associated uncertainties in the following steps:

- (1) Tracks are selected based on impact parameters relative to the beam spot and number of hits in the tracker.
- (2) These tracks are clustered based on the distance from the beam spot along the  $z$  axis. The deterministic annealing (DA) algorithm [53] is used to optimize clustering such that it can resolve very close vertices while not splitting true single vertex into multiple vertices.
- (3) Vertex positions are fitted with the help of tracks associated to them using adaptive vertex filter [54].

Once all the interaction vertices are reconstructed, the primary vertex (PV) is defined as an estimate of the interaction vertex of the hard process. The PV is identified as the one with the largest  $p_T^2$  sum of all physics objects associated with that vertex.

**4.2.4.2 Electrons** The electrons reconstructed using the GSF algorithm are used as a base to build high purity isolated electrons. Electron identification checks for refined tracks consistent with ECAL clusters in multiple steps. Electron candidates are defined as reconstructed electron tracks initiated by the reconstruction of superclusters in the ECAL matched with hits in the pixel detector. Electron candidate identification includes refitting tracks and reclustering ECAL clusters. The EGamma POG provides standard definitions of electron candidates by tuning various parameters and classify them into loose, medium, and tight working points based on the electron efficiencies and misidentification rates [55]. In this analysis, electron candidates are chosen based on the EGamma POG's cut-based medium working point with an average efficiency of 95%. In addition, an isolation requirement is imposed on electron candidates. The electron isolation requirement is based on  $I_{\text{mini}}$  defined by:

$$I_{\text{mini}} = \frac{\sum_R p_T(h^\pm) + \max\left(0, \sum_R p_T(h^0) + p_T(\gamma) - \rho EA \left(\frac{R}{0.3}\right)^2\right)}{p_T(e)}, \quad (4.1)$$

where  $h^\pm$ ,  $h^0$ , and  $\gamma$  represent charged hadron, neutral hadron, and photon, respectively. Energy density of the event is given by  $\rho$ , and EA is the effective area. The size of the cone ( $R$ ) used for the isolation calculation in this analysis varies with  $p_T$  of the lepton (electron in this case) as shown in Table 4.1. The decrease in cone size with increasing electron  $p_T$  accounts for the increased collimation of the decay

products from the electron’s parent particle as the Lorentz boost of the parent particle increases [56, 57]. Additionally, electrons are required to have  $p_T > 10$  GeV and  $|\eta| < 2.5$  to qualify as an electron candidate in this analysis.

Table 4.1: Radius of the cone with respect to  $p_T$  of the lepton

$R$	$p_T$ range
0.2	$p_T < 50$ GeV
$10 \text{ GeV}/p_T$	$50 < p_T < 200$ GeV
0.05	$p_T > 200$ GeV

**4.2.4.3 Muons** Muon candidates are reconstructed from the tracks in muon system (DT, CSC, and RPC) and inner tracker, after checking they are compatible. A *global muon* track fitting is performed from muon hits in the tracker and muon systems. Muons reconstructed only from hits in the muon system are called *standalone muons*. *Tracker muons* are defined when at least one muon segment matches extrapolated tracks from the inner tracker. Global muon performance is better than the tracker muon or the standalone muon as it uses information from the muon system and inner tracker.

Muon candidates with varying reconstruction efficiencies and misidentification rates are defined with additional requirements on top of PF muons, and are classified into multiple working points as also done for electrons. The POG-recommended “medium” working point is used in this analysis, which takes muon candidates from global and tracker muons and apply additional track-quality and muon-quality requirements. This analysis also require  $p_T^{\text{muon}} > 10$  GeV and  $|\eta^{\text{muon}}| < 2.4$  in order to identify a muon candidate. In order to reconstruct prompt muons (that arise from

the hard process), the isolation condition of  $I_{\text{mini}} < 0.2$  is also imposed, where  $I_{\text{mini}}$  is defined by Eq. (4.1) but for a muon instead for an electron. This is required in order to separate them from the muons produced in b hadron decays which are typically produced at low muon  $p_T$ .

**4.2.4.4 Photons** ECAL superclusters with little associated energies in HCAL mostly arise from electrons and photons, and out of which if the ECAL superclusters have no track associate, then they are qualified to be photon candidates. Events with a photon reconstructed with POG recommended “loose” selections are rejected in this analysis. The photon candidates are required to be well isolated from electrons and with  $p_T > 100$  GeV in the barrel ( $|\eta| < 1.4442$ ) or endcap ( $1.566 < |\eta| < 2.5$ ) regions. Also, photon candidates are not allowed to have any associated pixel seed.

**4.2.4.5 Jets** The quarks and gluons produced in hard scattering processes decay or radiate particles, which continues as a chain reaction until the particles have reached some threshold momentum. These decayed particles form a spray of mesons and baryons through a process of hadronization. The spray of particles can be clustered into a cone-like shape called a jet. In this analysis which searches for bosons that decay hadronically, jets play a vital role in achieving results in this analysis.

Jets are reconstructed by clustering PF candidates using anti- $k_T$  algorithm [58]. The algorithm is based on two specific measures of distance also from the beam which



can be described by:

$$d_{ij} = \min \left( \frac{1}{p_{T,i}^2}, \frac{1}{p_{T,j}^2} \right) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.2)$$

$$d_{iB} = \frac{1}{p_{T,i}^2}, \quad (4.3)$$

where  $i$  and  $j$  are two different particles, and  $B$  refers to the beam.  $\Delta R_{ij}^2$  is the spatial separation of these particles in  $(y, \phi)$  phase space defined as:

$$\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2, \quad (4.4)$$

where  $y$  is the rapidity as defined in Eq. (3.4). The size of a jet is tunable size parameter defined by  $R$ . Typically this parameter is set to 0.4 and the resultant jets are called AK4 jets. When a hadronic decay of boosted (high  $p_T$ ) bosons occurs, its spread is often wider than the size of AK4 jets. Therefore, the size parameter is set to 0.8 for reconstructing jets arising from boosted bosons, and the resultant jets are referred to as AK8 jets or fat jets.

The jet clustering proceeds iteratively by comparing the smallest value of  $d_{ij}$  and  $d_{iB}$ . Until the values of  $d_{ij}$  is less than or equal to  $d_{iB}$ , the algorithm merges those entities into a jet. While, in the case of AK8 jets, the pileup per particle identification (PUPPI) algorithm is used for better reconstruction of jet substructure that applies more constraints on the charged hadrons and rescales the energy of neutral hadrons by a factor representing their probabilities to originate from the primary interaction. Looking closely into the jet energy fraction, most of the jet energy is carried by charged hadrons ( $\sim 65\%$ ), then followed by photons ( $\sim 25\%$ ), and the reminder by neutral hadrons ( $\sim 10\%$ ) as shown in Fig. 4.2.

The bottom plot of Fig. 4.2 shows some differences between data and MC simulation especially at high  $p_T$ . The differences in the jet four-momentum between

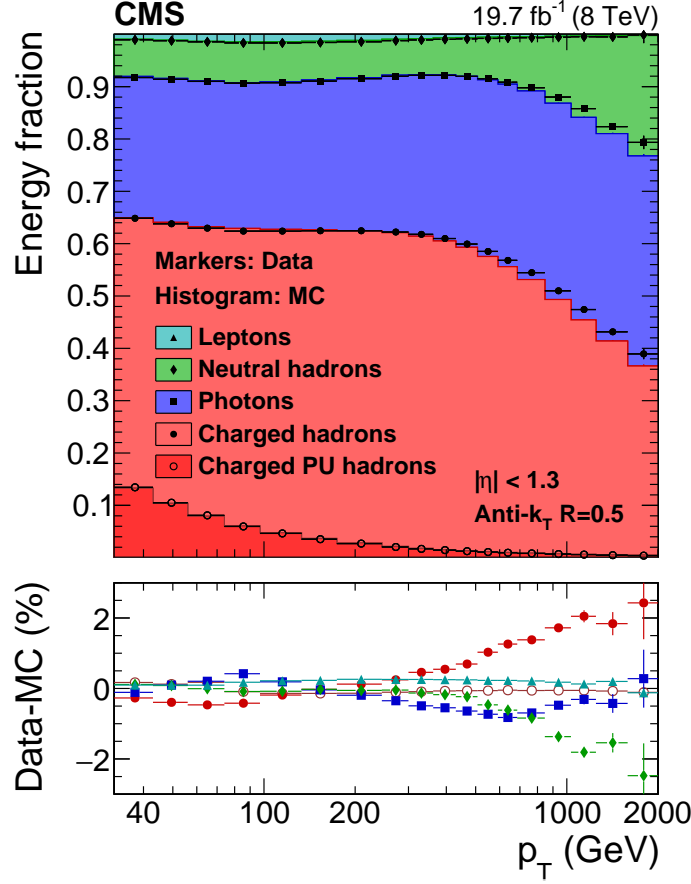


Figure 4.2: Energy fraction in jets as a function of  $p_T$  in data and MC simulation. Figure source [49].

data and MC simulation can occur since low  $p_T$  particles are not used in jet reconstruction, and due to non-linear detector response. Therefore, the  $p_T$  and mass of the jets needs to be corrected in order to accommodate these differences. A factorized approach is used to correct jets, which is called the jet energy correction (JEC). The flow diagram in Fig. 4.3 represents the various corrections applied at different levels as discussed below.

Starting from the reconstructed jets, several correction factors are applied to correct the four-momentum of a jet [60]:

$$p_\mu^{corr} = \mathcal{C} p_\mu^{raw}, \quad \mathcal{C} = C_{\text{offset}}(p_T^{raw}, \eta, N_{PV}) \cdot C_{MC}(p_T', \eta) \cdot C_{rel}(\eta) \cdot C_{abs}(p_T''), \quad (4.5)$$

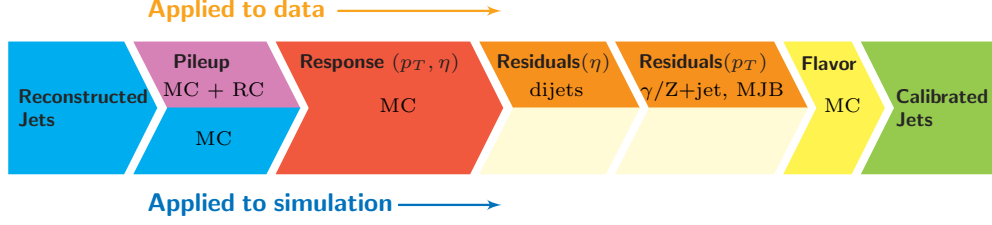


Figure 4.3: Flowchart showing the JEC application steps on data and MC simulation. RC stands for random cone, MJB refers to analysis of multijet events. Figure source [59].

where  $p'_T$  is the  $p_T$  of the jet after offset correction, and  $p''_T$  is the  $p_T$  of the jet after all the corrections excluding  $C_{abs}$ . The overall correction factor  $\mathcal{C}$  consists of the following individual corrections:

- $C_{offset}$  removes extra energy which is estimated to be higher than the actual value due to pileup contributions and noise. For data events, based on the data vs MC simulation differences, an additional correction is obtained using the random cone (RC) method.
- $C_{MC}$  is the correction factor applied after the offset corrections are applied. Due to the non-uniformity in  $\eta$  and non-linearity in  $p_T$ , the response varies for jets w.r.t. its  $p_T$ . This correction factor aims to apply the correction as a function of jet  $p_T$  and  $\eta$ .
- $C_{rel}$  is the relative energy scale correction w.r.t.  $\eta$ . Forward region jets to corrected back to the level of jets in the barrel region.
- $C_{abs}$  is absolute energy scale correction which matches the reconstructed jets to particle-level jets and in the next step residual corrections are applied.

In simulated samples, jets can also be reconstructed directly from stable particles emerging from fragmentation and hadronization processes, which are referred to as particle-level jets. The  $p_T$  response of the jet, defined as the ratio of the measured

jet  $p_T$  to the particle-level jet  $p_T$  is shown in Fig. 4.4 at different stages of JEC application.

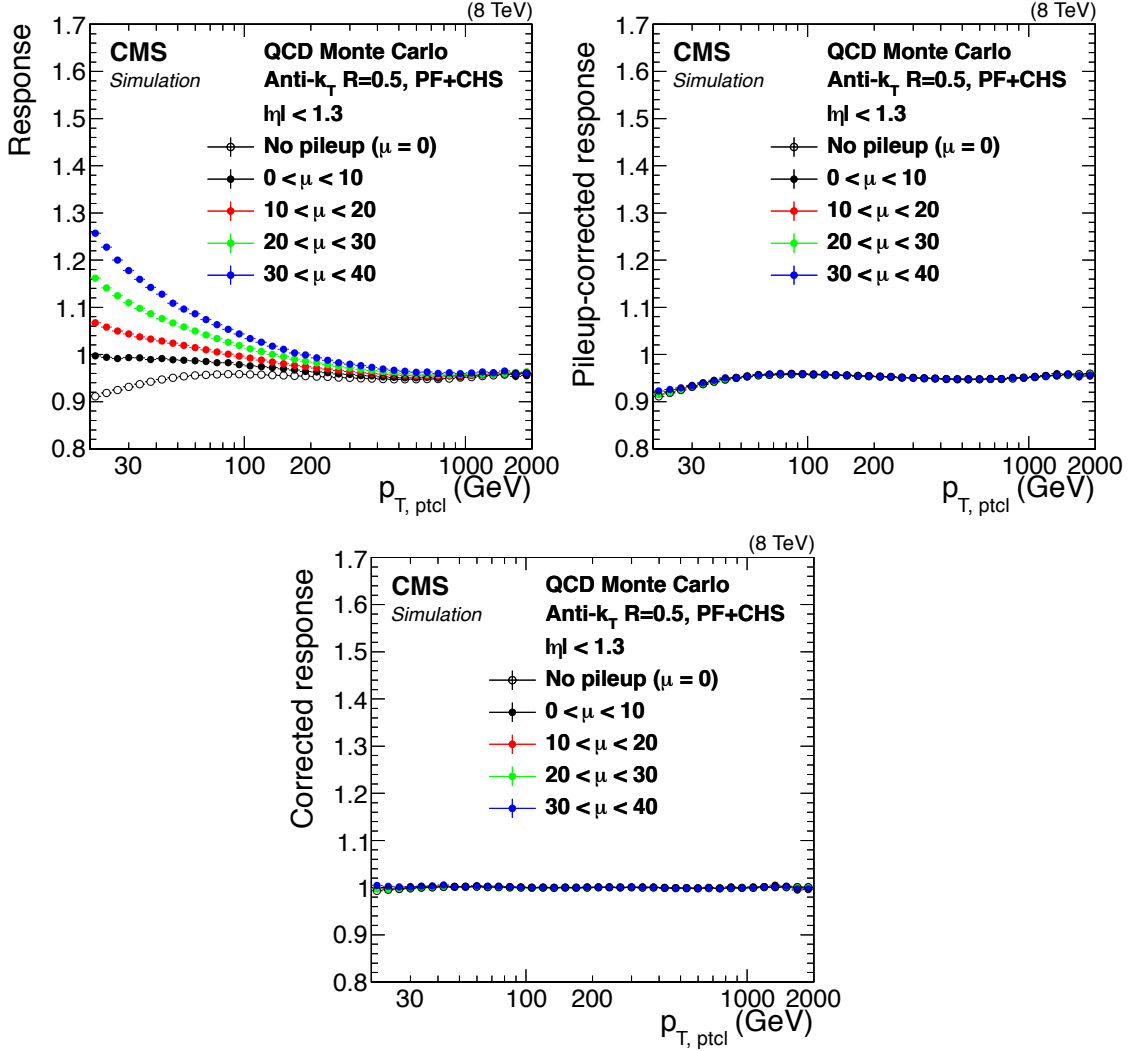


Figure 4.4: The top left plot shows the raw  $p_T$  response of the jets, the top right plot shows the improvement after pileup corrections, and the bottom plot shows good response after all the residual corrections are applied. Figure source [59].

In addition of energy scale corrections to jets, a few other constraints are imposed on jets so that jets from spurious sources, such as electronics noise, can be eliminated. AK4 jets are required to have  $p_T > 30$  GeV and  $|\eta| < 2.4$  along with jet

identification (JetID) criteria [61]. For 2016 data, the loose working point of JetID is chosen, while for 2017 and 2018 years, the tight working point of JetID is being used. When any jet in the event fails the JetID criteria, the event is not considered for the further analysis. In this search, AK8 jets are required to have a high transverse momentum of  $p_T > 200$  GeV and  $|\eta| < 2.0$ .

In general the mass resolution in simulation is good as compared to data, therefore the mass distribution of the AK8 jets passing the Deep-W tagger and having its mass between 65–105 GeV are studied in  $t\bar{t}$ +jets enriched single lepton control regions. Scale factors for jet mass scale (JMS) and jet mass resolution (JMR) are derived for each data-taking year. The correction to JMS is applied on MC simulation yield by scaling the mass and JMR by smearing the MC simulation yield mass distribution. Similarly, the JMS and JMR scale factors are applied for AK8 jets passing Deep-WMD and having its mass between 65–105 GeV. As it is difficult to sample  $H/Z \rightarrow b\bar{b}$  in data events, no JMS and JMR uncertainties are applied for bb-vs-light tagger.

**4.2.4.6 *b quark jets*** Jets that originate from the hadronization of a bottom quark play a special role in the classification of this analysis. The b hadrons have masses about 5–6 GeV and relatively longer lifetime than light-flavored hadrons with lifetime of 1.5 ps. Depending on the Lorentz boost, the b hadrons travel hundreds of microns from the primary vertex before decaying, creating a secondary vertex (SV) as shown in Fig. 4.5. This process gives an important handle to identify (tag) jets originating from b quarks, also known as b jets.

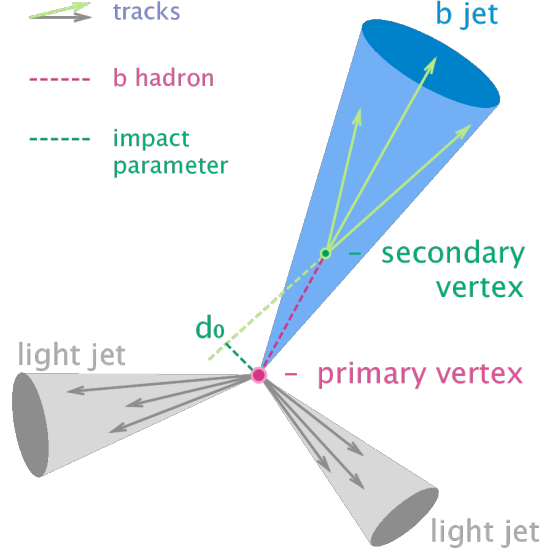


Figure 4.5: Diagram showing the common principle of identification of b jets initiated by b hadron decays.

For the b jet tagging, various algorithms are developed by the CMS Collaboration. One of the recent algorithms for identification of AK4 b jets is the combined secondary vertex algorithm based on deep-learning techniques (DeepCSV [62]) that is used in this search. The DeepCSV tagger takes in many inputs in the training including number of SV, number of tracks from SV, 2D flight distance from PV, track information, and several jet variables, and then output a score ranging from zero (least likely to be a b-tagged jet) to one (most likely to be a b-tagged jet). The efficiency of tagging affects the rate of misidentification of the b jet, as they are inversely proportional. To characterize the tagging efficiency and misidentification rates, several working points are specified. The medium working point is used in this search to identify the b jets depends on Run 2 data taking period. Any jet that exceeds the cutoff value determined for that year are defined as a b-quark jet. In this case, the threshold values for the DeepCSV jet tagger score are 0.6321, 0.4941, 0.4184 in the

years 2016, 2017, and 2018, respectively. In this analysis, the b jets are required to fulfill  $p_T > 30$  GeV and  $|\eta| < 2.4$  in this analysis.

*4.2.4.7 Heavy object jets* At high  $p_T$ , decay products from heavy SM particles, such as the top quark and Higgs, Z, and W bosons, may be collimated enough to be well contained in a AK8 jet. For this analysis, a deep neural network (DNN) based algorithm called DeepAK8 is used in order to find candidate AK8 jets that are consistent with the decay, fragmentation, and hadronization of heavy SM particles. The DNN is made to classify AK8 jets produced by hadronically decaying particles into one of five main groups: W, Z, H, top, and “other”. Jets are further divided into subcategories based on how likely it is that they came from a certain decay of these unstable particles, such as  $Z(b\bar{b})$ ,  $Z(c\bar{c})$ , etc. The scores assigned to each classification are utilized to create jet taggers.

Three jet taggers—the W tagger, the V tagger, and the  $b\bar{b}$  tagger—that were developed using the aforementioned classifiers are used in this analysis to categorize AK8 jets. These taggers are also interchangeably referred to as Deep-W, Deep-WMD, and bb-vs-light taggers in this dissertation. All, or none of the three taggers may tag a single AK8 jet. The mass of the jets ( $65 < m_J < 105$  GeV) and a DNN score designed to distinguish between hadronic W boson decays and QCD jets are used by the W tagger to identify jets consistent with a  $W(q\bar{q}')$  decay. The V tagger’s DNN score is optimized similarly for the W tagger, which uses adversarial training to de-correlate the DNN score and the jet mass. This de-correlation technique is referred to as mass decorrelation (MD). This mass decorrelation (MD) technique allows the V tagger (also known as the Deep-WMD tagger) to be sensitive to hadronic decays

sensitive of both the W and Z bosons despite having a higher misidentification rate. By using the jet mass ( $75 < m_J < 140$  GeV) and a DNN score tailored for identifying Lorentz-boosted  $b\bar{b}$  topologies, the  $b\bar{b}$  tagger (also known as the bb-vs-light tagger) detects AK8 jets consistent with a  $Z(b\bar{b})$  or  $H(b\bar{b})$  decay. Adversarial training is another technique the  $b\bar{b}$  tagger uses to associate the DNN score and the jet mass. The DeepAK8 [63] tagger outperforms several other taggers as shown in Fig. 4.6 where the tagging efficiency is higher and the misidentification rate is lower for the DeepAK8 tagger compared to the others for the same.

This analysis also involves the tagging of a highly boosted AK8 jet with  $p_T > 200$  GeV that contains two b quarks and correlates with a boson decay. A neural network is trained using the variables involving sub-structure variables, SV and other jet information to tag the AK8 jets using DNNs. Several working points are defined for the DeepAK8 tagger in a similar way as done for the DeepCSV tagger. The DeepAK8 algorithm that can discriminate between jets arising from  $b\bar{b}$  decays and jets arising from light flavor jets is called the DeepAK8 bb-vs-light (bbvsL) tagger, as described in the previous subsection, and it is used to identify the decays of Z or Higgs bosons into a pair of b quarks. Tagger efficiency in simulation is defined as the fraction of simulated hadronically decaying W, Z, or Higgs bosons that are identified by the appropriate tagger discussed above, and it is used to correct the simulation yields to better match the data. The scale factors for the bb-vs-light tagger correction are obtained using a collection of AK8 jets enriched in gluon splitting to a  $b\bar{b}$  pair based on the presence of secondary vertices. Similarly, for the Deep-W tagger, the  $W(q\bar{q}')$  signal is obtained from a sample enriched in top quark decays.



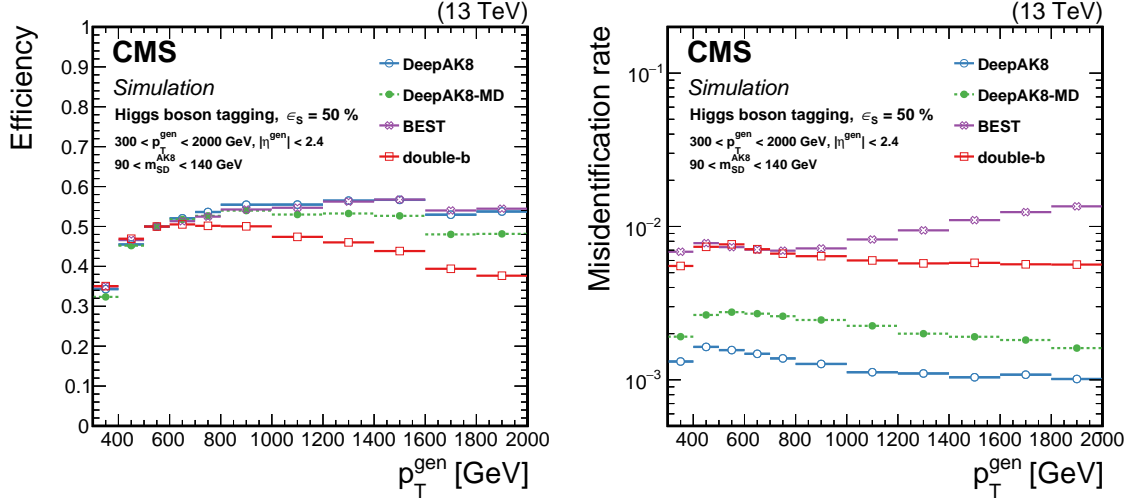


Figure 4.6: Left plot shows the efficiency as a function of the generated particle  $p_T$  for a given working point for Higgs boson identification. Right plot shows the misidentification rate as a function of the generated particle  $p_T$ . Figure source [63].

**4.2.4.8 Missing transverse momentum** Another prominent kinematic variable used in this analysis is missing transverse momentum or missing transverse energy (denoted by MET or  $p_T^{\text{miss}}$ ). Since the protons are accelerated and have initial momentum along the  $z$  axis, the momentum in the transverse  $x$ - $y$  plane is zero. Therefore, the final momentum in the transverse plane after collision must be also zero from the momentum conservation principles. After all the visible PF candidates are reconstructed, a significant momentum imbalance in the transverse plane is likely due to particles that pass through the detector layers undetected. Among SM particles, neutrinos contribute to  $p_T^{\text{miss}}$ . In this analysis, the signals involve the production of lightest neutralinos (SUSY particles), which can also escape the CMS detector and contribute to  $p_T^{\text{miss}}$ . This quantity can be mathematically defined as:

$$p_T^{\text{miss}} = \left| - \sum_i^n \vec{p}_T^i \right|, \quad (4.6)$$

where  $\sum_i^n \vec{p}_T^i$  denotes the sum  $\vec{p}_T$  of all the PF candidates. Measuring this quantity correctly is crucial in this analysis since significant  $p_T^{\text{miss}}$  is an important requirement to enhance signal-like events.

#### 4.2.5 MET event filters and further event cleaning

The CMS detector is a very complex detector, and consequently the algorithm to reconstruct its readout information is also complicated. Unfortunately, issues such as beam-induced background, electronics noise, nonfunctional detector readout channels, and occasional poor reconstruction may lead to large spurious  $p_T^{\text{miss}}$  measurements. This analysis has search bins in  $p_T^{\text{miss}}$ , therefore it is crucial to reject events that impact the measurement of  $p_T^{\text{miss}}$ . The CMS JetMET POG identifies and prescribes a list of MET filters for the Run 2 physics analyses. The list of MET event filters used in this analysis are:

- *Good vertex filter*: This filter removes events that fail to have a good primary vertex.
- *Beam halo filter*: The collisions of the beam with the residual gas inside the LHC vacuum chamber produces secondary particles and contribute towards beam background, known as beam halo. This filter helps to remove events contaminated by beam halo particles.
- *EE bad supercrystal filter*: Occasional anomalous pulses in the faulty ECAL endcap crystals cause spurious  $p_T^{\text{miss}}$ , and this filter removes this kind of events from data.

- *ECAL bad calibration filter*: Some faulty calibrations in 2017 and 2018 data-taking years led to random pulses in ECAL. Events with such random pulses are removed using this filter.
- *ECAL trigger primitive filter*: This filter removes events where a significant amount of energy appears to be lost in nonfunctional readout cells according to independent trigger primitive readout information.
- *Bad PF muon filter*: Events with high  $p_T$  and low quality PF muons are rejected.
- *HBHE isolation noise filter*: Events that have clusters of noise in HCAL are filtered using this topological filter.
- *HBHE noise filter*: Anomalous HCAL noise due to the instrumentation from the readout box in hadron barrel and endcap calorimeters are removed using this filter.

In addition, the following requirements are utilized in this analysis in order to ensure robust measurements of  $p_T^{\text{miss}}$  and other physics objects:

- *HEM veto*: In 2018, sectors of the HCAL endcap subdetector on the minus side, “HEM”, became inoperable. The impacted region corresponds to  $-3.0 \leq \eta \leq -1.4$ , and  $-1.57 \leq \phi \leq -0.87$ . This HEM issue may lead to spurious jets and wrong measurement of  $p_T^{\text{miss}}$  along with copious amounts of reconstructed isolated electrons. Events are vetoed if any isolated electrons are in the detector region of  $-3.0 \leq \eta \leq -1.4$  and  $-1.57 \leq \phi \leq -0.87$ , or if any AK4 jets are in the detector region of  $-3.2 \leq \eta \leq -1.2$  and  $-1.77 \leq \phi \leq -0.87$ . As the simulated samples do not have the HEM veto

issue modeled, a proportional fractional of MC simulation events are removed according to the data event loss due to this incident.

- $p_T^{\text{miss}}$  *ratio*: Additionally, events are required to pass the condition:  $\frac{\text{PF-}p_T^{\text{miss}}}{\text{Calor-}p_T^{\text{miss}}} < 2$ , in order to ensure high quality  $p_T^{\text{miss}}$  measurement.

## CHAPTER FIVE

Search for electroweak supersymmetry in final states with hadronic decays of a WW, WZ, or WH boson pair and large missing transverse momentum

### 5.1 Introduction

There have been numerous searches to find SUSY at the LHC, but no evidence has been found of a major departure from the SM backgrounds. This chapter discusses a hunt for electroweak production of SUSY particles as a part of our quest for SUSY. This analysis focuses on a particular kind of SUSY production model where the  $R$ -parity is conserved, which means that the lightest supersymmetric particle (LSP) is stable and has a very weak interaction with the detector and may be able to evade it, possibly leading to large  $p_T^{\text{miss}}$ . The reason for conducting this search stems from the fact that the gluon and top squark, which are currently subjected to have limits in the order of TeV, whereas winos and higgsinos are subjected to very weak constraints because of their small cross sections, as shown in Fig. 2.7. Additionally, based on the naturalness idea described in Chapter Two, it is anticipated that the electroweakinos (charginos or neutralinos) will be within the current LHC energy limits. On the plus side, the findings of this search are made possible by the addition of data from Run 2 LHC operation and advancements in jet identification techniques.

Under the assumption that strongly-coupled SUSY particles are too massive to be produced at the LHC, this analysis focuses on the search for electroweak production of SUSY particles. The bino and wino (superpartners of gauge bosons of unbroken U(1) and SU(2) symmetries) combine with higgsinos (super partner of the Higgs

boson) form mass eigen states of charginos ( $\tilde{\chi}_i^\pm$ , where  $i = 1, 2$ ) and neutralinos ( $\tilde{\chi}_i^0$ , where  $i = 1, 2, 3, 4$ ). This electroweakino search targets the production of the lightest chargino ( $\tilde{\chi}_1^\pm$ ), neutralino ( $\tilde{\chi}_2^0$ ). The decays of these lightest chargino and neutralino proceed through W, Z, or the Higgs boson depending on the charge of the electroweakino. Branching fraction of the neutralinos to Z or H bosons depends on the mixing among the wino, bino, and higgsinos.

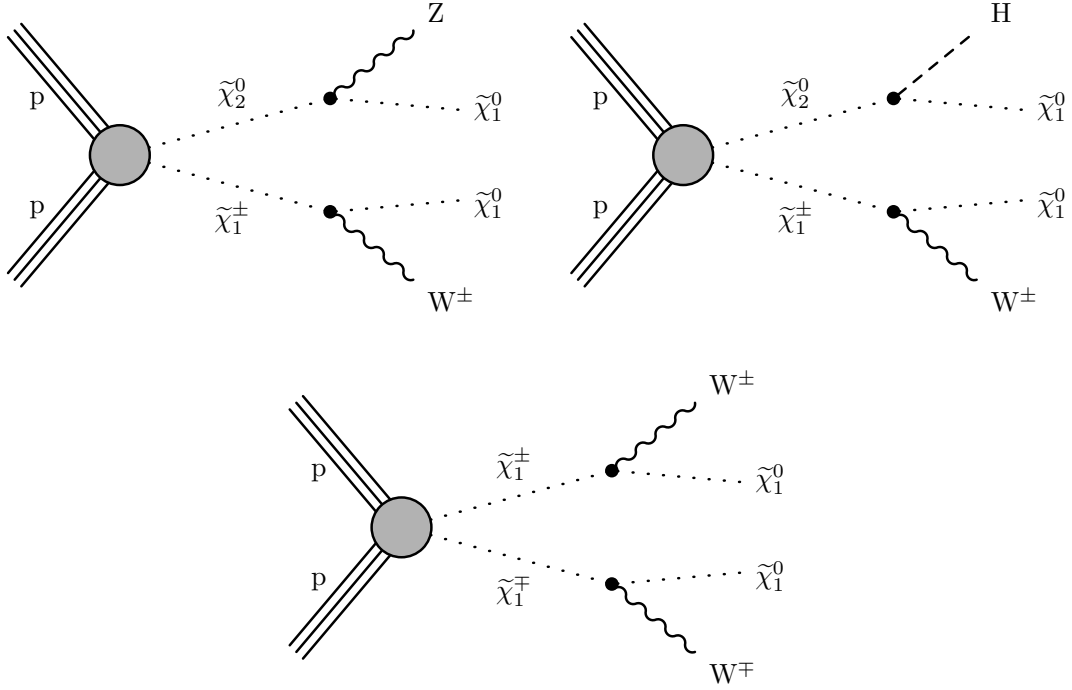


Figure 5.1: Simplified SUSY models showing the production of  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  with  $\tilde{\chi}_1^\pm$  decaying to a W boson and the  $\tilde{\chi}_1^0$ , and the  $\tilde{\chi}_2^0$  decaying to either a Z boson (left) and the  $\tilde{\chi}_1^0$  or a H boson (right) and the  $\tilde{\chi}_1^0$ .

As shown in Fig. 5.1, direct production of  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  and  $\tilde{\chi}_1^0 \tilde{\chi}_1^0$  is considered, with the  $\tilde{\chi}_1^\pm$  always decaying to a W boson and the  $\tilde{\chi}_1^0$ , while the  $\tilde{\chi}_2^0$  can decay to either a Z or H boson and  $\tilde{\chi}_1^0$ . The search bins are established in this study with the assumption that  $\tilde{\chi}_1^0$  is a bino-like LSP that will evade the CMS detector without leaving any traces

and produce large  $p_T^{\text{miss}}$ . The MSSM signal models are intricate and have numerous parameters. The main objectives of this analysis are the targeted simplified models displayed above. According to the bosons in the final state in Fig. 5.1, they are referred to as TChiWZ (left), TChiWH (right), and TChiWW (bottom) models. The final states also include the hadronic decay of the vector boson that is captured within a large radius (AK8) jet.

