

ABSTRACT

Title of dissertation: Search for Pair Production of
Third-Generation Scalar Leptoquarks
and R-Parity Violating Top Squarks
in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV

Kevin Pedro, Doctor of Philosophy, 2014

Dissertation directed by: Professor Sarah C. Eno
Department of Physics

This dissertation presents a search for pair production of third-generation scalar leptoquarks or top squarks in R-parity violating supersymmetry decaying through the coupling λ'_{333} , with the new scalar particle decaying to a tau lepton and a bottom quark. The data used in this search were collected by the Compact Muon Solenoid experiment and comprise 19.7 fb^{-1} of proton-proton collision data from the Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 8$ TeV. No excess of events above the standard model background prediction is observed. The existence of third-generation scalar leptoquarks or top squarks with masses below 740 GeV is excluded at the 95% confidence level, assuming the branching fraction for the decay to a tau lepton and a bottom quark is 100%. This is currently the strongest limit on this kind of leptoquark or top squark. The search is extended to address a chargino-mediated decay of the top squark involving the R-parity violating coupling λ'_{3kj} , producing a tau lepton, a bottom quark, and two light quarks. The existence of top squarks in this model with masses below 580 GeV is excluded at the 95% confidence level. This is the first search for the pair production of top squarks with the decay involving the R-parity violating coupling λ'_{3kj} .

Search for Pair Production of Third-Generation Scalar Leptoquarks
and R-Parity Violating Top Squarks in Proton-Proton Collisions at
 $\sqrt{s} = 8 \text{ TeV}$

by

Kevin Pedro

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Sarah C. Eno, Chair/Advisor
Professor Nick Hadley
Assistant Professor Alberto Belloni
Associate Professor Zackaria Chacko
Professor Alice Mignerey, Dean's Representative

© Copyright by
Kevin Pedro
2014

Dedication

To my parents, Philip and Lisa

Acknowledgments

In today's world of massive experiments and collaborations, every physicist owes his results in part to the hard work of many others. For me, the first among those is my advisor, Professor Sarah Eno. Her advice and instruction have been invaluable, not only for improving my understanding of the intricacies of high energy physics, but also for navigating the communities and bureaucracies of the University of Maryland, the CMS collaboration, Fermilab, and CERN. I have been able to excel in the CMS collaboration in large part because of Sarah's guidance and suggestions when choosing projects and solving problems.

I am grateful to my analysis partners Matthieu Marionneau and Ketino Kaadze for their invaluable contributions to the analysis presented in this dissertation. I learned a great deal from both of them about data analysis and high energy physics. It was a superb experience to be part of a group in which everyone treated each other as equals and everyone took responsibility for their share of the work. I must also thank Youngdo Oh and Anirban Saha for generating the top squark signal samples, and the various members of the Exotica group for their many helpful comments.

I have learned and grown as a physicist while at the University of Maryland, thanks to the encouragement and support of the CMS group here. My gratitude goes to Professors Alberto Belloni, Nick Hadley, Drew Baden, Andris Skuja, and Tom Ferbel; postdocs Matthieu Marionneau, Ted Kolberg, Jeff Temple, and Ellie Twedt; scientist Dick Kellogg; and engineer Tom O'Bannon. I may never recapture the magic of this group of CMS grad students: Brian Calvert, Young Ho Shin, and

Chris Anelli. I also thank Marguerite Tonjes, Brian Calvert, and Chris Ferraioli for keeping our Tier 3 cluster running, and former grad student Dinko Ferenčec for his help in preparing this dissertation. I am grateful to my dissertation committee members for volunteering their time and effort to help me improve this dissertation.

This section is too narrow to contain the full list of CMS collaborators who have aided me on the path to my doctoral degree. I have been privileged to work on many areas of CMS, including trigger operations, HCAL, jets, long-term upgrades, fast simulation, and offline software. My thanks go to all of my CMS colleagues from these groups for their help and support. I also thank the many professors, teaching assistants, and administrators in the University of Maryland physics department for their help with the courses and logistics of graduate student life.

Socially and personally, graduate school has been an unexpected delight for me. This is largely due to my housemates in Physics House, official and honorary, old and new: Nat Steinsultz, Joe Garrett, George Hine, Kiersten Ruisard, Meredith Lukow, Ginny Garrett, Neil Anderson, and Lexi Parsagian. I will dearly miss all of them when I leave Maryland. I extend that gratitude to the many friends I have made at both Maryland and CERN during my graduate career.

Finally, I thank my family – Pedros, Gravels, Mougalian, Wrights – for the myriad ways they have supported me during my time as a graduate student. This dissertation is dedicated to my father Philip, from whom I inherited a love of science and technology, and to the memory of my mother Lisa, whose love and support will remain with me throughout my life.

Table of Contents

Table of Contents	v
List of Tables	vii
List of Figures	xi
List of Abbreviations	xvi
1 Introduction	1
2 Theoretical Motivations	5
2.1 The Standard Model	5
2.2 Beyond the Standard Model	10
2.3 Leptoquarks	13
2.4 R-Parity Violating Supersymmetry	20
3 Compact Muon Solenoid Experiment	31
3.1 The Large Hadron Collider	32
3.2 Tracker	35
3.3 Electromagnetic Calorimeter	37
3.4 Hadron Calorimeter	40
3.5 Solenoid	44
3.6 Muon System	45
3.7 Trigger	48
3.8 Luminosity Measurement	51
4 Event Reconstruction	53
4.1 Event Generation	55
4.2 Detector Simulation	58
4.3 Tracks and Vertices	59
4.4 Particle Flow	62
4.5 Electrons	65
4.6 Muons	67

4.7	Jets	69
4.8	Taus	72
4.9	b-tagging	76
4.10	Missing Transverse Energy	79
5	Data Analysis	81
5.1	Data Samples	81
5.1.1	Data	81
5.1.2	Monte Carlo	82
5.2	Selection and Optimization	86
5.2.1	Preselection	86
5.2.2	Main Selection	88
5.2.3	Final Selections	91
5.3	Background Estimations	96
5.3.1	$t\bar{t}$ Irreducible Background	97
5.3.2	Major Reducible Background	106
5.3.2.1	QCD Multijet Reducible Background	117
5.3.3	Other Backgrounds	126
5.4	Systematic Uncertainties	127
5.5	Results	131
6	Conclusions	139
A	Full CLs Shape-Based Limits	141
B	Event Displays	144
B.1	Leptoquark Search	144
B.2	Top Squark Search	147
C	Table of Monte Carlo Datasets	150
D	CMS Collaboration	153
	Bibliography	154

List of Tables

2.1	The quantum numbers of each category of fermions, based on chirality and particle type: up-type quarks, down-type quarks, charged leptons, and neutrinos. The various flavors of each category, also called the first, second, and third generations of matter, possess the same quantum numbers and differ only in their masses.	10
2.2	The quantum numbers of the different types of scalar and vector leptoquarks in the mBRW model.	16
2.3	The pair production cross sections for a range of scalar leptoquark masses at $\sqrt{s} = 8 \text{ TeV}$. Theoretical uncertainties from the PDFs and from varying the factorization/renormalization scale μ from $\mu = M_{\text{LQ}}/2$ to $\mu = 2M_{\text{LQ}}$ are indicated.	19
4.1	The sequence of tracking iterations used during the 2012 run, including information on the seeds and selection criteria used in each step [67].	61
4.2	The quality cuts for the medium and loose working points of the electron identification.	67
4.3	The quality cuts for the tight and loose working points of the muon identification.	69
4.4	The quality cuts for the loose working point of the jet identification.	72
4.5	The leading hadronic decays of tau leptons, including branching fractions and intermediate hadronic resonances [8]. The symbol h^- can be either π^- or K^-	73
4.6	The working points for the different tau discriminators used in the tau lepton identification.	76
5.1	The list of data samples used in each channel.	83
5.2	A summary of the generation of signal samples for the chargino-mediated λ'_{3jk} decay of the top squark. The chargino mass is determined by the value of μ	85

5.3	The simulated background and observed event yields after the preselection in the $e\tau_h$ and $\mu\tau_h$ channels. The statistical uncertainties are given for each simulated background. The contribution from the QCD multijets process is not taken into account in the total background yield.	88
5.4	The simulated background and observed event yields after the main selection in the $e\tau_h$ and $\mu\tau_h$ channels. The statistical uncertainties are given for each simulated background. The contribution from the QCD multijets process is not taken into account in the total background yield.	94
5.5	Acceptances (in %) for the $t\bar{t}$ process in each leptonic decay channel of the W boson produced by the top quark decays. Variations in the acceptances are mostly due to differences in the p_T distributions for each type of lepton.	99
5.6	Lepton identification efficiencies ε^{ID} measured in simulated $t\bar{t}$ events and associated scale factors ρ measured in Z events.	100
5.7	The relative selection efficiencies (in %) and data/MC scale factors ρ for the $t\bar{t}$ process in each channel $e\tau_h$, $\mu\tau_h$, and $e\mu$. The total selection efficiency is the product of all the relative selection efficiencies for a given search.	102
5.8	The estimated $t\bar{t}$ irreducible yields after the leptoquark final selection, from: direct simulation, closure test from the $e\mu$ simulation, and calculation from the observed $e\mu$ data. In the direct simulation, only the statistical uncertainty is given. In the closure test and the data result, the first uncertainty value corresponds to the statistical uncertainty in the simulation and observed data, while the second uncertainty value corresponds to the propagation of the uncertainties in the acceptances, efficiencies and scale factors.	105
5.9	The estimated $t\bar{t}$ irreducible yields after the top squark final selection, from: direct simulation, closure test from the $e\mu$ simulation, and calculation from the observed $e\mu$ data. In the direct simulation, only the statistical uncertainty is given. In the closure test and the data result, the first uncertainty value corresponds to the statistical uncertainty in the simulation and observed data, while the second uncertainty value corresponds to the propagation of the uncertainties in the acceptances, efficiencies and scale factors.	106
5.10	The results of the data-driven major reducible background estimation for the leptoquark search in the $\mu\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.	117

5.11	The results of the data-driven major reducible background estimation for the top squark search in the $\mu\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.	118
5.12	The results of the data-driven major reducible background estimation for the leptoquark search in the $e\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.	119
5.13	The results of the data-driven major reducible background estimation for the top squark search in the $e\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.	120
5.14	The relative systematic uncertainties on the yields of the simulated signal and the simulated backgrounds. The magnitude of the effects may vary within indicated ranges for different signal masses, for different background processes within a given category, or for the different searches.	128
5.15	The relative systematic uncertainties on the yields of the major backgrounds estimated from the observed data for each channel of each search.	128
5.16	The estimated backgrounds, observed event yields, and expected number of signal events for the leptoquark search. For the simulation-based entries, the statistical and systematic uncertainties are shown separately, in that order. Only the systematic uncertainties are shown for the data-driven background entries.	133
5.17	The estimated backgrounds, observed event yields, and expected number of signal events for the top squark search. For the simulation-based entries, the statistical and systematic uncertainties are shown separately, in that order. Only the systematic uncertainties are shown for the data-driven background entries.	133
5.18	Selection efficiencies in % for the signal in the leptoquark search, estimated from the simulation.	133
5.19	Selection efficiencies in % for the signal in the top squark search, estimated from the simulation.	134

B.1	The kinematic properties of the selected particles for the two highest S_T observed events in the $\mu\tau_h$ channel of the leptoquark search. . . .	144
B.2	The kinematic properties of the selected particles for the two highest S_T observed events in the $\mu\tau_h$ channel of the top squark search. . . .	147
C.1	The full list of MC samples for the LQ signal used in the analysis. Each dataset name should be followed by <i>/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</i> . Cross sections σ are given at the specified perturbative order in α_s for each process.	151
C.2	The full list of MC samples for the SM background used in the analysis. Each dataset name should be followed by <i>/Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM</i> . Cross sections σ are given at the specified perturbative order in α_s for each process.	152

List of Figures

2.1	A table of all the elementary particles in the standard model, with the spin, electric charge, and mass values of each particle [1]. The faint gray lines indicate which gauge bosons interact with which fermions.	6
2.2	A diagram illustrating the leading order interactions between particles in the standard model, including self-interactions [4].	11
2.3	The LO Feynman diagrams for leptoquark pair production from quark-antiquark annihilation (top) and gluon-gluon fusion (middle, bottom).	17
2.4	The LO diagrams for single leptoquark production from quark-gluon scattering. The LQ is produced in association with a lepton.	18
2.5	One-loop diagrams for contributions to the Higgs mass parameter m_{H}^2 from a fermion f (left) and a scalar S (right) [30].	22
2.6	Summary of CMS exclusion limits for the masses of various SUSY particles with $\sqrt{s} = 8$ TeV [36]. These limits assume simplified models and unity branching fractions for the specified decays. Two scenarios are presented: the dark shades show $m(\text{LSP}) = 0$ and the light shades show $m(\text{mother}) - m(\text{LSP}) = 0$	28
2.7	The LO diagrams for the three trilinear RPV couplings λ (left), λ' (middle), and λ'' (right) [31]. These couplings violate lepton number, lepton number, and baryon number, respectively.	29
3.1	The layout of the CMS detector, with the subdetectors labeled and two humans shown for a height reference.	32
3.2	A diagram of the LHC with the major experiments labeled [41].	33
3.3	A diagram of the CERN accelerators which form the LHC injector [42].	34
3.4	A diagram of an LHC dipole magnet, with the major components labeled [43].	35
3.5	The layout of the CMS tracker, with subsystems labeled.	36
3.6	A diagram of the CMS ECAL, with subsystems and η ranges labeled.	38
3.7	The layout of the HCAL subsystems HB, HE, HO, and HF in the CMS detector.	41
3.8	A diagram of the depth segmentation scheme in the HB and the HE.	44
3.9	An artistic rendering of the CMS solenoid, showing the five rings placed inside the cryostat, along with the support structure.	45

3.10	The layout of the muon system, with the three subsystems labeled.	46
3.11	The architecture of the L1 Trigger.	50
4.1	A cross-sectional view of the CMS detector with all subdetectors labeled and examples of signals left by muons, electrons, charged hadrons, neutral hadrons, and photons [54].	54
4.2	PDFs calculated at NLO by MSTW, plotted against momentum fraction x for two different values of the momentum scale Q^2 [56].	56
4.3	An illustration of a proton-proton collision, showing the hard scattering, ISR, FSR, and the underlying event [57].	58
4.4	An example of jets reconstructed with the anti- k_T algorithm, demonstrating the characteristic circular shape [80].	71
4.5	A simple diagram of the different hadronic tau decay mode topologies reconstructed by the HPS algorithm [87].	75
5.1	A plot of the total integrated luminosity over time for the 2012 run of the LHC, showing the luminosity delivered by the LHC (blue), recorded by the CMS detector (dark orange), and validated as good for physics analysis (light orange) [92].	82
5.2	Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $e\tau_h$ channel after the preselection: the visible mass of the electron and hadronic tau (top left), the number of b-tagged jets (top right), the p_T (left) and η (right) of the electron (middle) and hadronic tau (bottom). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.	89
5.3	Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $\mu\tau_h$ channel after the preselection: the visible mass of the muon and hadronic tau (top left), the number of b-tagged jets (top right), the p_T (left) and η (right) of the muon (middle) and hadronic tau (bottom). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.	90
5.4	Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $e\tau_h$ channel after the main selection: the p_T spectra (top) of the first (left) and second (right) selected jets, $M(\tau_h, \text{jet})$ (bottom left), and S_T (bottom right). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.	92
5.5	Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $\mu\tau_h$ channel after the main selection: the p_T spectra (top) of the first (left) and second (right) selected jets, $M(\tau_h, \text{jet})$ (bottom left), and S_T (bottom right). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.	93

5.6	$M(\tau_h, \text{jet})$ (left) and N_{jets} (right) before the respective final selection cuts on these variables, with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for leptoquarks with a mass of 500 GeV (left) or top squarks with a mass of 300 GeV (right) is added on top of the background prediction. Both plots include the cut $p_T(\tau_h) > 50$ GeV and the major backgrounds are estimated from observed data.	95
5.7	The S_T distribution in the $e\mu$ control region after the leptoquark final selection. The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.	98
5.8	Comparison of observed data and simulation in the $Z \rightarrow \mu\mu + \text{jets}$ control region for $M_{\mu\mu}$ (left) and multiplicity of extra jets, which are not misidentified as hadronic taus (right).	108
5.9	The fake rates measured in the $Z \rightarrow \mu\mu + \text{jets}$ control region, using all events (left), only events with one or more extra jets (middle), only events with two or more extra jets (right).	109
5.10	Comparison of the fake rates between simulated $Z \rightarrow \mu\mu + \text{jets}$ events with at least one extra jet and simulated semileptonic $t\bar{t}$ events (left). The difference in the two fake rates is found to be 18%, independent of the hadronic tau p_T (right).	110
5.11	Plots of the anti-isolated control region after the leptoquark final selection, showing the hadronic tau p_T spectrum (left) and multiplicity (right) in the $\mu\tau_h$ channel (top) and the $e\tau_h$ channel (bottom).	112
5.12	Plots of the anti-isolated control region after the top squark final selection, showing the hadronic tau p_T spectrum (left) and multiplicity (right) in the $\mu\tau_h$ channel (top) and the $e\tau_h$ channel (bottom).	113
5.13	The transverse mass of the electron and missing transverse energy system for the same-sign anti-isolated control region, after the leptoquark final selection. The overall excess in the observed data, not localized in a specific range of M_T values, indicates the presence of QCD events.	123
5.14	The missing transverse energy spectrum for the same-sign control region selected before the $M(\tau_h, \text{jet}) > 250$ GeV requirement for the leptoquark search. The excess of observed events at low \cancel{E}_T is an indication of the presence of QCD events.	125
5.15	The QCD S_T distribution estimated by subtracting the simulated distribution from the observed distribution in the same-sign control region after the leptoquark final selection. The negative values are set to zero when using the distribution.	126
5.16	The final S_T distribution for the leptoquark search with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for leptoquarks with $M_{LQ} = 500$ GeV is added on top of the background prediction. The last bin contains the overflow events. The horizontal bar on each observed data point indicates the width of the bin in S_T	134

5.17	The final S_T distribution for the top squark search with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for top squarks with $M_{\tilde{t}} = 300$ GeV is added on top of the background prediction. The last bin contains the overflow events. The horizontal bar on each observed data point indicates the width of the bin in S_T	135
5.18	The expected and observed combined upper limits on the third-generation LQ pair production cross section σ times the square of the branching fraction, \mathcal{B}^2 , at the 95% CL, as a function of the LQ mass. These limits also apply to top squarks decaying directly via the coupling λ'_{333} . The green (darker) and yellow (lighter) uncertainty bands represent the 68% and 95% CL intervals on the expected limit. The dark blue curve and the hatched light blue band represent the theoretical LQ pair production cross section, assuming $\mathcal{B} = 100\%$, and the uncertainties due to the choice of PDF and renormalization/factorization scales.	136
5.19	The expected and observed combined upper limits on the top squark pair production cross section σ times the square of the branching fraction, \mathcal{B}^2 , at the 95% CL, as a function of the top squark mass. These limits apply to top squarks with a chargino-mediated decay through the coupling λ'_{3kj} . The green (darker) and yellow (lighter) uncertainty bands represent the 68% and 95% CL intervals on the expected limit. The dark blue curve and the hatched light blue band represent the theoretical top squark pair production cross section, assuming $\mathcal{B} = 100\%$, and the uncertainties due to the choice of PDF and renormalization/factorization scales.	137
5.20	The expected (dashed black) and observed (green solid) 95% CL upper limits on the branching fraction for the leptoquark decay to a tau lepton and a bottom quark, as a function of the leptoquark mass. A search for top squark pair production [119] has the same kinematic signature as the leptoquark decay to a tau neutrino and a top quark. This search is reinterpreted to provide the expected (blue hatched) and observed (blue open) 95% CL upper limits for low values of \mathcal{B} , assuming the leptoquark only decays to third-generation fermions. . .	138
A.1	Comparison of the observed value (red line) to the probability densities for H_0 (background only, blue line) and H_1 (signal + background, brown line) as a function of the log likelihood ratio. Green area: CL_{s+b} , yellow area: $1 - CL_b$. From [116].	143

B.1	A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the highest S_T observed event in the $\mu\tau_h$ channel of the leptoquark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.	145
B.2	A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the second-highest S_T observed event in the $\mu\tau_h$ channel of the leptoquark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.	146
B.3	A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the highest S_T observed event in the $\mu\tau_h$ channel of the top squark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.	148
B.4	A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the second-highest S_T observed event in the $\mu\tau_h$ channel of the top squark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.	149

List of Abbreviations

AOD	Analysis Object Data
APD	Avalanche Photodiode
ASIC	Application-Specific Integrated Circuits
ATLAS	A Toroidal LHC ApparatuS
BDT	Boosted Decision Tree
BRW	Buchmüller-Rückl-Wyler
BSM	Beyond Standard Model
CERN	European Organization for Nuclear Research
CH	Charged Hadron
CKM	Cabibbo-Kobayashi-Maskawa
CL	Confidence Level
CMS	Compact Muon Solenoid
CP	Charge-Parity
CPU	Central Processing Unit
CSC	Cathode Strip Chamber
CSCTF	Cathode Strip Chamber Track Finder
CTEQ	Coordinated Theoretical-Experimental Project on QCD
CTF	Combinatorial Track Finder
DAS	Data Aggregation System
DGLAP	Dokshitzer-Gribov-Lipatov-Altarelli-Parisi
DT	Drift Tube
DTTF	Drift Tube Track Finder
EB	ECAL Barrel
ECAL	Electromagnetic Calorimeter
EE	ECAL Endcap
EM	Electromagnetic
ES	ECAL Preshower
EWSB	Electroweak Symmetry Breaking
FCNC	Flavor-Changing Neutral Current
FPGA	Field-Programmable Gate Array
FSR	Final-State Radiation
GSF	Gaussian Sum Filter
GUT	Grand Unified Theory
HB	HCAL Barrel
HCAL	Hadron Calorimeter
HE	HCAL Endcap
HF	HCAL Forward
HO	HCAL Outer
HPD	Hybrid Photodiode
HLT	High-Level Trigger
IP	Interaction Point
ISR	Initial-State Radiation
L1	Level 1
L1A	Level-1 Accept

LEP	Large Electron-Positron Collider
LHC	Large Hadron Collider
LHCb	Large Hadron Collider beauty
LO	Leading Order
LQ	Leptoquark
LSP	Lightest Supersymmetric Particle
LUT	Lookup Table
mBRW	minimal Buchmüller-Rückl-Wyler
MB	Muon Barrel
MC	Monte Carlo
ME	Muon Endcap
MET	Missing Transverse Energy
MIP	Minimum Ionizing Particle
MSSM	Minimal Supersymmetric Model
MSTW	Martin-Stirling-Thorne-Watt
MVA	Multivariate
NH	Neutral Hadron
NLO	Next-to-Leading Order
NNLO	Next-to-Next-to-Leading Order
NNLL	Next-to-Next-to-Leading Logarithmic
PD	Primary Dataset
PF	Particle Flow
PDF	Parton Distribution Function
PMNS	Pontecorvo-Maki-Nakagawa-Sakata
PMT	Photomultiplier Tube
PS	Proton Synchrotron
PSB	Proton Synchrotron Booster
PU	Pileup
QED	Quantum Electrodynamics
QCD	Quantum Chromodynamics
RF	Radio Frequency
RG	Renormalization Group
RMS	Root Mean Square
RPC	Resistive Plate Chamber
RPC	R-Parity Conserving
RPV	R-Parity Violating
SiPM	Silicon Photomultiplier
SM	Standard Model
SPS	Super Proton Synchrotron
SS/OS	Same-Sign/Opposite-Sign
SUSY	Supersymmetry
TCS	Trigger Control System
TEC	Tracker End Cap
TIB	Tracker Inner Barrel
TID	Tracker Inner Disks

TOB	Tracker Outer Barrel
TPG	Trigger Primitive Generator
TTC	Timing, Trigger and Control
VEV	Vacuum Expectation Value
VPT	Vacuum Phototriode
WIMP	Weakly Interacting Massive Particle
WLS	Wavelength-Shifting

Chapter 1: Introduction

Just over a century ago, Ernest Rutherford and Niels Bohr developed a revolutionary model of the atom: a nucleus containing positively charged protons and electrically neutral neutrons, surrounded by negatively charged electrons in quantized orbitals. This model and other related models, along with the empirical results which motivated them, led to the development of quantum mechanics to describe physics at the subatomic level. As physicists investigated further, accelerating particles closer and closer to the speed of light in order to reach higher energies and smaller distances, it became necessary to merge the formulations of quantum mechanics with Einstein's special relativity to develop quantum field theories. The results of decades of these experimental tests are collected in the unified framework of the standard model of particle physics. The standard model describes all observed elementary particles and three of the fundamental forces, including electromagnetism and the weak and strong nuclear interactions. The weak nuclear interaction is responsible for most radioactivity and for nuclear fusion, while the strong force holds quarks and gluons together as protons and neutrons, and residually holds protons and neutrons together as atomic nuclei.

The last missing piece of the standard model, the Higgs boson, was discovered

in 2012 by the Compact Muon Solenoid experiment and the ATLAS experiment at the Large Hadron Collider. This collider and the several detectors placed around it are the largest scientific experiment yet undertaken by humanity. Two beams of protons are each accelerated to 99.999997% of the speed of light, achieving an energy level never before produced in the laboratory, and then collided together. The constituent quarks and gluons of the protons interact with each other through the fundamental forces, producing any of the particles in the standard model. Just as a radioactive element will decay into a different element by emitting radiation, the produced particles decay into the lightest stable particles, which are detected by the Compact Muon Solenoid. Each subsystem of the detector is optimized to measure certain types of particles, including photons, electrons, muons, charged hadrons such as protons, and neutral hadrons such as neutrons. The presence of non-interacting particles such as neutrinos can be inferred by balancing the momentum measured by the detector in each collision event. The unprecedented energy scale and collision rate of these experiments enabled the discovery of the Higgs boson, the first observed scalar elementary particle. The Higgs boson is an excitation of the Higgs field, which was postulated and is now confirmed to provide masses to the various elementary particles.

However, there are some phenomena which remain beyond the capability of the standard model to explain. This dissertation presents a search for several types of new particles, leptoquarks and top squarks, which are predicted by theories that might supersede the standard model. Leptoquarks are bosons which have properties of both leptons (such as electrons) and quarks (such as the constituents of

protons and neutrons). The existence of leptoquarks could indicate new relationships between the leptons and quarks in the standard model. James Maxwell unified electricity and magnetism into electromagnetism, and similarly, Abdus Salam, Sheldon Glashow, and Steven Weinberg unified electromagnetism and the weak nuclear interaction into the electroweak interaction. Leptoquarks could point the way to possible solutions for the grand unification of all three fundamental forces. The theory of supersymmetry postulates a partner for each standard model particle, such as the top squark, which is the supersymmetric partner of the top quark, a heavy version of the up quarks contained in protons and neutrons. As the heaviest elementary particle, the top quark is the most likely to interact with the Higgs boson. The discovery of its supersymmetric partner the top squark would indicate how that interaction is balanced to produce the relatively light Higgs boson that has been observed.