In addition to the signal model targeted, there are several sources of backgrounds from the standard model events, which includes Z + jets, W + jets, QCD multijet, and top backgrounds as listed in Section 5.2. Several optimizations are performed in order to reduce these background contributions as described in Section 5.3. Later, in order to test for the traces of SUSY signatures in data, a method to estimate the amount of SM backgrounds in data through a background estimation procedure is required, which is described in Section 5.4. Uncertainties on signal and background yields arising from theoretical and experimental sources are discussed in Section 5.5. In order to determine whether the background estimation technique is effective, it is validated at the end, and the validation results are shown in Section 5.6. The search results and interpretations are shown in Section 5.7 using simplified models [64].

## 5.2 Data and simulation

### 5.2.1 Data and triggers

The search in this analysis uses the LHC proton-proton collision data set recorded with the CMS detector at center-of-mass energy of 13 TeV collected from the Run 2 (2016–2018), corresponding to an integrated luminosity of up to  $137 \text{ fb}^{-1}$ . The LHC runs are divided into eras for each year and the integrated luminosities for

each era are given in Table 5.1. There is data quality monitoring in place by several object groups to certify the data. Then, certified data blocks are recommended for use in analyses.

Table 5.1: MET primary dataset integrated luminosities considered by run era in each of the data-taking years. All luminosities are in  $\text{pb}^{-1}$ .

Run Era	2016	2017	2018
A	—	—	14024.2
B	5746.4	4793.3	6907.9
C	2572.9	9632.7	6894.8
D	4242.3	4247.7	31719.5
E	3924.3	9314.0	—
F	3104.5	13498.4	—
G	7575.8	—	—
H	8649.0	—	—
<b>Total</b>	35815.2	41486.1	59546.4

The data are classified into primary datasets based on the higher level event selections that are required by the HLT paths. In this analysis, the MET primary dataset is used for the signal regions. The events of interest in the signal region are chosen from the MET dataset using a set of  $p_{\text{T}}^{\text{miss}}\text{-}H_{\text{T}}^{\text{miss}}$ -based triggers (signal triggers):

- `HLT_PFMETX_PFMHTX_IDTight_v*` ( $X=90, 100, 110, 120, 130, 140$ ).

where  $X$  indicates the threshold applied to the online calculation of  $p_{\text{T}}^{\text{miss}}$  and  $H_{\text{T}}^{\text{miss}}$  ( $H_{\text{T}}^{\text{miss}} = \left| \sum_{\text{jets}} \vec{p}_T \right|$ ), both based on the particle flow (PF) algorithm; the asterisks indicates more than one version of the same trigger using the ‘OR’ logical operation has been used.

Most of the SM backgrounds contributes towards real- $p_{\text{T}}^{\text{miss}}$  arising from the neutrinos. Data events from the single-electron or the single-muon primary datasets



are used to derive trigger efficiencies. The events in these datasets are required to pass the following HLT paths. The events are required to pass the following reference triggers:

- HLT\_Ele20\_(eta2p1\_)WPLoose\_Gsf\_v,
- HLT\_Ele23\_Ele12\_CaloIdL\_TrackIdL\_IsoVL\_DZ\_v,
- HLT\_Ele25\_eta2p1\_WPTight\_Gsf\_v,
- HLT\_EleX\_WPTight\_Gsf\_v (X=27,32,35),
- HLT\_Ele27\_eta2p1\_WPLoose\_Gsf\_v,
- HLT\_Ele45\_WPLoose\_Gsf\_v,
- HLT\_Photon165\_HE10\_v,
- HLT\_PhotonX\_v (X=175,200),

for events from the single-electron primary dataset, and

- HLT\_IsoMuX\_v (X=20,22,24,27),
- HLT\_IsoMuX\_eta2p1\_v (X=22,24),
- HLT\_IsoTkMuX\_v (X=22,24),
- HLT\_Mu17\_TrkIsoVVL\_Mu8\_TrkIsoVVL\_v,
- HLT\_Mu17\_TrkIsoVVL\_TkMu8\_TrkIsoVVL\_v,
- HLT\_Mu45\_eta2p1\_v,
- HLT\_MuX\_v (X=50,55)

for events from the single-muon primary dataset.

The JetHT primary dataset is also used as a reference to measure the trigger efficiency as large  $p_T^{\text{miss}}$  and  $H_T$  are typical of chargino and neutralino production when a high momentum boson is present. Trigger efficiency is defined as the ratio of the number of events passing both the signal and reference triggers to the number of

events passing only the reference triggers. Figure 5.2 shows the trigger efficiency and the fit parameterization with an error function.

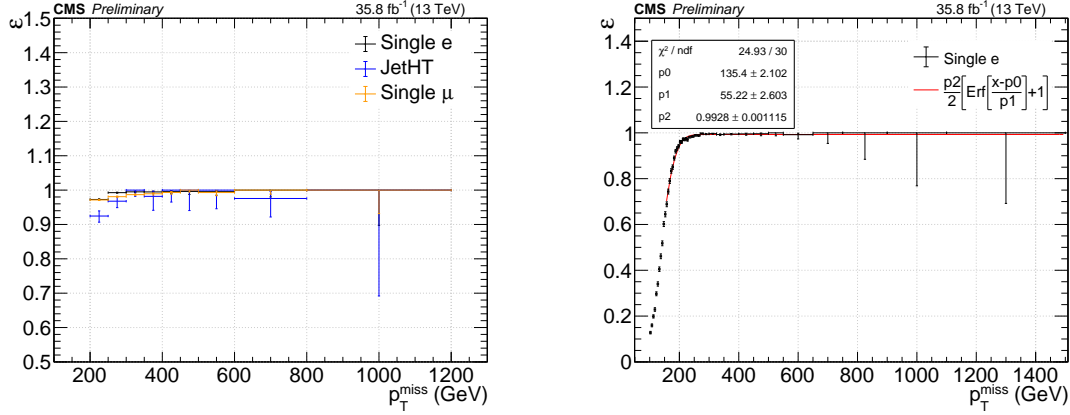


Figure 5.2: (Left) Trigger efficiency versus  $p_T^{\text{miss}}$  for different dataset in 2016. (Right) Fitted parameterization of signal trigger efficiencies versus  $p_T^{\text{miss}}$  in single-electron dataset in 2016.

The trigger efficiency at the baseline selection of  $p_T^{\text{miss}} > 200$  GeV is 95%, 78%, and 74% for data collected in 2016, 2017, and 2018, respectively. The efficiency parameterizations derived from the single-electron dataset are used to apply weights to the simulated events and the efficiencies derived from ratio of single-electron to single-muon datasets are used to calculate trigger uncertainties. These uncertainties are applied to final yield estimates of signal and rare backgrounds, which are derived from simulation directly and are 3% for the first MET bin (200–250 GeV) and  $< 2\%$  for all other bins in this analysis. The dilepton control sample events are selected using a set of electron or muon triggers in this analysis.

### 5.2.2 Signal and background simulation

The event simulation involving the steps described in Section 4.1 are performed using event generators such as MADGRAPH [65] or POWHEG [66] for simulating the hard process. The PYTHIA program is used for parton showering, fragmentation, and hadronization. Lastly, GEANT4 is used to model the full detector response for the SM samples, while the fastsim package [48] is used for SUSY signal samples. The samples listed in table 5.2 are generated in order to compare against the data. To model the quark and gluon distributions, a set of parton distribution functions, NNPDF [43,44] is being used.

*5.2.2.1 Signal MC samples* The simplified model spectra (SMS) SUSY scenarios examined in this analysis include TChiWW, TChiWZ, and TChiWH models generated using MADGRAPH5\_aMC@NLO 2 generator at leading order using the NNPDF 3 PDF set and PYTHIA 8.2 for parton showering. Since there are thousands of signal mass points that need to be tested in this analysis, the fastsim package described in Section 4.1.2 is used to generate them. Any differences in the fast simulation are corrected by applying scale factors to resemble the fullsim conditions.

In the case of chargino-chargino ( $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ) production, it is assumed that each chargino decays to a W boson and  $\tilde{\chi}_1^0$ . For chargino-neutralino ( $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ) production, assuming that the chargino and  $\tilde{\chi}_2^0$  are mass degenerate. In this scenario, two cases are considered, when  $\tilde{\chi}_1^\pm$  decays 100% into a Z boson or a H boson plus  $\tilde{\chi}_1^0$ , and the  $\tilde{\chi}_2^0$  always decays into a W boson and  $\tilde{\chi}_1^0$ . All possible SM bosons are allowed in the production. A two-dimensional scan is performed over particle masses in which the

masses of the lightest supersymmetric particle (LSP) and the next-to-LSP (NLSP) vary. The list of SMS signal MC fastsim samples are listed in Table 5.2.

Table 5.2: List of SUSY signal samples used in this analysis with the decay mode, signal model name, and the production cross section for 1 TeV mass of the parent particle.

Decay Mode	SMS model name	$\sigma[\text{pb}]$ for 1 TeV $\tilde{\chi}_1^\pm/\tilde{\chi}_2^0$
$\tilde{\chi}_2^0\tilde{\chi}_2^0 \rightarrow \text{WW} + \tilde{\chi}_1^0\tilde{\chi}_1^0$	TChiWW	$1343 \pm 176$
$\tilde{\chi}_2^0\tilde{\chi}_1^\pm \rightarrow \text{WZ} + \tilde{\chi}_1^0\tilde{\chi}_1^0$	TChiWZ	$1343 \pm 176$
$\tilde{\chi}_2^0\tilde{\chi}_1^\pm \rightarrow \text{WH} + \tilde{\chi}_1^0\tilde{\chi}_1^0$	TChiWH	$622 \pm 77$

**5.2.2.2 Standard model MC samples** Standard model backgrounds are generated with the help of MC simulations which are used to optimize the event selections to get maximum signal sensitivity, derive and apply any scale factors to match the data, validate the data-driven background estimations. In general, the background samples have much high event count than data and are weighted according to the year and luminosity. Majority of the SM samples that include  $t\bar{t} + \text{jets}$ ,  $Z + \text{jets}$ ,  $W + \text{jets}$  and QCD multijet events are simulated with MADGRAPH5\_aMC@NLO 2 at leading-order (LO). The events with single top production are produced at next-to-leading order (NLO) with MADGRAPH5\_aMC@NLO generator for the  $s$ -channel and POWHEG v2 for the  $t$ -channel with associated W boson production. Small SM backgrounds that include the associated production of  $t\bar{t}$  with bosons, diboson and triboson processes are produced at NLO cross sections. Cross sections used to normalize most of SM process samples correspond to next-to-next-to-leading-order precision in QCD. Signal events for the simplified models explored in this analysis are generated in a manner similar to that of the SM backgrounds, with the MADGRAPH5\_aMC@NLO generator

at LO precision. Similar to the signal samples, NNPDF 3.0 (3.1) [43, 44] set of parton distribution functions are used to generate the SM samples. PYTHIA 8.2 is used for parton showering, fragmentation and modeling the underlying event. Additional contributions from the adjacent beam crossings (PU interactions) are modeled by PYTHIA 8.2 generator using multiple minimum bias events. Detector simulation is performed using GEANT4. The list of SM background samples used in this analysis is shown in Tables 5.3 and 5.4.

### 5.3 Event selection

The final states in this search include the two AK8 jets from the boson decays and large  $p_T^{\text{miss}}$ . Keeping this in consideration, the signal yield is maximized while reducing the SM background yields using specific selection criteria as discussed in this section. Four signal regions (each targeting specific signal topology) are used to maximize the sensitivity (or improve the signal to background ratio), broadly classified into b-veto signal region and three b-tagged signal regions. Certain event selections are common for all the signal regions and some control regions, known as baseline sections, outlined in Section 5.3.1. Each signal region (SR) has its corresponding control region (CR) which is used to constrain SM background events. Background estimation techniques are developed using dedicated CRs in order to estimate the amount of background in data. Validation regions (VRs) are used to validate the background estimation procedure in a region that is orthogonal to the SR and CR, which are discussed in Section 5.6.

Table 5.3: List of standard model MC samples used in this analysis with the process name, selection condition in the generator step along with the cross section ( $\sigma$ ) for each Run 2 year for statistically dominant backgrounds. Some SM processes are generated in  $H_T$  bins ( $H_T$  bins are shown in GeV).

Process	Generator selection	2016	$\sigma$ [pb] 2017	2018
$t\bar{t}, \bar{t} \rightarrow 2q, t \rightarrow 2q$	Inclusive, $600 \leq H_T \leq 800$	2.6154	1.7966	1.7966
$t\bar{t}, \bar{t} \rightarrow 2q, t \rightarrow 2q$	Inclusive, $800 \leq H_T \leq 1200$	1.0772	0.7378	0.7378
$t\bar{t}, \bar{t} \rightarrow 2q, t \rightarrow 2q$	Inclusive, $1200 \leq H_T \leq 2500$	0.1950	0.1326	0.1326
$t\bar{t}, \bar{t} \rightarrow 2q, t \rightarrow 2q$	Inclusive, $H_T \geq 2500$	0.0023	0.0015	0.0015
$t\bar{t}, t \rightarrow \ell\nu, \bar{t} \rightarrow 2q$	—	182.16	182.16	182.16
$t\bar{t}, \bar{t} \rightarrow \ell\nu, t \rightarrow 2q$	—	182.16	182.16	182.16
$t\bar{t}, 2\ell$	—	87.34	87.34	87.34
$W \rightarrow \ell\nu_\ell$	$200 \leq H_T \leq 400$	440.76	476.23	478.94
	$400 \leq H_T \leq 600$	59.73	67.69	68.08
	$600 \leq H_T \leq 800$	14.77	16.85	16.93
	$800 \leq H_T \leq 1200$	6.73	7.55	7.66
	$1200 \leq H_T \leq 2500$	1.627	1.685	1.696
	$H_T \geq 2500$	0.0393	0.0407	0.0404
$Z \rightarrow \nu\bar{\nu}$	$200 \leq H_T \leq 400$	95.11	103.59	102.89
	$400 \leq H_T \leq 600$	13.14	14.71	14.67
	$600 \leq H_T \leq 800$	3.134	3.633	3.627
	$800 \leq H_T \leq 1200$	1.444	1.565	1.574
	$1200 \leq H_T \leq 2500$	0.3531	0.3469	0.3472
	$H_T \geq 2500$	0.0085	0.0078	0.0079
$DY + \text{Jets}$	$200 \leq H_T \leq 400$	52.58	56.51	56.51
	$400 \leq H_T \leq 600$	6.761	8.046	8.046
	$600 \leq H_T \leq 800$	1.676	1.967	1.967
	$800 \leq H_T \leq 1200$	0.8314	0.8556	0.8556
	$1200 \leq H_T \leq 2500$	0.1427	0.1926	0.1926
	$H_T \geq 2500$	0.0032	0.0027	0.0027
QCD	$300 \leq H_T \leq 500$	347500	323196	323761
	$500 \leq H_T \leq 700$	32060	30069	30193
	$700 \leq H_T \leq 1000$	6829	6361.49	6325.76
	$1000 \leq H_T \leq 1500$	1207	1095.27	1098.15
	$1500 \leq H_T \leq 2000$	120	100.19	99.99
	$H_T > 2000$	25.25	20.63	20.43

Table 5.4: List of standard model MC samples used in this analysis with the process name, selection condition in the generator step along with the cross section ( $\sigma$ ) for each Run 2 year.

Process	Generator selection	$\sigma[\text{pb}]$		
		2016	2017	2018
t	s-channel	5.401	5.996	5.993
t	t-channel	136.02	113.30	123.47
t	tW-channel	80.95	67.91	73.62
$\bar{t}$	t-channel	20.22	18.67	18.61
$\bar{t}$	tW-channel	20.20	18.71	18.65
WW	$WW \rightarrow \ell\nu q\bar{q}$	50	50	50
WW	$WW \rightarrow \ell\nu\ell\nu$	12.18	12.18	12.18
WZ	$W \rightarrow \ell\nu, Z \rightarrow q\bar{q}$	10.73	11.66	11.76
WZ	$W \rightarrow \ell\nu, Z \rightarrow \nu\bar{\nu}$	3.054	3.294	3.322
WZ	$W \rightarrow \ell\ell, Z \rightarrow q\bar{q}$	5.595	5.606	5.606
WZ	$W \rightarrow \ell\nu, Z \rightarrow \ell\ell$	4.430	4.430	4.430
WH	$W \rightarrow \ell\nu, H \rightarrow b\bar{b}$	0.2616	0.2616	0.2616
ZH	$Z \rightarrow \nu\bar{\nu}, H \rightarrow b\bar{b}$	0.1029	0.1029	0.1029
ZZ	Inclusive	16.523	16.523	16.523

### 5.3.1 Baseline selection

All events in our signal regions are required to satisfy a common set of baseline selection criteria. The baseline selection is defined to retain events from all the targeted signal models, and reduce main backgrounds in this search such as  $W/Z + \text{jets}$  production with  $W \rightarrow \ell\nu$  or  $Z \rightarrow \nu\bar{\nu}$ ,  $t\bar{t}$  production, and QCD multijet production. The baseline selections in this search are:

- *Lepton veto:* The events in which electrons or muons reconstructed using the conditions mentioned in Sections 4.2.4.2–4.2.4.3 are rejected. This search specifically target the fully-hadronic final states; therefore, events with a reconstructed electron or muon are vetoed. The events with a lepton can still enter the search region if they are lost or misidentified.
- *Isolated track veto:* In case the electron or muon fails to get reconstructed, they could be passing the above condition. In order to remove these types of leptons, events with isolated tracks are rejected. Isolation is defined as the ratio of the scalar  $p_T$  sum of all charged-hadron, neutral-hadron, and photon PF candidates in a cone to the track  $p_T$ . Leptonic tracks must have  $p_T > 5$  GeV and isolation value less than 0.2 for a cone with  $\Delta R = 0.3$ . Hadronic tracks are required to have  $p_T > 10$  GeV and isolation value less than 0.1. Additional condition to veto isolated track is if they satisfy:

$$m_T(\text{tk}, \vec{p}_T^{\text{miss}}) = \sqrt{2p_T^{\text{tk}}p_T^{\text{miss}}(1 - \cos \Delta\phi)} < 100 \text{ GeV}, \quad (5.1)$$

where  $p_T^{\text{tk}}$  is the transverse momentum of the track and  $\Delta\phi$  is the azimuthal separation between the track and  $\vec{p}_T^{\text{miss}}$ .



- *Photon veto:* The search focuses on events with hadronic decays of bosons; therefore, any event in which a photon is reconstructed using the selections described in Section 4.2.4.4 are vetoed.
- *Missing transverse energy:* Low values of  $p_T^{\text{miss}}$  are dominated with events from QCD multijet background. Also, in order to extend the search for massive LSPs, the search needs to be performed at higher values of  $p_T^{\text{miss}}$ . This search requires  $p_T^{\text{miss}} > 200$  GeV.
- *$H_T$ :*  $H_T$  is defined as the sum  $p_T$  of all the AK4 jets in an event. This analysis searches for boosted boson hadronic decay and requires  $H_T > 300$  GeV.
- *AK4 jets:* The number of AK4 jets,  $n_j$ , must satisfy  $2 \leq n_j \leq 6$ .
- *AK8 jets:* The hadronic decay of each boosted boson can be captured within an AK8 jet. Number of AK8 jets ( $n_J$ ), must satisfy  $n_J \geq 1$  and at least one AK8 jet with its mass  $m_J > 50$  GeV.
- *Angular separation:* If the azimuthal angle ( $\Delta\phi$ ) between the  $p_T^{\text{miss}}$  and any of the four leading AK4 jets is very small, it is indicative of QCD multijet background. To suppress QCD multijet background,  $\Delta\phi_{1,2,3,4} > 1.5, 0.5, 0.3, 0.3$  conditions are applied for four leading AK4 jets. Also, the azimuthal angular separation between two leading AK8 jets and  $p_T^{\text{miss}}$ ,  $\Delta\phi$  should satisfy  $\Delta\phi_{1,2} > 1.5, 0.5$ .
- *Trigger pass:* The search is performed in bins of  $p_T^{\text{miss}}$  and each event must pass at least one of the signal triggers mentioned in Section 5.2.

After the baseline selections, the events are then classified broadly into the two signal regions which are elaborated in Sections 5.3.2 and 5.3.4. Figure 5.3 shows the event-level kinematics after the baseline selection. Some plots showing the distributions of

Table 5.5: Summary of the baseline event selections. The  $i$ -th highest- $p_T$  AK4 (AK8) jet is denoted by  $j_i(J_i)$ .

Physics object	Selection
Electrons	vetoed, if $p_T > 10$ GeV & $ \eta  < 2.5$
Muons	vetoed, if $p_T > 10$ GeV & $ \eta  < 2.4$
Leptonic isolated tracks	vetoed, if $m_T < 100$ GeV & $p_T > 5$ GeV & $I_{tk} < 0.2$
Hadronic isolated tracks	vetoed, if $m_T < 100$ GeV & $p_T > 10$ GeV & $I_{tk} < 0.1$
Photons	vetoed, if $p_T > 100$ GeV & $ \eta  < 2.5$
$p_T^{\text{miss}}$	$> 200$ GeV
$H_T$	$> 300$ GeV
AK4 jets ( $n_j$ )	2–6
AK8 jets ( $n_J$ )	$\geq 1$ , $m_J > 50$ GeV
$\Delta\phi(j_i, p_T^{\text{miss}})$	$> 1.5, 0.5, 0.3, 0.3$
$\Delta\phi(J_i, p_T^{\text{miss}})$	$> 1.5, 0.5$
Passing $p_T^{\text{miss}}$ trigger	True

jet properties are presented in Figs. 5.4 and 5.5 for the SM backgrounds and three signal models after the baseline selections are applied.

### 5.3.2 $b$ -veto signal region

As the name suggests, this signal region is composed of events with zero  $b$  jets or none of the AK4 jets that are  $b$ -tagged. The definition of  $b$  jets is given in Section 4.2.4.6. This signal region mainly targets TChiWW and TChiWZ signal models where the  $Z$  boson can decay into any pair of  $q\bar{q}$  pair other than  $b\bar{b}$ . In addition to the above baseline selections, this signal region requires at least one AK8 jet has softdrop mass [67] that satisfies the WZ mass requirement ( $65 < m_J < 105$  GeV) and passes the Deep-W threshold cut (W-tag), and at least one jet (other than the W-tagged jet) to satisfy the same WZ mass requirement and pass the Deep-WMD tagging (called as V-tag) requirement. The Deep-WMD tagger is trained on events with no information on the jet mass; therefore, it is used here to identify hadronic decays of both  $W$  and  $Z$  bosons.

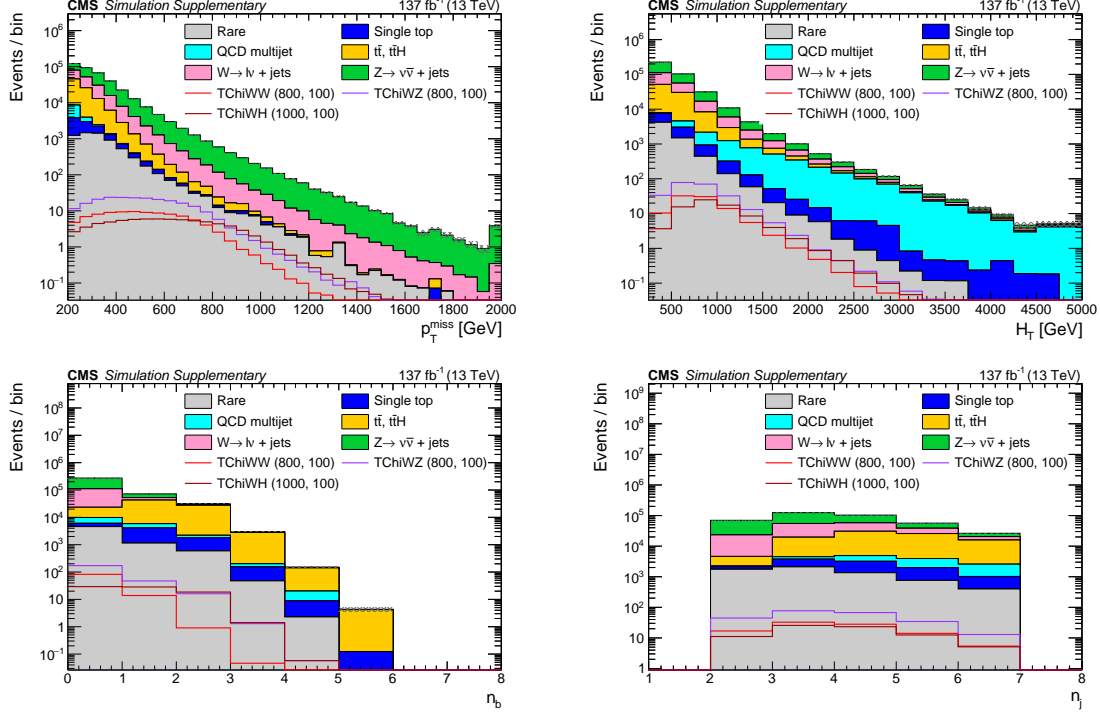


Figure 5.3: Distributions of  $p_T^{\text{miss}}$  (upper left),  $H_T$  (upper right), the number of b tags (bottom left), and the number of AK4 jets (bottom right) with the baseline selection. The filled histograms show the SM predictions based on simulation, and the open histograms show the expectations for selected signal models, which are denoted in the legend by the name of the model followed by the assumed masses of the NLSP and LSP.

The main background in this SR arises from  $W/Z + \text{jets}$  production with  $W \rightarrow l\nu$  or  $Z \rightarrow \nu\bar{\nu}$ . The  $W \rightarrow l\nu$  background is substantially reduced by requiring the number of reconstructed charged leptons and isolated tracks to be zero. These events still satisfy the event selection criteria when the charged lepton lies outside the lepton acceptance, is not reconstructed, or is not isolated. In  $W/Z + \text{jets}$  background events, both W- and V-tagged AK8 jets arise from misidentification of jets not originating from hadronic W or Z boson decays. These events together with background events arising from QCD multijet production do not contain any resonance reconstructed as a single AK8 jet, and are referred to as the “0-res” background.

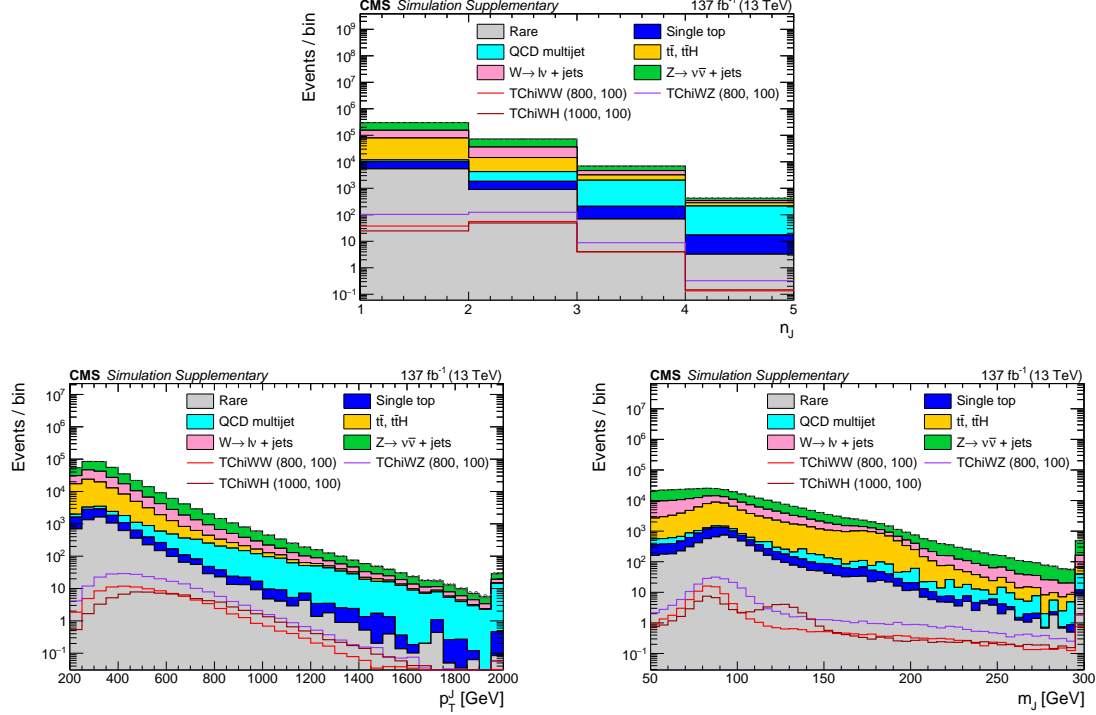


Figure 5.4: Distributions of the number of AK8 jets (upper left), the leading AK8 jet  $p_T$  (upper right), and the soft-drop mass of the leading AK8 jet (bottom) with the baseline selection. The filled histograms show the SM predictions based on simulation, and the open histograms show the expectations for selected signal models, which are denoted in the legend by the name of the model followed by the assumed masses of the NLSP and LSP.

The next largest background contributions come from  $t\bar{t}$ , single top quark, and diboson production. These events typically have one leptonically decaying vector boson and one hadronically decaying vector boson. Therefore, one W- or V-tagged AK8 jet arises from a hadronic W or Z boson decay, and the other tag arises from misidentification. The backgrounds which contain only one resonance reconstructed as a single AK8 jet are referred to as the “1-res” backgrounds. The remaining minor background contributions, which constitute less than 10% of the expected event yield in any  $p_T^{\text{miss}}$  bin of the SR, are expected from rare processes such as triboson production and  $t\bar{t}$  pairs produced in association with a W, Z, or Higgs boson are referred to as

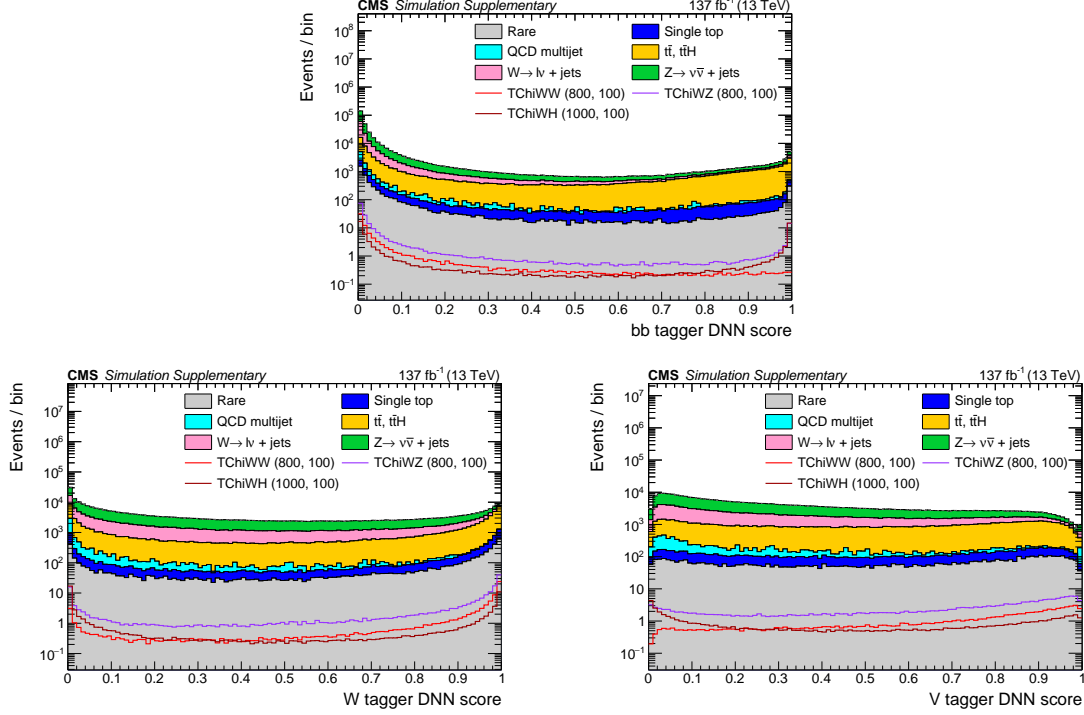


Figure 5.5: Distributions of the leading AK8 jet’s DNN scores for the  $b\bar{b}$  tagger (upper left), the W tagger (upper right), and the V tagger (bottom) with the baseline selection. The filled histograms show the SM predictions based on simulation, and the open histograms show the expectations for selected signal models, which are denoted in the legend by the name of the model followed by the assumed masses of the NLSP and LSP.

rare background. The search region is binned in  $p_T^{\text{miss}}$ . Distributions of  $p_T^{\text{miss}}$  with this binning are shown for SM backgrounds as well as for three signal mass points in Fig. 5.6.

### 5.3.3 $b$ -veto control regions

Control regions (CRs) are defined such that they are dominated with specific types of background in SR to be estimated. Corresponding to the  $b$ -veto SR, two  $b$ -veto CRs are defined. All the  $b$ -veto SR and CRs are summarized in Table 5.6.

- V-tag (or wmdT2M) CR: The terms V-tag CR and wmdT2M CR are used interchangeably to refer this CR and requires almost the same selections as

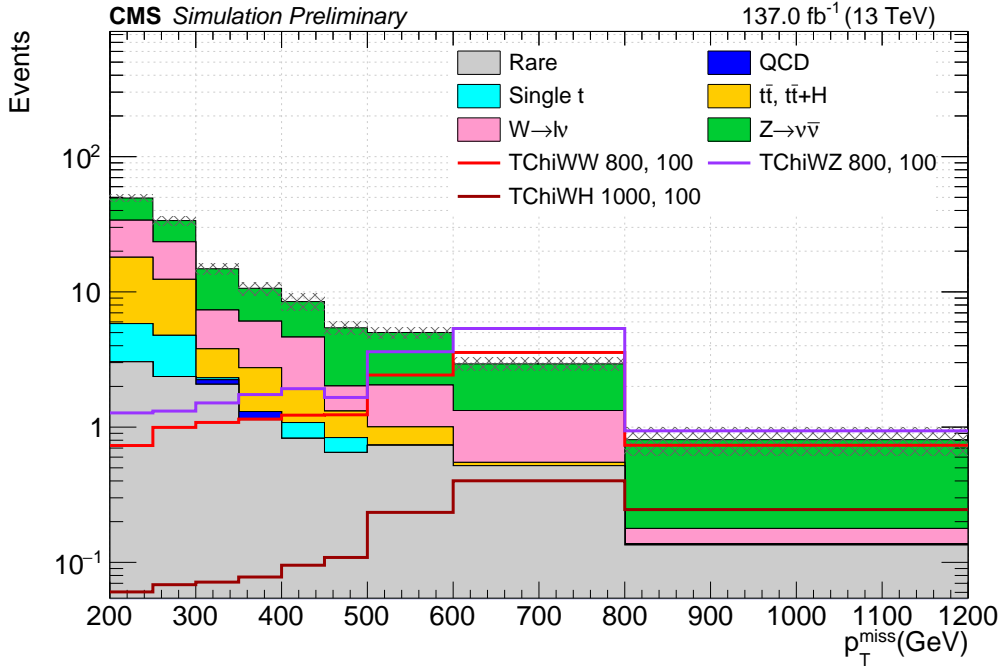


Figure 5.6: Histogram showing the  $p_T^{\text{miss}}$  distribution for various backgrounds (stacked) and signals in the b-veto SR. Filled histograms show predictions of SM background, and the hatched regions show statistical uncertainties for the total SM MC yields. Open histograms show electroweak SMS models. All yields are from direct simulation.