In this dissertation, a search is conducted for scalar leptoquarks coupling to the third generation of elementary particles, including tau leptons and bottom quarks. The search for top squarks considers R-parity violating supersymmetry, in order to get around experimental constraints on R-parity conserving supersymmetry. The R-parity violating model eliminates a proposed separation between standard model and supersymmetric particles, which allows the supersymmetric particles to decay to final states involving only standard model particles. There are two cases considered for top squarks. In the first case, the top squarks decay identically to third-generation scalar leptoquarks. In the second case, the top squarks decay differently, producing a final state similar to the first case: a tau lepton, a bottom quark, and

two light quarks. The existence of third-generation scalar leptoquarks and the first case of top squarks is already excluded if their mass is below values probed by earlier searches, but the high energy of the 2012 run of the Large Hadron Collider will extend the experimental reach of this search. Previously, no direct search for the second case of top squarks had ever been performed.

Chapter 2: Theoretical Motivations

2.1 The Standard Model

The standard model (SM) of particle physics includes three fundamental forces and all of the particles of ordinary matter using a locally gauge-invariant quantum field theory framework. The three fundamental forces are electromagnetism, the weak force, and the strong force. These forces are carried by spin-1 gauge bosons, while the matter particles consist of quarks and leptons, two categories of spin-1/2 fermions. When these forces are weak enough, quantum field theory calculations use a perturbative method in which the leading order (LO) term is calculated, then the next-to-leading order (NLO) correction is added, then the next-to-next-to-leading order correction (NNLO), and so forth. In addition, the standard model contains a scalar spin-0 boson, the Higgs boson, which is part of the Higgs field that provides masses to certain gauge bosons and fermions. Figure 2.1 summarizes the particles in the standard model, including the spin, electric charge, and mass values of each particle. The standard model and quantum field theory are described in more detail in many standard textbooks.

The electromagnetic force is mediated by photons (γ), spin-1 gauge bosons which have no mass or electric charge Q . This force causes interactions between

10^{-18} m. The charged current weak interaction, mediated by the W^\pm bosons, is sensitive to the chirality of fermions; only left-handed fermions and right-handed antifermions participate in this interaction. The neutral current weak interaction, mediated by the Z boson, acts on fermions of all chiralities, with different coupling strengths depending on the chirality.

As suggested by the inclusion of both forces in the previous paragraph, the electromagnetic and weak forces can be unified to form the electroweak force, represented by the symmetry group $SU(2)_L \times U(1)_Y$. In this unification, the quantum numbers of electromagnetism and the weak force are related by a new conserved quantum number, weak hypercharge $Y = 2(Q - T_3)$. The Higgs mechanism is responsible for electroweak symmetry breaking (EWSB). In order for the electroweak theory to be gauge invariant, the gauge bosons must be massless, but the W^\pm and Z bosons are observed to have mass. The Higgs mechanism solves this dilemma via spontaneous EWSB due to its non-zero vacuum expectation value (VEV). The Higgs field consists of a doublet, with two charged particles and two neutral particles, all scalar bosons. The two charged particles and one of the neutral particles act as Goldstone bosons, combining with the W^\pm and Z bosons to produce their masses. The remaining neutral particle is the Higgs boson, which was discovered at the LHC in 2012 [2, 3].

The strong force, quantum chromodynamics (QCD), is mediated by gluons (g) and can be represented by an $SU(3)$ symmetry group. Gluons, like photons, are spin-1 gauge bosons without mass or electric charge. However, gluons do possess color charge, the quantum number on which the strong force acts. Color charge

is so named because the charge has three possible values, which are labeled red, green, or blue. The strong force between quarks does not decrease as they become spatially separated, a phenomenon known as confinement. The energy in the gluon field between the separated quarks can become large enough to form one or more quark-antiquark pairs, which prevents quarks or gluons from existing in a bare state. Correspondingly, the range of the strong force is limited to $\sim 10^{-15}$ m. Quarks and gluons are always observed in nature as bound states called hadrons, and the formation of those bound states is called hadronization. States with one quark and one antiquark are mesons, while states with three quarks are baryons. Mesons and baryons are the two allowed types of bound states because they represent color singlets. Complementarily, as quarks get closer together, the strong force between them weakens. This behavior is known as asymptotic freedom; because short distances are equivalent to high energies, the strong interactions of quarks at a high-energy collider like the LHC can be calculated perturbatively. A residual form of the strong force acts on nucleons, protons and neutrons, to form atomic nuclei.

As mentioned, fermions are the particles of matter, which are separated into two groups: quarks and leptons. Quarks have fractional electric charge, weak isospin, and color charge, so they are affected by all three fundamental forces. There are two types of quarks: up-type quarks that have $Q = 2/3$ and down-type quarks that have $Q = -1/3$. Leptons consist of charged leptons and neutrinos. Charged leptons possess electric charge and weak isospin, while neutrinos only possess weak isospin. In total, there are six flavors of leptons and six flavors of quarks, arranged into three generations. The flavors of up-type quarks are the up, charm, and top

quarks; the flavors of down-type quarks are the down, strange, and bottom quarks; the flavors of charged leptons are the electron, muon, and tau lepton; and the flavors of the neutrinos are the electron, muon, and tau neutrinos. The top quark is the heaviest elementary particle and is so heavy that it decays before hadronizing, making it an exception to the rule that quarks are only observed in bound states. Quarks possess an additively conserved quantum number called baryon number B , which is defined as $B = \frac{1}{3}(n_q - n_{\bar{q}})$. The conservation of baryon number results in the stability of the proton, as it is the lightest baryon. Similarly, lepton number L is defined for leptons as $L = n_\ell - n_{\bar{\ell}}$. Specific lepton flavor numbers L_e , L_μ , L_τ are defined for each flavor pair of leptons. Baryon number and lepton number are accidental symmetries of the standard model, as only higher-dimensional terms excluded from the SM Lagrangian would break them.

The fermions are arranged into multiplets based on their chirality. The left-handed up- and down-type quarks are grouped together in a weak doublet q_L , as are the left-handed charged leptons and neutrinos in ℓ_L . The right-handed particles are weak singlets. It is important to note that right-handed neutrinos, and correspondingly left-handed antineutrinos, do not exist in the standard model. The Higgs field spontaneously provides masses to the quarks and charged leptons through a Yukawa interaction which couples the left- and right-handed versions of each flavor of particle. For a fermion f , this interaction takes the form $-y_f \bar{f}_L H f_R$, where y_f is the Yukawa coupling. The mass eigenstates formed by this interaction are mixtures of the weak eigenstates, leading to flavor-changing charged weak currents for the quarks, with the amount of mixing between any two flavors given by the unitary

Cabibbo-Kobayashi-Maskawa (CKM) matrix. The quantum numbers of each type of particle are summarized in Table 2.1, and the interactions among all the particles are illustrated in Fig. 2.2.

	Particle	Q	T_3	Y	B	L
Quarks	$q_L = \begin{pmatrix} u \\ d \end{pmatrix}_L$	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	1/3	1/3	0
	u_R	2/3	0	4/3	1/3	0
	d_R	-1/3	0	-2/3	1/3	0
Leptons	$\ell_L = \begin{pmatrix} \nu \\ e \end{pmatrix}_L$	$\begin{pmatrix} 0 \\ -1 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	-1	0	1
	e_R	-1	0	-2	0	1

Table 2.1: The quantum numbers of each category of fermions, based on chirality and particle type: up-type quarks, down-type quarks, charged leptons, and neutrinos. The various flavors of each category, also called the first, second, and third generations of matter, possess the same quantum numbers and differ only in their masses.

2.2 Beyond the Standard Model

The predictions of the standard model have been confirmed by decades of precise experimental tests. However, as an effective field theory, its domain of applicability is ultimately limited; at a high enough energy, the theory will break down. Further, some laboratory and cosmological observations are difficult to accommodate in the standard model. These limitations and indications of new phenomena, as well as aesthetic considerations, motivate various searches for physics beyond the standard model (BSM), including the searches which will be presented in this dissertation.

Unless unnatural fine-tuning occurs, the calculation of the Higgs mass produces

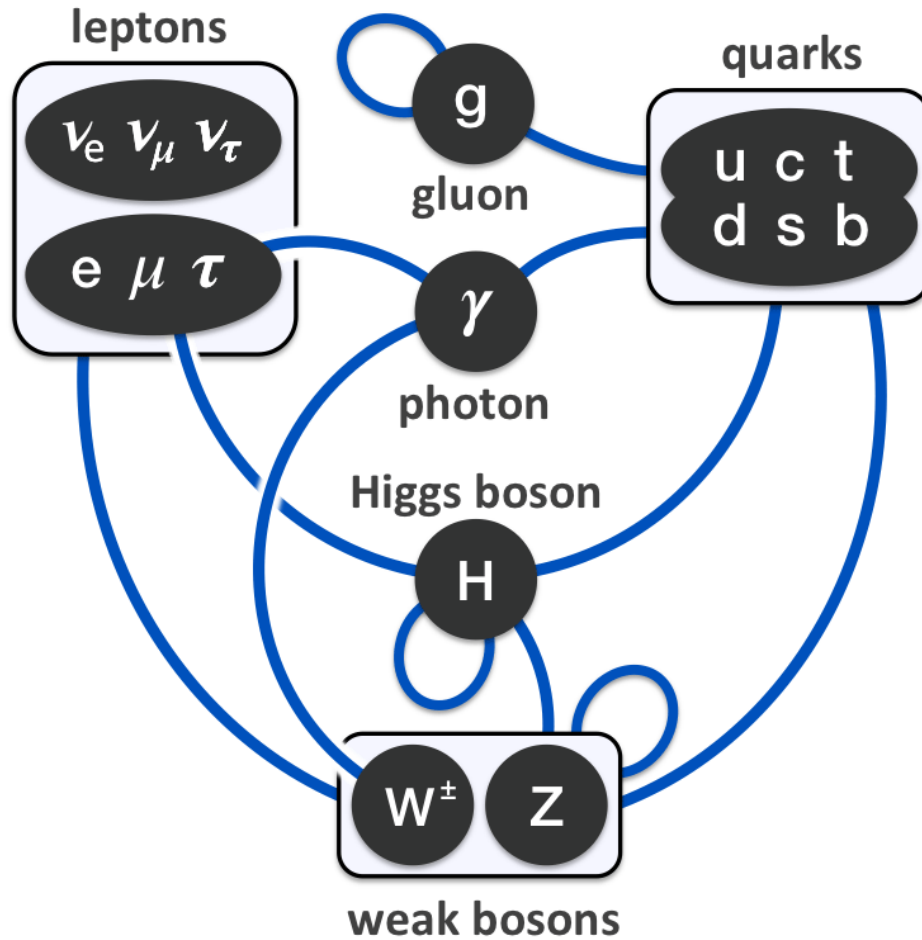


Figure 2.2: A diagram illustrating the leading order interactions between particles in the standard model, including self-interactions [4].

a value near the Planck scale of 10^{19} GeV, orders of magnitude higher than the observed value of 125 GeV, which characterizes the electroweak scale [5]. This is known as the hierarchy problem. To construct a natural theory which avoids such fine-tuning while making predictions that agree with observation, it is necessary to cancel divergent contributions to the Higgs mass. The hierarchy problem is an important motivation to search for new physics at the LHC [6].

Astrophysics provides numerous indications of the need for BSM theories.

Most notably, the measurement of galactic rotation curves and galaxy cluster collisions [7] indicates that dark matter makes up $\sim 85\%$ of the matter in the universe. Dark matter interacts gravitationally but not electromagnetically or strongly; it is currently unknown if dark matter interacts weakly. Dark matter is likely to be a new particle not present in the standard model. The most popular type of dark matter candidate is a weakly interacting massive particle (WIMP) [6], but many other candidates have been proposed [8].

Electroweak unification and charge quantization suggest that a Grand Unified Theory (GUT) could unify all three fundamental forces, at an expected energy scale of $\sim 10^{16}$ GeV [9]. Further, gravity is not included in the standard model. A successful unification of general relativity and quantum field theory has not been achieved, due to the difficulty of constructing a renormalizable theory for the spin-2 graviton. At the Planck scale, gravitational effects become comparable to SM interactions, which calls for a new theory.

The following sections discuss the theories of leptoquarks and R-parity violating supersymmetry. The existence of leptoquarks can be a consequence of grand unification or other theories that address the parallels between leptons and quarks in the standard model. Supersymmetry is motivated by basic considerations of quantum field theory, the hierarchy problem, grand unification, and dark matter [10,11]. The introduction of R-parity violation in supersymmetry evades existing limits on signatures with large missing transverse energy due to the stability of the lightest supersymmetric particle (LSP). Without a stable LSP, R-parity violating supersymmetry lacks a dark matter candidate. However, it retains other desirable character-

istics of supersymmetry, including a solution to the hierarchy problem and grand unification. R-parity violating supersymmetry can also act as a signature generator to suggest novel searches which might discover or rule out other BSM theories [12].

2.3 Leptoquarks

Many BSM theories include a deeper relationship between leptons and quarks. Such a relationship is indicated by the cancellation of SM gauge anomalies from triangle diagrams, which requires each generation of matter to consist of quarks and leptons with the specific weak hypercharge values and multiplet arrangements that they possess in the SM [13]. Such theories introduce a class of particles called leptoquarks (LQs), which possess lepton number, baryon number, color charge, and fractional electric charge. Leptoquarks are bosons, either scalar with spin 0 or vector with spin 1. The values of the LQ quantum numbers are model-dependent. A total fermion number F can be defined as $F = 3B + L$ to characterize the combinations of lepton and baryon numbers found in LQs. The possible values are $F = 0$ for LQs which couple to $\bar{\ell}q$ or $\ell\bar{q}$ pairs and $|F| = 2$ for coupling to ℓq or $\bar{\ell}\bar{q}$ pairs.

In general, grand unified theories group leptons and quarks together in multiplets. The first BSM theory to include leptoquarks was the $SU(4)$ model by Pati and Salam [14], a GUT which casts lepton number as the fourth type of color charge, hence the $SU(4)$ symmetry instead of the SM $SU(3)$. Another GUT, the $SU(5)$ model by Georgi and Glashow [15], also contains leptoquarks. However, the symmetries in these theories are typically assumed to break at the GUT scale, rendering

the expected LQ masses very large and therefore unable to be directly produced at colliders. E_6 superstring theory [16] can also contain LQs, as it behaves similarly to GUTs below the string compactification scale, which is typically near the Planck scale. In other models, leptoquarks may be composite particles [17]. These include extended technicolor theories [18], which postulate a new strong interaction similar to QCD, and provide spontaneous masses to SM fermions using technifermions. A techniquark and anti-technilepton can bind together to form a technimeson which interacts with SM fermions as a leptoquark with a model-dependent coupling. Technicolor, though, has become disfavored after experimental data confirmed the Higgs mechanism as the solution to EWSB and spontaneous fermion masses.

The Buchmüller-Rückl-Wyler (BRW) model of leptoquarks includes all renormalizable Lagrangian terms compatible with the standard model [19]. There are several constraints imposed in the BRW model:

1. LQ interactions are dimensionless, in order to be renormalizable.
2. LQ interactions are invariant under the overall SM symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$.
3. LQ interactions conserve B and L , to avoid contributions to proton decay.
4. LQs couple only to SM particles.

Further consideration of experimental limits imposes two additional constraints, creating the minimal BRW (mBRW) model:

5. LQ couplings are chiral, involving either only left-handed fermions or only

right-handed fermions, due to limits on otherwise chirally-suppressed decays such as $\pi^+ \rightarrow e^+ \nu_e$.

6. LQ couplings involve only a single generation of leptons and quarks, to evade limits on flavor-changing neutral currents (FCNCs).

The Lagrangian terms are given for scalar LQs in Eq. (2.1) and for vector LQs in Eq. (2.2) below, using the ‘‘Aachen’’ notation as specified in Ref. [20].

$$\begin{aligned}
\mathcal{L}_S = & (\lambda_{L,S_0} \bar{q}_L^c i\sigma_2 \ell_L + \lambda_{R,S_0} \bar{u}_R^c e_R) S_0^\dagger + \lambda_{R,\tilde{S}_0} \bar{d}_R^c e_R \tilde{S}_0^\dagger \\
& + (\lambda_{L,S_{1/2}} \bar{u}_R \ell_L + \lambda_{R,S_{1/2}} \bar{q}_L i\sigma_2 e_R) S_{1/2}^\dagger + \lambda_{L,\tilde{S}_{1/2}} \bar{d}_R \ell_L \tilde{S}_{1/2}^\dagger \\
& + \lambda_{L,S_1} \bar{q}_L^c i\sigma_2 \boldsymbol{\sigma} \ell_L \cdot \mathbf{S}_1^\dagger + \text{h.c.}
\end{aligned} \tag{2.1}$$

$$\begin{aligned}
\mathcal{L}_V = & (\lambda_{L,V_0} \bar{q}_L \gamma_\mu \ell_L + \lambda_{R,V_0} \bar{d}_R \gamma_\mu e_R) V_0^{\mu\dagger} + \lambda_{R,\tilde{V}_0} \bar{u}_R \gamma_\mu e_R \tilde{V}_0^{\mu\dagger} \\
& + (\lambda_{L,V_{1/2}} \bar{d}_R^c \gamma_\mu \ell_L + \lambda_{R,V_{1/2}} \bar{q}_L^c \gamma_\mu e_R) V_{1/2}^{\mu\dagger} + \lambda_{L,\tilde{V}_{1/2}} \bar{u}_R^c \gamma_\mu \ell_L \tilde{V}_{1/2}^{\mu\dagger} \\
& + \lambda_{L,V_1} \bar{q}_L \gamma_\mu \boldsymbol{\sigma} \ell_L \cdot \mathbf{V}_1^{\mu\dagger} + \text{h.c.}
\end{aligned} \tag{2.2}$$

In these equations, scalar leptoquarks are denoted by S and vector leptoquarks are denoted by V . The subscripts 0, 1/2, and 1 denote singlet, doublet, and triplet states, respectively. Similar states with different quantum numbers are separated in the notation by the presence or absence of a tilde $\tilde{}$. The coupling constants for the Yukawa couplings between leptoquarks, leptons, and quarks are represented by λ , with the chirality L or R and the leptoquark type indicated in the subscript. The generation indices for the couplings and fermion multiplets are suppressed. The Pauli matrices are denoted by σ_i and the Dirac matrices by γ_μ . The Hermitian conjugate terms are indicated as ‘‘h.c.’’ The quantum numbers for the different

types of mBRW leptoquarks are listed in Table 2.2.

	Particle	Q	T_3	Y	F
Scalar	S_0	$-1/3$	0	$-2/3$	2
	\tilde{S}_0	$-4/3$	0	$-8/3$	2
	$S_{1/2}$	$\begin{pmatrix} -2/3 \\ -5/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$-7/3$	0
	$\tilde{S}_{1/2}$	$\begin{pmatrix} 1/3 \\ -2/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$-1/3$	0
	S_1	$\begin{pmatrix} 2/3 \\ -1/3 \\ -4/3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$	$-2/3$	2
Vector	V_0	$-2/3$	0	$-4/3$	0
	\tilde{V}_0	$-5/3$	0	$-10/3$	0
	$V_{1/2}$	$\begin{pmatrix} -1/3 \\ -4/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$-5/3$	2
	$\tilde{V}_{1/2}$	$\begin{pmatrix} 2/3 \\ -1/3 \end{pmatrix}$	$\begin{pmatrix} 1/2 \\ -1/2 \end{pmatrix}$	$1/3$	2
	V_1	$\begin{pmatrix} 1/3 \\ -2/3 \\ -5/3 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$	$-4/3$	0

Table 2.2: The quantum numbers of the different types of scalar and vector leptoquarks in the mBRW model.

As color-charged particles, leptoquarks are primarily produced by strong interactions in pp collisions. For pair production of leptoquarks, these interactions include gluon-gluon fusion and quark-antiquark annihilation, whose LO Feynman diagrams are shown in Fig. 2.3. An additional contribution to quark-antiquark annihilation may proceed through the Yukawa coupling λ of the leptoquark to the quark and lepton pair. However, the ratio M_{LQ}/λ , where M_{LQ} is the leptoquark mass, is restricted by limits from low-energy processes including $\pi^+ \rightarrow e^+\nu_e$ and atomic parity violation. The limits on M_{LQ}/λ range from 1800–6400 GeV/ c^2 , depending on

the type of leptoquark [21, 22, 23]. The expected accessible mass for leptoquarks at the LHC with $\sqrt{s} = 14 \text{ TeV}$ ranges from 900–1200 GeV/c^2 for scalar LQs and 1200–1500 GeV/c^2 for vector LQs [24]. Up to these masses, λ will be small enough that its contribution to leptoquark production can be neglected. This applies both to the Yukawa-based pair production diagram in Fig. 2.3 and the single production diagrams from quark-gluon scattering in Fig. 2.4.

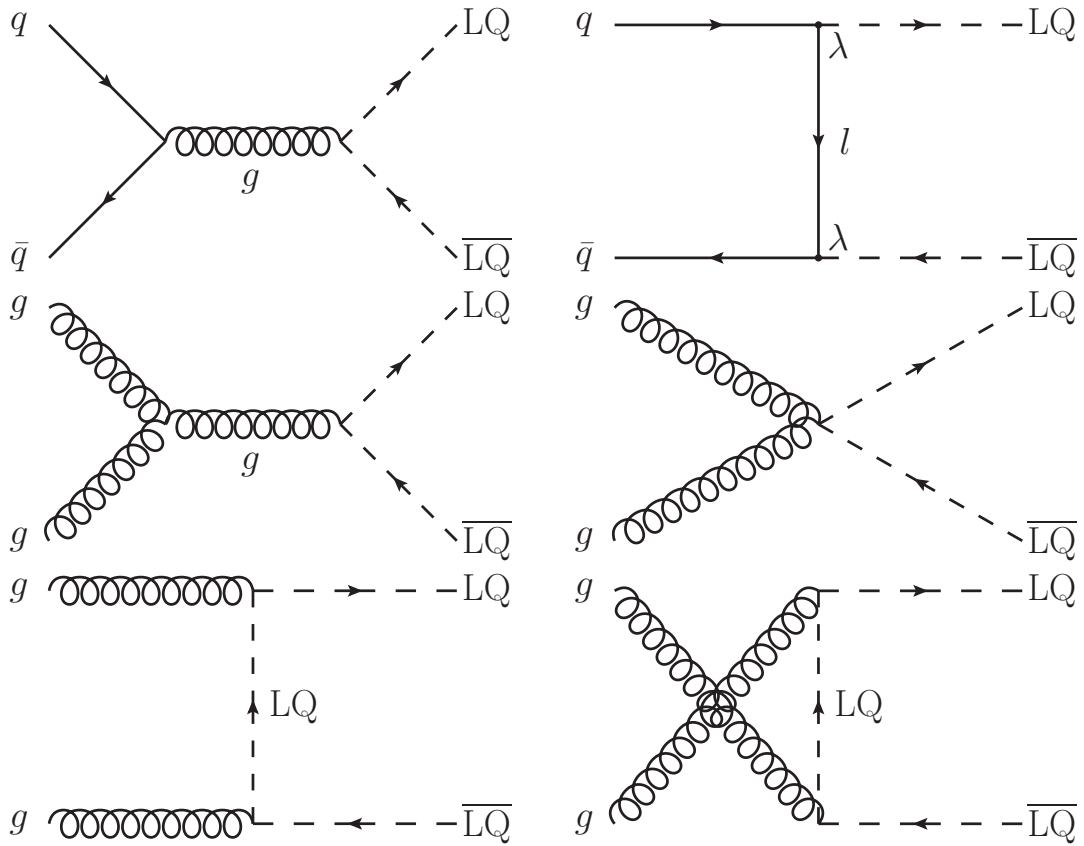


Figure 2.3: The LO Feynman diagrams for leptoquark pair production from quark-antiquark annihilation (top) and gluon-gluon fusion (middle, bottom).

Table 2.3 lists the pair production cross sections for a range of scalar leptoquark masses. They have been calculated using the CTEQ6 parton distribution functions (PDFs) [25, 26] with K -factors applied to include NLO corrections from QCD [27].

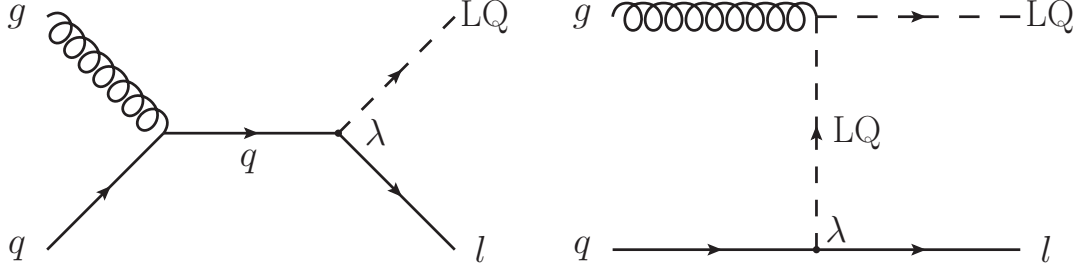


Figure 2.4: The LO diagrams for single leptoquark production from quark-gluon scattering. The LQ is produced in association with a lepton.

Theoretical uncertainties are calculated by propagating the PDF uncertainty and varying the factorization/renormalization scale μ from $\mu = M_{\text{LQ}}/2$ to $\mu = 2M_{\text{LQ}}$. These cross sections are sensitive only to the leptoquark mass and spin, so they are largely model-independent. For vector leptoquarks, the cross sections may be modified by anomalous triple and quartic gauge couplings. The cross sections for vector leptoquarks are typically larger than the cross sections for scalar leptoquarks. The decay widths for scalar and vector leptoquarks can be calculated according to Eqs. (2.3) and (2.4), respectively [19]:

$$\Gamma_S = \sum_i \frac{\lambda_i^2}{16\pi} M_{\text{LQ}}, \quad (2.3)$$

$$\Gamma_V = \sum_i \frac{\lambda_i^2}{24\pi} M_{\text{LQ}}. \quad (2.4)$$

Equations (2.3) and (2.4) sum over all Yukawa couplings for a given leptoquark. Given the limits on λ discussed above, leptoquarks accessible at the LHC with a single Yukawa coupling can be expected to have a fractional decay width of less than 0.1–0.2%.