Table 5.6: Summary of tagging requirements for the b-veto SR and CRs. Each of these regions includes the baseline selection described in Table 5.5 and requires zero b-tagged AK4 jets and at least two AK8 jets satisfying  $65 < m_J < 105$  GeV.

Region	Requirements
b-veto (or 2T2M) SR	$\geq 1$ V-tagged jet
	$\geq 1$ W-tagged jet
	$\geq 2$ V- or W-tagged jets
b-veto 0-tag (or 0T2M) CR	0 V-tagged jets
	0 W-tagged jets
b-veto V-tag (or wmdT2M) CR	1 V-tagged jet
	0 other W-tagged jets

SR except inverting the W-tagging requirement. In this CR, at least two AK8 jets with the WZ mass requirement are required, but only one jet out of them is tagged by the Deep-WMD tagger and the other jets are not tagged.

- 0-tag (or 0T2M) CR: The terms 0-tag CR and 0T2M CR are used interchangeably to refer this CR. In this CR, all the jets satisfying the WZ mass requirement are required to fail both the Deep-W and Deep-WMD tagger requirements.

In these control regions, the background composition from  $W + \text{jets}$  and  $Z + \text{jets}$  are enhanced. The control regions are designed so that they are orthogonal to the SR. In addition, a set of validation regions (VRs) is also defined to validate the background estimation procedure, which is discussed in Section 5.6.2. The  $p_T^{\text{miss}}$  distributions in the two CRs are shown in Fig. 5.7. The composition of various backgrounds are shown in Fig. 5.8.

For the purpose of deriving fake scale factors for Deep-W and Deep-WMD taggers, backgrounds that can fake the Deep-W and Deep-WMD tags are required (more on this in Section 5.4). One such control region is the dilepton CR, which

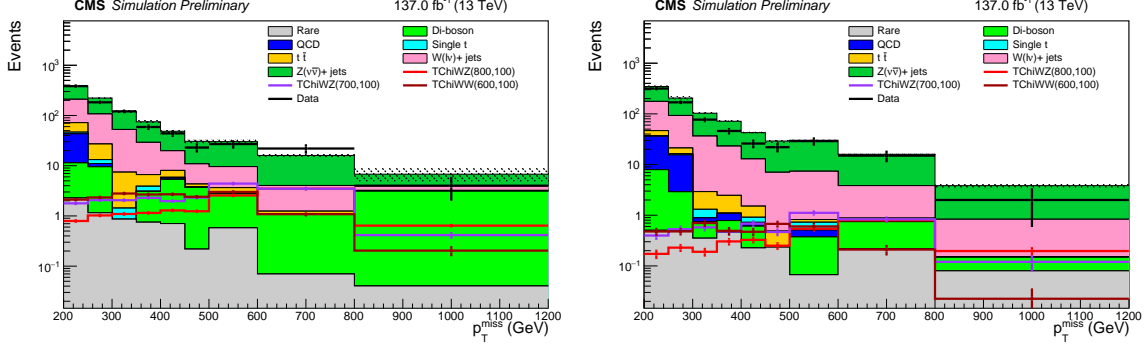


Figure 5.7: Distribution of  $p_T^{\text{miss}}$  in two b-veto control regions: wmdT2M CR (left) and OT2M CR (right). Filled histograms depict SM background predictions. Electroweak SMS models are shown in open histograms. Black markers indicate data yields. All yields are direct simulation predictions. The hatched areas represent statistical uncertainties for total SM MC yields.

requires two opposite-sign leptons, either both electrons or both muons. Drell-Yan (DY) events dominate this CR as shown in Fig. 5.9. The event selection here include:

- Baseline selections except the requirement of either two electrons or two muons, both the leptons are required to satisfy the electron/ muon candidate selections, leading lepton:  $p_T > 60$  GeV and sub-leading lepton:  $p_T > 20$  GeV with opposite charges
- $p_T^{\text{miss}} < 100$  GeV
- At least one of leptons satisfy tight ID
- No b-tagged AK4 jets
- Events with one or more photons are vetoed
- Invariant mass of dilepton,  $75 < m_{\ell\ell} < 105$  GeV
- Number of AK8 jets,  $n_J \geq 1$  with its softdrop mass  $60 < m_J < 105$  GeV
- The AK8 jet should be away from the leptons  $\Delta R(\ell, \text{AK8-jet}) > 0.6$ .



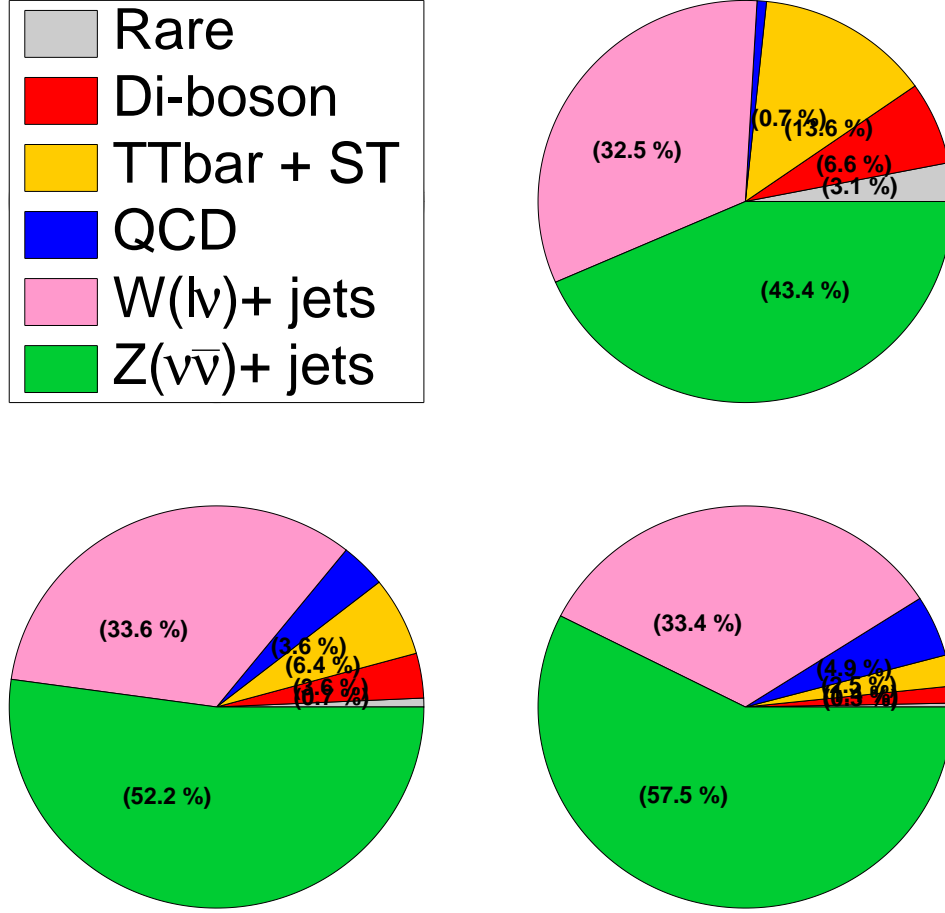


Figure 5.8: The background process compositions in the b-veto signal region (2T2M) (top), and wmdT2M and 0T2M control regions (bottom left and right) as predicted from simulated events. They are evaluated for event yields integrated over all  $p_T^{\text{miss}}$  bins.

#### 5.3.4 *b*-tag signal regions

The b-tag signal region is designed to maximize the acceptance of signal events. An event should have at least one b-tagged AK4 jet to be considered in the b-tag SR, subdivided into three signal regions, WH region, the W region and H region. The AK8 jets are classified into Higgs candidate jets based on AK8 jet proximity to the b-tagged AK4 jet ( $\Delta R < 0.8$ ). These Higgs candidate jet is one of the criteria

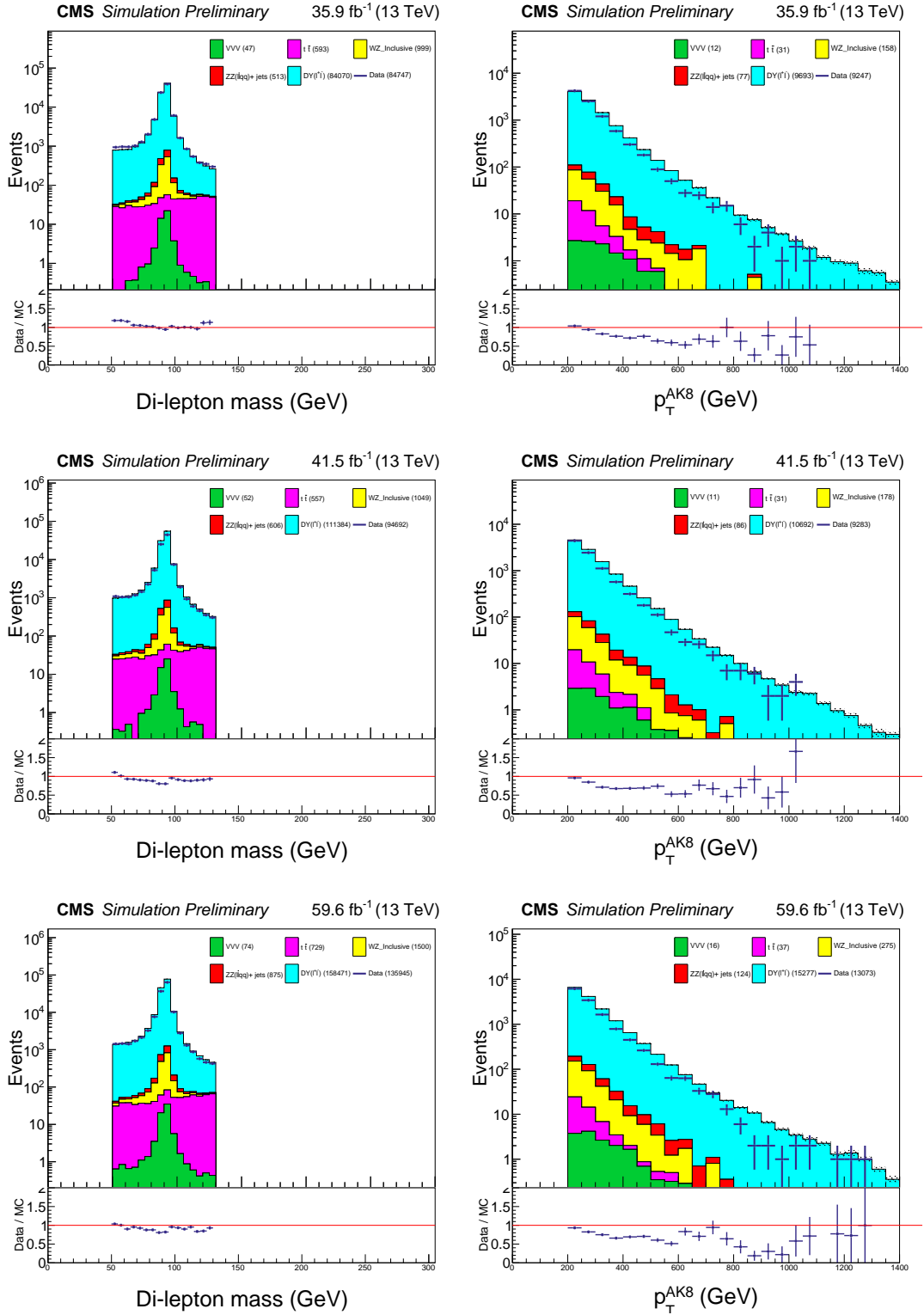


Figure 5.9: Data (points) and SM simulation (filled histograms) distributions of dilepton invariant mass (left) and leading AK8 jet  $p_T$  (right) in the b-veto dilepton CR for 2016, 2017, and 2018.

required to identify both H ( $b\bar{b}$ ) and Z ( $b\bar{b}$ ) decays. All other AK8 jets far away ( $\Delta R > 0.8$ ) from b-tagged AK4 jet are referred to as W candidate jets.

Apart from the jet candidate selections, the selections that define each of the signal regions including the baseline selections are:

- WH (or WHTag) signal region: The W candidate jet is required to be in the WZ mass window (65–105 GeV) and pass the Deep-W tagging requirement. The Higgs candidate jet is required to be in the ZH mass window (75–140 GeV) and pass the bb-vs-light tagging requirement.
- W (WTag) signal region: Only the W candidate jet is required to be in the WZ mass window (65–105 GeV) and pass the Deep-W tagging requirement. Any other Higgs candidate is required to either fail the ZH mass window (75–140 GeV) criteria or fail the bb-vs-light tagging requirement or both.
- H (HTag) signal region: Only the Higgs candidate jet is required to be in the ZH mass window (75–140 GeV) and pass the bb-vs-light tagging requirement. Any other W candidate is required to either fail the WZ mass window (65–105 GeV) criteria or fail the Deep-W tagging requirement or both.

The distribution of various backgrounds in the b-tagged signal regions is shown in Fig. 5.10 and the background fractions are similar in each of the SRs. In these SRs, the top background is dominating with 60% of total background and followed by  $W \rightarrow \ell\nu$  and  $Z \rightarrow \nu\bar{\nu}$  processes contributing about 35% of the total background. Rest of the background includes single top and rare background which account for about 5% of total background. The WH signal region is prominent in obtaining sensitivity for the TChiWH model, and even contributes to 20% of sensitivity for the TChiWZ model.

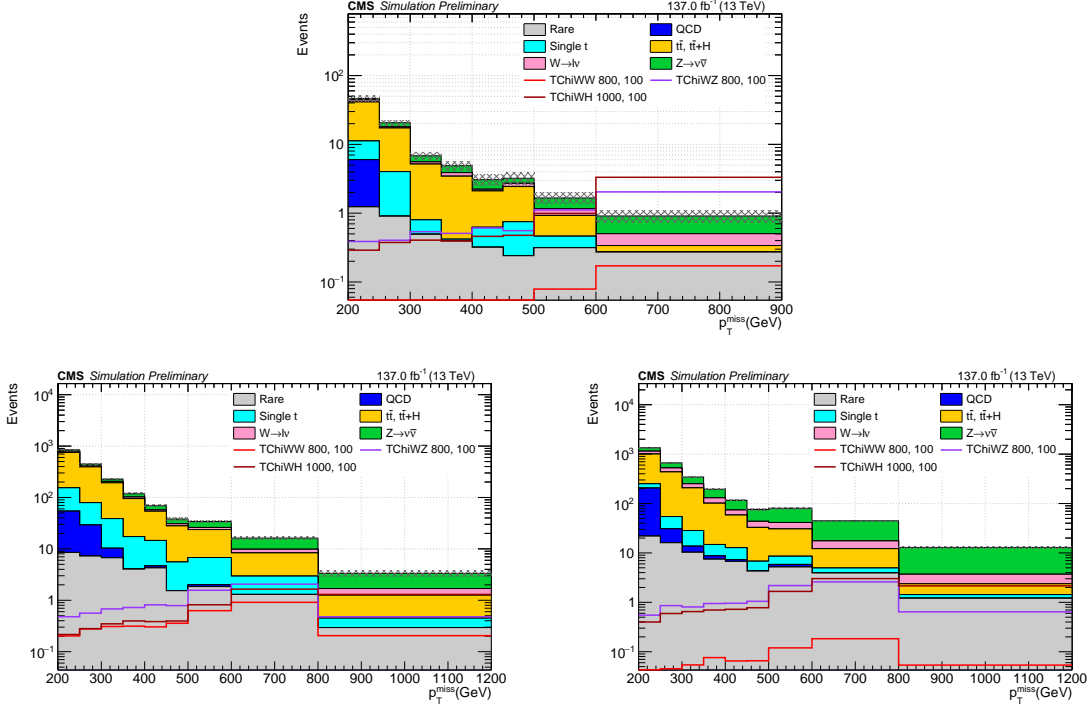


Figure 5.10: Distributions of  $p_T^{\text{miss}}$  for three different signal models and the SM background in three b-tagged signal regions. The events in the top plot pass the WH SR selections. The events that passed the W (H) SR selections are displayed in the bottom left (right) plot. Filled histograms display SM background predictions. Electroweak SMS models are visible in open histograms. All yields are directly predicted by simulation.

### 5.3.5 *b*-tag control regions

Similar to the b-veto CR, three CRs corresponding to three b-tag SRs with the tagger requirement inverted are defined. Here is a summary of all the  $0\ell$  control regions:

- $0\ell$  WH control region: The W candidate jet is required to be in the WZ mass window (65–105 GeV) and fail the Deep-W tagging requirement. The Higgs candidate jet is required to be in the ZH mass window (75–140 GeV) and fail the bb-vs-light tagging requirement.

- $0\ell$  W control region: Only the W candidate jet is required to be in the WZ mass window (65–105 GeV) and fail the Deep-W tagging requirement. Any other Higgs candidate is required to fail the ZH mass window (75–140 GeV) criteria.
- $0\ell$  H control region: Only the Higgs candidate jet is required to be in the ZH mass window (75–140 GeV) and fail the bb-vs-light tagging requirement. Any other W candidate is required to fail the WZ mass window (65–105 GeV) criteria.

The set of these three control regions are referred to as the  $0\ell$  WH,  $0\ell$  W, and  $0\ell$  H antitag CRs, respectively. The distribution of  $p_T^{\text{miss}}$  in each of these CRs is shown in Fig. 5.11. Generally, CRs are defined such that they have no overlap with SRs. Since there are multiple SRs and CRs, there are a few cases where events having three or more AK8 jets could satisfy both the SR and CR requirements, a priority is given to WH regions followed by the W regions. Various  $1\ell$  SRs and CRs are also considered in addition to the six SRs and CRs as shown in Table 5.7, which are identical except that exactly one charged lepton is required.

Since majority of the background is from top decays, it is useful to design control regions enriched in top background. The 1-lepton ( $1\ell$ ) control region are constructed for each of the corresponding  $0\ell$  SR and CR. Therefore, a total of six  $1\ell$  control regions using the same selections as their corresponding  $0\ell$  counter parts are defined, except that exactly 1-lepton is required.

The distributions of the lepton  $p_T$ ,  $\eta$ , and  $m_T$  for the 1-lepton CR events is shown in Fig. 5.12. There is good agreement between the shapes for data and prediction and difference in normalization of 20%. AK8 jet properties for the events

Table 5.7: Summary of tagging requirements for the b-tag SRs and CRs. Each of these regions includes the baseline requirements described in Table 5.5 and requires at least one b-tagged AK4 jet and at least two AK8 jets. The definitions of W and Higgs boson candidates are given in Section 5.3.4. In addition to the six regions described in this table, the b-tag predictions also use six single-lepton CRs that are identical except that exactly one charged lepton is required. A dash (—) indicates that no requirement is imposed.

	W boson candidate		Higgs boson candidate	
	W tagged	not W tagged	$b\bar{b}$ tagged	not $b\bar{b}$ tagged
WH (or WHTag) SR	$\geq 1$	—	$\geq 1$	—
W (or WTag) SR	$\geq 1$	—	0	—
H (or HTag) SR	0	—	$\geq 1$	—
WH antitag CR	0	$\geq 1$	0	$\geq 1$
W antitag CR	0	$\geq 1$	0	0
H antitag CR	0	0	0	$\geq 1$

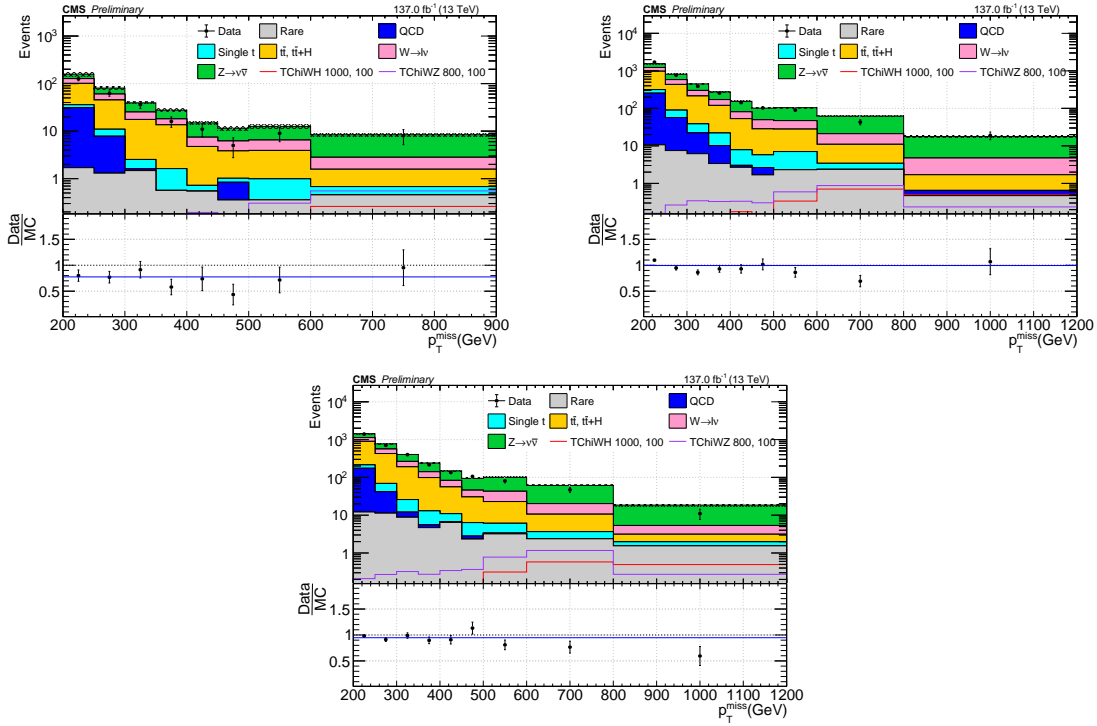


Figure 5.11:  $p_T^{\text{miss}}$  distribution for the events passing the antitag control region selections. The top left plot shows the WH antitag CR. The top right plot shows the W antitag CR. The bottom plot shows the H antitag CR.

in these 1-lepton CR are shown in Fig. 5.13. The kinematic distributions for the 1-lepton CR are shown in Fig. 5.14. All these plots include the baseline and 1 lepton (e or mu) selection requirement.

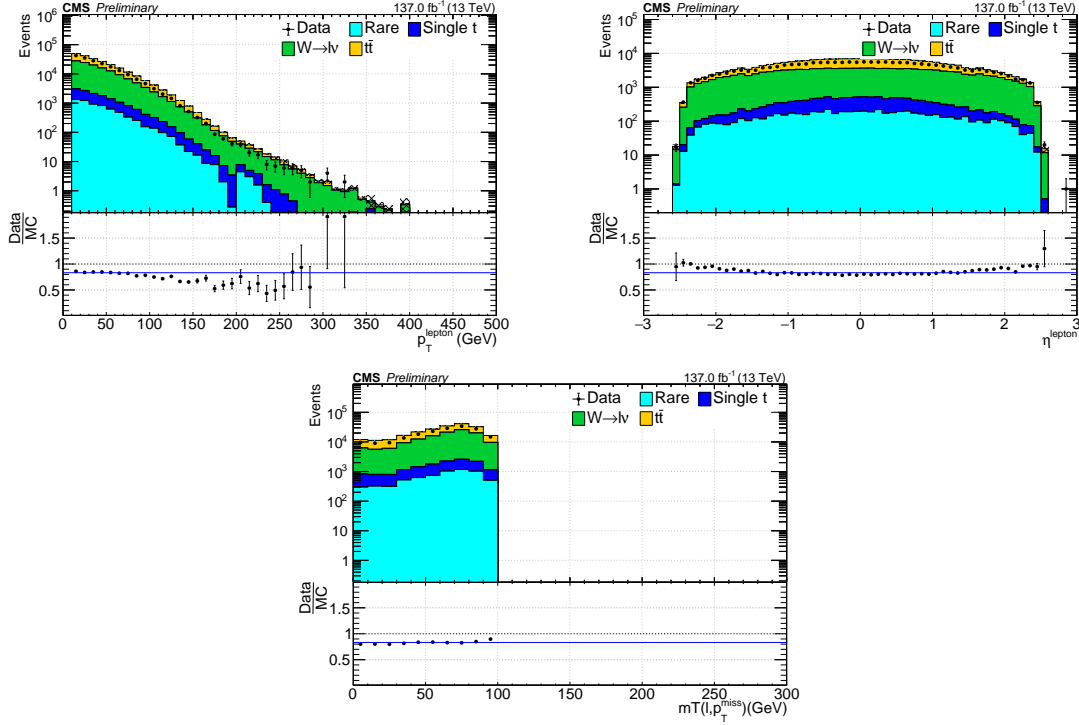


Figure 5.12: The distributions of lepton  $p_T$  (left),  $\eta$  (middle), and  $m_T$  (right) in single lepton events.

The distributions of  $p_T^{\text{miss}}$  in each one of the six  $1\ell$  control regions are shown in Fig. 5.15 which have dominant background from the  $t\bar{t}$  process because of single lepton requirement.

Similar to the one for the dilepton CR defined in the b-veto analysis, another dilepton CR is also defined for the b-tagged analysis except for b-veto requirement. Additional requirements include:

- Separation of b jet from the identified leptons ( $\Delta R > 0.4$ ),

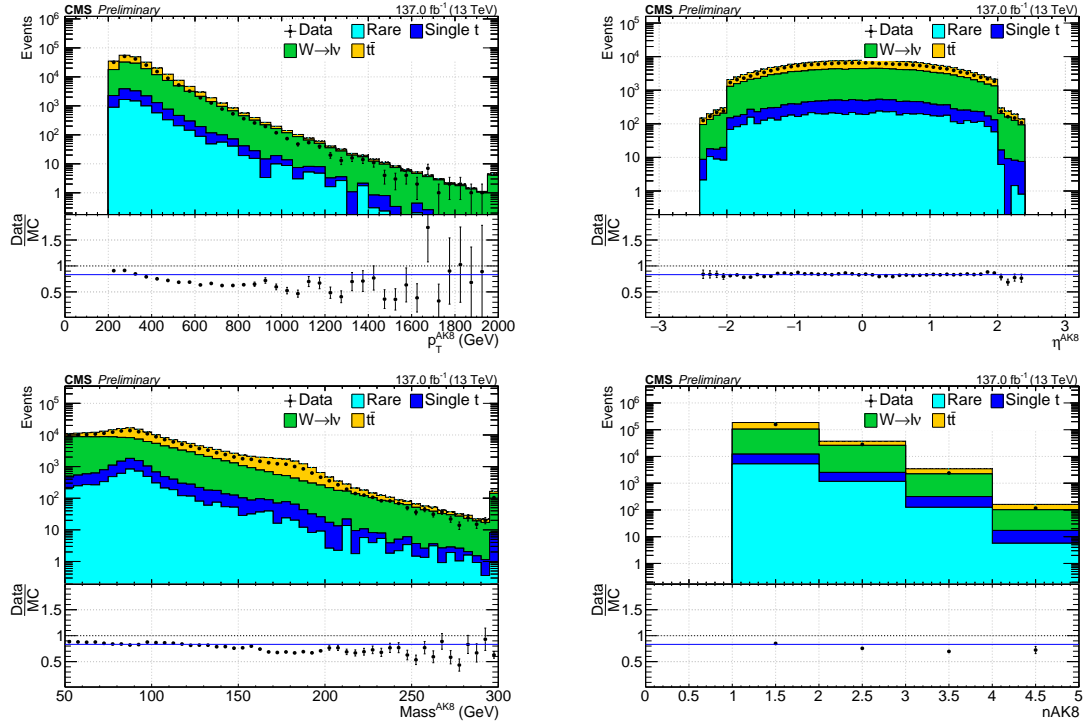


Figure 5.13: Top plots show the leading AK8  $p_T$  (left),  $\eta$  (right) in single lepton events. Bottom plots show the AK8 jet mass (left) and the multiplicity of AK8 jets (right) with  $p_T > 200$  GeV and  $|\eta| < 2.0$ .



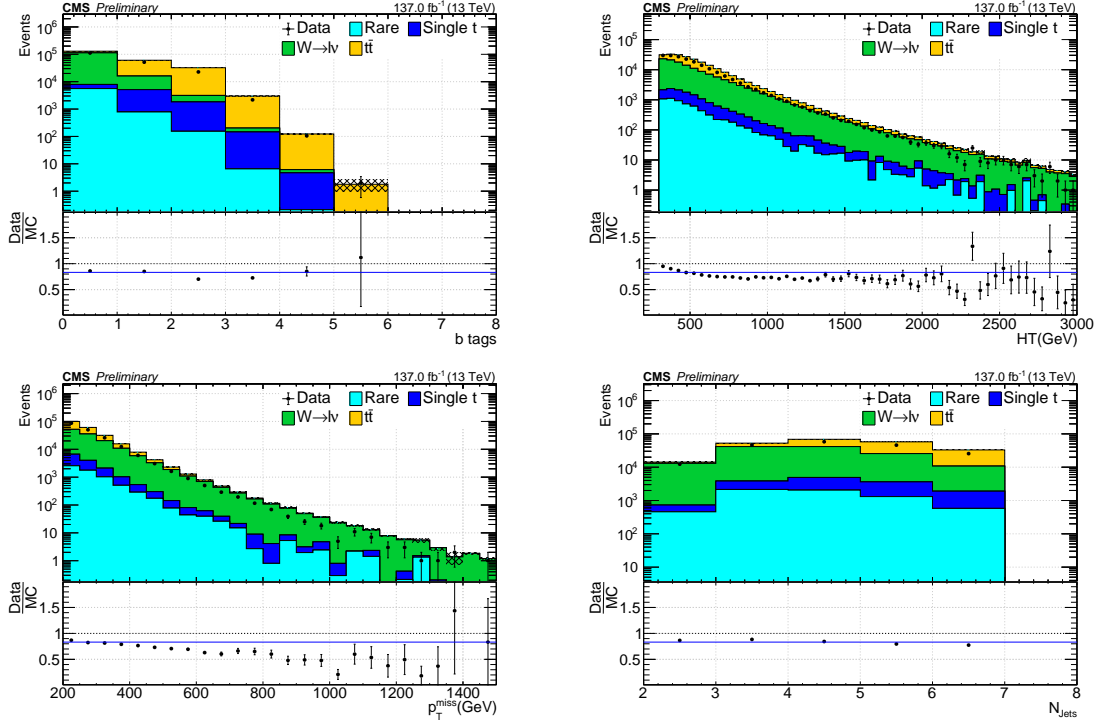


Figure 5.14: Kinematic distributions of b jet multiplicity (top left),  $H_T$  (top right),  $p_T^{\text{miss}}$  (bottom left), and AK4 jet multiplicity (bottom right) in single lepton events.

- Dilepton mass requirement in the range 60–120 GeV,
- $p_T$  of the dilepton system is required to be greater than 200 GeV.

Fig. 5.16 shows the distribution of dilepton invariant mass in the b-tag region.

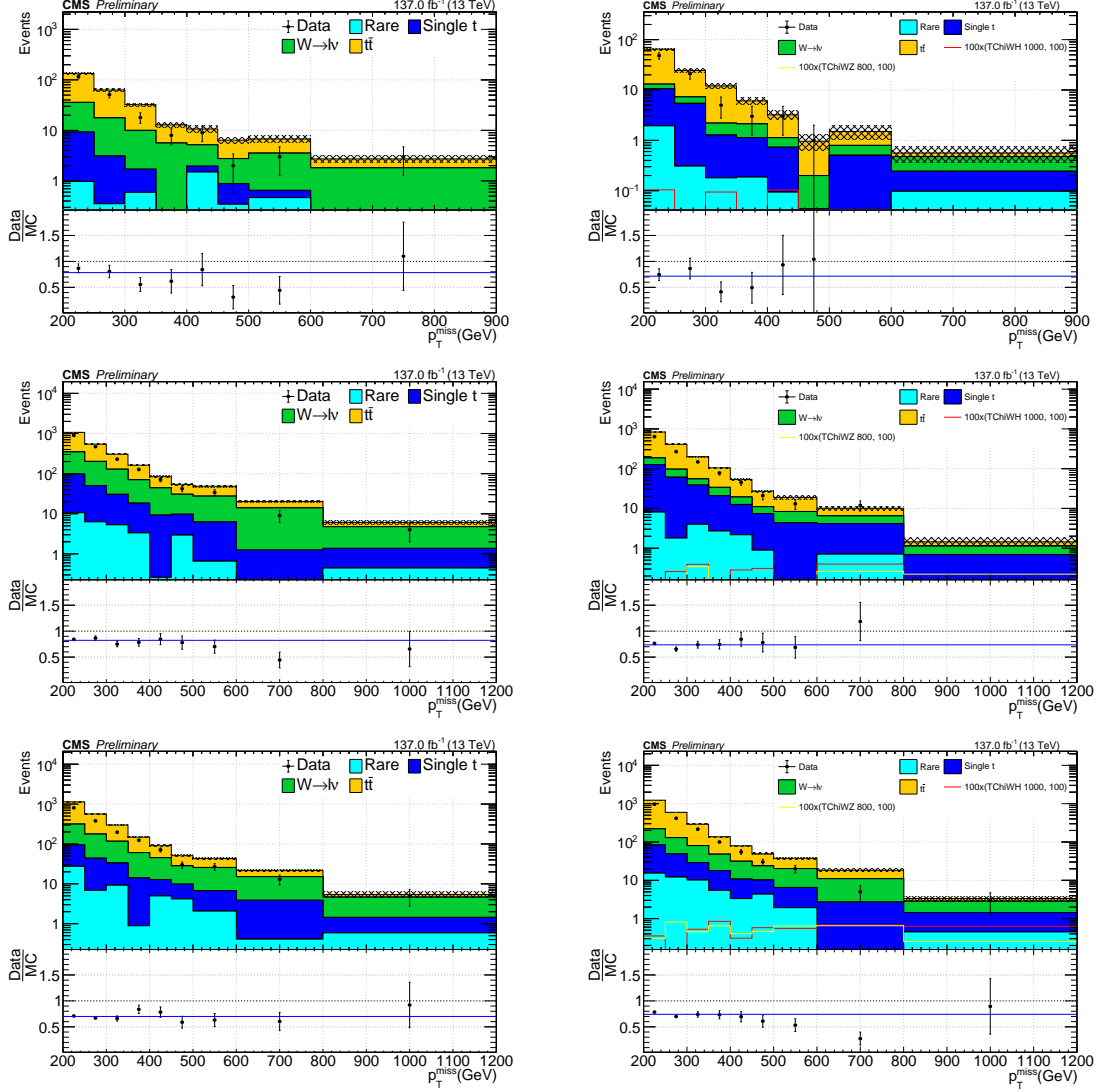


Figure 5.15: Events passing the  $1\ell$  antitag CR definitions (left) and the  $1\ell$  tag CR definitions (right) for b-tagged events. Top:  $1\ell$  antitag CR &  $1\ell$  tag CR for the WH SR, middle:  $1\ell$  antitag CR &  $1\ell$  tag CR for the W SR, bottom:  $1\ell$  antitag CR &  $1\ell$  tag CR for the H SR. Solid open lines in the right plots correspond to the signal events with a lepton, and they are scaled up by 100 for better visibility.