In this dissertation, a search is performed for pair production of scalar leptoquarks decaying to third generation fermions. The symbol \mathcal{B} is used for the

M_{LQ} [GeV]	$\sigma(\mu = M_{LQ})$ [pb]	$\delta(\text{PDF})$ [pb]	$\sigma(\mu = M_{LQ}/2)$ [pb]	$\sigma(\mu = 2M_{LQ})$ [pb]
200	17.4	1.24	15.0	19.7
250	5.26	0.487	4.54	5.94
300	1.89	0.214	1.63	2.13
350	0.77	0.102	0.663	0.866
400	0.342	0.052	0.295	0.385
450	0.163	0.0278	0.14	0.183
500	0.082	0.0155	0.0704	0.0922
550	0.0431	0.00893	0.037	0.0485
600	0.0235	0.0053	0.0201	0.0265
650	0.0132	0.00322	0.0113	0.0149
700	0.00761	0.002	0.00648	0.00858
750	0.00448	0.00126	0.00381	0.00506
800	0.00269	0.00081	0.00228	0.00304
850	0.00164	0.000527	0.00139	0.00186
900	0.00101	0.000347	0.000856	0.00115
950	0.000634	0.000231	0.000534	0.000722
1000	0.000401	0.000155	0.000337	0.000458

Table 2.3: The pair production cross sections for a range of scalar leptoquark masses at $\sqrt{s} = 8$ TeV. Theoretical uncertainties from the PDFs and from varying the factorization/renormalization scale μ from $\mu = M_{LQ}/2$ to $\mu = 2M_{LQ}$ are indicated.

branching fraction for the decay $LQ \rightarrow \tau b$. Because they are pair produced, this results in a final state with two tau leptons and two bottom quarks. One tau lepton is required to decay leptonically: $\tau \rightarrow \ell \bar{\nu}_\ell \nu_\tau$, where ℓ can be a muon or an electron, which are collectively called light leptons. The other tau lepton is required to decay hadronically, denoted as τ_h ; see Sec. 4.8 for more information about hadronic decays of tau leptons. These decays result in two channels based on the leptonic decay of the tau, which are labeled as $e\tau_h$ and $\mu\tau_h$, or collectively $\ell\tau_h$ when the light lepton flavor is unimportant. Both bottom quarks hadronize into b-jets, as described in Sec. 4.9. Currently, the strongest mass limits on such third-generation scalar leptoquarks come from direct searches. Assuming $\mathcal{B} = 100\%$, the lower limit is approximately 530 GeV, set by both the CMS [28] and ATLAS [29] experiments using 4.7–4.8 fb⁻¹ of data from pp collisions with $\sqrt{s} = 7$ TeV. Indirect limits from

low-energy processes are discussed in Refs. [20, 21, 22, 23]. Though this search does not consider vector leptoquarks explicitly, scalar and vector leptoquarks tend to have similar angular distributions. Combined with the larger cross section for vector leptoquarks, this implies that the same search can set higher mass limits on vector leptoquarks, as shown in Ref. [28] for the 7 TeV data.

2.4 R-Parity Violating Supersymmetry

Supersymmetry proposes a symmetry between bosons and fermions. Such a symmetry appeals to mathematical considerations in quantum field theory by simplifying many calculations [13]. However, the primary motivation for a theory of SUSY with effects at the electroweak scale and therefore accessible at the LHC is the hierarchy problem. The mass of the Higgs boson, experimentally measured as $M_H = 125 \text{ GeV}$, is theoretically sensitive to quantum corrections via loop diagrams from any particle that couples to it. This sensitivity arises due to the scalar nature of the Higgs boson, which means that there is no symmetry available to protect its mass value. Therefore, precise cancellations must occur among the various quantum corrections to produce such a small mass value, relative to the expected energy scales of new physics. Any BSM theory lacking a simple mechanism to produce such cancellations must undergo unnatural fine-tuning in order to arrive at the correct value of M_H . The following discussion of supersymmetry is drawn primarily from Ref. [30].

The Higgs mass parameter m_{H}^2 appears in the Higgs potential:

$$V = m_{\text{H}}^2 |\text{H}|^2 + \lambda_{\text{H}} |\text{H}|^4. \quad (2.5)$$

Consider the one-loop contribution to m_{H}^2 from a fermion f which has a Yukawa coupling to the Higgs field, $-y_f \bar{f}_L \text{H} f_R$. The correction term in the mass parameter calculation can be written as follows:

$$\Delta_f m_{\text{H}}^2 = -\frac{|y_f|^2}{8\pi^2} \Lambda_{\text{UV}}^2 + \dots. \quad (2.6)$$

The factor Λ_{UV} is the cutoff scale used to handle the ultraviolet divergence in the loop integral through regularization. This cutoff scale is typically related to the energy scale of new physics, e.g. the GUT scale or Planck scale. Even if the cutoff scale is small, the contribution to the Higgs mass from any new heavy fermion will be proportional to its Yukawa coupling y , which could itself be large. Similarly, consider the one-loop contribution from a scalar boson S with a coupling to the Higgs written as $-y_S |\text{H}|^2 |S|^2$, which produces a correction term:

$$\Delta_S m_{\text{H}}^2 = \frac{y_S}{16\pi^2} \Lambda_{\text{UV}}^2 + \dots. \quad (2.7)$$

Both one-loop diagrams are shown in Fig. 2.5. The leading terms of Eqs. (2.6) and (2.7) have opposite signs, which suggests that the two terms could cancel if there were two scalars, one left-handed and one right-handed, for each fermion, with $y_S = |y_f|^2$.

This observation motivates SUSY as an extension of the standard model. While many formulations of SUSY exist, it is instructive to consider the minimal supersymmetric model (MSSM) to understand the fundamental components of the



Figure 2.5: One-loop diagrams for contributions to the Higgs mass parameter m_{H}^2 from a fermion f (left) and a scalar S (right) [30].

class of SUSY theories. SM particles are paired with new particles called superpartners with a spin difference of $1/2$; SM fermions have scalar boson superpartners and SM bosons have spin- $1/2$ fermion superpartners. The scalar superpartners are named after the corresponding SM fermions with the prefix “s-”, e.g. squarks and sleptons. The fermionic superpartners are named after the corresponding SM bosons with the suffix “-ino”, e.g. higgsinos and gauginos. The symbols for superpartners are indicated by a tilde \sim . SM particles and their superpartners are arranged together in supermultiplets which extend the multiplets listed in Table 2.1. Chiral supermultiplets contain scalar bosons and spin- $1/2$ fermions, while gauge supermultiplets contain spin- $1/2$ fermions and vector bosons. The superpartners possess the same quantum numbers as the corresponding SM particles, except for spin.

The Higgs boson, as a scalar, therefore becomes part of a chiral supermultiplet with weak hypercharge values $Y = \pm 1/2$. However, the existence of the corresponding chiral higgsinos creates new triangle gauge anomalies which must be cancelled to preserve the consistency of the theory. In addition, the weak hypercharge values of the Higgs supermultiplet restrict the possible Yukawa interactions with the other

chiral supermultiplets, which are necessary for spontaneous fermion masses. For both of these reasons, the MSSM contains two Higgs doublets, conventionally labeled H_u and H_d to indicate their couplings to up-type quarks or down-type quarks and charged leptons, respectively. Each Higgs doublet has its own VEV, labeled v_u and v_d , respectively. The angle β is defined as $\tan(\beta) = v_u/v_d$, with each VEV taken as a component of a single value v .

Given these details, the superpotential of the MSSM, W_{MSSM} , contains these terms:

$$W_{\text{MSSM}} = y_{u,ij} U_i^c Q_j H_u + y_{d,ij} D_i^c Q_j H_d + y_{e,ij} E_i^c L_j H_d + \mu H_u H_d. \quad (2.8)$$

Equation (2.8) is written in terms of superfields. U is the up-type quark singlet superfield; D is the down-type quark singlet superfield; Q is the quark doublet superfield; E is the charged lepton singlet superfield; and L is the lepton doublet superfield. The Yukawa couplings y are labeled by the fermion type u , d , or e with generation indices i, j . As an example, the top quark Yukawa coupling is $y_{u,33} = y_t$. The higgsino mass parameter is denoted as μ .

The superpotential is limited to these terms by requiring the conservation of R-parity. The quantity R is related to the baryon and lepton numbers as well as the particle spin S and may be defined in two equivalent ways [31], as shown in Eq. (2.9). Equation (2.10) defines R-parity based on R .

$$R = 3B + L + 2S = 3(B - L) + 2S, \quad (2.9)$$

$$R_p = (-1)^R. \quad (2.10)$$

All SM particles have $R_p = +1$, while all superpartners have $R_p = -1$. If R-parity

is conserved, interactions which violate lepton or baryon number are not allowed. Notably, the lightest supersymmetric particle must be stable in order to conserve R-parity in the decays of SUSY particles. In many models, the LSP is the lightest neutralino, a state created by mixing between the neutral higgsinos and gauginos due to EWSB. As a weakly interacting massive particle, the neutralino LSP is a promising candidate for WIMP dark matter.

If supersymmetry were unbroken, superpartners would have the same masses as their corresponding SM particles and would already have been detected. Therefore, there must be some mechanism responsible for breaking supersymmetry. In order for broken SUSY to continue to solve the hierarchy problem, it must preserve the conditions such as $y_S = |y_f|^2$ that lead to natural cancellations in the Higgs mass correction terms. So-called “soft” supersymmetry breaking accomplishes this by separating the Lagrangian into two separate terms:

$$\mathcal{L} = \mathcal{L}_{\text{SUSY}} + \mathcal{L}_{\text{soft}}, \quad (2.11)$$

where $\mathcal{L}_{\text{SUSY}}$ is made up of the SUSY-preserving terms, including the Yukawa interactions from the superpotential as well as the gauge interactions, and $\mathcal{L}_{\text{soft}}$ is made up of the SUSY-violating terms. The contributions to $\mathcal{L}_{\text{soft}}$ must be only mass terms and couplings whose parameters have positive mass dimension. These include triple-scalar interactions between two sfermions and a Higgs, such as $a_{u,ij} \tilde{u}_i^c \tilde{q}_j H_u$, where $a_{u,ij}$ are the soft couplings for those interactions. All such soft couplings are either at or below the mass scale m_{soft} , which is expected to be around the TeV scale from hierarchy problem considerations.

In general, the soft SUSY breaking terms can introduce FCNCs and charge-parity (CP) violation that are limited by precision measurements. These limits can be avoided by assuming the sfermion masses are universal; in other words, all sfermions should have the same mass regardless of flavor. This “flavor-blindness” eliminates mixing between flavors, and, as a bonus, greatly reduces the number of free parameters in the MSSM. The first- and second-generation sfermions possess flavor-blind universal masses. However, in the third generation, the sfermion masses can differ from the other two generations. In quantum field theory calculations, the couplings and mass parameters are treated as running factors which change based on renormalization group (RG) equations. These RG equations lead to corrections which tend to be small for Yukawa interactions, except for third generation particles, which have relatively large Yukawa couplings. In addition, at low energy scales, the slepton masses may deviate from the squark masses due to the different running behaviors of their respective couplings. The effects of RG evolution in SUSY on the gauge couplings for the three fundamental forces significantly aid grand unification, which is another motivation for SUSY.

The concept of naturalness, described in Ref. [32], requires the third-generation sfermion masses to be relatively light, while the other superpartners may be heavy. The tree level relationship $M_Z^2 \propto |\mu|^2 + |m_{H_u}^2|$ implies that the superpartners with the largest contributions to μ and $m_{H_u}^2$ must have masses near the electroweak scale. This applies especially to the higgsinos, whose masses are constrained by μ , and to the top squark, whose partner the top quark generates the largest correction to $m_{H_u}^2$. This consideration also applies to some extent to the other third-generation

sfermions and the gluino. In addition, because the top squark masses are not flavor-blind, the mass eigenstates can involve significant mixing between the chiral eigenstates. The top squark mass matrix is written as follows:

$$\mathbf{m}_{\tilde{t}}^2 = \begin{pmatrix} m_{Q_3}^2 + m_t^2 + \Delta_{\tilde{u}_L} & v(a_t^* \sin(\beta) - \mu y_t \cos(\beta)) \\ v(a_t \sin(\beta) - \mu^* y_t \cos(\beta)) & m_{U_3}^2 + m_t^2 + \Delta_{\tilde{u}_R} \end{pmatrix}. \quad (2.12)$$

The terms $\Delta_{\tilde{u}_L}$ and $\Delta_{\tilde{u}_R}$ arise via hyperfine splitting from quartic interactions and are defined in Ref. [30]. When this mass matrix is diagonalized, a large mixing angle typically occurs because of the off-diagonal entries, which contain terms involving the large top Yukawa coupling and soft coupling. This means that one top squark mass eigenstate will be lighter than the other, and in fact it will be the lightest squark. This mass splitting, together with the general consideration of naturalness that implies a light top squark mass, suggests that top squarks are likely to be accessible at the LHC.

In R-parity conserving (RPC) SUSY, all decays of superpartners eventually produce at least one LSP. As a weakly interacting particle, the LSP will escape a particle detector without interacting, leading to events with signatures including large missing transverse energy or \cancel{E}_T . The precise details of the production and decay of SUSY particles depend on which model is considered. This section has focused on the MSSM, but many variations of this model exist. These include, but are not limited to [8]: the constrained MSSM, in which there are only five parameters: the sfermion mass m_0 , the gaugino mass $m_{1/2}$, the soft parameter scale A_0 , $\tan(\beta)$, and μ ; the phenomenological MSSM, which uses experimental limits to restrict the number of free parameters to ~ 19 ; and the next-to-MSSM,

which introduces a gauge singlet field to explain the electroweak-scale value of μ [33]. In order to produce results with the broadest possible applicability, the CMS experiment uses simplified models, sometimes called decoupled models, in which the SUSY particles not under direct consideration are assumed to have masses too large to contribute significantly to the interactions. The latest results from the broad program of SUSY searches at the CMS experiment with $\sqrt{s} = 8 \text{ TeV}$ are summarized in Fig. 2.6. The requirements for naturalness introduced in Ref. [32], which include top squark and bottom squark masses less than 500–700 GeV and gluino mass less than 900–1500 GeV, are very nearly excluded at this point. In addition, recent measurements of the decay $B_s^0 \rightarrow \mu^+ \mu^-$ from the CMS [34] and LHCb [35] experiments are in agreement with the SM prediction, further limiting the possible forms of SUSY, which would enhance this decay.

The introduction of R-parity violating (RPV) terms in the SUSY Lagrangian is one way to evade these limits. If R-parity is violated, SUSY particles can decay to final states containing only SM particles, avoiding the characteristically large \cancel{E}_T from the LSP, which is no longer stable. This generally eliminates the LSP as a dark matter candidate, making it a less popular option. However, RPV SUSY still solves the hierarchy problem and assists in grand unification. The possible RPV terms in the superpotential are [31]:

$$W_{\text{RPV}} = \frac{1}{2} \lambda_{ijk} L_i L_j E_k^c + \lambda'_{ijk} L_i Q_j D_k^c + \frac{1}{2} \lambda''_{ijk} U_i^c D_j^c D_k^c + \mu_i L_i H_u. \quad (2.13)$$

The different RPV coupling constants are denoted as λ_{ijk} , λ'_{ijk} , λ''_{ijk} , and μ_i , where i , j , and k are generation indices. Figure 2.7 shows the tree-level diagrams for

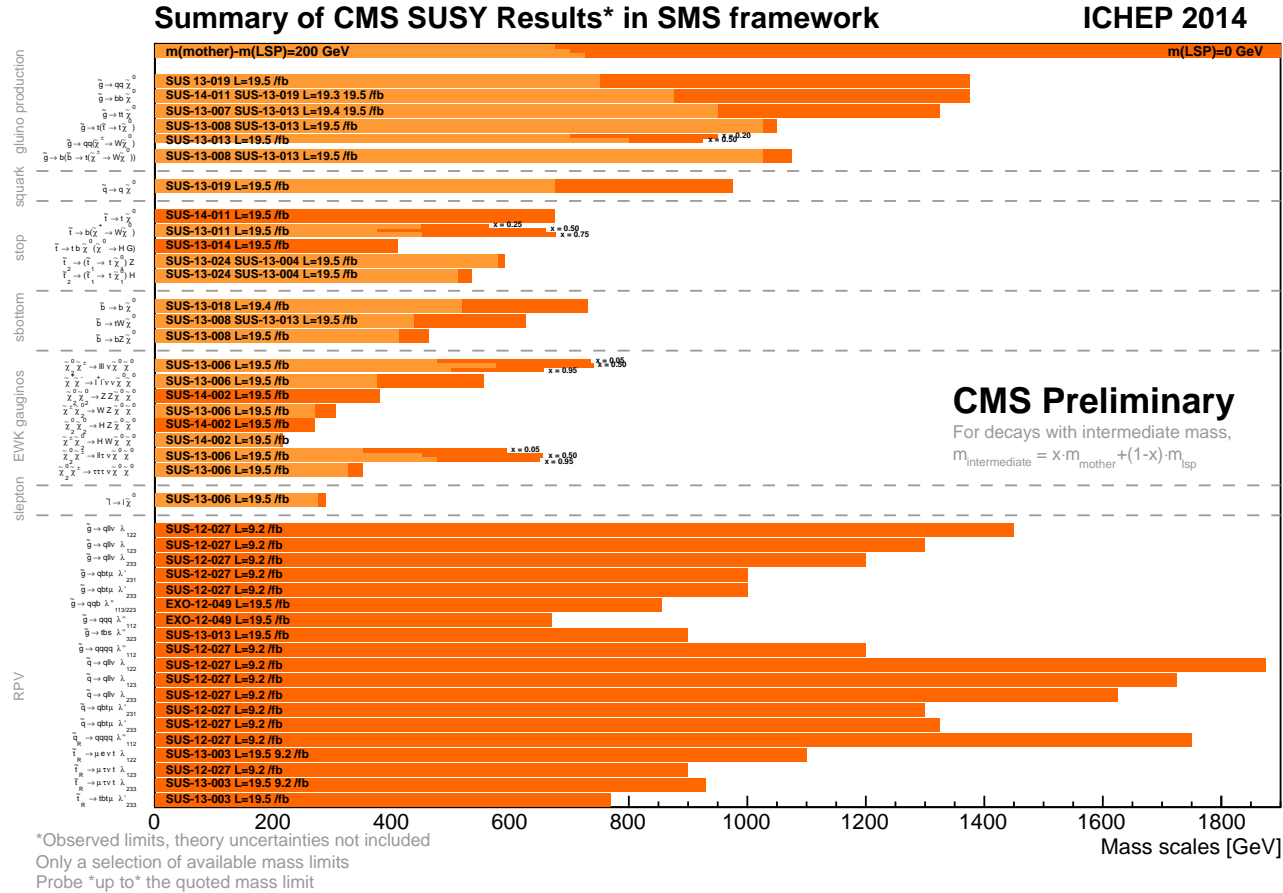


Figure 2.6: Summary of CMS exclusion limits for the masses of various SUSY particles with $\sqrt{s} = 8 \text{ TeV}$ [36]. These limits assume simplified models and unity branching fractions for the specified decays. Two scenarios are presented: the dark shades show $m(\text{LSP}) = 0$ and the light shades show $m(\text{mother}) - m(\text{LSP}) = 0$.

the trilinear RPV couplings. As indicated by Fig. 2.7 and the definition of R in Eq. (2.9), violating R is equivalent to violating lepton number or baryon number. Scenarios which allow all of the RPV couplings produce large contributions to proton decay, due to the violation of both B and L , and are therefore ruled out. Scenarios in which only a certain type of RPV coupling is allowed can still be viable.

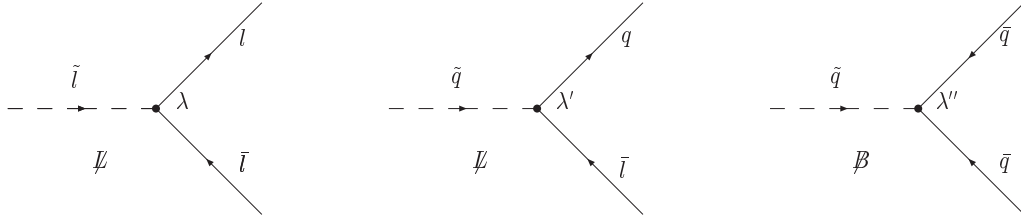


Figure 2.7: The LO diagrams for the three trilinear RPV couplings λ (left), λ' (middle), and λ'' (right) [31]. These couplings violate lepton number, lepton number, and baryon number, respectively.

This dissertation considers the pair production of top squarks with two different RPV decays, using a decoupled model as described above. As a strongly-interacting scalar particle, the top squark has a production cross section that is identical to the leptoquark production cross section at leading order. In decoupled models, the cross section depends on the squark mass scale and the top squark mixing angle only at higher order, and the corrections from these terms amount to less than 2% [37]. The first decay considered is $\tilde{t} \rightarrow \tau b$ via the coupling λ'_{333} . The final-state signature and kinematic distributions of this signal are identical to those from the pair production of third-generation scalar leptoquarks, as described in Sec. 2.3. Therefore, the results of the leptoquark search can be directly reinterpreted to apply to the λ'_{333} decay of top squarks.

In some natural SUSY models, if the higgsinos are lighter than the top squark, or if the RPV couplings that allow direct decays to SM particles are sufficiently small, the top squark decay may preferentially proceed via superpartners [38]. This motivates the second search in the dissertation, which focuses on a scenario in which the dominant RPC decay of the top squark is $\tilde{t} \rightarrow \tilde{\chi}^{\pm} b$. This requires the mass splitting between the top squark and the chargino to be less than the mass of the top quark, so it is chosen to be 100 GeV. The chargino is assumed to be a pure higgsino and to be nearly degenerate in mass with the neutralino, with a decay $\tilde{\chi}^{\pm} \rightarrow \tilde{\nu} \tau^{\pm} \rightarrow qq\tau^{\pm}$. The decay of the sneutrino occurs via the RPV coupling λ'_{3jk} , where the cases $j, k = 1, 2$ are considered. Such a signal can only be probed by searches that do not require large \cancel{E}_T , as chiral suppression prevents the other possible decay of the chargino, $\tilde{\chi}^{\pm} \rightarrow \nu \tilde{\tau}$, from contributing to scenarios involving the λ'_{3jk} coupling.

The final state from pair production of top squarks undergoing this chargino-mediated RPV decay contains two tau leptons, two b-jets, and at least four additional jets. The search for the signal with this final state is called the top squark search, to distinguish it from the similar but not identical final state in the leptoquark search. As in the leptoquark search, the analysis is divided into two channels, $e\tau_h$ and $\mu\tau_h$, based on the required leptonic decay of one of the tau leptons. The symbol \mathcal{B} in the top squark search is used to represent the branching fraction for the decay $\tilde{t} \rightarrow \tilde{\chi}^{\pm} b, \tilde{\chi}^{\pm} \rightarrow \tilde{\nu} \tau^{\pm} \rightarrow qq\tau^{\pm}$. This dissertation presents the first search for the chargino-mediated λ'_{3jk} decay of the top squark.

Chapter 3: Compact Muon Solenoid Experiment

The Compact Muon Solenoid (CMS) experiment is one of two general-purpose detectors at the LHC. It is located about 100 m underground on the LHC ring, near Cessy, France. The detector is cylindrically shaped, with a total length of 22 m, a diameter of 15 m, and a weight of 14000 tons. Figure 3.1 shows the overall layout of the detector. The following sections describe the LHC (based on Ref. [39]) and the CMS subdetector systems (based on Ref. [40]).

The center of the detector, the interaction point (IP), is used as the origin of the right-handed coordinate system that describes locations and directions within the detector. The z -axis is assigned to the direction of the LHC beam line. The polar angle θ is often transformed into pseudorapidity, defined as $\eta = -\ln[\tan(\theta/2)]$. Differences in pseudorapidity are Lorentz invariant for boosts in the z direction, and particle production is approximately uniform in η . The plane transverse to the z -axis comprises the x - and y -axes, with the x -axis pointing toward the center of the LHC ring and the y -axis pointing upward in the normal direction. The azimuthal angle ϕ and the radial coordinate r are also defined in the transverse plane. The magnitude of the component of momentum in the transverse plane is labeled p_T .

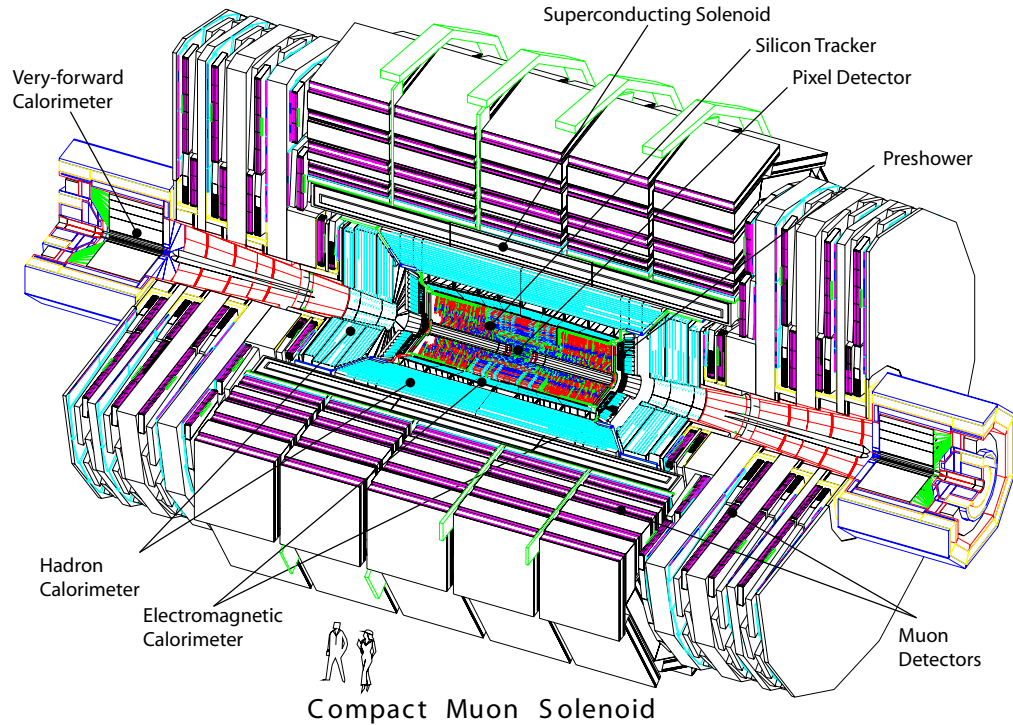


Figure 3.1: The layout of the CMS detector, with the subdetectors labeled and two humans shown for a height reference.

3.1 The Large Hadron Collider

The Large Hadron Collider is the largest machine ever built and the highest-energy collider in the world. It uses the tunnel originally constructed for the Large Electron-Positron Collider (LEP), with a circumference of 26.7 km. The tunnel is located underground in Switzerland and France, near Geneva. Figure 3.2 shows a diagram of the LHC. The location of the CMS experiment is provided and the other major experiments are also indicated.