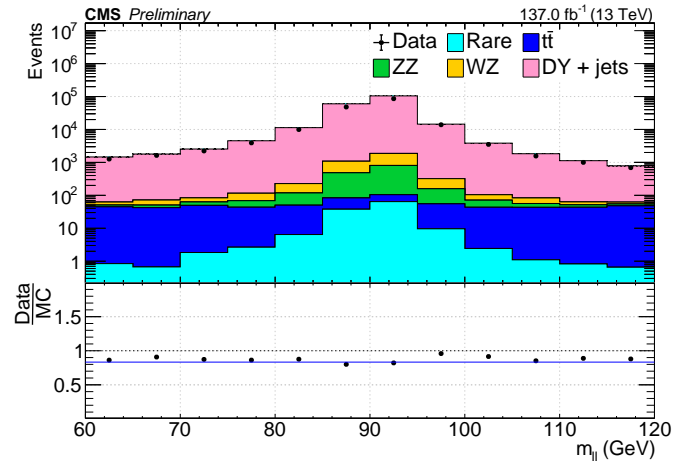


Figure 5.16: Distribution of dilepton invariant mass for data (points) and SM simulation in the b-tag dilepton CR (filled histograms).

## 5.4 Background estimation

The accurate estimation of SM background is crucial in searches for physics beyond the SM. For this analysis, data-driven background estimation techniques are used along with the control regions defined above to predict the main backgrounds in the signal regions. This procedure is different for the b-veto and b-tag analyses as described below.

### 5.4.1 Background estimation for the b-veto SR

Based on the background process in terms of number of resonances with W or Z boson decays, the backgrounds are classified into:

- Non-resonant (or 0-res) backgrounds:  $W(\ell\nu) + \text{jets}$ ,  $Z(\nu\bar{\nu}) + \text{jets}$ , and QCD multijets,
- 1-resonant (1-res) backgrounds:  $t\bar{t}$ , single top, and diboson,
- Rare backgrounds: triboson and  $t\bar{t}+X$  ( $X = W, Z, H$ ).

If any background process has exactly one AK8 jet tagged by of Deep-W and/or Deep-WMD taggers from real  $W/Z \rightarrow q\bar{q}$  decays, it is said to have one resonance. The 0-res backgrounds enter the SR when at least two AK8 jets are misidentified by Deep-W and/or Deep-WMD but contain no real  $W/Z \rightarrow q\bar{q}$ , meaning both jets are misidentified. In this case, the W and Z bosons decay leptonically and the large  $p_T^{\text{miss}}$  is from the neutrinos. As seen in Fig. 5.8, the dominant backgrounds are the 0-res backgrounds, followed by 1-res background ( $\sim 20\%$ ) and the remaining is the rare background. The 1-res backgrounds enter the SR when one of the AK8 jet originates from  $W/Z \rightarrow q\bar{q}$  and tagged by either Deep-W or Deep-WMD taggers, and the other jet is also tagged but not from real  $W/Z \rightarrow q\bar{q}$ , hence there is one misidentified jet.

Events in the rare background category often have two resonances,  $W$  ( $q\bar{q}'$ ) or  $Z$  ( $q\bar{q}$ ), that are tagged by the Deep-W and/or Deep-WMD taggers and fall into the b-veto SR.

To quantify, around 96% of the events categorized as 1-res backgrounds have AK8 jets that are matched to generator level  $W/Z \rightarrow q\bar{q}$  within  $\Delta R(\text{AK8}, W/Z) < 0.6$ . For rare backgrounds, around 84% of the events have two AK8 jets matched to generator level  $W/Z \rightarrow q\bar{q}$  with the same matching condition, and the rest of the events have only one AK8 jet that is matched to  $W/Z \rightarrow q\bar{q}$ . The misidentification rate for the Deep-W and Deep-WMD taggers is much lower compared to identification of  $W/Z \rightarrow q\bar{q}$  decays. The 0-res and 1-res backgrounds are estimated using the background estimation procedure, while the rare background is estimated directly from the MC simulation.

Mathematically, the number of background events in each of the  $p_T^{\text{miss}}$  search bin are given by:

$$N_{\text{SR}}^{\text{pred}} = \mathcal{R}_i(N_{\text{CR}_i}^{\text{data}} - N_{\text{CR}_i, \text{rare}}^{\text{MC}}) + N_{\text{SR}, \text{rare}}^{\text{MC}},$$

where  $N$  denotes the number of events expected (or observed in data) in regions and from processes indicated by the superscripts and subscripts,  $\mathcal{R}_i = N_{\text{SR}, 0\&1\text{-res}}^{\text{MC}}/N_{\text{CR}_i, 0\&1\text{-res}}^{\text{MC}}$ , and  $\text{CR}_i$  is either the 0T2M or wmdT2M CR.

Since there are two control regions (0T2M and wmdT2M), two transfer factors ( $\mathcal{R}_i$ ) are used to make the predictions for 0-res and 1-res backgrounds. The rare background yields are directly taken from the MC simulation. Transfer factor is only applied on events in data control region with the rare background fraction subtracted. The background composition of the wmdT2M CR is very similar to the 2T2M SR, as

shown in Fig. 5.8. However, some non-negligible amount of signal events enters this wmdT2M CR, which is referred to as signal contamination.

Here, the 0T2M CR helps in mitigating the signal contamination as shown in Fig. 5.7, since none of the AK8 jet candidates (W/Z) is allowed to be tagged by either of the taggers. The 1-res background fraction in this CR is less than 5% which is much smaller than about 20% in the case of the b-veto SR, and this deficit makes the TF sensitive to simulation of 1-res background. A combination of both the control regions are used as discussed in Section 5.6.1.

The Deep-W and Deep-WMD tagger scale factors are applied to simulated events to minimize the tagging efficiencies between data and MC simulation. For the real  $W \rightarrow q\bar{q}'$  AK8 jet decays, the data/MC scale factors are measured by the JetMET POG [68] to match the W tagging rates in data. The corrections for jets matched to generator-level W bosons are obtained from a  $t\bar{t}$ -enriched sample in which one of the W bosons from the top quark decays leptonically and the other decays hadronically. The same corrections are also applied to jets matched to generator-level Z bosons. Since the Deep-WMD tagger efficiencies are similar for W and Z bosons, the same scale factors are used for  $Z \rightarrow q\bar{q}$  AK8 jets in this analysis. No uncertainty is assigned for tagging either the W or Z decay (have  $< 5\%$  effect on total yield), since the tagger uncertainty dominates with around 10%. Since most of the jets in the 0-res and 1-res backgrounds come from a misidentified jet, deriving the scale factors for misidentification forms a crucial part of this analysis. These scale factors are measured using the dilepton control regions as discussed below, and verified in validation regions as discussed in Section 5.6.2.

Table 5.8: Summary of W-tagging data/MC scale factor measurements.

$p_T$ [ GeV ]	200–250	250–300	300–400	400–500	500–600	600–800
2016	$1.25 \pm 0.05$	$1.25 \pm 0.07$	$1.52 \pm 0.10$	$1.42 \pm 0.17$	$1.73 \pm 0.40$	$1.75 \pm 0.47$
2017	$1.12 \pm 0.04$	$1.20 \pm 0.06$	$1.24 \pm 0.08$	$1.42 \pm 0.16$	$1.17 \pm 0.26$	$1.25 \pm 0.40$
2018	$1.08 \pm 0.04$	$1.16 \pm 0.05$	$1.26 \pm 0.07$	$1.34 \pm 0.14$	$1.43 \pm 0.26$	$1.47 \pm 0.33$

#### 5.4.1.1 Deep-W/WMD data/MC scale factor measurements

The scale factors are derived in the dilepton CR, as discussed in Section 5.3.3.

The scale factors are measured as the ratio of tagging efficiencies of 0-res component in data to those in MC simulation. The tagging efficiency is defined as the ratio of number of events tagged (or passing the discriminator cut) to total number of events.

The scale factors are measured by:

$$SF(\text{faketag}) = \frac{Data_{\text{pass}} - (\text{resonant}_{\text{pass}}^{\text{MC}} \cdot SF)}{Data_{\text{all}} - (\text{resonant}_{\text{all}}^{\text{MC}})} \bigg/ \frac{0\text{-res}_{\text{pass}}^{\text{MC}}}{0\text{-res}_{\text{all}}^{\text{MC}}}. \quad (5.2)$$

Here, the terms resonant and 0-res refer to an event containing resonant or 0-res AK8 jets. A non-resonant AK8 jet here refers to an AK8 jet that is not matched with generator level  $W/Z \rightarrow q\bar{q}$  within  $\Delta R(\text{AK8}, W/Z) < 0.6$ , whereas resonant AK8 jet refers to an AK8 jet that is matched. The subscript “pass” denotes all the MC events that pass the tagger threshold, while “all” denotes all the events that pass the baseline selections for the dilepton CR. Also, the AK8 jets in this CR are required to be away from the leptons,  $\Delta R(\ell, \text{AK8 jet}) > 0.6$ . The measured scale factors in this CR are higher than 1, which are consistent with the trend observed by the JetMET POG (as shown in Ref. [63], Fig. 43). The scale factors are derived for each year separately for  $p_T$  of jet in specific range intervals as shown in Tables 5.8 and 5.9.

The distributions of the Deep-W (Deep-WMD) tagger score for the AK8 jet with highest  $p_T$  in each event are shown in Fig. 5.17.

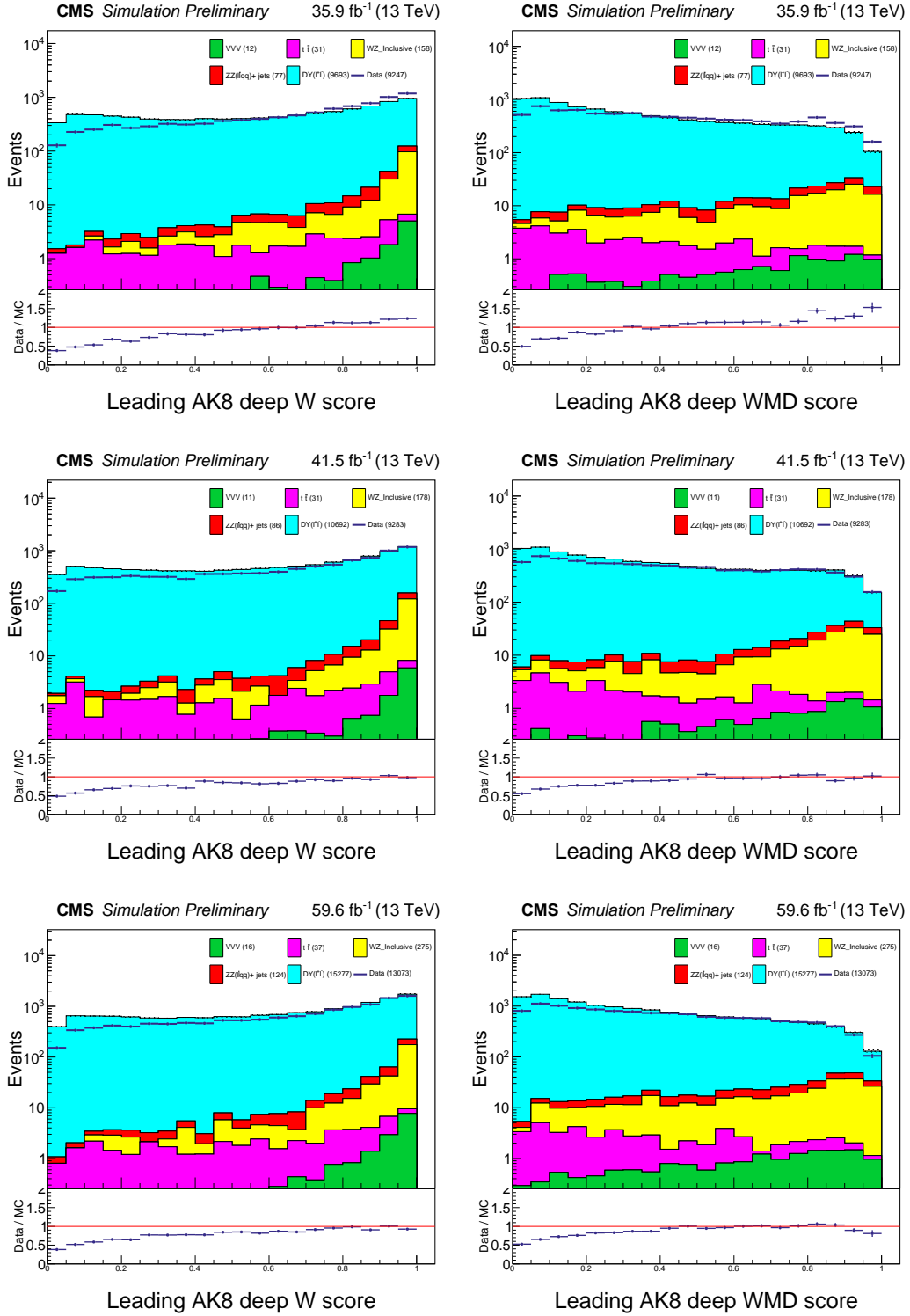


Figure 5.17: Distributions of Deep-W (left) and Deep-WMD (right) discriminators measured in the dilepton control regions for 2016, 2017, and 2018 data. The points are for data, and the filled histograms are for SM simulation events.



Table 5.9: Summary of V-tagging data/MC scale factor measurements.

$p_T$ [ GeV ]	200–250	250–300	300–400	400–500	500–600	600–800
2016	$1.24 \pm 0.03$	$1.31 \pm 0.05$	$1.41 \pm 0.06$	$1.26 \pm 0.11$	$1.36 \pm 0.21$	$1.55 \pm 0.31$
2017	$1.14 \pm 0.03$	$1.15 \pm 0.04$	$1.25 \pm 0.05$	$1.32 \pm 0.10$	$1.52 \pm 0.20$	$1.36 \pm 0.28$
2018	$1.14 \pm 0.03$	$1.20 \pm 0.04$	$1.29 \pm 0.05$	$1.28 \pm 0.09$	$1.23 \pm 0.17$	$1.47 \pm 0.22$

For the AK8 jet candidates that pass the WZ mass requirements ( $65 < m_J < 105$  GeV), the scale factors can be applied as an event weight, where the event weight is a number multiplied to the MC yield:

$$w_W = \prod_{\text{tagged by W}} SF_W \prod_{\text{not tagged}} \frac{(1 - \epsilon_W \cdot SF_W)}{1 - \epsilon_W}, \quad (5.3)$$

$$w_{WMD} = \prod_{\text{tagged by WMD}} SF_{WMD} \prod_{\text{not tagged}} \frac{(1 - \epsilon_{WMD} \cdot SF_{WMD})}{1 - \epsilon_{WMD}}. \quad (5.4)$$

Depending on the matching of the AK8 jet with generator-level  $W/Z \rightarrow q\bar{q}'$  within  $\Delta R(\text{AK8}, W/Z) < 0.6$ , different scale factors are applied. In case of matching, the scale factors for  $W \rightarrow q\bar{q}'$  provided by the JetMET POG are used. If no match, then the fake scale factors derived above are used. The symbols  $SF_W$  and  $SF_{WMD}$  are the data/MC scale factors for Deep-W and Deep-WMD taggers, and  $\epsilon_W$  and  $\epsilon_{WMD}$  are the efficiencies of Deep-W and Deep-WMD tagging. These efficiencies are derived from MC simulations for both  $W \rightarrow q\bar{q}'$ ,  $Z \rightarrow q\bar{q}$  processes with the same matching conditions shown above as fractions of AK8 jets that pass  $p_T$ ,  $\eta$ , WZ mass window, and the tagger requirements. An AK8 jet can have two tagger scores and can result in four scenarios:

- tagged by both Deep-W and Deep-WMD,
- tagged by Deep-W, but not by Deep-WMD,
- tagged by Deep-WMD, but not by Deep-W, and

- tagged by neither Deep-W not by Deep-WMD.

The event weight used in this analysis that provides good description of the tagging rates is given by:

$$weight = \prod_{\text{tagged by W}} SF_W \prod_{\substack{\text{tagged by WMD,} \\ \text{not by W}}} SF_{WMD} \prod_{\text{not tagged}} \frac{(1 - \epsilon_W \cdot SF_W - \epsilon_{WMD}^{\text{not W}} \cdot SF_{WMD})}{(1 - \epsilon_W - \epsilon_{WMD}^{\text{not W}})} \quad (5.5)$$

where  $\epsilon_{WMD}^{\text{not W}}$  is the efficiency of AK8 jets tagged only by Deep-WMD.

#### 5.4.2 Background estimation for the $b$ -tag SRs

Similar to the background classification in the  $b$ -veto analysis, the backgrounds for  $b$ -tag analysis are classified as:

- 0-res backgrounds: No real hadronic decays of W, Z, or Higgs boson are present, processes such as  $W \rightarrow \ell\nu$  and  $Z \rightarrow \nu\bar{\nu}$  plus jets contribute to this category,
- Top backgrounds: In these backgrounds, at least one hadronic W is produced, mostly from the top pair production,
- Rare backgrounds: These processes have two resonances which includes top quark pair with a W, Z, or Higgs boson, diboson and triboson processes.

The 0-res backgrounds are estimated using a control region that has tagging requirements inverted w.r.t. signal region, similar to that of  $b$ -veto background estimations. For top backgrounds, 1-lepton ( $1\ell$ ) control regions are used (subdivided again into tag and antitag regions). The rare backgrounds are taken directly from the MC simulations. The schematic diagram showing SR, various CRs, and the relevant transfer factors is shown in Fig. 5.18.

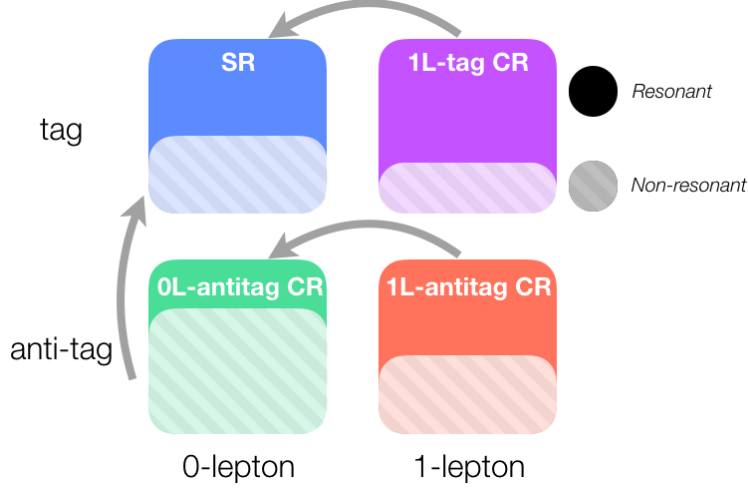


Figure 5.18: Schematic diagram showing how control regions are used to extrapolate both top and 0-res backgrounds into a b-tag signal region. Solid areas denote relative contributions of top backgrounds. Striped regions represent the relative contribution of 0-res backgrounds.

First, the estimation is made for the top backgrounds in  $0\ell$  tag and antitag regions using the  $0\ell/1\ell$  transfer factors shown by the horizontal arrows in the schematic diagram. The  $0\ell/1\ell$  TF are computed as the ratio of top background yield in  $0\ell$  region to the top background yields in  $1\ell$  control region. These TF are derived for each of the three b-tag signal regions. The top background prediction in each SR is done by applying this TF on the observed yields in the  $1\ell$  region as expressed in Eq. (5.6):

$$N_{0\ell,\text{res}}^{\text{pred}} = N_{0\ell,\text{res}}^{\text{MC}} N_{1\ell}^{\text{data}} / N_{1\ell,\text{all}}^{\text{MC}}. \quad (5.6)$$

The  $0\ell/1\ell$  TFs are calculated as a function of  $p_T^{\text{miss}}$  and in total there are six  $0\ell/1\ell$  TFs (three for the SRs and three for antitag CRs) and they are shown in Fig. 5.19.

Once the top background prediction is calculated in the SR, the 0-res background contribution is predicted using the  $0\ell$ -antitag CR. A pass-fail transfer factor,  $\mathcal{R}_{p/f}$ , is defined as the ratio of number of 0-res events passing the SR selections to

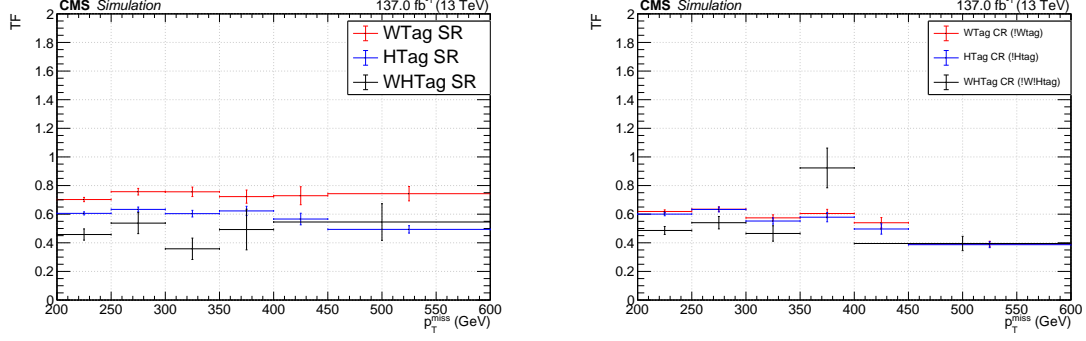


Figure 5.19: Transfer factors for extrapolating from  $1\ell$  to  $0\ell$  regions as computed using simulations in each of the three b-tagged SRs and for each  $p_T^{\text{miss}}$  bin. The left plot shows the TFs for the tag regions. The right plot shows the TFs for the antitag regions. The WH, W, and H tag topological regions are represented with different colors in each plot.

the number of 0-res events passing the  $0\ell$ -antitag CR selections. This TF is derived for each of the SR separately as shown in Fig. 5.20.

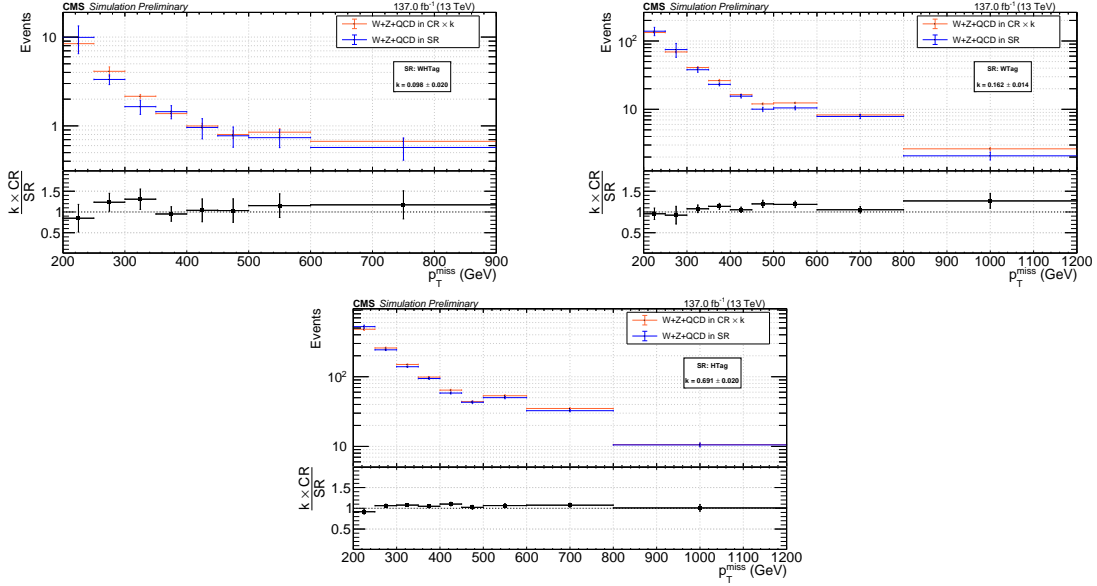


Figure 5.20: Pass-fail ratio as computed from simulation is shown for each b-tagged SRs and for each  $p_T^{\text{miss}}$  bin. The left plot shows this ratio for the WH SR. The right (bottom) plot shows this ratio for the W (H) SR. The top panels show the normalized distribution of the pass-fail ratio. The average value of the pass fail ratio in each region is reported as  $\kappa$ . The lower panels show the pass-fail ratio versus  $p_T^{\text{miss}}$  after dividing out the average value.

The obtained TF is multiplied with the observed data events in the  $0\ell$  antitag CR that has no contributions from the top and rare backgrounds in this region. So the  $0\ell/1\ell$  TF helps in making the top background prediction in  $0\ell$  antitag CR ( $N_{!tag,res}^{pred}$ ) and the rare background ( $N_{!tag,rare}^{MC}$ ) is taken directly from MC simulation. These two contributions are subtracted from the observed data yields. The 0-res background predictions are shown by Eq. (5.7):

$$N_{non-res}^{pred} = \mathcal{R}_{p/f}(N_{!tag}^{data,0\ell} - N_{!tag,res}^{pred} - N_{!tag,rare}^{MC}) = \frac{N_{tag,non-res}^{MC}}{N_{!tag,non-res}^{MC}}(N_{!tag}^{data,0\ell} - N_{!tag,res}^{pred} - N_{!tag,rare}^{MC}). \quad (5.7)$$

The  $0\ell/1\ell$  TF is mostly flat for dominant 0-res backgrounds with around 30% variation over the full  $p_T^{miss}$  with an average value, denotes as  $\kappa$ , which is around 0.17. To correct for the tagging efficiency differences in the SR to the  $0\ell$ -antitag CR, the scale factors are derived for this pass-fail ratios. Using the dilepton control region selections similar to that in Section 5.3.3 except that here it requires at least one b-tagged jet, which are dominated by Drell-Yan + jets processes, the pass-fail scale factors are derived for the Deep-W and bb-vs-light taggers as shown in Table 5.10.

Table 5.10: W-tag and bb-vs-light pass-fail ratio SFs derived from dilepton b-tag regions.

P/F SF for	2016	2017	2018	Run 2
W-tag (1 AK8)	$1.69 \pm 0.29$	$1.43 \pm 0.23$	$1.38 \pm 0.19$	$1.48 \pm 0.13$
W-tag (2 AK8)	$1.56 \pm 0.45$	$1.72 \pm 0.47$	$1.32 \pm 0.29$	$1.51 \pm 0.22$
bb-vs-light (1 AK8)	$1.02 \pm 0.08$	$1.00 \pm 0.07$	$0.83 \pm 0.05$	$0.93 \pm 0.04$
W-tag (2 AK8)	$0.98 \pm 0.20$	$0.89 \pm 0.17$	$1.75 \pm 0.12$	$0.86 \pm 0.09$

## 5.5 Systematic uncertainties

The uncertainties related to each of the correction or method applied to the background events are discussed in this section. The b-veto and b-tag analyses are affected by a different set of systematic uncertainties, so systematic uncertainties in the two SRs are discussed separately below. The uncertainties are evaluated in  $p_T^{\text{miss}}$  bins generally, though constant uncertainties are used when no systematic trend is observed in  $p_T^{\text{miss}}$ . There are uncertainties associated with the limited amount of MC simulation events and the data events in the CR which are categorized under statistical uncertainties and listed in the summary table below.

### 5.5.1 Systematic uncertainties for the b-veto SR

The background estimation section shows that the  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$  transfer factor is crucial to predict a large fraction of SM background. Sources of systematic uncertainties arise from the procedures mostly involving event reconstruction and are generally not random in nature. Also the low number of data events or MC simulation yields in the control regions lead to several sources of uncertainties as outlined below. The systematic uncertainty is evaluated by varying one of the source with its systematic uncertainty up or down.

The uncertainties on the  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$  transfer factor and rare backgrounds that are considered are:

- Deep-W and Deep-WMD tagger uncertainties: The tagger uncertainties for  $W/Z \rightarrow q\bar{q}$  and misidentification are large compared to the rest of the uncertainties. The uncertainties on the TF from the 0T2M region are greater compared to those obtained from the wmdT2M region since the mistag tag

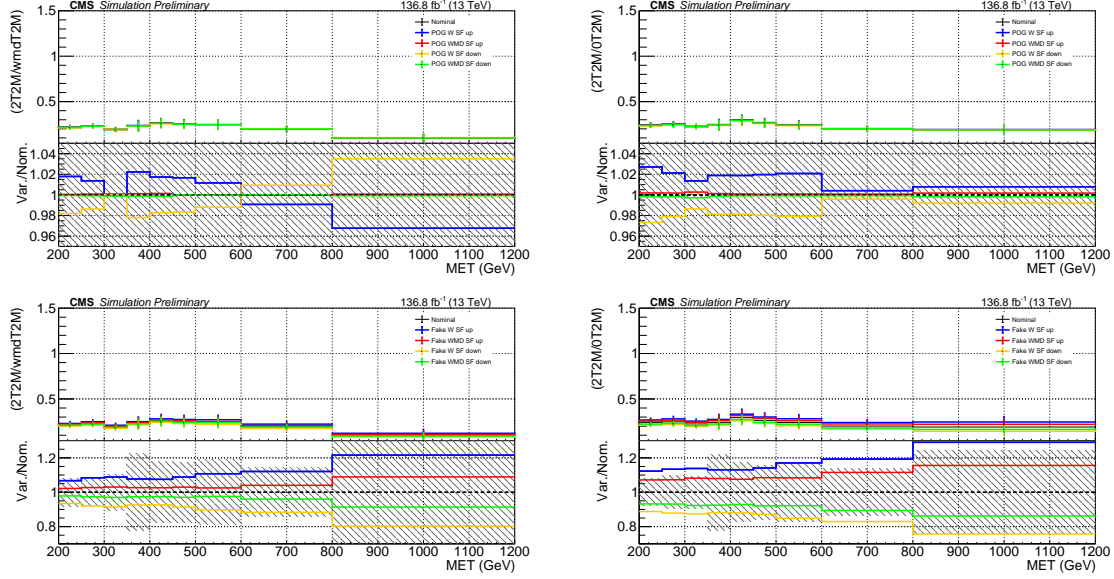


Figure 5.21: The systematic uncertainties on the transfer factors  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$ , used for the non-resonant plus 1-resonant background prediction in the signal region, due to the Deep-W and Deep-WMD data/MC scale factors for  $W/Z \rightarrow q\bar{q}$  (top) and misidentification (bottom). The uncertainties are evaluated separately for two associated control regions, wmdT2M (left) and 0T2M (right).

uncertainties are applied twice on the events in 0T2M regions while it's applied once on the wmdT2M region events. As signal contamination in the 0T2M CR is much smaller compared to the wmdT2M CR, there is an advantage to include the 0T2M CR and the large uncertainties from this region. For the rare backgrounds, typically containing two  $W/Z$  resonances, the Deep-W and Deep-WMD uncertainties are around 11–16% while the mistag SF are less than 1% since a majority of the events in these backgrounds enter the SR. Figure 5.21 shows variations in the TFs due to the Deep-W and Deep-WMD taggers due to the Deep-W and Deep-WMD tagger uncertainties for identifying  $W/Z \rightarrow q\bar{q}$  or misidentification.

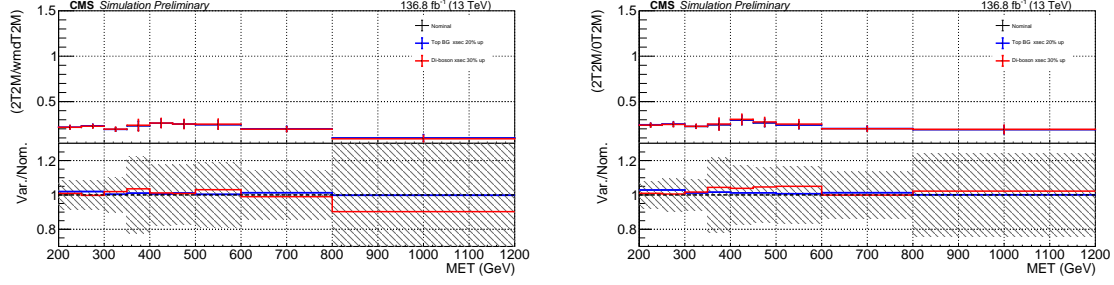


Figure 5.22: The systematic uncertainties on the transfer factors  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$ , used for the non-resonant plus 1-resonant background prediction in the signal region, due to uncertainties in the  $t\bar{t}$ , single top, and diboson production cross section. The uncertainties are evaluated separately for two associated control regions, wmdT2M (left) and 0T2M (right).