The LHC is designed to accelerate two beams of protons up to energies of 7 TeV each, at instantaneous luminosities up to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. The use of supercooled su-

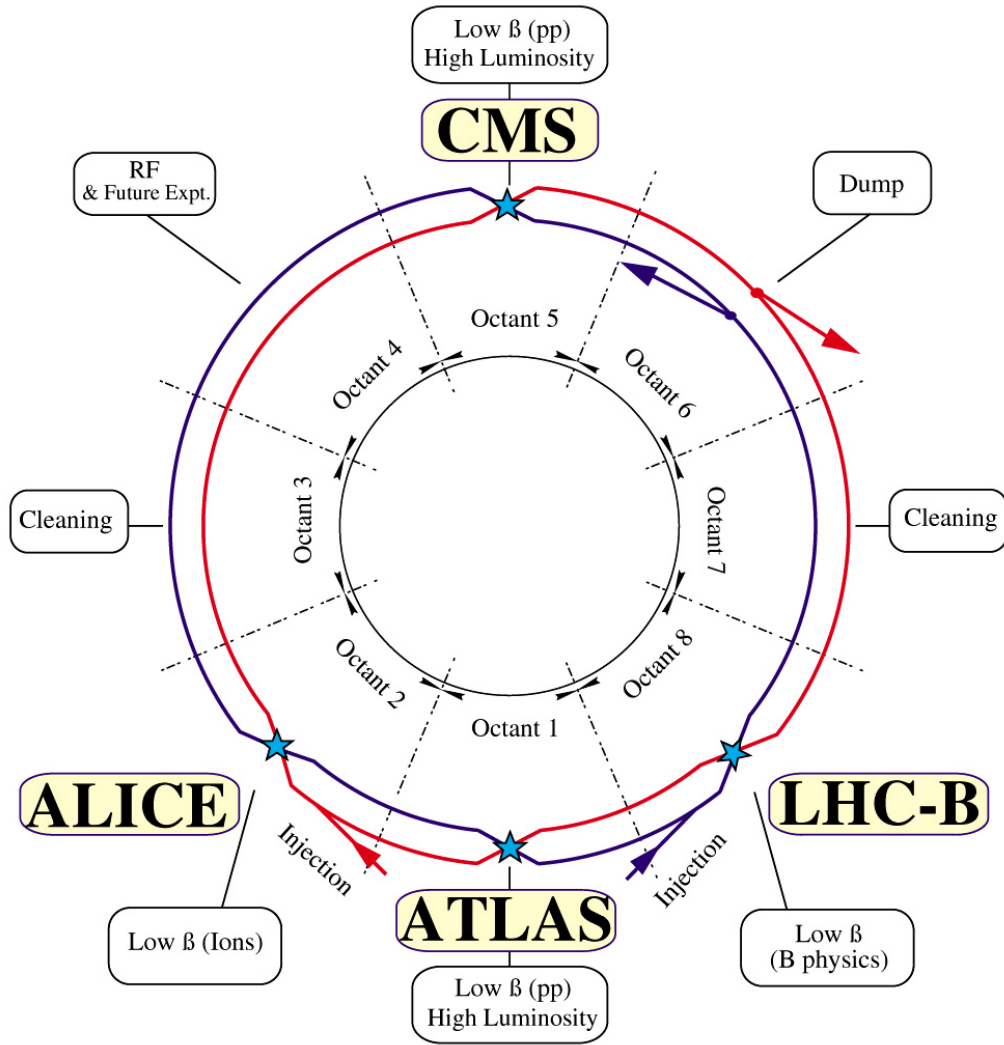


Figure 3.2: A diagram of the LHC with the major experiments labeled [41].

perconducting magnets, discussed below, is crucial. Several stages of CERN accelerators are used to inject proton beams into the LHC, as shown in Fig. 3.3. These include the linear accelerator Linac2, the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS), and the Super Proton Synchrotron (SPS). The accelerated protons are grouped into bunches using radio frequency (RF) electromagnetic fields. The LHC is designed to accommodate a bunch spacing of 25 ns, with 10^{11} protons per bunch and 2808 bunches per beam.

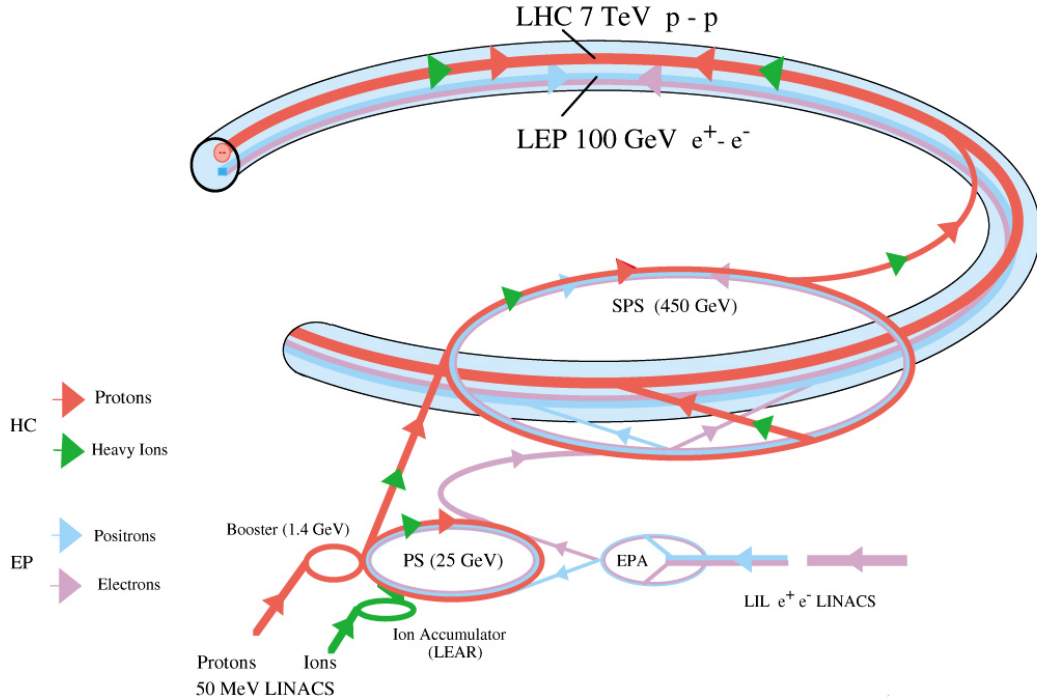


Figure 3.3: A diagram of the CERN accelerators which form the LHC injector [42].

Due to size limitations in the tunnel, the two rings used to accelerate the two proton beams are formed by twin bore magnets. Each magnet has a single mechanical structure and cryostat, in which are placed two coils and two beam channels. The dipole magnet coils use superconducting NbTi Rutherford cables cooled to 1.9 K, as shown in Fig. 3.4, with a design field strength of 8.33 T for acceleration of protons up to 7 TeV. This extreme cooling is accomplished using superfluid helium. In total, the LHC contains 1232 dipole magnets. Thousands of quadrupole, sextupole, octupole, and decapole magnets are used to correct and focus the beam.

In 2012, the LHC accelerated proton beams to energies of 4 TeV each, with a peak instantaneous luminosity of $7.67 \times 10^{33} \text{ cm}^{-2} \text{ s}^{-1}$ and a bunch spacing of

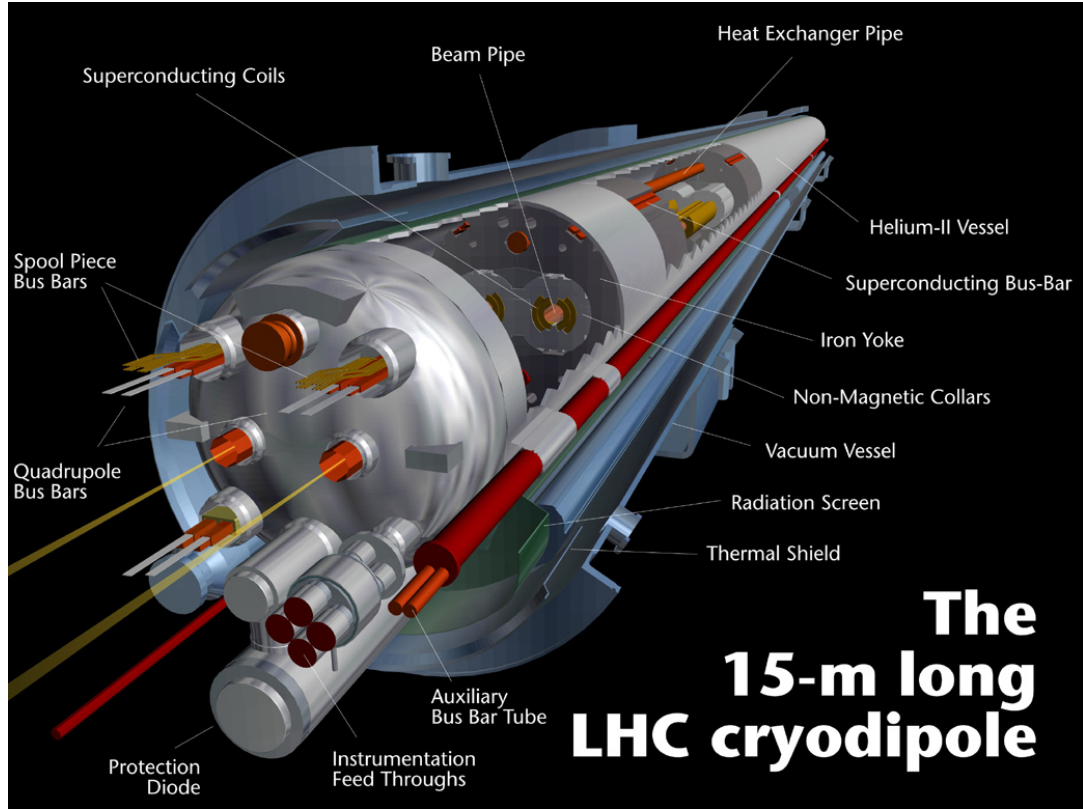


Figure 3.4: A diagram of an LHC dipole magnet, with the major components labeled [43].

50 ns. During that year, it delivered 23.30 fb^{-1} of integrated luminosity to the CMS detector, of which 21.79 fb^{-1} was recorded [44]. In the upcoming 2015 run, the LHC will achieve its design energy, instantaneous luminosity, and bunch spacing.

3.2 Tracker

The CMS tracker is the first subdetector to measure charged particles produced in collisions at the IP. It is 5.8 m long and 2.5 m in diameter, covering the pseudorapidity range $-2.5 < \eta < 2.5$. Two subsystems make up the tracker: the pixel detector and the silicon strip tracker. The layout of the tracker, with these

subsystems labeled, is shown in Fig. 3.5. Due to the tracker’s location close to the IP, it experiences severe radiation doses that are expected to range from 0.18 to 84 Mrad after 500 fb^{-1} of data. Hence, the tracker must be robust against radiation damage, requiring operation at -10°C and influencing the design of the sensors and electronics. For tracks with momentum of 100 GeV, the tracker has a transverse momentum resolution of 1–2% for $|\eta| < 1.6$; at higher η , the reduced transverse depth of the tracker degrades the resolution.

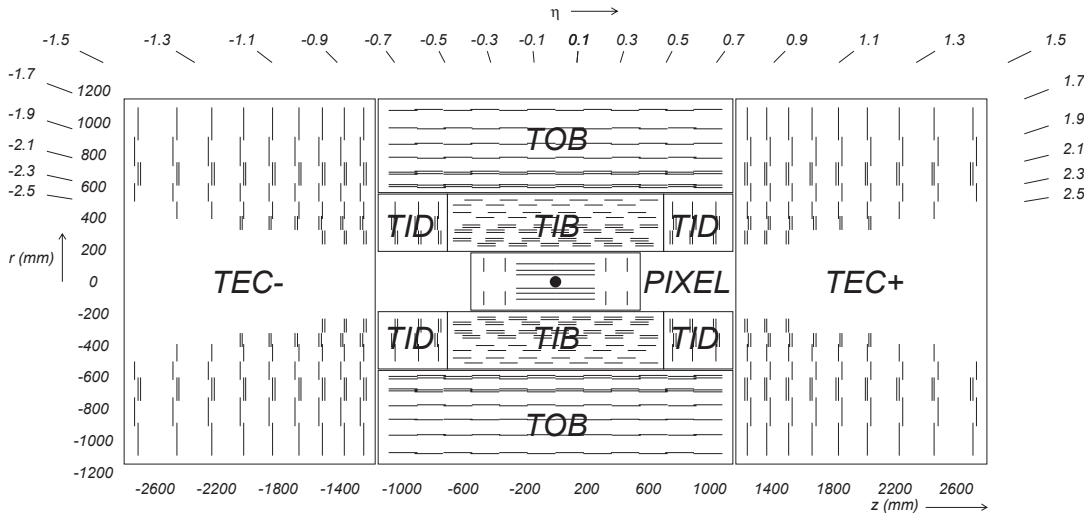


Figure 3.5: The layout of the CMS tracker, with subsystems labeled.

The pixel detector is the innermost portion of the tracker. It consists of three barrel layers, collectively called BPIX, and two endcap layers, called FPIX. Each pixel cell is a hybrid silicon detector with dimensions $100 \times 150 \mu\text{m}^2$. The small pixel size enables precise track resolutions of $10 \mu\text{m}$ in the $r-\phi$ direction and $20 \mu\text{m}$ in the z direction. In total, the BPIX comprises 48 million pixels and the FPIX comprises 18 million pixels. The pixel detector is important for many key components of CMS physics analysis. These include the reconstruction of secondary vertices from

decays of tau leptons and bottom quarks, as well as producing seed tracks for the strip tracker and the high level trigger.