- 1-resonant background normalization: Based on the differential cross section measurements of the  $t\bar{t}$  and single top production [69, 70], uncertainties of 20% are assigned to the top background normalization ( $t\bar{t}$  and single top production). The diboson background normalization is also based on differential cross section measurements [71, 72] and uncertainty of 30% are assigned. The systematic variations in the transfer factors are evaluated by varying the top and diboson background normalizations by 20% and 30%, respectively, as shown in Fig. 5.22. Since there is no systematic shape dependence, constant uncertainties of 0.7% and 2.7% obtained by integration over all the  $p_{\text{T}}^{\text{miss}}$  bins are applied for the top and diboson uncertainties, respectively.
- JEC/JER uncertainties: The jet energy corrections and jet energy resolution uncertainties are evaluated by applying the variations to the transfer factors. Some random variations are due to the limited number of MC sample events. In the case of variations in the TF, less than 1% of uncertainty is observed for JEC and JER uncertainty, while the rare backgrounds have 2.4% and 2.5%



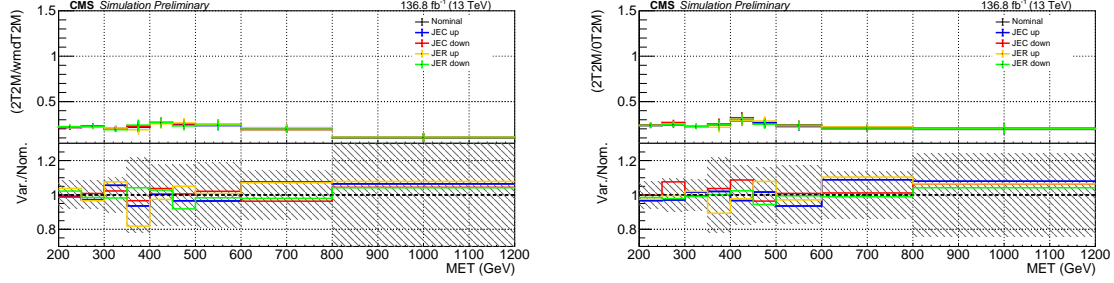


Figure 5.23: The systematic uncertainties on the transfer factors  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$ , used for the non-resonant plus 1-resonant background prediction in the signal region, due to the jet energy scale and resolution uncertainties. The uncertainties are evaluated separately for two associated control regions, wmdT2M (left) and 0T2M (right).

uncertainties, respectively. The variations in the TF due to these systematic uncertainties are shown in Fig. 5.23.

- QCD scale uncertainties: The QCD renormalization and factorization scale uncertainties are evaluated on the transfer factors and the rare backgrounds by varying these scales up by a factor of two, up and down. Figure 5.24 shows nine variations of QCD scale uncertainties in total for nominal, renormalization and factorization scales by varying each of them with either up, down, or nominal values of uncertainty, these uncertainties are within a few percent for the TF. The variation for rare background is 8% when integrated over  $p_{\text{T}}^{\text{miss}}$ .
- PDF uncertainties: The proton PDF used for MC sample production is provided by the NNPDF Collaboration [43, 44] with nominal distribution and 100 uncertainty variations for PDF choice. The uncertainty related to the choice of PDF is evaluated by reweighting the simulation using the variations in the corresponding NNPDF [43, 44] sets. These variations are propagated

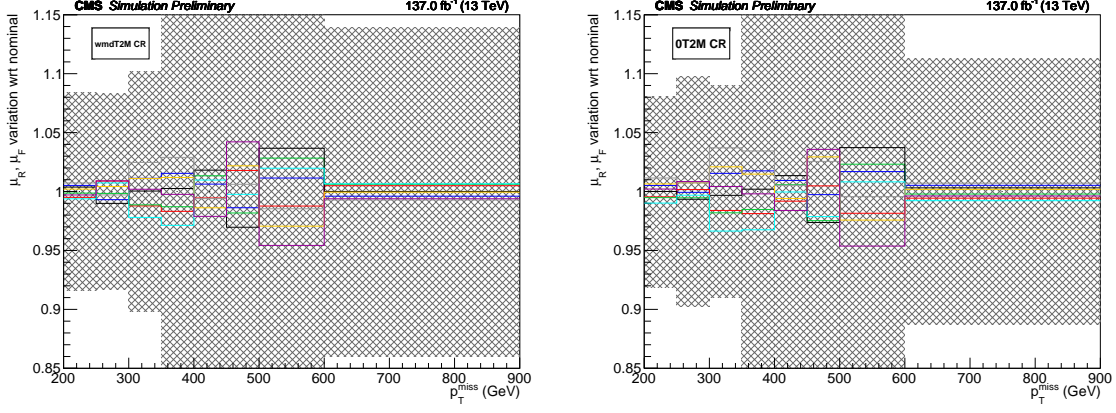


Figure 5.24: The systematic uncertainties on the transfer factors  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$ , used for the non-resonant plus 1-resonant background prediction in the signal region, due to the QCD renormalization and factorization scale uncertainties. The uncertainties are evaluated separately for two associated control regions, wmdT2M (left) and OT2M (right).

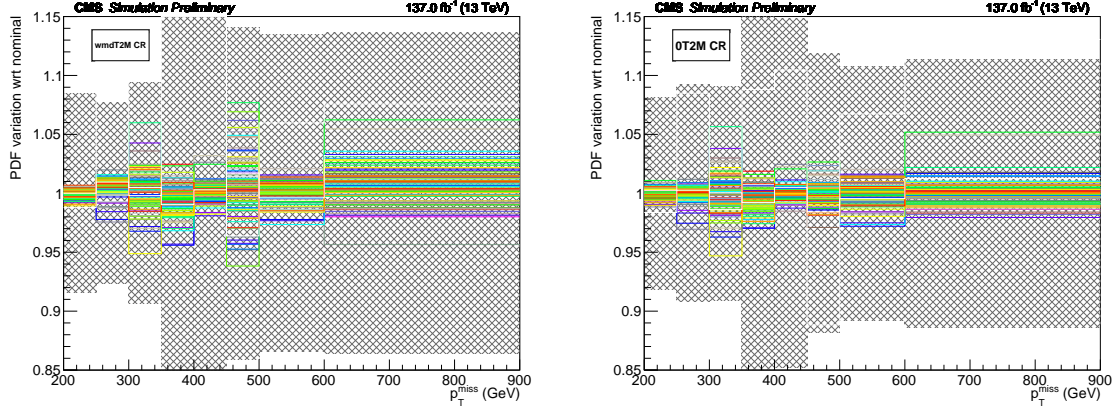


Figure 5.25: The systematic uncertainties on the transfer factors  $N_{\text{SR},0\&1\text{-res}}^{\text{MC}}/N_{\text{CR},0\&1\text{-res}}^{\text{MC}}$ , used for the non-resonant plus 1-resonant background prediction in the signal region, due to PDF uncertainties. The uncertainties are evaluated separately for two associated control regions, wmdT2M (left) and OT2M (right).

to the transfer factors as shown in Fig. 5.25. This uncertainty comes out to be around 1% on the TF and 2% on the rare backgrounds.

- b-tag uncertainty: Since the b tagger is used to veto b-tagged AK4 jets, the b-tag SFs are propagated to the signal region. The uncertainties due to the

tag SF and mistag SF are added in quadratures. This uncertainty is found to be less than 1% for all components of the b-veto background.

- Pileup uncertainty: The recommended cross-section for Run 2 is 69.2 mb and uncertainty in this measurement is 4.6%. The uncertainty is varied up/down from the nominal values and there is no dependence on  $p_T^{\text{miss}}$  while varying the PU uncertainty on the transfer factors and rare backgrounds. A flat uncertainty of 2% for the TF and 6.8% for rare background is assigned.
- Trigger uncertainty: The trigger efficiency has some uncertainties as discussed in Section 5.2. These uncertainties are propagated to rare backgrounds and result in 1–3% uncertainty, depending on  $p_T^{\text{miss}}$ .
- Non-closure uncertainty: Section 5.6.2 defines a correction to the non-closure of the background predictions. The full value of this correction is taken as the non-closure uncertainty.

The summary of all the uncertainties in the b-veto signal region are shown in Table 5.11.

### 5.5.2 Systematic uncertainties for the b-tag SRs

Most of the systematic uncertainties for the b-tag SRs are similar to those for the b-veto SR. In the case of b-tag SRs, the uncertainty is evaluated on the  $\mathcal{R}_{p/f}$  TF and also the  $0\ell/1\ell$  TF for each of the three signal regions. The considered uncertainties are:

- Deep-W and bb-vs-light SF uncertainties: The uncertainty in the Deep-W and bb-vs-light tagger efficiencies are propagated to the TFs and rare backgrounds. The Deep-W tagger uncertainty leads to 1% uncertainty on the

Table 5.11: Summary of systematic uncertainties (in %) on background estimation in the signal region for the b-veto analysis. A dash (—) indicates that the source of uncertainty is not applicable.

Source	0- and 1-resonant		Rare
	w/ wmdT2M CR	w/ 0T2M CR	
CR data size	5–50	6–71	—
MC sample size	8–30	8–25	14–24
Deep-W & WMD SF ( $W/Z \rightarrow q\bar{q}$ )	< 2	< 2	11–16
Deep-W & WMD SF (miss ID)	7–22	14–28	< 1
Top bkg norm.	0–2	0–3	—
Diboson norm.	0.7	2.7	—
Luminosity	—	—	1.6
JEC	1	1	2.4
JER	1	1	0.5
PDF	1	1	2
$\mu_R, \mu_F$ scales	0.4	1.2	8
Pileup	1	1	1
b tag (veto) SF	< 1	< 1	< 1
Trigger	—	—	2–3
Non-closure (Deep-W)	3–48	3–48	—
Non-closure (Deep-WMD)	—	1–27	—

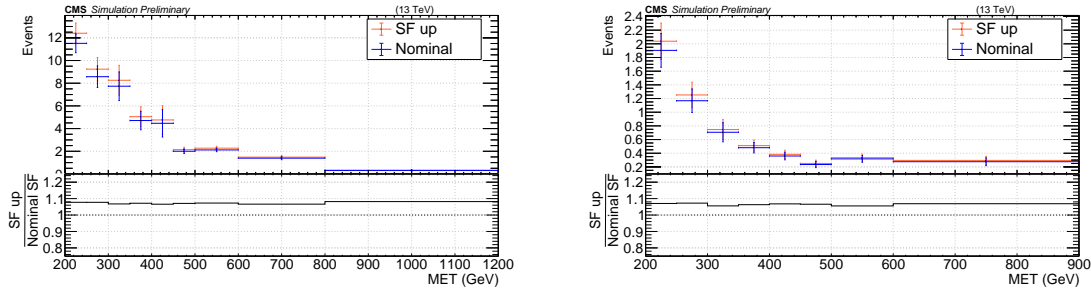


Figure 5.26: Comparison of rare background yields with nominal Deep-W tagger SFs and up variation of the SFs for WTag SR (left) and WHTag SR (right).

$0\ell/1\ell$  TF and a 7% uncertainty for rare backgrounds as shown in Fig. 5.26.

These uncertainties are correlated across rare SRs and anti-correlated with  $0\ell/1\ell$  TFs. The bb-vs-light SF uncertainties translate to an uncertainty of 2% for WHTag SR and 7% for WTag SR.

- Deep-W and bb-vs-light fake SF uncertainties: The uncertainty on the Deep-W fake SF and the bb-vs-light fake SF amount to 8.8% and 4.3% effect on  $\mathcal{R}_{p/f}$ , respectively. These uncertainties are derived from the dilepton control region events. These are some of the dominating uncertainties in the b-tag SR.
- Luminosity uncertainty: Luminosity uncertainty of 1.6% are applied to the predicted rare background yields.
- MC simulation statistical uncertainties: Due to the limited size of background samples, the uncertainties on the  $0\ell/1\ell$  TFs, pass-fail ratios, and rare background yields are evaluated and listed in the summary table 5.12.
- Lepton SF uncertainties: For the  $0\ell/1\ell$  TF, the  $1\ell$  control regions require the lepton identification and hence the lepton SF. The lepton SF uncertainties are correlated across different years and account for 1% uncertainty on the TFs.
- b-tag SF uncertainties: These uncertainties are propagated to the TFs and rare background yields using the method described in method 1a [73]. The uncertainties due to tag SF and SF are added in quadrature. These effects amount to a 0.5–1% uncertainty on  $0\ell/1\ell$  TF, and between 0.5–3% on pass-fail ratios and rare background yields.
- Other uncertainties: PDF uncertainties vary from 1% at low  $p_T^{\text{miss}}$  bins to 5% at high  $p_T^{\text{miss}}$  bins and are within the statistical uncertainties on simulations. A flat uncertainty of 0.5% on the normalization is applied on all backgrounds obtained by integrating yields across all the search bins. Effects of pileup mismodeling is estimated to be 1% on the normalization. Trigger efficiency

uncertainty is similar to that of b-tag SR. Luminosity uncertainty of 1.6% is applied to the rare backgrounds.

- JEC/JER uncertainties: The  $p_T$  of the jets are varied according to scale uncertainties and smeared to obtain JEC and JER uncertainties, respectively. These uncertainties are evaluated for  $\mathcal{R}_{p/f}$  and the  $0\ell/1\ell$  TFs as shown in Figs. 5.27–5.28. These uncertainties are within statistical uncertainties and have random variations. A flat 1% uncertainty is assigned to both JEC and JER uncertainty.
- QCD scale uncertainties: Similar to the b-veto SR, these uncertainties are evaluated on pass-fail ratios and  $0\ell/1\ell$  TFs as shown in Fig. 5.29. The effect on  $\mathcal{R}_{p/f}$  is much smaller than statistical uncertainties (1–4%), while the uncertainty on  $0/1\ell$  TF is between 2–5% which is applied as flat 0.5% uncertainty. Uncertainty on normalization of rare background accounts for 11%.

The summary of all the uncertainties in the b-tag signal regions are shown in Table 5.12.

### 5.5.3 Systematic uncertainties on signal models

The SMS signal models (TChiWW, TChiWZ, and TChiWH) are generated using the fast simulation framework as outlined in Section 4.1.2 for numerous mass points. Since the fast simulation may not have the best efficiency in terms of event reconstruction, several corrections are applied to the fastsim signal samples to account for observed differences. The uncertainties that affect the shapes of signal distributions are the QCD scale and ISR  $p_T$  modeling. The rest of the uncertainties

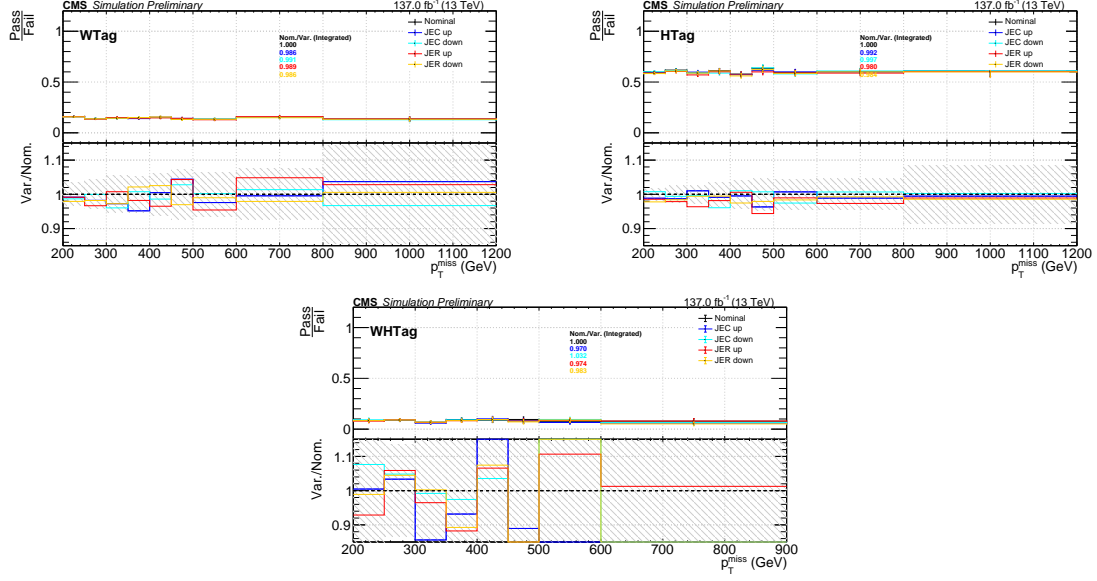


Figure 5.27: Effect of JEC and JER uncertainties on pass-fail ratios for WTag (top left), HTag (top right), and WHTag (bottom). The hashed bands in the bottom panels show the statistical uncertainty on nominal pass-fail ratios.

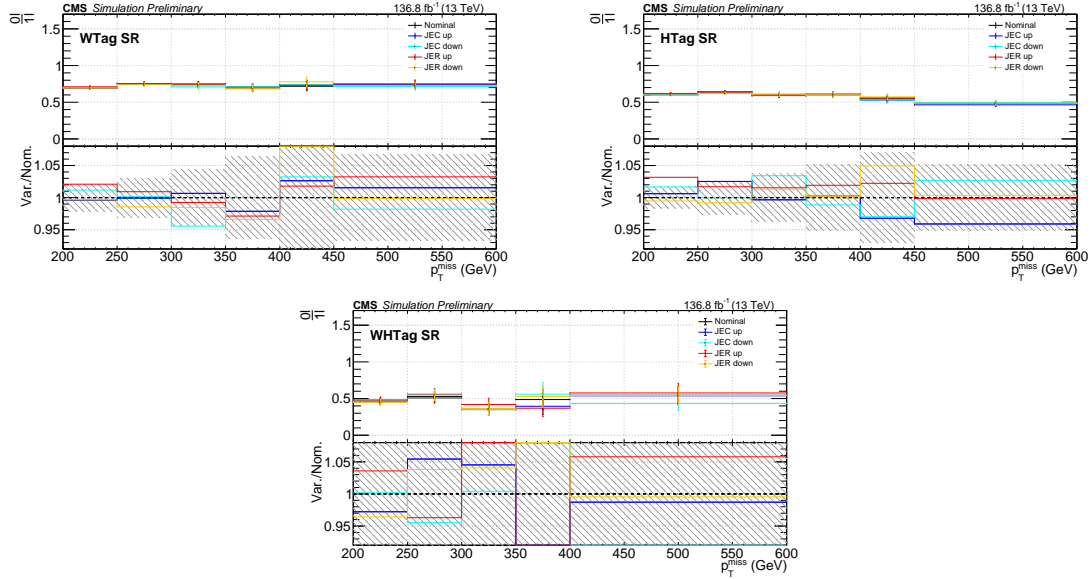


Figure 5.28: Effect of JEC and JER uncertainties on  $0/1\ell$  TFs for tag regions of WTag (top left), HTag (top right), and WHTag (bottom). The hashed bands in the bottom panels show the statistical uncertainty on nominal TFs.

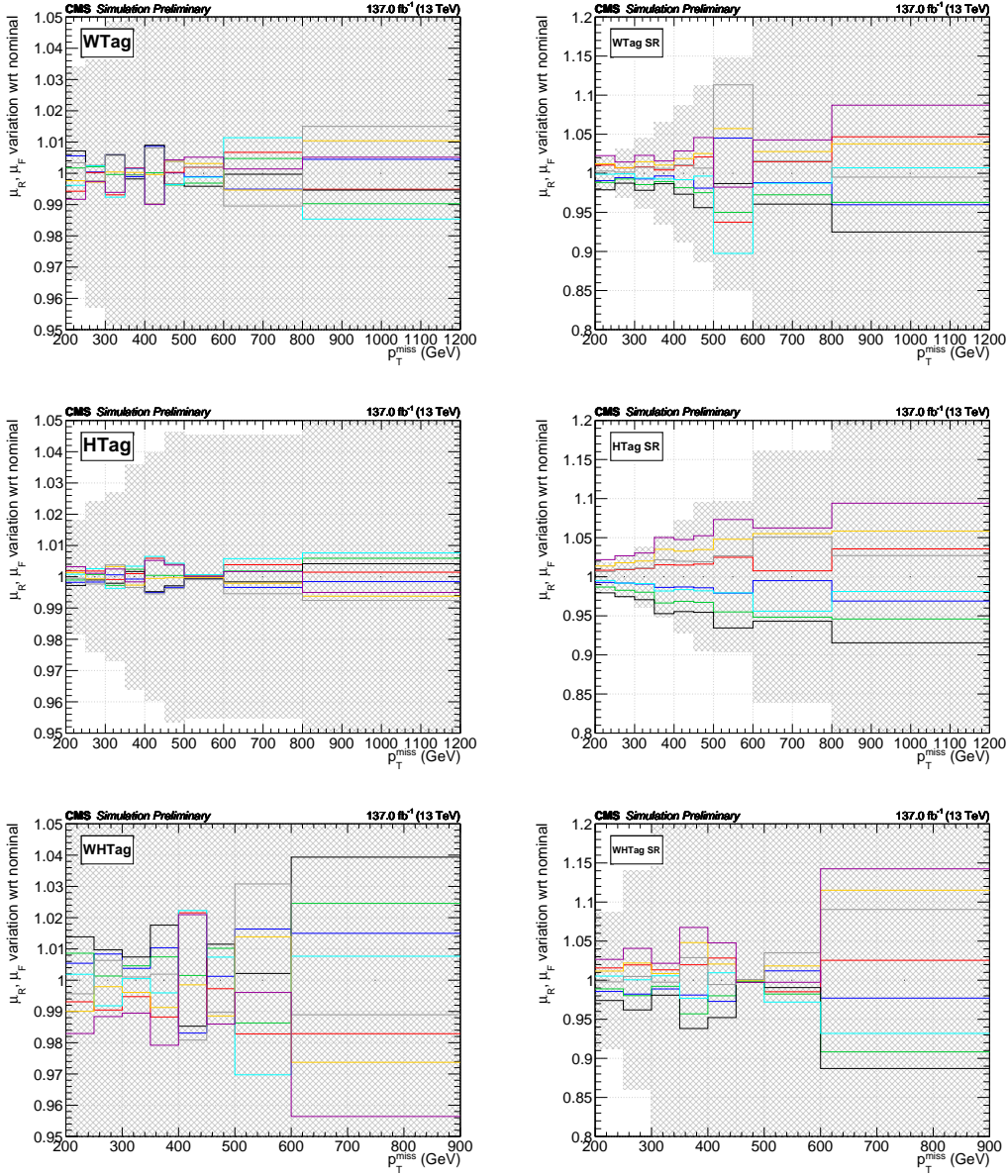


Figure 5.29: Effects of QCD renormalization and factorization scale variations on  $\mathcal{R}_{p/f}$  ratios (left) and  $0\ell/1\ell$  transfer factors (right) vs  $p_T^{\text{miss}}$  in the W SR (top), H SR (middle), and WH SR (bottom).



Table 5.12: Summary of systematic uncertainties (in %) in the b-tag SRs on background estimations. A dash (—) indicates that the source of uncertainty is not applicable.

Source	Top, $t\bar{t}H$	Nonres	Rare
Luminosity	—	—	1.6
Pileup	1.0	1.0	1.0
b-tag SF	1.0	< 3.0	< 3.0
Trigger	—	—	2–3
Deep-W SF	1	—	7.0
Deep-W fake SF	—	8.8	—
bb-vs-light SF	—	—	2–7
bb-vs-light fake SF	—	4.3	—
Lepton SF	1.0	—	—
PDF	0.5	0.5	0.5
QCD scale	2–10	0.5	11.0
JEC	1.0	1.0	1.0
JER	1.0	1.0	1.0
MC sample size	2–28	3–40	4–27
CR data size	3–100	2–35	—

only affect overall normalization. Most of the uncertainties do not depend on the signal mass points except a few uncertainties. The uncertainties applied on the signal yields are listed below.

- QCD scale uncertainty: For various signal models and  $(m_{\text{NLSP}}, m_{\text{LSP}})$  mass combinations, the impact of varying renormalization scales ( $\mu_R$ ) and factorization scales ( $\mu_F$ ) on the shape of the signal is investigated. When these scales are changed, signal samples with lower NLSP masses typically exhibit larger shape differences. Fig. 5.30 displays the variations in signal shape relative to the nominal for various models and mass points. Depending on the SR category,  $p_{\text{T}}^{\text{miss}}$  bin, and mass point under consideration, the overall uncertainty varies from 0.25 to 5%.

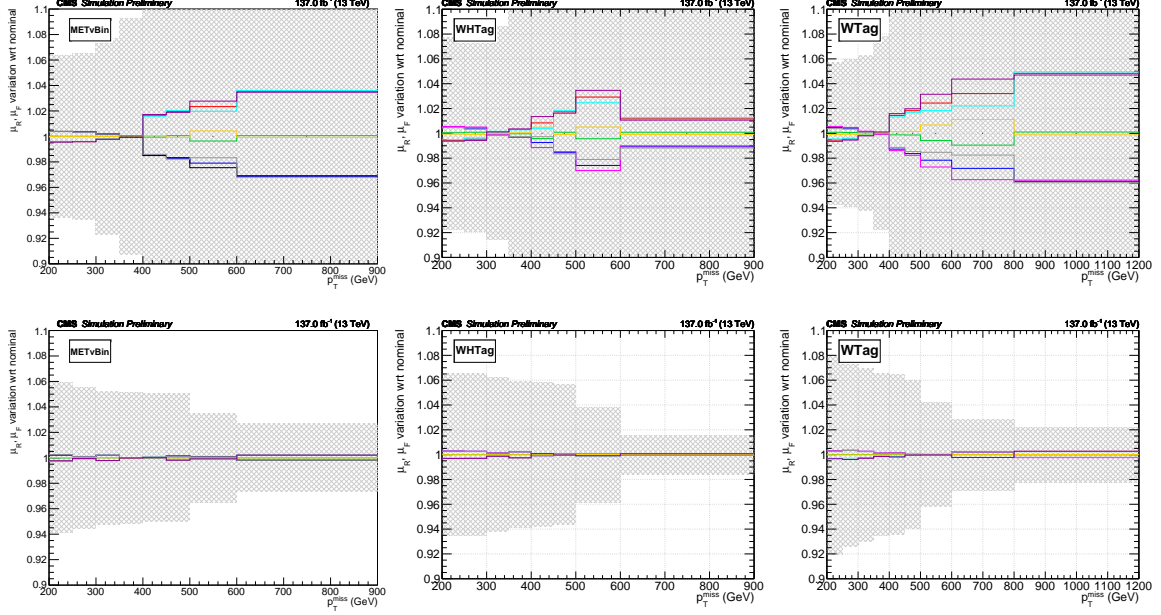


Figure 5.30: Signal shape uncertainty for the b-veto (left), WHTag (middle), and WTag (right) SRs at QCD scale. The top row plots are for signal samples corresponding to low NLSP masses (400 GeV), while the bottom row represents the uncertainty for high NLSP masses (800 GeV for b-veto and 1200 GeV for b-tag SRs). The b-tag SRs use the TChiWH model, while b-veto regions use the TChiWZ model. The LSP mass in each of these plots is 1 GeV. The gray hashed band represents relative statistical uncertainties on nominal distributions.

- ISR  $p_T$  modeling: To assess the degree of uncertainty associated with ISR  $p_T$  modeling in simulation, the standard SUSY recommendation by CMS is used. Events are re-weighted while maintaining the same cross section based on the  $p_T$  of the system of pair produced electroweakinos. This re-weighting has little impact in the non-compressed region where  $m_{\text{NLSP}} \gg m_{\text{LSP}}$  and can have  $\approx 22\%$  impact in the compressed region for signal events in high  $p_T^{\text{miss}}$  bins. Uncertainty is attributed to the full size of the correction, which is the difference in relation to unweighted yields. The impact of ISR  $p_T$  re-weighting for a mass point in the compressed and non-compressed regions is depicted

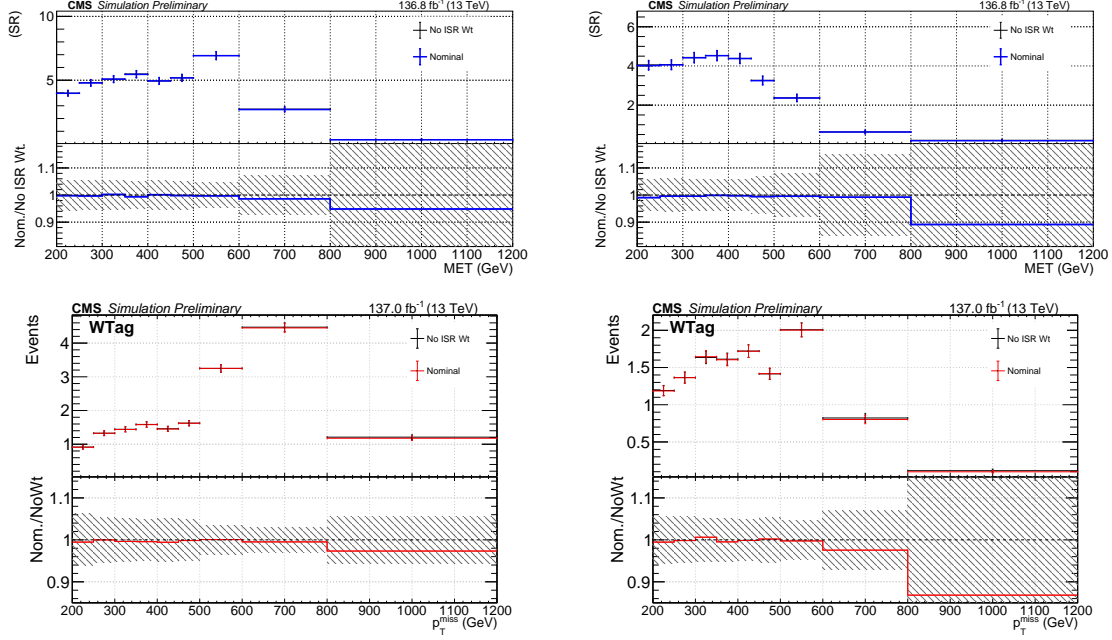


Figure 5.31: ISR weighted (nominal) and unweighted distributions are compared. Top left: b-veto SR for TChiWZ(600,1), top right: b-veto SR for TChiWZ(600,200), bottom left: WTag SR for TChiWH(800,1), bottom right: WTag SR for TChiWH(800,400). The bottom ratio panel's hashed bands represent the degree of statistical uncertainty associated with unweighted yields.

in Fig. 5.31. The WHTag and HTag SR categories exhibit similar behavior to the WTag SR.

- b-tagging: The b-tag scale factors are derived by the bTag POG using *CP5 Tune*, while the signal samples are generated with *CP2 Tune*. Scale factors are applied to account for these tune differences in MC sample generation and also with data. To assess the uncertainties brought on by tagging and mistagging, the fast-to-full-simulation and data-to-full-simulation SFs are both changed simultaneously. This uncertainty is approximately 1% for b-veto, approximately 2% for WTag, approximately 2.5% for WHTag, and approximately 3% for HTag SRs. Tagging SFs makes the larger contribution to the overall uncertainty compared to the mistagging SFs.

- JEC, JER, and pileup: Similar procedure to the one used for rare backgrounds is used to evaluate these uncertainties. For b-veto SRs, these correspond to 1% each for JEC & JER and 2% for pileup. For b-tag SRs, each of these have  $\approx 1\%$  uncertainty.
- Jet mass scale and resolution: These uncertainties are negligible for the b-tag SRs while it has  $\approx 1.2\%$  uncertainty for the JMR correction in the b-veto SR. Since the effect of JMR is small, only the uncertainty of 1.2% is applied to the b-veto SR.
- Fastsim  $p_T^{\text{miss}}$  modeling: Any mismodeling of  $p_T^{\text{miss}}$  in fastsim samples is accounted by comparing it against gen- $p_T^{\text{miss}}$ . Comparing the gen- $p_T^{\text{miss}}$  and nominal- $p_T^{\text{miss}}$  in Fig. 5.32, half the difference between them is taken as the uncertainty and is correlated across all the bins. This uncertainty varies depending on the SR category, signal model, and the  $p_T^{\text{miss}}$  bin and can reach up to  $\approx 10\%$ .
- Fastsim boson tagging: Differences in tagger performance in fullsim and fastsim due to the tune setting is addressed in this uncertainty. These corrections are broken down into factors associated with the efficiency of cutting on the signal soft-drop mass window and the efficiency of cutting on the applicable POG-derived tagger working points. For the soft-drop mass corrections, the scale factors are roughly 0.92–1.10 with an uncertainty around 0.05. The Deep AK8 tagger scale factors are roughly 0.99–1.05 with an uncertainty around 0.01.
- Fastsim jet response: In the b-tag SRs, the b-tagged jet response evaluated for TChiWZ signal model is 2–3% higher than unity in 2017 and 2018 MC

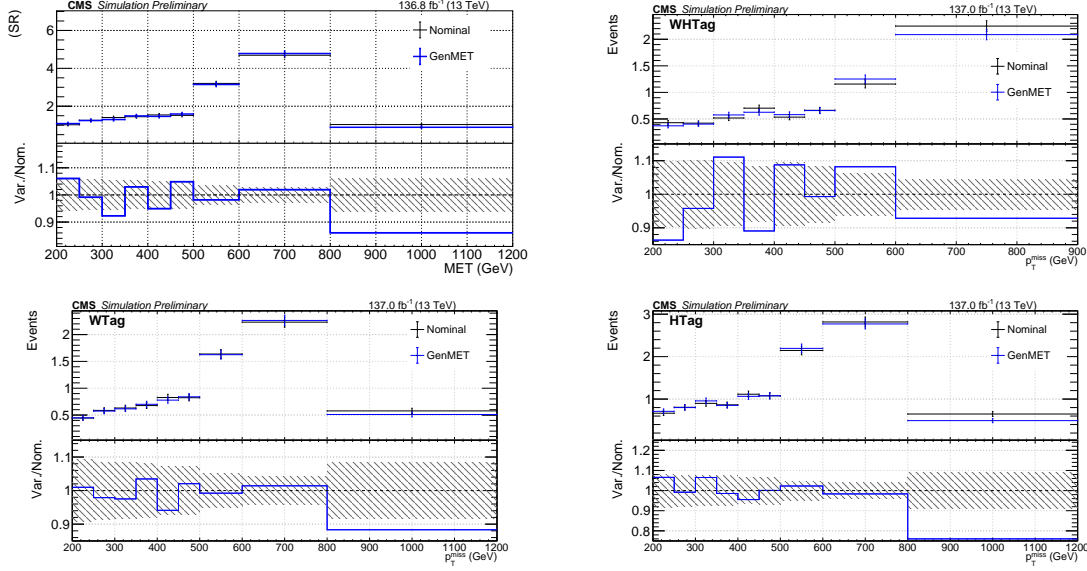


Figure 5.32: Comparison of the nominal  $p_T^{\text{miss}}$ -based and gen- $p_T^{\text{miss}}$ -based distributions for the TChiWZ(800,1) signal sample for the b-veto (top left), WHTag (top right), WTag (bottom left), and HTag (bottom right) SRs. The systematic uncertainty is calculated as half the difference between these two. The bottom panel's hashed bands display the relative statistical uncertainty in nominal terms.

samples. Correction in response is done by scaling the momenta of b jets down by 3% and propagate the effect to  $p_T^{\text{miss}}$ ,  $H_T$ , and number of AK4 jets. The difference between response corrected and uncorrected yields, which amounts to 2% uncertainty, is applied for all models and mass points. Figure 5.33 shows the response of b jets in fullsim and fastsim samples.

- Other uncertainties: Uncertainty in luminosity measurement is 1.6% across all the  $p_T^{\text{miss}}$  bins. Trigger efficiency uncertainties, similar to backgrounds have been included. Data to fullsim AK8 tagger SFs vary from a few percent to  $\approx 15\%$  depending on the tagger and the SR. The PDF uncertainty is included in the signal cross section uncertainty. The jet mass resolution in MC simulation is better than in data, therefore the resolution of the jets is worsened in MC simulation by applying smearing so that the resolution

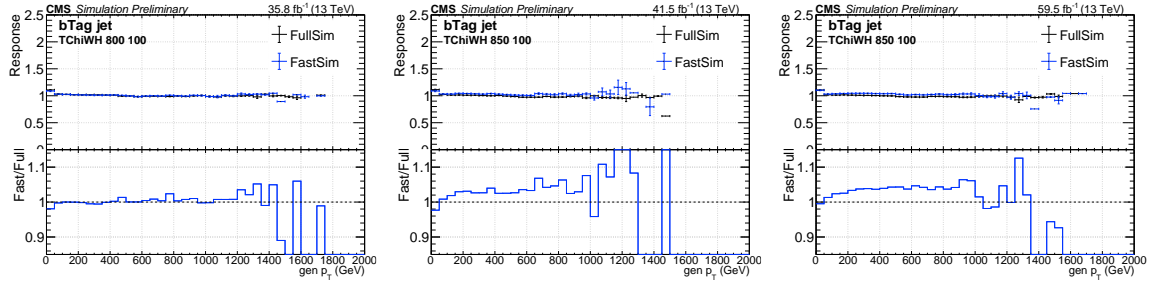


Figure 5.33: Response of b jets in fullsim and fastsim for 2016 (left), 2017 (middle), and 2018 (right) MC simulation samples as a function of closest gen-jet  $p_T$ . All fullsim samples and 2016 fastsim samples show response close to unity. Response in 2017 and 2018 fastsim is 2–3% higher than unity.

matches with data and the uncertainty on the JMR scale factor is applied as systematic uncertainty.

Table 5.13 lists all the systematic uncertainties applicable on the signal yields.

Table 5.13: Summary of systematic uncertainties (in %) on signal yields. A dash (—) indicates that the source of uncertainty is not applicable.

Source	b-veto	b-tag
Pileup	1.0	1.0
b-tag SF	1.0	2–3
Deep-W SF	15	9.0
Deep-WMD SF	2.0	—
bb-vs-light SF	—	4.0
JMR	1.2	—
JEC		1.0
JER		1.0
Luminosity		1.6
Trigger		2–3
MC sample size		2–5
QCD scale (shape)		0.25–5
ISR $p_T$ (shape)		1–22
Fastsim related systematic uncertainties		
bb-vs-light SF	—	4.0
AK8 mass ( $b\bar{b}$ )	—	4.0
b jet $p_T$ response	—	2.0
AK8 near b jet $p_T$ response	—	3.0
Deep-WMD SF	< 1	—
Deep-W SF		< 1
AK8 mass ( $q\bar{q}$ )		< 1
$p_T^{\text{miss}}$ modeling		1–10

## 5.6 Validation

### 5.6.1 Statistical modeling of backgrounds in the $b$ -veto SR

In the  $b$ -veto analysis, there is one SR and two CRs (0T2M and wmdT2M) binned in  $p_T^{\text{miss}}$  as discussed in Section 5.4.1. Four contributions to each of SR and CR are considered: signal, and 0-res, 1-res, and rare backgrounds. The 0-res and 1-res backgrounds are treated together by using the same rate parameters for both. The rate parameters are placed in order to maintain the relative rates fixed in the SR and CRs which is the same as multiplying the CR data yields by data yields.

In each of the  $p_T^{\text{miss}}$  bins, the nominal value is described by a collection of three Poisson distributions. Various unconstrained parameters are used to constrain the SM backgrounds in the SR based on the observations in the CRs. Few correction factors as described in Section 5.6.2 are applied through rate parameters on the 0-res and 1-res backgrounds. All the nuisances are modeled with log-normal prior distributions. Figure 5.34 shows the prefit and postfit nuisance parameters from the fit.

### 5.6.2 Validation regions for the $b$ -veto SR

Validation regions are designed for  $b$ -veto analysis to check compatibility of the tagger mistag scale factors that were derived and also the background estimation procedure. All the validation regions and the corresponding CRs are required to pass the baseline selections except the AK8 jet requirements as listed for each of the region. The validation regions are identified by “xTyMzJ”, where “T” stands for tagger, “M” stands for mass requirement and “J” stands for number of AK8 jets. The rest of the terminology is explained below. The symbol “z” refers to the multiplicity of AK8 jets satisfying  $p_T > 200$  GeV,  $|\eta| < 2.0$ . Out of “z” AK8 jets, if they also have its softdrop



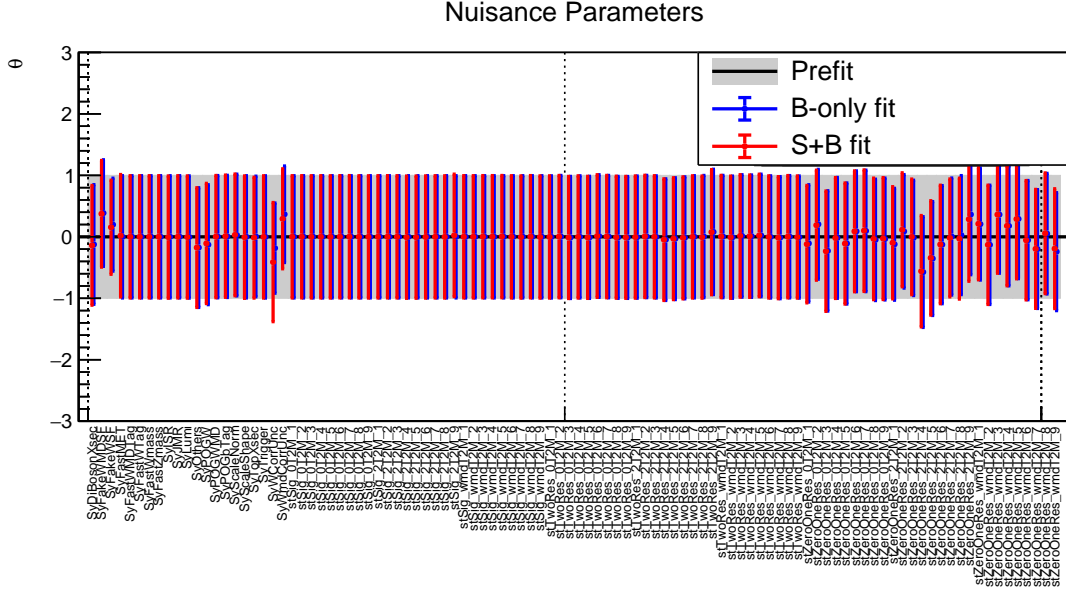


Figure 5.34: A comparison of prefit-postfit nuisance parameters from the fit to the b-veto signal region and control regions. The TChiWZ(800,100) signal is used in this fit. The yields are the actual data yields.

mass within  $65 < m_J < 105$  GeV (WZ mass requirement), then they are counted for "y". The "x" refers to the tagger type of only one AK8 jet that is tagged by either Deep-W or Deep-WMD tagger.

The wT2M validation region requires almost the same selections as b-veto SR except that WMD-tagging requirement is inverted. The wT2M VR requires at least two AK8 jets with the WZ mass requirement but only one jet out of them is tagged by the Deep-W tagger and the other jets are not tagged. This is the final validation region that is close to the SR. The list of validation regions and their corresponding control regions (inverted tag requirement from the VR) are described in Table 5.14. The overview of the validation procedure is shown by a schematic in Fig. 5.35.

The first step is to test the Deep-W and Deep-WMD mistag scale factors which were derived in Section 5.4.1.1. The wT1M1J and wmdT1M1J regions are

Table 5.14: Summary of mass and tagging requirements used for the b-veto validation regions (VRs). All of them veto on AK4 b-tagged jets. “xTyMzJ” below stands for “x” tagged jet ( $x = \text{Deep-W or Deep-WMD tagger}$ ), “y” stands for number of AK8 jets with WZ mass out of total number of AK8 jets (“z”). The WZ mass requirement is  $65 < m_{J} < 105 \text{ GeV}$ .

“x”T1M1J regions (1 AK8 jet passing WZ mass requirement)	
VR (wT1M1J)	1 jet passing Deep-W tag
VR (wmdT1M1J)	1 jet passing Deep-WMD tag
CR (for wT1M1J)	no jet passing Deep-W tag
CR (for wmdT1M1J)	no jet passing Deep-WMD tag
“x”T1M2J regions ( $\geq 2$ AK8 jets, one of them passing WZ mass requirement)	
VR (wT1M2J)	1 jet passing Deep-W tag
VR (wmdT1M2J)	1 jet passing Deep-WMD tag
CR (for wT1M2J)	no jet passing Deep-W tag
CR (for wmdT1M2J)	no jet passing Deep-WMD tag

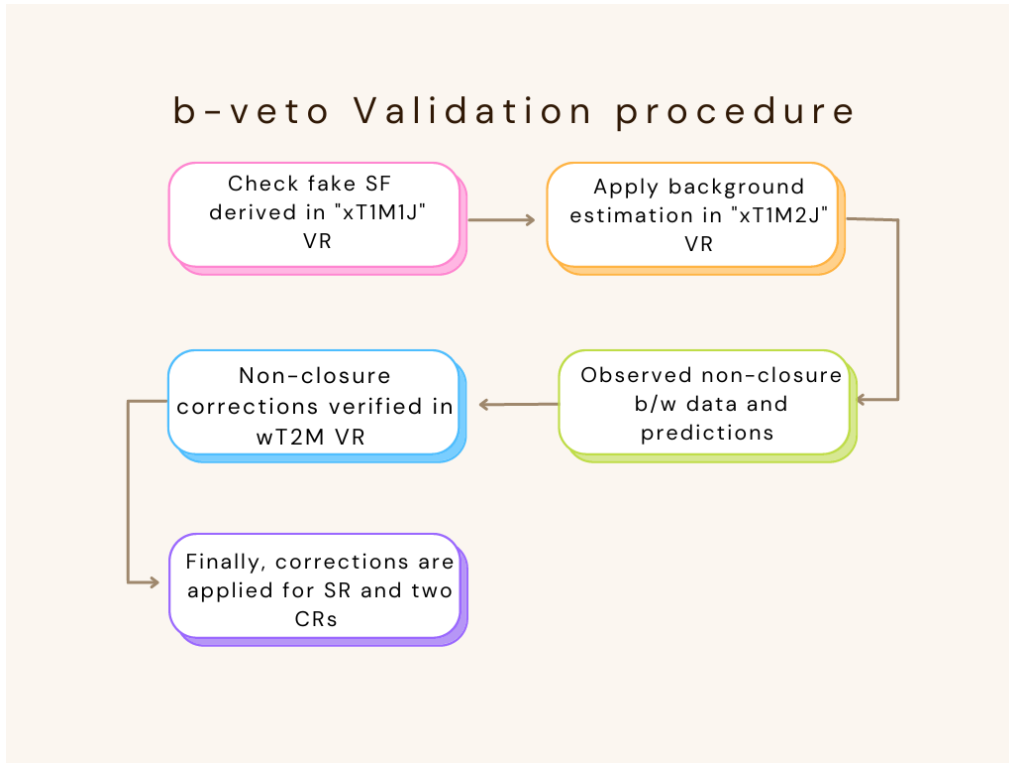


Figure 5.35: Schematic showing the validation steps involved in b-veto analysis.

well-suited to validate these scale factors since events in this region contains one AK8 jet tagged by either Deep-W or Deep-WMD tagger. The CRs for these “1J” validation

regions are defined by inverting the tagger requirement. The background prediction is performed, and the data are found to agree with predictions very well in this region as shown in Fig. 5.36.

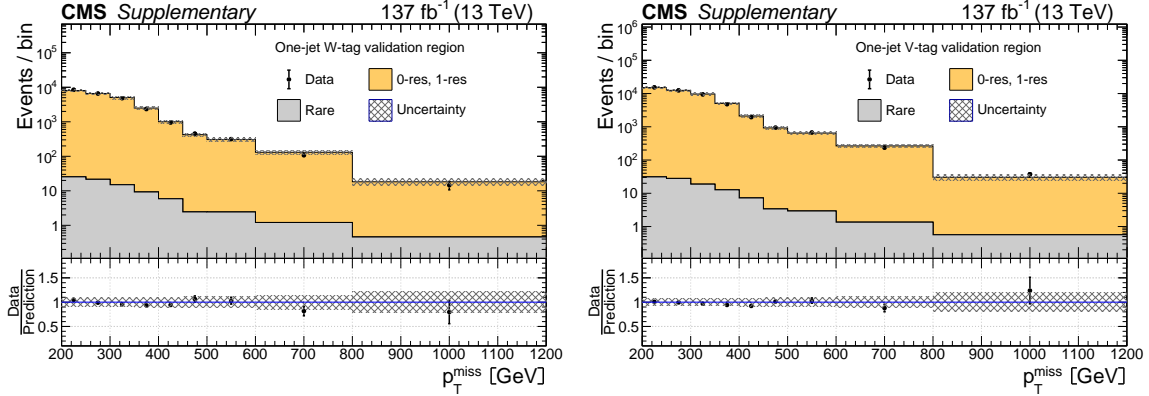


Figure 5.36: Data yields in the 1-jet W-tag (left) and V-tag (right) validation regions compared to the predicted SM backgrounds as functions of  $p_T^{\text{miss}}$ . In these validation regions, events are required to pass the baseline selection except for the second AK8 jet requirement, and to have only one AK8 jet, which is W or V tagged. The predicted SM 0-res and 1-res backgrounds in each validation region are based on an extrapolation from control regions defined by the inversion of the DNN requirement for the W or V tag. The rare background is taken directly from simulation. The hatched gray bands correspond to the total uncertainty in the prediction.

Moving one step at a time, now the events in this validation regions, wT1M2J and wmdT1M2J, have an extra AK8 jet that does not satisfy the WZ mass requirement. The CRs are defined for these validation regions as shown in Table 5.6. After the background predictions are done in these VRs, the data agree well with the predictions at low  $p_T^{\text{miss}}$ , while at higher  $p_T^{\text{miss}}$ , some differences are seen as shown in Fig. 5.37. One source of these differences could be that the quark-gluon mixture differences in the “1M1J” and “1M2J” regions. In order to account for these differences, a “non-closure” correction is obtained from the wT1M2J and wmdT1M2J

regions for Deep-W and Deep-WMD taggers, respectively, as shown in Fig. 5.37. The non-closure correction can be expressed as:

$$\text{non-closure correction} = \frac{\text{data} - \text{prediction (rare)}}{\text{prediction (01-res)}}. \quad (5.8)$$

This correction is a systematic trend versus  $p_T^{\text{miss}}$  and is applied to all the “2M” CRs, VRs, and signal regions. The 100% of the correction is assigned as the systematic uncertainty for this correction. The correction factors are obtained and a linear fit is performed as shown in Fig. 5.38.

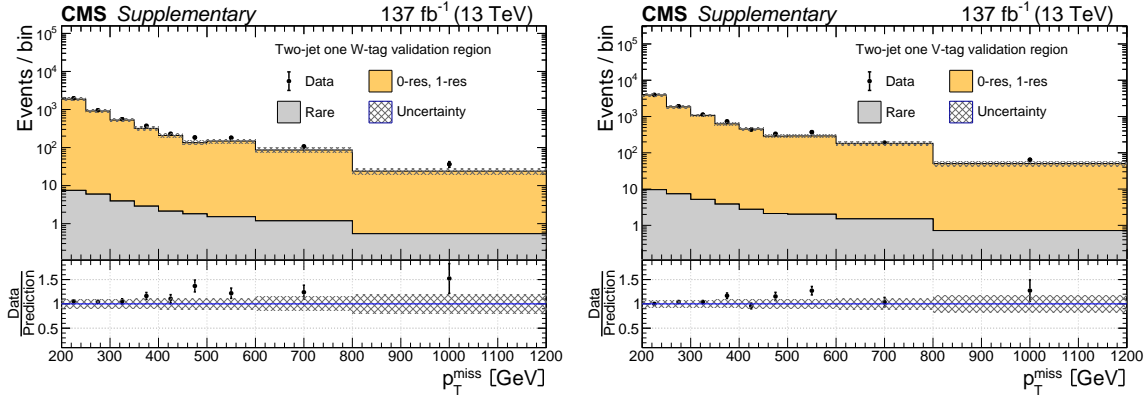


Figure 5.37: Data yields in the 2-jet, 1-tag validation regions, the W-tag region (left) and the V-tag region (right), compared to the predicted SM backgrounds as functions of  $p_T^{\text{miss}}$ . In these validation regions, events are required to pass the baseline selection including the requirement of having at least two AK8 jets, but only one AK8 jet may pass the WZ mass requirement of  $65 < m_J < 105$  GeV, and only one AK8 jet is permitted to be W or V tagged. The predicted SM 0-res and 1-res backgrounds in each validation region are based on an extrapolation from control regions defined by the inversion of the DNN requirement for the W or V tag. The rare background is taken directly from simulation. The deviation from unity of the ratio of the data yields to the SM predictions in each of these validation regions is used to extract the non-closure corrections for the W and V taggers as discussed in the main text, and the full size of the correction is assigned as its uncertainty. The hatched gray bands correspond to the total uncertainty in the prediction.

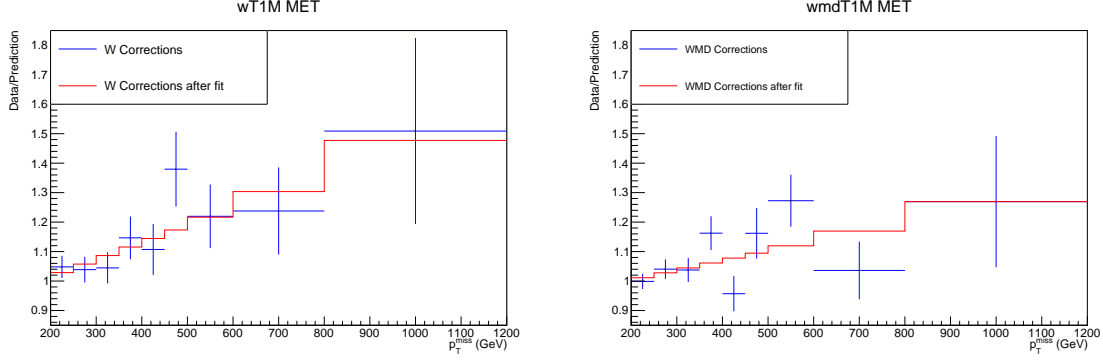


Figure 5.38: The Deep-W and Deep-WMD tagging rate correction factors extracted from data-prediction differences observed in wT1M2J and wmdT1M2J validation regions (see Fig. 5.37) and a linear fit.

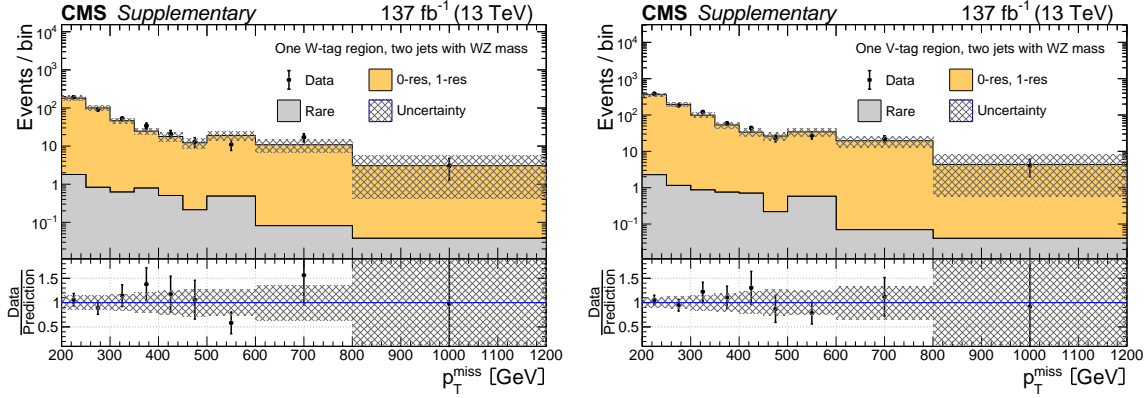


Figure 5.39: Data yields in the one W-tag region (left) and one V-tag region (right) compared to the predicted SM backgrounds as functions of  $p_T^{\text{miss}}$ . In these regions, events are required to pass the baseline selection including the requirement of having at least two AK8 jets, at least two AK8 jets must pass the WZ mass requirement of  $65 < m_J < 105$  GeV, but only one AK8 jet is permitted to be W or V tagged. The predicted SM 0-res and 1-res backgrounds in each validation region are based on an extrapolation from the 0-tag control region, in which events have at least two AK8 jets passing the WZ mass requirement of  $65 < m_J < 105$  GeV but none of them pass the DNN requirements for the W or V taggers. The rare background is taken directly from simulation. The hatched gray bands correspond to the total uncertainty in the prediction.

The final step before moving to the SR is to test the non-closure corrections on the wT2M VR and the wmdT2M CR (defined in Section 5.3.3). The CR considered for both these 2M validation regions is the 0T2M CR. After the background predictions

are performed for the 0-res and 1-res backgrounds, the data are found to be in fair agreement with the predictions as shown in Fig. 5.39. These validation regions are close to the SR, in the sense that there are two or more AK8 jets passing the WZ mass requirement as also in the SR, but only one of them pass either one of the tagger requirements.

### 5.6.3 Statistical modeling of backgrounds in the b-tag SRs

Similar to the modeling of backgrounds in the b-veto SR, the likelihood is a binned likelihood with each  $p_T^{\text{miss}}$  bin of each SR and the  $0\ell$ -antitag CR explicitly modeled. Each of the SR and CR bins have contributions from non-resonant, top, and rare backgrounds. The observed yields in the SR and  $0\ell$ -antitag CR are modeled by two Poisson distributions. All the nuisance parameters from the systematic sources are modeled.

The  $0\ell/1\ell$  transfer factors, used in deriving the top backgrounds (both for the tag and anti-tag CRs), is modeled by a gamma function that can constrain the expected yields of resonant backgrounds. The non-resonant backgrounds are modeled using rate parameters that correlate expected yields across different regions. The transfer factors are set by simulation which control the relative rates between the SR and CR, but the yields themselves can float with rate parameters. To mitigate the tagging differences between data and simulation, pass-fail ratio SF are used to adjust the expected yields in the  $0\ell$ -antitag CR so that the transfer factors are the same across several  $p_T^{\text{miss}}$  bins. All the systematic uncertainties are modeled by log-normal distributions and are uncorrelated across various b-tag regions and  $p_T^{\text{miss}}$  bins except

the pass-fail ratio SF nuisance parameter which is correlated across all  $p_T^{\text{miss}}$  bins. The pulls for each nuisance parameter is shown in Fig. 5.40.

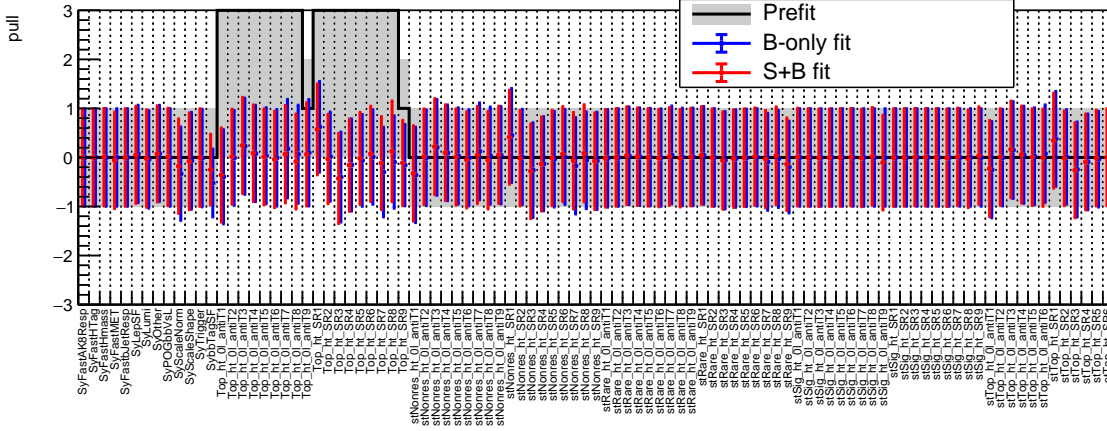


Figure 5.40: Prefit and postfit comparison of nuisance parameters and their uncertainties for fits to the H tag validation region. Uncertainties and any shifts in the central value of a nuisance parameter are plot relative to the corresponding nuisance parameters prefit uncertainty.

#### 5.6.4 Validation regions for the b-tag SRs

The background estimation procedure for the b-tag analysis is validated in so-called pseudo signal regions. The validation regions are orthogonal to the SR and  $0\ell$ -antitag CR by requiring exactly one AK8 jet. So two validation regions, 1J W tag and 1J H tag validation regions, are defined by the following selections and are summarized in Table 5.15.

The 1J W-tag pseudo-SR derives the predictions based on the CR where there is exactly one AK8 jet with WZ mass requirements but fails the Deep-W discriminator. Based on Eqs. (5.6) and (5.7), the predictions are made directly from the simulations and the distributions are shown in Fig. 5.41 (left). The right plot shows

Table 5.15: Selections in validation pseudo-SR and CRs for b-tagged events after the baseline selections have been applied. Each of the pseudo SR and VR have single-lepton regions that are identical except that exactly one charged lepton is required. A dash (—) indicates that no requirement is imposed.

	W boson candidate		Higgs boson candidate	
	W tagged	not W tagged	$b\bar{b}$ tagged	not $b\bar{b}$ tagged
1J W tag pseudo SR	1	0	—	—
1J W tag VR	0	1	—	—
1J H tag pseudo SR	—	—	1	0
1J H tag VR	—	—	0	1

the postfit predictions using only the CRs after the likelihood parameterization is minimized with respect to the background-only hypothesis. Various nuisance parameters are compared before and after the fit as shown in Fig. 5.42, suggesting some mild pulls, where pull is defined as:

$$\text{pull} = \frac{\text{data} - \text{pred}}{\sqrt{\text{pred} + \Delta_{\text{pred}}^2}} \quad (5.9)$$

where  $\Delta_{\text{pred}}$  is the total uncertainty in the prediction.

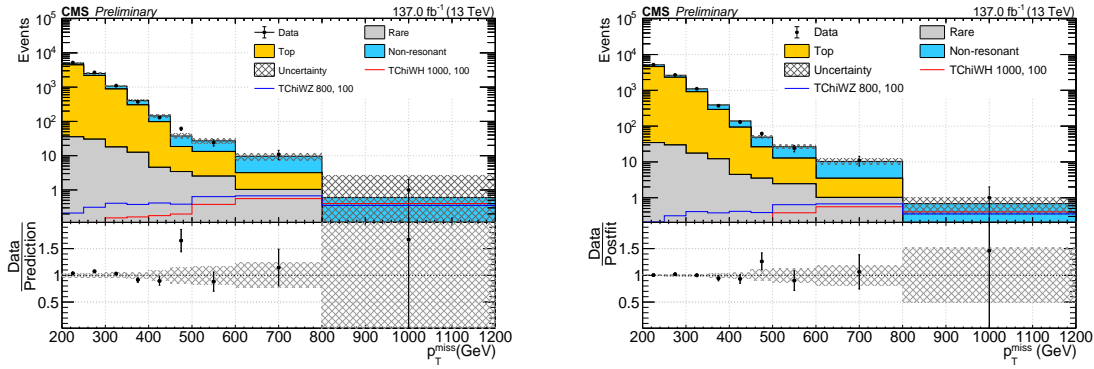


Figure 5.41: Predicted yields in the 1J W tag validation region. The left plot shows the prediction based on Eqs. (5.6) and (5.7). The right plot shows the prediction based on a fit to the pseudo signal region and the control regions.



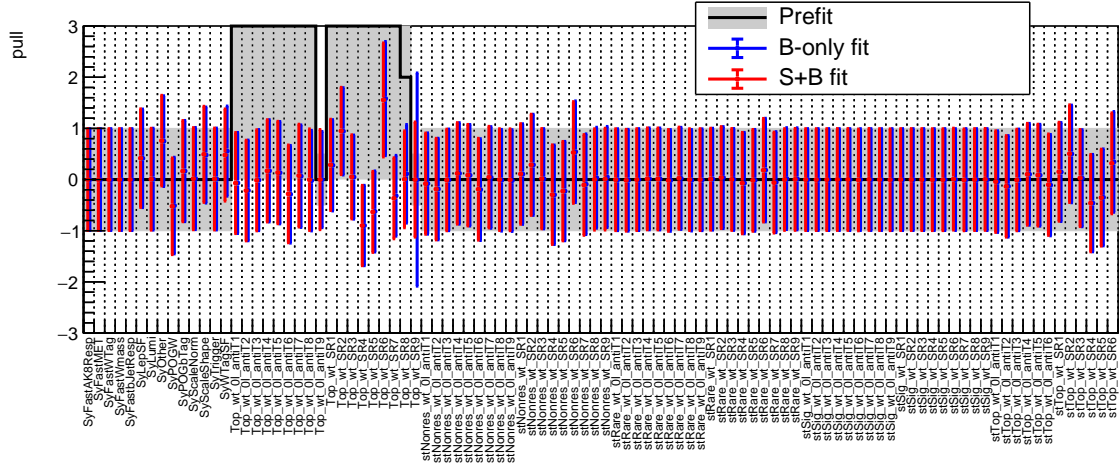


Figure 5.42: Prefit and postfit comparison of nuisance parameters and their uncertainties for fits to the W tag validation region. Uncertainties and any shifts in the central value of a nuisance parameter are plot relative to the corresponding nuisance parameters prefit uncertainty.

The 1J H-tag pseudo-SR follows the equivalent testing as the 1J W-tag pseudo signal region except that the AK8 jet is considered for ZH mass requirements and bb-vs-light tagger. Similar comparisons are made between equation based predictions and postfit predictions as shown in Fig. 5.43.

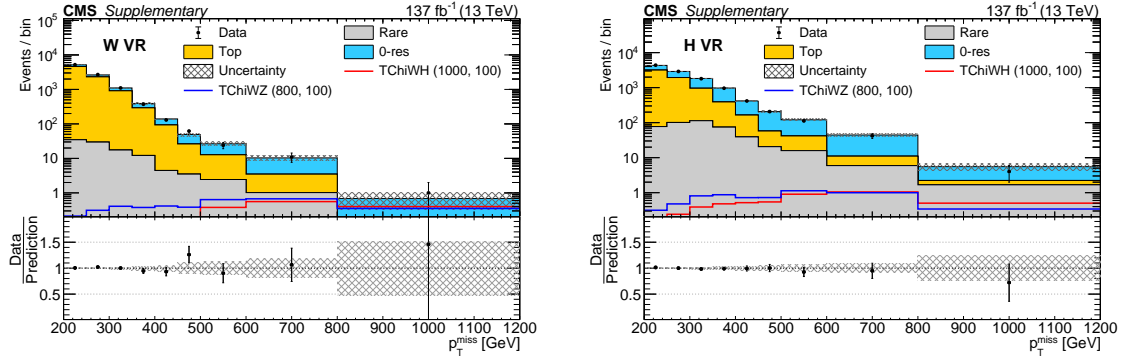


Figure 5.43: Data yields in the W (left) and H (right) b-tag validation regions compared to the predicted SM backgrounds as functions of  $p_T^{\text{miss}}$ . In these regions, events are required to pass the baseline selection except for the second AK8 jet requirement. Events are required to contain exactly one AK8 jet and at least one b-tagged AK4 jet. The hatched gray bands correspond to the total uncertainty in the prediction. The open histograms show the expectations from selected signal models, which are denoted in the legend by the name of the model followed by the assumed masses of the NLSP and LSP.

## 5.7 Results and interpretations

### 5.7.1 Results

The background estimation procedure described in the previous section is applied on b-veto and b-tag search regions. Utilizing a statistical model of our SM background predictions, simultaneous fits to our SRs and CRs are performed to further constrain our predictions and their uncertainties.

**5.7.1.1 Results in b-veto SR** The background predictions in the signal region are obtained after simultaneous fits to two control regions only, and the predictions obtained are called post-fit predictions for 0- and 1-res backgrounds. The observed data yields and predictions are consistent with each other as shown in Fig. 5.44. Table 5.16 shows the number of events in each  $p_T^{\text{miss}}$  bin for the observed yields, signal and different backgrounds. No statistically significant excess of events is observed in the data with respect to the SM background predictions.

Table 5.16: Observations and SM background predictions derived from the fit to the data in the CRs in the b-veto SR. The column labeled “Signal” provides the expected signal yields from the TChiWZ model with an NLSP mass of 600 GeV and an LSP mass of 100 GeV.

$p_T^{\text{miss}}$ (GeV)	0- and 1-res bkg.	Rare	Total bkg.	Obs.	Signal
200 – 250	$86.7 \pm 9.2$	$1.5 \pm 0.3$	$88.1 \pm 9.2$	82	$4.4 \pm 0.2$
250 – 300	$46.8 \pm 6.7$	$1.4 \pm 0.3$	$48.2 \pm 6.7$	48	$5.3 \pm 0.3$
300 – 350	$24.2 \pm 4.3$	$1.1 \pm 0.3$	$25.4 \pm 4.4$	24	$5.6 \pm 0.3$
350 – 400	$14.8 \pm 3.7$	$0.7 \pm 0.2$	$15.5 \pm 3.7$	9	$6.0 \pm 0.3$
400 – 450	$11.7 \pm 3.1$	$0.6 \pm 0.2$	$12.3 \pm 3.1$	8	$5.5 \pm 0.3$
450 – 500	$7.4 \pm 2.1$	$0.4 \pm 0.1$	$7.9 \pm 2.1$	6	$5.7 \pm 0.3$
500 – 600	$9.1 \pm 2.7$	$0.4 \pm 0.1$	$9.6 \pm 2.8$	9	$7.6 \pm 0.3$
600 – 800	$5.7 \pm 2.2$	$0.4 \pm 0.1$	$6.2 \pm 2.2$	6	$3.0 \pm 0.2$
$\geq 800$	$0.8 \pm 0.7$	$0.1 \pm 0.1$	$0.9 \pm 0.7$	3	$0.4 \pm 0.1$

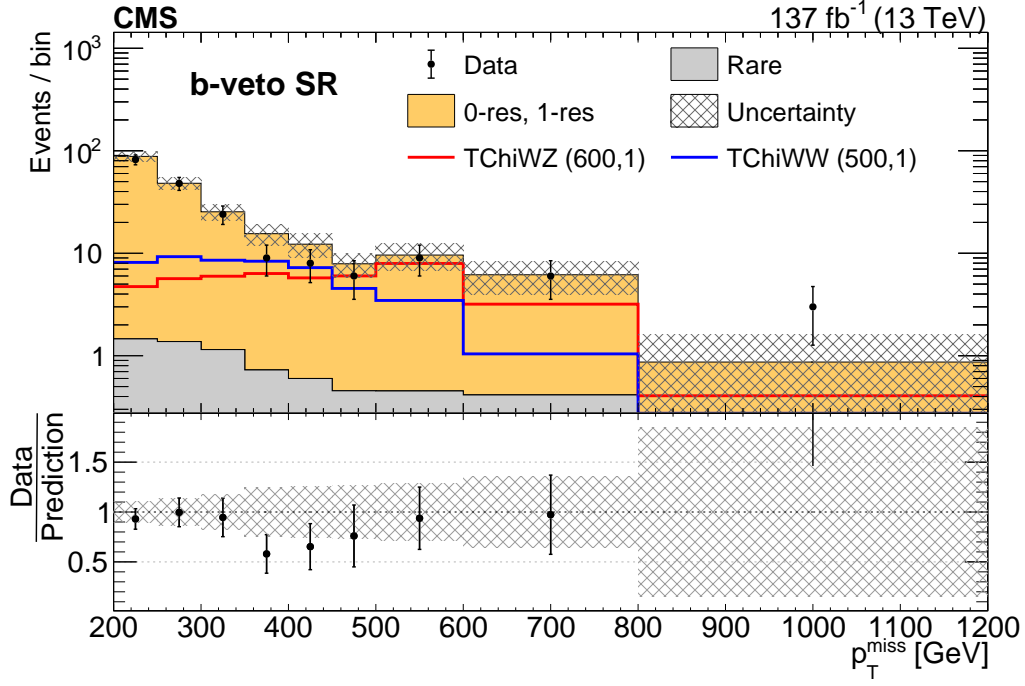


Figure 5.44: Data yields and the post-fit predictions in the b-veto SR

**5.7.1.2 Results in  $b$ -tag SR** After the good agreement between the data and predictions is observed in the  $b$ -tag validation regions, similar background estimation procedure is used to predict the backgrounds in each of the three  $b$ -tag signal regions. Table 5.17 shows the observed yields, background predictions and pulls in each of the  $b$ -tag signal regions. The yields of the SM backgrounds, determined from the fit applied only to the CRs under the background-only hypothesis, are shown along with the predicted yields of the signals and the observations in Fig. 5.45. No statistically significant excess of events is observed in the data with respect to the SM background predictions.

## 5.7.2 Interpretations

A careful comparison is performed to statistically interpret the deviation of data from SM backgrounds in each of the search bin. This section aims to discuss

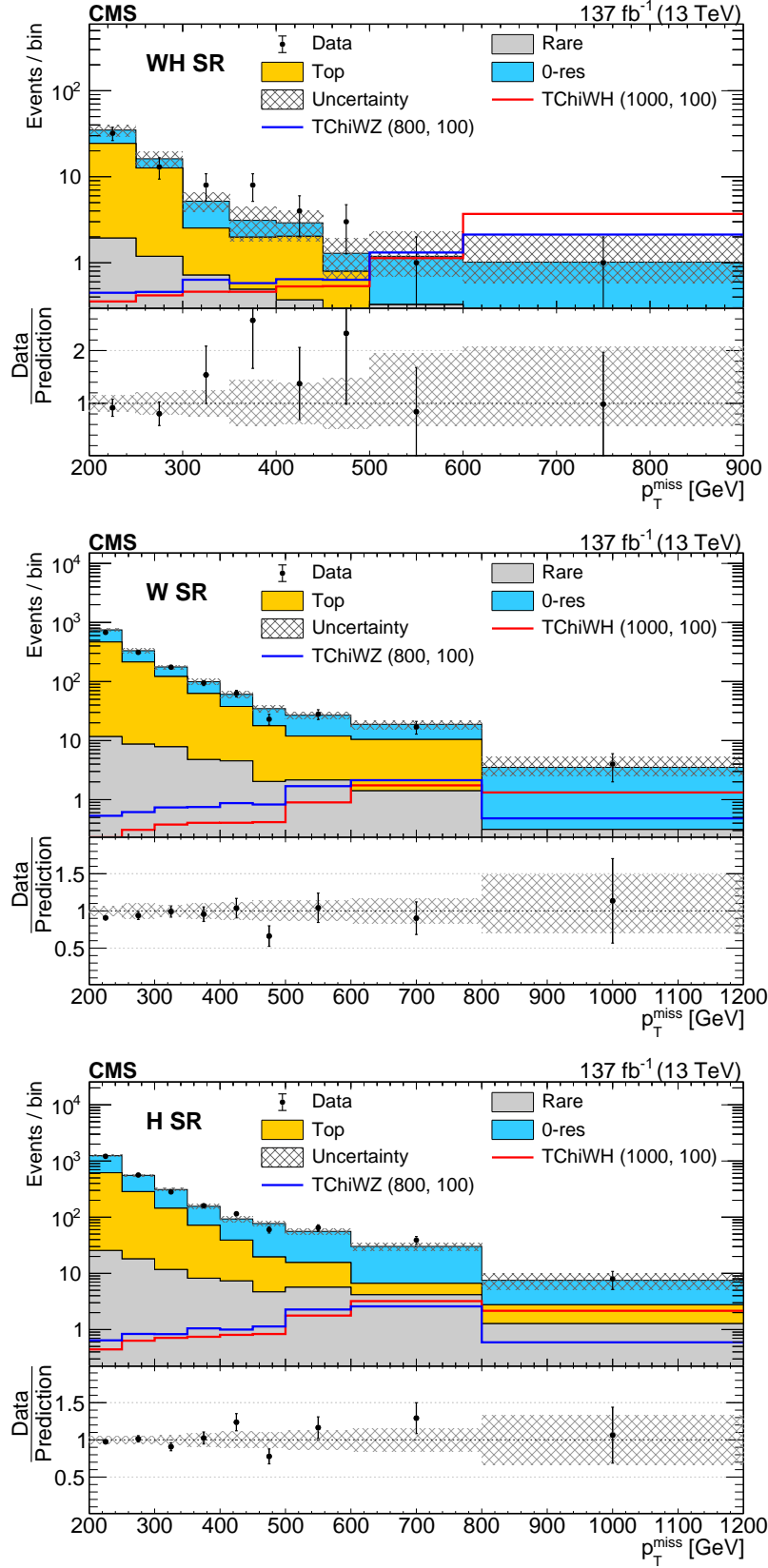


Figure 5.45: The observed numbers of events and post-fit SM background predictions in three b-tag signal regions: (top) WHTag SR, (middle) WTag SR, (bottom) HTag SR

Table 5.17: Observations and SM background predictions derived from the fit to the data in the CRs in each of the b-tag SRs. The column labeled “Signal” provides the expected signal yields from the TChiWH model with an NLSP mass of 1000 GeV and an LSP mass of 100 GeV.

$p_T^{\text{miss}}$ (GeV)	Rare	Top	0-res	Total	Obs.	Signal
WH SR						
200–250	$1.93 \pm 0.38$	$22.4 \pm 3.9$	$10.5 \pm 4.0$	$34.9 \pm 5.4$	32	$0.36 \pm 0.02$
250–300	$1.19 \pm 0.22$	$11.4 \pm 3.0$	$3.5 \pm 1.2$	$16.2 \pm 3.3$	13	$0.42 \pm 0.02$
300–350	$0.72 \pm 0.16$	$1.81 \pm 0.91$	$2.67 \pm 0.83$	$5.2 \pm 1.3$	8	$0.46 \pm 0.02$
350–400	$0.49 \pm 0.10$	$1.49 \pm 0.98$	$1.13 \pm 0.78$	$3.1 \pm 1.3$	8	$0.46 \pm 0.02$
400–450	$0.37 \pm 0.07$	$1.7 \pm 1.1$	$0.88 \pm 0.52$	$2.9 \pm 1.2$	4	$0.53 \pm 0.03$
450–500	$0.24 \pm 0.06$	$0.55 \pm 0.52$	$0.49 \pm 0.34$	$1.29 \pm 0.62$	3	$0.53 \pm 0.03$
500–600	$0.33 \pm 0.06$	$0.0^{+1.0}_{-0.0}$	$0.86 \pm 0.50$	$1.19^{+1.1}_{-0.50}$	1	$1.13 \pm 0.04$
$\geq 600$	$0.29 \pm 0.07$	$0.0^{+1.0}_{-0.0}$	$0.73 \pm 0.42$	$1.02^{+1.1}_{-0.43}$	1	$3.70 \pm 0.07$
W SR						
200–250	$11.6 \pm 1.6$	$455 \pm 27$	$282 \pm 42$	$749.4 \pm 46.1$	680	$0.22 \pm 0.02$
250–300	$8.7 \pm 1.4$	$205 \pm 16$	$118 \pm 28$	$332.4 \pm 32.5$	312	$0.31 \pm 0.02$
300–350	$7.8 \pm 1.4$	$114 \pm 11$	$54.9 \pm 8.0$	$176.4 \pm 13.5$	175	$0.38 \pm 0.02$
350–400	$4.77 \pm 0.94$	$57.8 \pm 7.9$	$36.7 \pm 5.4$	$99.4 \pm 9.9$	95	$0.41 \pm 0.02$
400–450	$4.5 \pm 1.2$	$33.3 \pm 5.5$	$22.9 \pm 4.0$	$60.7 \pm 6.9$	63	$0.41 \pm 0.02$
450–500	$2.03 \pm 0.32$	$15.8 \pm 3.5$	$16.9 \pm 2.8$	$34.7 \pm 4.7$	23	$0.42 \pm 0.02$
500–600	$2.14 \pm 0.32$	$9.8 \pm 2.7$	$15.0 \pm 2.4$	$26.9 \pm 3.6$	28	$0.90 \pm 0.03$
600–800	$1.41 \pm 0.21$	$9.0 \pm 2.7$	$8.4 \pm 1.7$	$18.8 \pm 3.2$	17	$1.74 \pm 0.05$
$\geq 800$	$0.31 \pm 0.06$	$0.00^{+1.4}_{-0.0}$	$3.2 \pm 1.0$	$3.5^{+1.7}_{-1.0}$	4	$1.32 \pm 0.04$
H SR						
200–250	$25.5 \pm 3.6$	$592 \pm 28$	$626 \pm 59$	$1244 \pm 61$	1212	$0.44 \pm 0.02$
250–300	$18.0 \pm 2.7$	$267 \pm 18$	$270 \pm 26$	$555 \pm 27$	563	$0.63 \pm 0.03$
300–350	$11.7 \pm 1.6$	$132 \pm 11$	$166 \pm 17$	$311 \pm 20$	282	$0.71 \pm 0.03$
350–400	$8.2 \pm 1.1$	$63.4 \pm 6.9$	$84 \pm 12$	$156 \pm 14$	160	$0.74 \pm 0.03$
400–450	$7.3 \pm 1.2$	$31.7 \pm 4.7$	$53.9 \pm 8.7$	$93 \pm 10$	115	$0.80 \pm 0.03$
450–500	$4.67 \pm 0.66$	$14.9 \pm 3.0$	$57.6 \pm 7.6$	$77.1 \pm 8.0$	60	$0.83 \pm 0.03$
500–600	$5.64 \pm 0.78$	$9.9 \pm 2.3$	$40.2 \pm 6.4$	$55.8 \pm 7.1$	65	$1.77 \pm 0.05$
600–800	$4.15 \pm 0.60$	$2.5 \pm 1.2$	$23.5 \pm 4.5$	$30.2 \pm 4.8$	39	$3.21 \pm 0.07$
$\geq 800$	$1.27 \pm 0.19$	$1.49 \pm 0.89$	$4.8 \pm 2.3$	$7.5 \pm 2.5$	8	$2.16 \pm 0.05$

a few statistical interpretations of the results. A modern frequentist method is used to compute the exclusion limits which defines the test statistic as ratio of likelihood ratio. A likelihood is defined as a probability of observed events given the expected number of events. A fit function is constructed to maximize the likelihood. Detailed introduction to the bayesian and frequentist approaches can be referred from [74].

The likelihood is a function dependent on parameters  $\mu$  and  $\theta$ . In this case, the parameter of interest is the signal strength modifier ( $\mu$ ), defined as the ratio of the observed cross-section divided by the nominal cross-section. The nuisance parameters whose values are not known are called *prior* nuisance parameters  $\tilde{\theta}$  that need to be fitted using the data. With around nine search bins in each of the four SRs defined in this analysis denoted by  $i$ , and the nuisance parameters denoted by  $j$ , the likelihood function can be described as:

$$\mathcal{L}(\text{data} | \mu, \theta_j) = \text{Poisson}(\text{data} | \mu s_i(\theta_j) + b_i(\theta_j)) p(\tilde{\theta}_j, \theta_j) \quad (5.10)$$

where  $p$  denotes the Poisson nuisance term. Predicted signal  $s_i$ , and predicted background  $b_i$  are also dependent on the nuisance parameters as shown below:

$$\text{Poisson}(\text{data} | \mu s_i + b_i) = \prod_i \frac{(\mu s_i + b_i)^{n_i}}{n_i!} e^{-(\mu s_i + b_i)}, \quad (5.11)$$

$$p(\tilde{\theta}_j, \theta_j) = \prod_j \frac{1}{\sqrt{2\pi}\sigma(\tilde{\theta}_j)} \frac{1}{\theta_j} e^{\left[ -\frac{(\ln(\theta_j/\tilde{\theta}_j))^2}{2(\sigma(\tilde{\theta}_j))^2} \right]}, \quad (5.12)$$

$$s_i(\theta_j) = s_{0,i} \prod_j R_{ij}^{\theta_j}(s), \quad (5.13)$$

$$b_i(\theta_j) = \sum_{x \in BG} \left( x_{0,i} \prod_j R_{ij}^{\theta_j}(x) \right). \quad (5.14)$$

In the case of limited number of events in the control region, as in some of the b-tag CRs, the uncertainty is modeled by a gamma distribution. All the nuisance parameters have a prior mean  $\tilde{\theta}_j$  of zero and sigma  $\sigma(\tilde{\theta}_j)$  equal to one.  $BG$  is a set of all backgrounds for the analysis, and  $x_{0,i}$  ( $s_{0,i}$ ) are the nominal prediction for a background (signal). The symbol  $R_{ij}$  denotes the signal and background dependent systematic uncertainty correction values for each bin and nuisance parameter, and  $n_i$  is the observed number of data events per bin.

As discussed above, the test statistic can be expressed mathematically as:

$$q_\mu = -2 \ln \frac{\mathcal{L}(\text{data} | \mu, \hat{\theta}_\mu)}{\mathcal{L}(\text{data} | \hat{\mu}, \hat{\theta})}, \text{ with a constraint } 0 \leq \hat{\mu} \leq \mu \quad (5.15)$$

where the numerator is for the *background-only* fit in which  $\mu$  is fixed to zero and the denominator is for the *signal + background* fit in which  $\mu$  is allowed to float freely. Both fits to likelihood functions are performed separately, and best fit is chosen that maximizes the likelihood function.

The level of agreement between the data and predicted yields with a value  $\mu$  can be quantified using the  $p$ -value. The  $p$ -value is found for a non-zero  $\mu$  value using the probability distribution function, given by:

$$f(t_\mu | \mu') = \frac{1}{2\sqrt{t_\mu}} \frac{1}{\sqrt{2\pi}} \left[ \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} + \frac{\mu - \mu'}{\sigma}\right)^2\right) + \exp\left(-\frac{1}{2} \left(\sqrt{t_\mu} - \frac{\mu - \mu'}{\sigma}\right)^2\right) \right] \quad (5.16)$$

Then, the  $p$ -values are given by:

$$p_\mu = \int_{q_\mu^{\text{obs}}}^{\infty} f(q_\mu | \mu, \hat{\theta}_\mu^{\text{obs}}) dq_\mu \quad (5.17)$$

$$p_b = \int_{-\infty}^{q_\mu^{\text{obs}}} f(q_\mu | 0, \hat{\theta}_0^{\text{obs}}) dq_\mu. \quad (5.18)$$

The symbol  $p_\mu$  denotes the  $p$ -value for the *signal + background* hypothesis with some floating value of  $\mu$  in the numerator of  $q_\mu$  with some signal injected. The symbol  $p_0$  denotes the  $p$ -value for *background-only* hypothesis with no signal injected. Finally, these  $p$ -values are used to construct confident levels  $\text{CL}_s$ , defined by:

$$\text{CL}_s(\mu) = \frac{p_\mu}{1 - p_b} \leq 1 - \alpha \quad (5.19)$$

The results are quoted for most of the CMS experiments at the 95% confidence level (*i.e.*,  $\alpha = 0.05$ ). Upper limit quoted at 95% confidence level on the value of  $\mu$  is denoted by  $\mu_{\text{up}}$ , which is often a gaussian distribution. Uncertainty on this value is



denoted by  $\pm 1\sigma$  (68%) and  $\pm 2\sigma$  (95%) as shown in Fig. 5.46, which is relevant for expected limits. Figure 5.46 shows the distribution of the probability distribution function and the calculation of  $p$ -value.

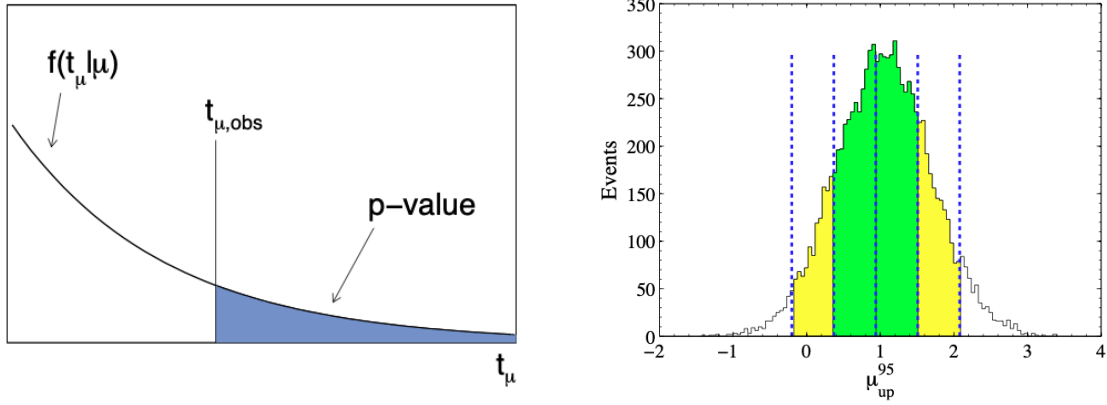


Figure 5.46: Distribution of the PDF and  $p$ -value calculation (left). Distribution of the upper limit on  $\mu$  at 95% CL (right).

In the context of this analysis, the fit obtained by the predictions to SM backgrounds are determined only from the fit to control regions under background-only hypothesis. All the b-veto and b-tag SR and CRs are used to derive limits for each model. The production of chargino pair or chargino-neutralino pair is the prime focus of this analysis; therefore, 95% CL upper limits are placed on these production cross sections. By comparing the upper limits on the production cross sections to the cross-sections predicted for chargino-pair production and for chargino-neutralino production, 95% CL mass exclusion contours are derived in the NLSP-LSP plane (LSP here is  $\tilde{\chi}_1^0$ ). For the case of NLSP being wino-like, the mass exclusion are obtained for three signal models:

- TChiWW model: Assuming a pair of chargino production, the masses of NLSP ( $\tilde{\chi}_1^\pm$ ) are excluded between 290 and 670 GeV as shown in Fig. 5.47 (top left). Since the observed events in  $300 < p_T^{\text{miss}} < 500$  GeV region of the b-tag SR are lower than the SM predictions in that bin, the observed limits are stronger than the expected limits.
- TChiWZ model: In the case of chargino-neutralino production, where the chargino decays 100% to a Z boson is denoted by TChiWZ model. Figure 5.47 (top right) shows the mass exclusions of NLSP between 230 and 760 GeV at low values of LSP. The last  $p_T^{\text{miss}}$  bin of b-tag SR has higher number of observed events compared to the predictions which make the observed limits weaker than expected limits.
- TChiWH model: When the chargino decays 100% to a H boson in chargino-neutralino production is denoted by TChiWH model. The limits on this model are mostly driven by the b-tag's WH signal region. The excess in a few intermediate  $p_T^{\text{miss}}$  bins in this SR makes the observed limits weaker than the expected limits as shown in Fig. 5.47 (bottom) and the NLSP masses between 240 and 970 GeV are excluded for low LSP masses.

In addition to these simplified models listed above, two more specific scenarios in which all the winos or all of the higgsinos are mass degenerate are considered, which are referred to as the wino- and higgsino-like NLSP scenarios. For the simultaneous production of  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ , the  $\tilde{\chi}_1^\pm$  and  $\tilde{\chi}_2^0$  are considered to be mass-degenerate wino-like NLSPs. The combination includes signal models of TChiWW along with either TChiWZ or TChiWH. The TChiWW + TChiWZ combination excludes the NLSP masses up to 870 GeV are excluded while the expected limit exclusion goes up

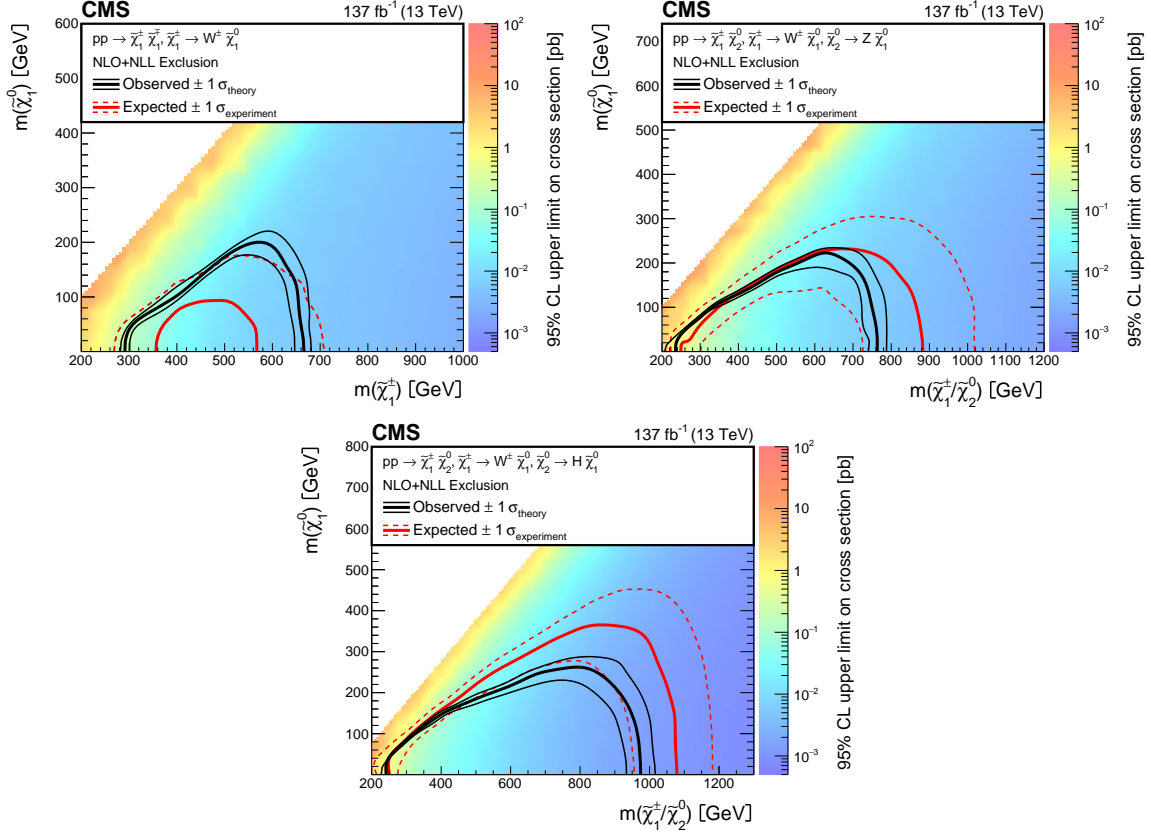


Figure 5.47: The 95% CL upper limits on the production cross sections for  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  assuming that each  $\tilde{\chi}_1^\pm$  decays to a W boson and  $\tilde{\chi}_1^0$  (upper left, the TChiWW model) and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production assuming that the  $\tilde{\chi}_1^\pm$  decays to a W boson and  $\tilde{\chi}_1^0$ , and that the  $\tilde{\chi}_2^0$  decays to a Z boson and  $\tilde{\chi}_1^0$  (upper right, the TChiWZ model) or that the  $\tilde{\chi}_2^0$  decays to a Higgs boson and  $\tilde{\chi}_1^0$  (lower, the TChiWH model). The black curves represent the observed exclusion contour and the change in this contour due to variation of these cross sections within their theoretical uncertainties ( $\sigma_{\text{theory}}$ ). The red curves indicate the mean expected exclusion contour and the region containing 68% ( $\pm 1 \sigma_{\text{experiment}}$ ) of the expected exclusion limits under the background-only hypothesis. The mass exclusion limits are computed assuming wino-like cross sections.

to 1010 GeV. The TChiWW + TChiWH combination excludes the NLSP masses up to 960 GeV are excluded while the expected limit exclusion goes up to 1110 GeV as shown in Fig. 5.48.

With the NLSP having higgsino-like cross sections and the LSP considered as lightest neutralino, the production cross sections are too low compared with wino-like cross sections. Since this analysis is sensitive to large class of electroweakino models as

shown above, it is the first analysis from CMS experiment to become sensitive to the direct production of higgsino-like NLSPs under the MSSM model with the neutralino LSP. The production processes include  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  productions. Fig. 5.49 shows the mass exclusions for higgsino-like NLSP masses that are excluded from 300 to 650 GeV and the expected exclusion limits from 320 to 810 GeV.

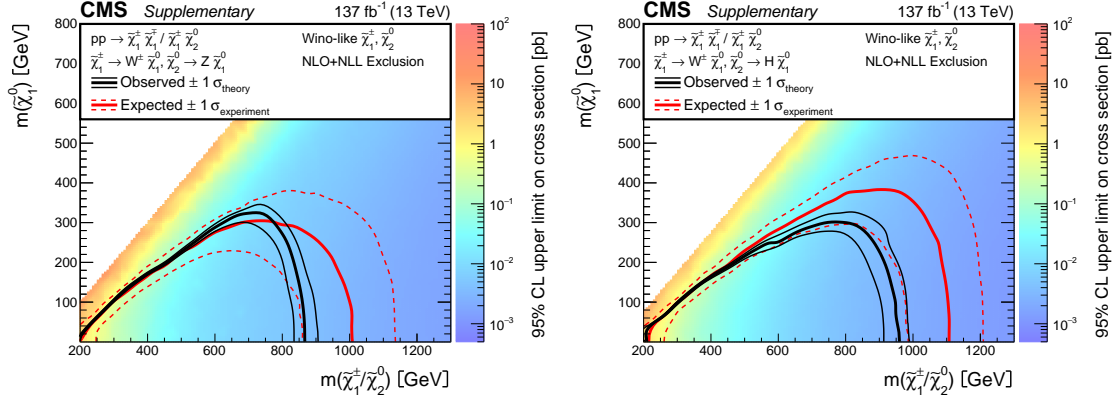


Figure 5.48: Expected and observed 95% CL upper limits on mass-degenerate wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production, assuming  $\mathcal{B}(\tilde{\chi}_1^\pm \rightarrow W\tilde{\chi}_1^0) = 100\%$ , and either  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0) = 100\%$  (left) or  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0) = 100\%$  (right). In each of these two plots, the red (black) contours represent the expected (observed) mass exclusion limits. Mass exclusion limits are computed assuming wino-like cross sections.

The projection from this Run 2 CMS search with  $137 \text{ fb}^{-1}$  of data for electroweak production of supersymmetric particles to the full expected integrated luminosity ( $3000 \text{ fb}^{-1}$ ) of the high-luminosity LHC (HL-LHC) operation is performed. The HL-LHC is a major upgrade for the LHC with more than 10 times increase in the luminosity [75] from the current value. The LHC has delivered only around 5% of the expected data within its life span. Therefore, the large amount of data are expected to improve the sensitivity towards the higgsino production (cross section of higgsino is many orders of magnitude smaller than the gluinos). The expected

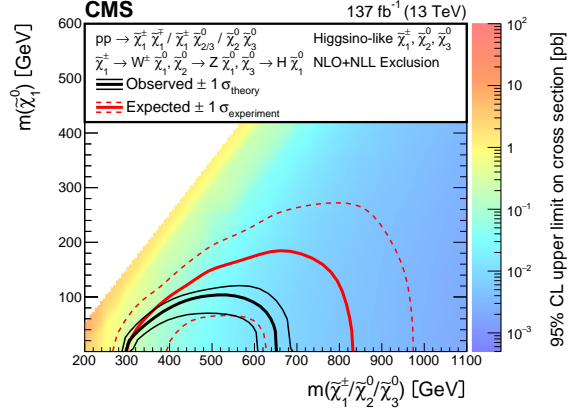


Figure 5.49: Expected and observed 95% CL exclusion for mass-degenerate higgsino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_3^0$ , and  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  production as functions of the NLSP and LSP masses with 95% CL upper limits on the production cross sections. The  $\tilde{\chi}_1^\pm$ ,  $\tilde{\chi}_2^0$ , and  $\tilde{\chi}_3^0$  are considered to be mass degenerate.

limits are projected with the higgsino cross sections in this analysis and are shown in Fig. 5.50. More details on the procedure used to derive these projections are described in Appendix A. The HL-LHC projection results are compared to the Run 2 results, where gains in sensitivity of around 500 GeV are expected, pushing higgsino (wino) sensitivity as high as 1390 (1590) GeV.

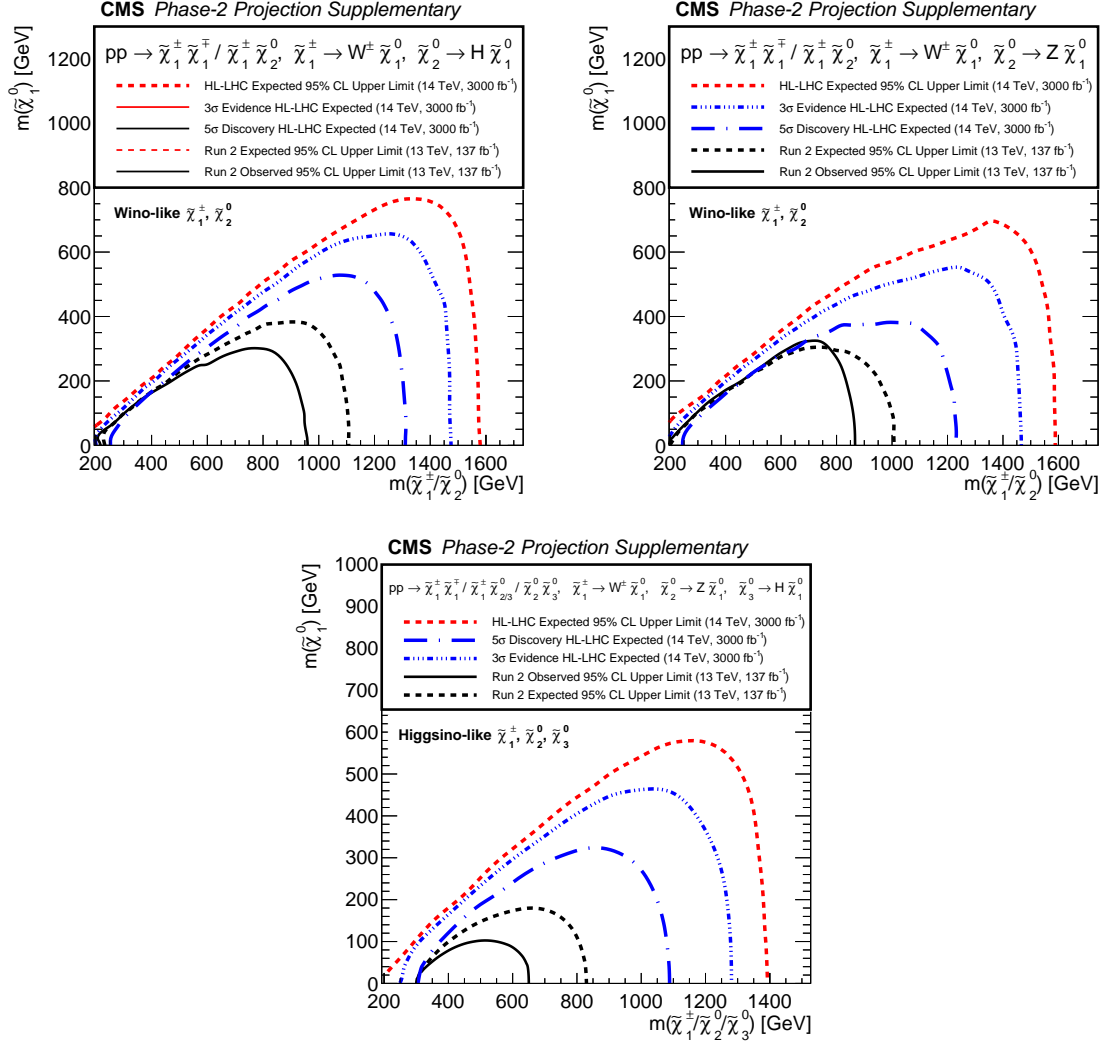


Figure 5.50: Projected 95% CL exclusion for 3 ab $^{-1}$  (red) for mass-degenerate wino-like  $\tilde{\chi}_1^\pm\tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$  production assuming  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0) = 100\%$  (upper left) or  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0) = 100\%$  (upper right) and higgsino-like  $\tilde{\chi}_1^\pm\tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm\tilde{\chi}_3^0$ ,  $\tilde{\chi}_1^\pm\tilde{\chi}_1^\mp$ , and  $\tilde{\chi}_2^0\tilde{\chi}_3^0$  production (lower) as functions of the NLSP and LSP masses. Projections are compared to results from LHC Run 2 with 137 fb $^{-1}$  (black). Projected 5 $\sigma$  and 3 $\sigma$  expected significance curves (blue) are also included.

## CHAPTER SIX

### Summary

Although the standard model (SM) of particle physics explains the majority of the empirically observed data, there are still a number of unanswered questions including those concerning the nature of dark matter, neutrinos, gauge hierarchy problem, and the mass of the Higgs boson. One of the proposed extensions of the standard model is supersymmetry (SUSY), which offers a candidate for dark matter that interacts weakly and can solve the significant corrections to the Higgs bare mass by introducing a superpartner for each SM particle. The hunt for supersymmetric particles called charginos and neutralinos, which decay to SM bosons and the lightest supersymmetric particle, is the foundation of this dissertation study. Two hadronically decaying bosons are reconstructed and a significant amount of missing transverse momentum is required in this search. Machine learning algorithms are utilized to identify hadronically decaying W, Z, and Higgs bosons reconstructed as large-size jets. The Run 2 data set from the CMS detector at the LHC is used for this search which corresponds to an integrated luminosity of  $137 \text{ fb}^{-1}$ .

The study is split into two orthogonal search regions to increase the sensitivity to the final states for three simplified SUSY models. Large-size (AK8) jets are used for identification of boson decays while small-size (AK4) jets are used to identify the b quark initiated jets in this analysis. The production of  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\pm$  or  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  signal with W and/or Z bosons in the final states, which decay into a pair of light quarks (other than b quark), is the focus of the b-veto search region, which requires events without

a b-tagged small-size jet. The production of  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\pm$  or  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  signal with W and Z or H bosons in the final state, in which the Z and H bosons decay into a pair of bottom quarks, is the target of the b-tag search region. The stable lightest supersymmetric particle (LSP) escapes the CMS detector as a result of an  $R$ -parity conserving SUSY, acting as a dark matter candidate. For each search region, the amount of background is estimated using a data-driven background estimation technique. The background estimation methods are well validated in orthogonal validation regions for both search regions. No statistically significant excess of events is observed in the data with respect to the expectation from the standard model.

Interpretations of this search are presented in terms of the upper limits on the cross section as well as the mass exclusion(s) in the two-dimensional plane of the parent next-to-lightest supersymmetric particle (NLSP) and the LSP. For all the interpretations, findings from both search regions are combined. The limits established by prior analyses that focus on the leptonic boson final states are exceeded by the mass exclusions achieved through this approach at high NLSP masses. For simplified SUSY models with WW, WZ, or WH final states, the NLSP mass exclusion limit set by this search for low-mass LSPs extends up to 670, 760, and 970 GeV, respectively. In addition, interpretations are also presented for more realistic SUSY models with multiple signal topologies combined as all these signals are more likely to be produced simultaneously.

With  $\tilde{\chi}_1^0$  as the LSP in the minimal supersymmetric extension to the standard model, this search yields a high sensitivity to the superpartner of the Higgs boson, higgsino, for the first time in CMS analyses. The predicted exclusion limits lie between 320 and 810 GeV, while the observation in data excludes NLSP masses between 300



and 650 GeV. Additionally, results of this research are utilized in the effort to combine multiple electroweak SUSY searches [76].

With more data from the Run 3 operations of the CMS experiment, combined with the latest improvement in deep tagging techniques, the mass exclusions can be pushed further towards high NLSP masses. Aside from the energy increase, the CMS detector is also undergoing a number of enhancements, including a new high-granularity calorimeter, which replaces the endcap section of the current electromagnetic and hadron calorimeters and offers improved timing capabilities. The sensitivities to higgsinos and other supersymmetric particles would still be improved with the numerous advancements in the use of machine learning techniques for particle identification, combined with the techniques used to better identify signals and reject backgrounds in various physics analyses. Attaining further sensitivities to higgsinos would be a great way to test the concept of naturalness in SUSY. With more data and detector upgrade efforts, there would be exciting results from the SUSY sector towards the end of present Run 3 and much more through the HL-LHC era.

## APPENDIX

## APPENDIX

### Projected future sensitivities for HL-LHC

In this appendix, the procedure used to extrapolate the Run 2 results to the full high-luminosity LHC (HL-LHC) scenario is described. A projection is performed based on the extrapolation of Run 2 results to access the future sensitivities of HL-LHC to higgsino-like and wino-like signals with  $3 \text{ ab}^{-1}$  of data at 14 TeV.

#### *A.1 Inputs to HL-LHC projection*

The same binning and background prediction techniques from the Run 2 result are applied to the projection. To simulate the  $3000 \text{ fb}^{-1}$  of data expected at the HL-LHC, the yields for both signal and background are scaled by the ratio of integrated luminosities for the HL-LHC and Run 2 ( $137 \text{ fb}^{-1}$ ). The HL-LHC is expected to collide protons at  $\sqrt{s} = 14 \text{ TeV}$ . The signal and background yields are also scaled to account for the difference in cross sections due to the increase in the LHC center-of-mass energy from 13 TeV to 14 TeV. The cross section ratio scale factors for signal are calculated using the Resummino prediction tool [77,78] for wino and higgsino pair production. They are mass dependent and range from around 1.1 to 1.6, increasing at higher masses. Cross section scale factors for standard model processes are taken from Table 1 of Ref. [79]. The main backgrounds for this search are  $t\bar{t}$ , with a 14/13 TeV ratio of 1.18, and  $W + \text{jets}$ ,  $Z + \text{jets}$ , and diboson processes, all with a 14/13 TeV ratio of 1.1.

In order to avoid possible contributions from signal events and statistical fluctuations in the signal regions, the observed data from the signal regions in the Run 2

results are not used. Instead, the projection is made based on a control-region-only fit of the SM background. Those best fit values for the SM contributions in the signal region are used and scaled up by the luminosity and cross section ratios.

The signal MC yields from the Run 2 result are used wherever possible. Since the projection sensitivities reach to higher masses, additional simulation samples are generated in a private production with the same configurations as used in the central CMS sample, but extended to higher masses. Kinematic distributions are checked between the private samples and the central CMS samples, and they agree well with each other. For simplicity, only 2018 signal MC samples are generated for the private production, and are scaled up to the full expected luminosity.

SUSY signal cross-sections are taken from our own Resummino studies [77,78] at 13 TeV to 14 TeV and compared to centrally-produced values on the LHC SUSY cross section working group twiki [80]. Comparisons between these generated cross sections and centrally-produced values are shown in Fig. A.1 for wino-like production and in Fig. A.2 for higgsino-like production. Aside from issues related to the comprehensiveness of the twiki values, such as the 13 TeV higgsino values not extending to high enough masses and 14 TeV cross sections not covering all of the needed production channels, especially in the case of higgsino cross sections, other limitations possibly due to rounding precision are also observed along with 5–10% differences in cross sections between the privately-computed and centrally-provided values as shown in Figs. A.1 and A.2. For these reasons, the privately computed signal cross-sections are used in this projection.

The Run 2 values for W-tagging uncertainties are halved and used for this projection study. The HL-HLC projection in general suffers from limited number of

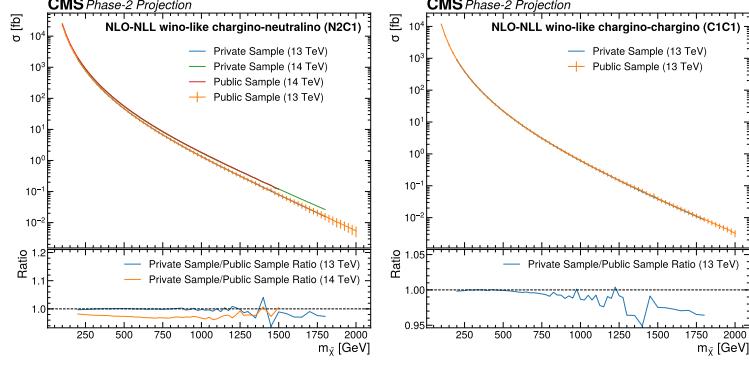


Figure A.1: Comparison between privately- and centrally-produced [77, 78] Resumino cross-sections for wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  (left) and  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  (right) production.

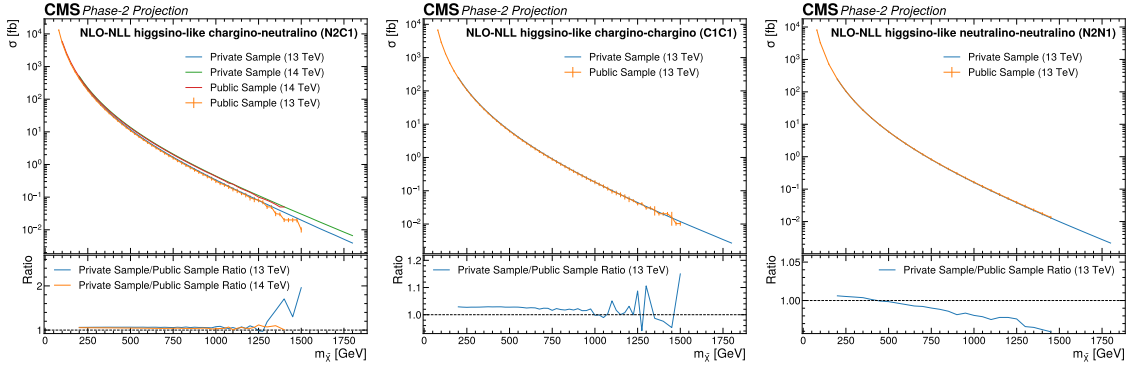


Figure A.2: Comparison between privately- and centrally-produced [77, 78] Resumino cross-sections for higgsino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  (left),  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  (middle), and  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  (right) production.

MC sample events. The larger sample of events that becomes available at the HL-LHC will allow us to reduce the tagging systematic uncertainties, and scale down these contributions by a factor of 2 from their Run 2 values for the projection. For Higgs tagging uncertainties, the same procedure as used for the W tagging uncertainties is used. The details of the treatment of the systematic uncertainties have a small effect on the final results.

## A.2 HL-LHC projection results

After applying this scaling to the analysis datacards, mass exclusions projected to the HL-LHC scenario are extracted for both wino-like and higgsino-like interpretations as shown in Fig. A.3.

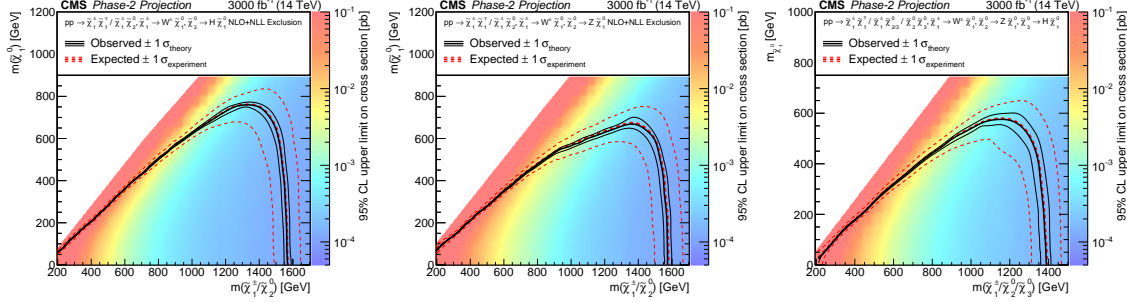


Figure A.3: Projected mass exclusions and 95% confidence level upper limits on cross sections for mass-degenerate wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production where either  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H \tilde{\chi}_1^0)$  (left) or  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z \tilde{\chi}_1^0)$  (middle) is assumed to be 100%, and higgsino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_3^0$ , and  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  production (right).

The wino-like interpretation combines  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production and  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  production. The individual channels contributing to the wino-like combinations are also projected to understand the sensitivity of each independently as shown in Fig. A.4.

The  $5\sigma$  and  $3\sigma$  expected significance curves are also obtained for wino-like and higgsino-like combinations that show thresholds for evidence and discovery in the HL-LHC dataset, with results in Fig. A.5. Significance is calculated using an a-posteriori model that accounts for the CR-only fit yields used to model projected observations.

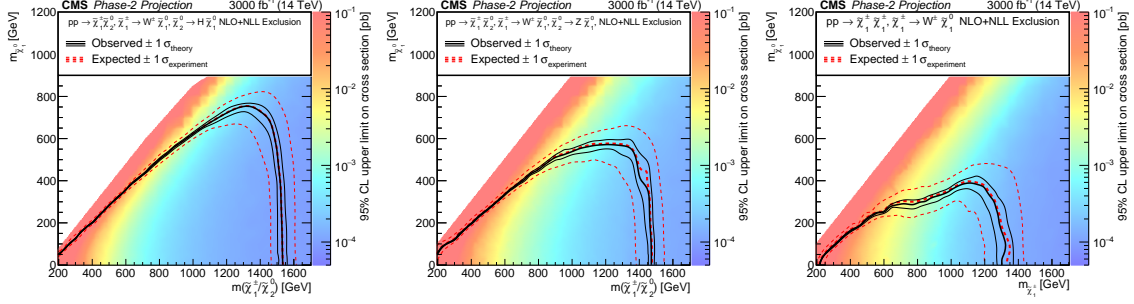


Figure A.4: Projected mass exclusions and 95% confidence level upper limits on cross sections for mass-degenerate wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production assuming that the  $\tilde{\chi}_1^\pm$  decays to a W boson and  $\tilde{\chi}_1^0$  and the  $\tilde{\chi}_2^0$  decays to a Higgs boson (TChiWH, left) or Z boson (TChiWZ, middle), as well as wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  assuming that each  $\tilde{\chi}_1^\pm$  decays to a W $^\pm$  boson and  $\tilde{\chi}_1^0$  (TChiWW, right).

Comparisons are made with the Run 2 results in Fig. A.6, showing increased sensitivity in all signal combinations. Projections made for higgsino and wino sensitivity reach masses as large as 1340 and 1590 GeV, respectively, roughly 500 GeV larger than the expected exclusion in the Run 2 results.

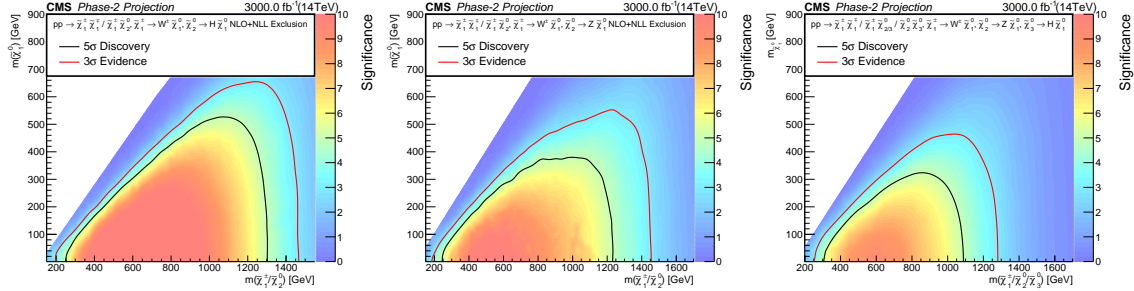


Figure A.5: 5 $\sigma$  (black) and 3 $\sigma$  (red) expected significance for mass-degenerate wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production assuming  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H\tilde{\chi}_1^0)$  (left) or  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z\tilde{\chi}_1^0)$  (middle) and higgsino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_3^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ , and  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  production (right) as functions of the NLSP and LSP masses.

In the wino-like scenario, the Run 2 results show better reach for the Higgs decay channel compared to the Z decay channel. This is due to the higher Higgs tagging efficiency and larger  $H \rightarrow b\bar{b}$  branching ratio. In the HL-LHC projection,

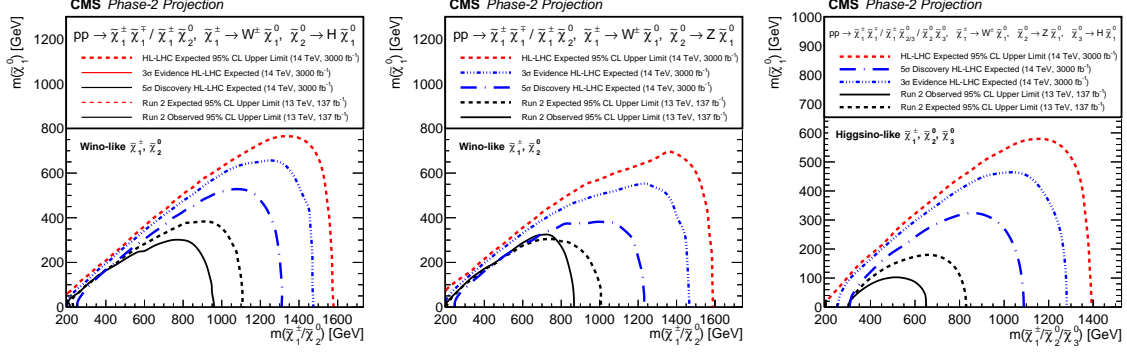


Figure A.6: Projected 95% CL exclusion for 3 ab<sup>-1</sup> (red) for mass-degenerate wino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$  and  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$  production assuming  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow H \tilde{\chi}_1^0)$  (left) or  $\mathcal{B}(\tilde{\chi}_2^0 \rightarrow Z \tilde{\chi}_1^0)$  (middle) and higgsino-like  $\tilde{\chi}_1^\pm \tilde{\chi}_2^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_3^0$ ,  $\tilde{\chi}_1^\pm \tilde{\chi}_1^\mp$ , and  $\tilde{\chi}_2^0 \tilde{\chi}_3^0$  production (right) as functions of the NLSP and LSP masses. Projections are compared to results from the LHC Run 2 with 137 fb<sup>-1</sup> (black). Projected 5σ and 3σ expected significance (blue) is also included.

however, we see that the mass reach for the Higgs and Z channels are comparable. The ratio of the expected cross section upper limits between the two channels is shown in Fig. A.7. In the region of the Run 2 limits around wino mass of 800–1000 GeV, the Higgs channel is better. In the region of the HL-LHC mass reach around wino mass of 1600 GeV the limits are comparable. This effect is understood to be the result of the different  $p_T^{\text{miss}}$  binning used in the b-tag and b-veto signal regions, where the tightest b-tag region is  $p_T^{\text{miss}} > 600$  GeV while the tightest b-veto signal region is  $p_T^{\text{miss}} > 800$  GeV. This is an artifact of the limited number of MC sample events and data events available in the control regions for the Run 2 analysis. An analysis fully optimized for the HL-LHC luminosity is expected to show a better mass reach for the Z channel and much more for the Higgs channel.



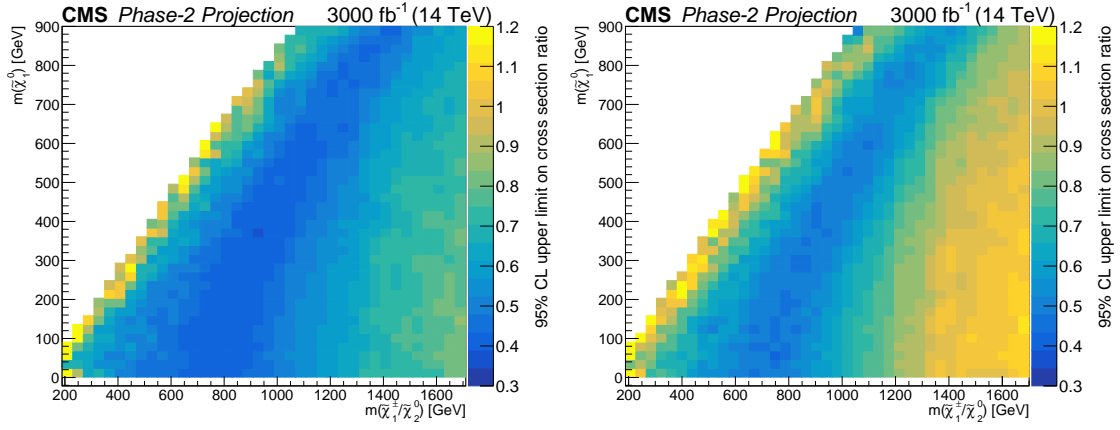


Figure A.7: Ratio of 95% CL upper limit cross sections for TChiWH and TChiWZ (left) and for combinations TChiWW+WH and TChiWW+WZ (right).

## BIBLIOGRAPHY

- [1] Foundation, CK-12, “Quarks ( read ) — physics”, 2017.  
<https://www.ck12.org/physics/quarks/lesson/Quarks-PHYS/>.
- [2] Cush, “File:standard model of elementary particles anti.svg - wikimedia commons”, 2018.  
<https://commons.wikimedia.org/w/index.php?curid=73606545>.
- [3] CMS Collaboration, “Dark matter search in CMS”, Technical Report CMS-CR-2017-147, 2017. <https://cds.cern.ch/record/2271096>.
- [4] H. Baer and X. Tata, “The Minimal Supersymmetric Standard Model”. Cambridge University Press, 2006. doi:10.1017/CB09780511617270.009.
- [5] D. J. Griffiths, “Introduction to elementary particles; 2nd rev. version”. Physics textbook. Wiley, New York, NY, 2008. ISBN 978-3-527-40601-2.
- [6] M. E. Peskin and D. V. Schroeder, “An introduction to quantum field theory”. Westview, 1995. <https://cds.cern.ch/record/257493>.
- [7] H. D. Politzer, “Reliable perturbative results for strong interactions”, *Phys. Rev. Lett.* **30** (1973) 1346, doi:10.1103/PhysRevLett.30.1346.
- [8] JabberWok, “A Feynman diagram of an electron and positron exchanging a photon. - wikimedia commons”, 2006. <http://en.wikipedia.org/wiki/Image:Electron-positron-scattering.svg>.
- [9] S. Weinberg, “A model of leptons”, *Phys. Rev. Lett.* **19** (1967) 1264, doi:10.1103/PhysRevLett.19.1264.
- [10] C. Schiller, “A conjecture on deducing general relativity and the standard model with its fundamental constants from rational tangles of strands”, *Physics of Particles and Nuclei* **50** (2019) 259, doi:10.1134/s1063779619030055.
- [11] R. L. Workman et al. (Particle Data Group), “Review of particle physics”, *Prog. Theor. Exp. Phys.* **2022** (2022) 083C01, doi:10.1093/ptep/ptac097.
- [12] M. Banner et al., “Observation of single isolated electrons of high transverse momentum in events with missing transverse energy at the CERN pp collider”, *Phys. Lett. B* **122** (1983) 476, doi:10.1016/0370-2693(83)91605-2.

- [13] F. Englert and R. Brout, “Broken symmetry and the mass of gauge vector mesons”, *Phys. Rev. Lett.* **13** (1964) 321, doi:10.1103/PhysRevLett.13.321.
- [14] P. W. Higgs, “Broken symmetries and the masses of gauge bosons”, *Phys. Rev. Lett.* **13** (1964) 508, doi:10.1103/PhysRevLett.13.508.
- [15] J. Ellis, “Higgs physics”, doi:10.5170/CERN-2015-004.117, arXiv:1312.5672.
- [16] A. I. Lonappan et al., “Bayesian evidences for dark energy models in light of current observational data”, *Phys. Rev. D* **97** (2018) 043524, doi:10.1103/PhysRevD.97.043524.
- [17] K. Kodama et al., “Observation of tau neutrino interactions”, *Phys. Lett. B* **504** (2001) 218, doi:10.1016/S0370-2693(01)00307-0.
- [18] M. Gonzalez-Garcia and M. Maltoni, “Phenomenology with massive neutrinos”, *Phys. Rep.* **460** (2008) 1, doi:10.1016/j.physrep.2007.12.004.
- [19] S. P. Martin, “A supersymmetry primer”, *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1, doi:10.1142/9789812839657\_0001, arXiv:hep-ph/9709356.
- [20] J. L. Feng, “Naturalness and the status of supersymmetry”, *Ann. Rev. Nucl. Part. Sci.* **63** (2013) 351, doi:10.1146/annurev-nucl-102010-130447, arXiv:1302.6587.
- [21] CMS Collaboration, “CMS SUS Physics Results”. TWiki: <https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSUS>.
- [22] C. Borschensky et al., “LHC SUSY Cross Section Working Group”. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections>.
- [23] CMS Collaboration, “Search for electroweak production of charginos and neutralinos at  $\sqrt{s} = 13$  TeV in final states containing hadronic decays of WW, WZ, or WH and missing transverse momentum”, arXiv:2205.09597.
- [24] CMS Collaboration, “Search for supersymmetry in proton-proton collisions at 13 TeV in final states with jets and missing transverse momentum”, *JHEP* **10** (2019) 244, doi:10.1007/JHEP10(2019)244, arXiv:1908.04722.
- [25] CMS Collaboration, “Search for chargino-neutralino production in events with Higgs and W bosons using 137 fb<sup>-1</sup> of proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **10** (2021) 045, doi:10.1007/JHEP10(2021)045, arXiv:2107.12553.
- [26] CMS Collaboration, “Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **04** (2021) 123, doi:10.1007/JHEP04(2021)123, arXiv:2012.08600.

- [27] CMS Collaboration, “Search for supersymmetry in final states with two or three soft leptons and missing transverse momentum in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **04** (2022) 091, doi:10.1007/JHEP04(2022)091, arXiv:2111.06296.
- [28] F. Marcastel, “CERN’s accelerator complex la chaîne des accélérateurs du CERN”, (2013). <https://cds.cern.ch/record/1621583>.
- [29] AC Team, “Diagram of an LHC dipole magnet schéma d’un aimant dipôle du LHC”, (1999). <https://cds.cern.ch/record/40524>.
- [30] CMS Collaboration, “CMS Luminosity Public Results”. TWiki: [https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi\\_year\\_plots](https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults#Multi_year_plots).
- [31] CMS Collaboration, “Cutaway diagrams of CMS detector”, (2019). <https://cds.cern.ch/record/2665537>.
- [32] CMS Collaboration, “Description and performance of track and primary-vertex reconstruction with the CMS tracker”, *JINST* **9** (2014) P10009, doi:10.1088/1748-0221/9/10/P10009, arXiv:1405.6569.
- [33] P. K. Siddireddy (for the CMS Collaboration), “The CMS ECAL trigger and DAQ system: electronics auto-recovery and monitoring”, arXiv:1806.09136.
- [34] F. Ferri (for the CMS Collaboration), “The CMS ECAL Phase-2 upgrade for high precision energy and timing measurements”, *Nucl. Instrum. Meth. A* **958** (2020) 162159, doi:10.1016/j.nima.2019.04.113.
- [35] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector performance and software”, CMS Technical Design Report, 2006. <https://cds.cern.ch/record/922757>.
- [36] M. Brice, L. Vaillet, and L. Lazic, “Images of the CMS HCAL barrel (HB)”, (2008). <https://cds.cern.ch/record/1431485>.
- [37] T. Virdee, M. Brice, and L. Veillet, “Images of CMS HCAL endcap (HE)”, (2008). <https://cds.cern.ch/record/1431487>.
- [38] M. Brice, T. Virdee, and T. Camporesi, “Images of CMS HCAL forward calorimeter (HF)”, (2008). <https://cds.cern.ch/record/1431489>.
- [39] J. R. Dittmann, “HCAL Tower Depth Segmentation Figures”, (2017). <https://cms-docdb.cern.ch/cgi-bin/DocDB/ShowDocument?docid=13342>.
- [40] CMS Collaboration, “Search for additional neutral MSSM Higgs bosons in the  $\tau\tau$  final state in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **09** (2018) 007, doi:10.1007/JHEP09(2018)007, arXiv:1803.06553.

- [41] CMS Collaboration, “Performance of the CMS Level-1 trigger in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JINST* **15** (2020) P10017, doi:10.1088/1748-0221/15/10/p10017, arXiv:2006.10165.
- [42] S. Raychaudhuri, “Introduction to Monte Carlo simulation”, doi:10.1109/WSC.2008.4736059.
- [43] NNPDF Collaboration, “Parton distributions for the LHC run II”, *JHEP* **04** (2015) 040, doi:10.1007/JHEP04(2015)040, arXiv:1410.8849.
- [44] NNPDF Collaboration, “Parton distributions from high-precision collider data”, *Eur. Phys. J. C* **77** (2017) 663, doi:10.1140/epjc/s10052-017-5199-5, arXiv:1706.00428.
- [45] CMS Collaboration, “Event generator tunes obtained from underlying event and multiparton scattering measurements”, *Eur. Phys. J. C* **76** (2016) 155, doi:10.1140/epjc/s10052-016-3988-x, arXiv:1512.00815.
- [46] CMS Collaboration, “Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements”, *Eur. Phys. J. C* **80** (2020) doi:10.1140/epjc/s10052-019-7499-4, arXiv:1903.12179.
- [47] S. Agostinelli et al., “GEANT4—a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003) 250, doi:10.1016/S0168-9002(03)01368-8.
- [48] S. Abdullin et al., “The fast simulation of the CMS detector at LHC”, *J. Phys. Conf. Ser.* **331** (2011) 032049, doi:10.1088/1742-6596/331/3/032049.
- [49] CMS Collaboration, “Particle-flow reconstruction and global event description with the CMS detector”, *JINST* **12** (2017) P10003, doi:10.1088/1748-0221/12/10/P10003, arXiv:1706.04965.
- [50] Jennifer Chu, “Latest experiment at large hadron collider reports first results”, 2015. <https://news.mit.edu/2015/large-hadron-collider-first-results-1014>.
- [51] CMS Collaboration, “Precision measurement of the structure of the CMS inner tracking system using nuclear interactions”, *JINST* **13** (2018) P10034, doi:10.1088/1748-0221/13/10/P10034, arXiv:1807.03289.
- [52] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the gaussian-sum filter in the CMS tracker at the LHC”, *Journal of Physics G: Nuclear and Particle Physics* **31** (2005) N9, doi:10.1088/0954-3899/31/9/N01.
- [53] K. Rose, “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”, *Proceedings of the IEEE* **86** (1998) 2210, doi:10.1109/5.726788.
- [54] W. Waltenberger, R. Frühwirth, and P. Vanlaer, “Adaptive vertex fitting”, *Journal of Physics G: Nuclear and Particle Physics* **34** (2007) N343, doi:10.1088/0954-3899/34/12/N01.

- [55] CMS E/Gamma POG, “Cut based electron ID for Run 2 legacy”. Twiki:  
<https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2>.
- [56] K. Rehermann and B. Tweedie, “Efficient Identification of Boosted Semileptonic Top Quarks at the LHC”, *JHEP* **03** (2011) 059, doi:10.1007/JHEP03(2011)059, arXiv:1007.2221.
- [57] CMS Collaboration, “Search for supersymmetry in pp collisions at  $\sqrt{s} = 13$  TeV in the single-lepton final state using the sum of masses of large-radius jets”, *JHEP* **08** (2016) 122, doi:10.1007/JHEP08(2016)122, arXiv:1605.04608.
- [58] M. Cacciari, G. P. Salam, and G. Soyez, “The anti- $k_T$  jet clustering algorithm”, *JHEP* **04** (2008) 063, doi:10.1088/1126-6708/2008/04/063, arXiv:0802.1189.
- [59] CMS Collaboration, “Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV”, *JINST* **12** (2017) P02014, doi:10.1088/1748-0221/12/02/P02014, arXiv:1607.03663.
- [60] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011) P11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277.
- [61] A. Harel and P. Schieferdecker, “Jet Identification”. TWiki:  
<https://twiki.cern.ch/twiki/bin/view/CMS/JetID>.
- [62] CMS Collaboration, “Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV”, *JINST* **13** (2018) P05011, doi:10.1088/1748-0221/13/05/P05011, arXiv:1712.07158.
- [63] CMS Collaboration, “Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques”, *JINST* **15** (2020) P06005, doi:10.1088/1748-0221/15/06/P06005, arXiv:2004.08262.
- [64] D. Alves et al., “Simplified models for LHC new physics searches”, *J. Phys. G* **39** (2012) 105005, doi:10.1088/0954-3899/39/10/105005, arXiv:1105.2838.
- [65] J. Alwall et al., “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”, *JHEP* **07** (2014) 079, doi:10.1007/JHEP07(2014)079, arXiv:1405.0301.
- [66] S. Alioli, P. Nason, C. Oleari, and E. Re, “A general framework for implementing NLO calculations in shower monte carlo programs: the POWHEG box”, *JHEP* **06** (2010) 043, doi:10.1007/JHEP06(2010)043, arXiv:1002.2581.
- [67] ATLAS Collaboration, “Jet substructure measurements with CMS and ATLAS”,.

- [68] L. Gouskos, P. Kontaxakis, “DeepAK8 v2 scale factors and working points”, (2020).  
<https://indico.cern.ch/event/877167/contributions/3744193/>.
- [69] CMS Collaboration, “Measurement of differential cross sections for the production of top quark pairs and of additional jets in lepton+jets events from pp collisions at  $\sqrt{s} = 13$  TeV”, *Phys. Rev. D* **97** (2018) 112003, doi:10.1103/PhysRevD.97.112003, arXiv:1803.08856.
- [70] CMS Collaboration, “Measurement of the production cross section for single top quarks in association with W bosons in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, *JHEP* **10** (2018) 117, doi:10.1007/JHEP10(2018)117, arXiv:1805.07399.
- [71] CMS Collaboration, “Measurements of the  $pp \rightarrow WZ$  inclusive and differential production cross section and constraints on charged anomalous triple gauge couplings at  $\sqrt{s} = 13$  TeV”, *JHEP* **04** (2019) 122, doi:10.1007/JHEP04(2019)122, arXiv:1901.03428.
- [72] CMS Collaboration, “Measurements of  $pp \rightarrow ZZ$  production cross sections and constraints on anomalous triple gauge couplings at  $\sqrt{s} = 13$  TeV”, *Eur. Phys. J. C* **81** (2020) doi:10.1140/epjc/s10052-020-08817-8, arXiv:2009.01186.
- [73] CMS B-tag POG, “Methods to apply b-tagging efficiency scale factors 1(a)”. TWiki: [https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods#1a\\_Event\\_reweighting\\_using\\_scale](https://twiki.cern.ch/twiki/bin/view/CMS/BTagSFMethods#1a_Event_reweighting_using_scale).
- [74] ATLAS and CMS Collaborations, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11, 2011. <https://cds.cern.ch/record/1379837>.
- [75] ATLAS and CMS Collaborations, “Snowmass white paper contribution: Physics with the Phase-2 ATLAS and CMS detectors”, Technical Report ATL-PHYS-PUB-2022-018, CMS-PAS-FTR-22-001, 2022. <https://cds.cern.ch/record/2805993>.
- [76] CMS Collaboration, “Combined search for electroweak production of winos, binos, higgsinos, and sleptons in proton-proton collisions at  $\sqrt{s} = 13$  TeV”, CMS Physics Analysis Summary CMS-PAS-SUS-21-008, 2023.
- [77] J. Debove, B. Fuks, and M. Klasen, “Threshold resummation for gaugino pair production at hadron colliders”, *Nucl. Phys. B* **842** (2011) 51, doi:10.1016/j.nuclphysb.2010.08.016, arXiv:1005.2909.
- [78] M. Klasen, B. Fuks, and M. Sunder, “Precision predictions for associated gluino-gaugino production at the LHC”, *PoS EPS-HEP2017* (2017) 298, doi:10.22323/1.314.0298, arXiv:1709.02680.

- [79] S. Sekmen et al., “HL-LHC and HE-LHC searches for new physics in hadronic final states with boosted  $W$  bosons or top quarks using razor variables”, CMS Analysis Note CMS-AN-18-251 v15, 2019.
- [80] LHC SUSY Cross Section Working Group, “SUSY Cross Sections for 13, 14, 33 and 100 TeV pp Collisions”. TWiki: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/SUSYCrossSections>.