The silicon strip detector consists of four subsystems. The Tracker Inner Barrel (TIB) has four layers with the three-layer Tracker Inner Disks (TID) at each end; both systems' strips are $320\ \mu\text{m}$ thick. Surrounding the TIB/TID is the Tracker Outer Barrel (TOB), which has six layers. The first four layers of the TOB use $500\ \mu\text{m}$ thick strips, while the last two layers use $122\ \mu\text{m}$ thick strips. The Tracker EndCaps (TEC) have nine disks with up to seven layers of strips, $320\ \mu\text{m}$ thick in the inner four rings and $500\ \mu\text{m}$ thick in the outer three rings. In total, all of these layers contain 9.3 million silicon strips.

The tracker maintained excellent performance during the 2012 run of the LHC. The pixel detector had 97.7% of channels operational in BPIX and 92.8% of channels operational in FPIX, while 97.5% of channels in the strip detector were active. The hit reconstruction efficiencies were greater than 99% for each layer of the strip detector and greater than 99.5% for each layer of the pixel except for the first layer of BPIX, which had an efficiency greater than 99.2% [45].

3.3 Electromagnetic Calorimeter

The electromagnetic calorimeter (ECAL) is a homogeneous calorimeter constructed entirely of lead tungstate (PbWO_4) crystals. The ECAL is divided into two subsystems: the ECAL barrel (EB) and the ECAL endcap (EE). In the endcap region, there is an additional ECAL preshower (ES) detector in front of the

EE. Figure 3.6 displays these subsystems. PbWO_4 has a peak emission wavelength of 425 nm and many desirable material properties. These properties include high density (8.28 g/cm^3), short radiation length (0.89 cm), short Molière radius (2.2 cm), and fast decay time (6 ns). The use of homogeneous PbWO_4 crystals enables precise energy resolution for electromagnetic objects. For photons with $E_T \approx 60 \text{ GeV}$, the energy resolution ranges from 1.1% to 2.6% for the EB and 2.2% to 5.0% for the EE. In general, the energy resolution σ varies as a function of energy E in GeV:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2. \quad (3.1)$$

In Eq. (3.1), S is the stochastic term, N is the noise term, and C is the constant term. Typical values for these terms were measured by a test beam to be $S = 2.8\%$, $N = 12\%$, $C = 0.30\%$.

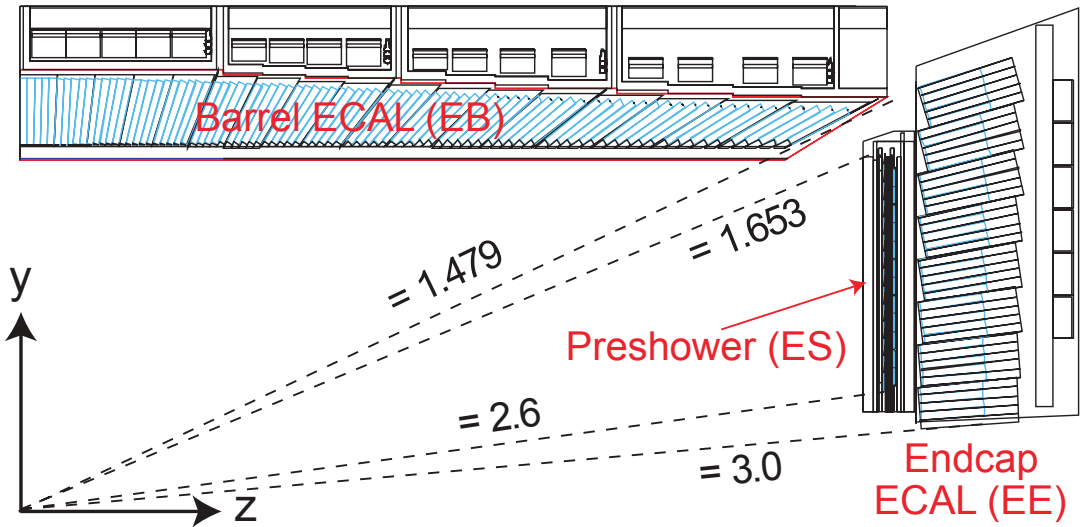


Figure 3.6: A diagram of the CMS ECAL, with subsystems and η ranges labeled.

The EB contains 61200 PbWO_4 crystals and covers the range $|\eta| < 1.479$. The crystals are arranged in a projective geometry with a tapered shape. The crystal

granularity is approximately 0.0174×0.0174 in η - ϕ , corresponding to dimensions of $22 \times 22 \text{ mm}^2$ at the front face and $26 \times 26 \text{ mm}^2$ at the back face. The EB has a depth of 230 mm or 25.8 radiation lengths (X_0). The scintillation light produced by the PbWO_4 crystals is read out using avalanche photodiodes (APDs). At 18°C , the APDs measure approximately 4.5 photoelectrons per MeV. The dark current of the APDs is sensitive to radiation exposure. Over the course of the 2012 run, the dark current ranged from 0.3–1.3 μA on average, corresponding to an average noise of 47–57 MeV [46].

The EE contains 14648 PbWO_4 crystals and covers the range $1.479 < |\eta| < 3.0$. The crystals are arranged in a non-projective x - y geometry, with dimensions of $28.62 \times 28.62 \text{ mm}^2$ at the front face and $30 \times 30 \text{ mm}^2$ at the back face. The EE has a depth of 220 mm or $24.7 X_0$. Vacuum phototriodes (VPTs) are used as the photodetectors to read out the scintillation light from the PbWO_4 crystals. They collect approximately 4.5 photoelectrons per MeV at 18°C . During the 2012 run, the average noise ranged from 180–220 MeV, with a more dramatic increase up to 600 MeV at high η because of the high radiation dose [46].

The ES is intended to identify neutral pions in the endcap region, covering the range $1.653 < |\eta| < 2.6$. It is a sampling calorimeter with two layers of lead absorber and silicon strip detectors. The first layer of lead absorber has a thickness of $2 X_0$, while the second layer has a thickness of $1 X_0$. Each layer of silicon strips is $320 \mu\text{m}$ thick and can collect 3.6 fC of charge from a minimum ionizing particle.

3.4 Hadron Calorimeter

The hadron calorimeter (HCAL) is a sampling calorimeter which measures the energy of hadrons. The HCAL is especially important for measuring neutral hadrons, which do not leave tracks in the tracker. In addition, by containing all hadronic activity in each event within $|n| < 5$, the HCAL enables the measurement of \cancel{E}_T caused by neutrinos and other theoretical weakly-interacting particles. The HCAL consists of four subsystems. Three of these subsystems use similar technology: the HCAL barrel (HB), the HCAL endcap (HE), and the HCAL outer (HO). The fourth subsystem, the HCAL forward (HF), uses an alternative technology necessary to survive the high radiation doses at its forward location. The locations of the various HCAL subsystems in the CMS detector are shown in Fig. 3.7. The calorimeter system, combining the ECAL and the HCAL, can measure charged pions with a resolution $\sigma/E \approx 100\%/\sqrt{E [\text{GeV}]} \oplus 5\%$ that varies with the jet energy E .

The HB is a 16-layer sampling calorimeter covering the range $|\eta| < 1.3$. The absorbing layers are made of C26000 cartridge brass, composed of 70% copper and 30% zinc. Cartridge brass has a density of 8.53 g/cm^3 , a radiation length of 1.49 cm, and a nuclear interaction length of 16.42 cm. The first absorbing layer in the HB is a 40-mm-thick steel plate. The next eight absorbing layers are 50.5-mm-thick brass plates, and the subsequent six absorbing layers are 56.5-mm-thick brass plates. The last absorbing layer is a 75-mm-thick steel plate. The overall thickness of the HB absorber ranges from 5.82 nuclear interaction lengths (λ_0) at $\eta = 0$ to $10.6 \lambda_0$ at

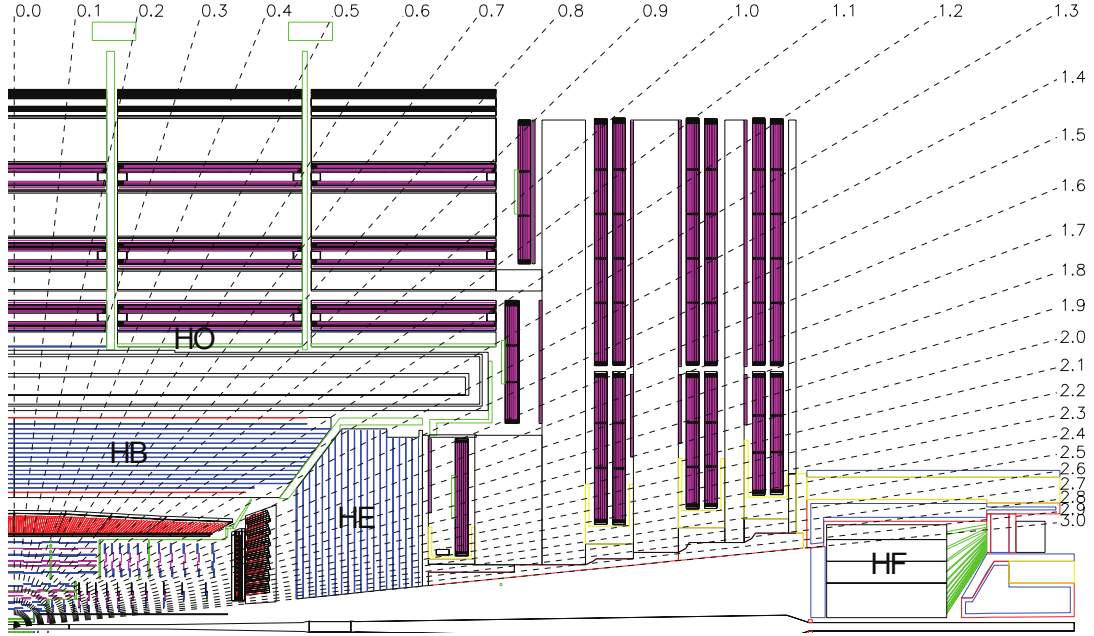


Figure 3.7: The layout of the HCAL subsystems HB, HE, HO, and HF in the CMS detector.

$|\eta| = 1.3$. The EB in front of the HB has a thickness of $1.1 \lambda_0$ and can measure the electromagnetic portions of early developing hadronic showers. The scintillating layers consist of 3.7-mm-thick Kuraray SCSN81 plastic scintillator, a polystyrene base doped with fluors. The exception is Layer 16, which has a thickness of 9 mm, in order to sample more from late developing hadronic showers. At the front of the HB, before the first absorbing layer, is the scintillator Layer 0, which is 9 mm of Bicron BC408 plastic scintillator, a polyvinyltoluene base doped with fluors. Layer 0 samples the energy deposited by hadronic showers in the dead material between the EB and the HB. The scintillator tiles are arranged in a projective geometry with a granularity of 0.087×0.087 in η - ϕ . In total, the HB has 16 η divisions, 36 ϕ divisions, and approximately 70000 tiles. The light from the scintillators is collected by Kuraray Y-11 green wavelength shifting (WLS) fiber, which is placed in a

groove shaped like the Greek letter σ in the scintillator tiles. The wavelength-shifted light from multiple layers is brought together and read out by hybrid photodiodes (HPDs). These photodetectors are used because of their large dynamical range and low sensitivity to magnetic fields.

The thinness of the HB at low η prevents it from fully containing hadronic showers, so the HO is added to act as an extension of the calorimeter system. The HO uses the same scintillator tile technology as the HB: 3.7-mm-thick SCSN81 with Y-11 WLS fiber and granularity 0.087×0.087 in η - ϕ , read out by HPDs. The residual magnetic field outside of the CMS solenoid is misaligned with the HPDs in the HO, causing random noise discharges [47, 48]. The HO is divided into five rings, each with a width of 2.536 m in the z direction, based on the structure of the iron return yoke outside of the solenoid. In the central Ring 0, the HO has two scintillating layers, one inside the solenoid and one outside of it. In the other rings, the HO has one scintillating layer outside of the solenoid. The thickness of the absorbing iron layer formed by the solenoid is 19.5 cm, extending the total depth of the calorimeter system to a minimum of $11.8 \lambda_0$.

The HE is a 17-layer sampling calorimeter covering the range $1.3 < |\eta| < 3.0$. Each absorbing layer consists of 79-mm-thick cartridge brass, the same material used for the HB absorbing layers. The scintillating layers use the same technology as the HB and the HO. In total, the HE contains 20916 scintillator tiles. The granularity of the tiles is the same as HB for $|\eta| < 1.6$; for higher η , they become coarser, with a granularity of approximately 0.17×0.17 in η - ϕ . Unlike the HB, the scintillating layers in each tower are split into multiple groups called depths before

being read out by HPDs. A diagram of the depth segmentation scheme is shown in Fig. 3.8. This depth segmentation allows for more precise recalibration of the HE, which experiences a higher radiation dose than the HB. Towers 27, 28, and 29, which are the closest to the beamline, have three readout depths, while the other towers have two readout depths. The crossover region between the HB and the HE, towers 15 and 16, also utilize depth segmentation. As in the HB, Layer 0 in the HE consists of 9-mm-thick BC408 to sample from the dead material between the EE and the HE. The combined calorimeter system, including both the EE and the HE, has an approximate thickness of $10 \lambda_0$.

The HF covers the range $3.0 < |\eta| < 5.0$, with no ECAL in front of it. It consists of a steel absorber structure with a thickness of 165 cm or $10 \lambda_0$. Polymer-cladded quartz fibers with diameter $800 \mu\text{m}$ are embedded in the steel absorber. The fibers are bundled together to form 13 towers in a non-projective x - y geometry with granularity 0.175×0.175 in η - ϕ . Over 1000 km of fiber is used in the HF. Half of the fibers run for the full 165 cm depth of the detector, while the other half start 22 cm into the detector. Electromagnetic showers deposit most of their energy in the first 22 cm of the HF, while hadronic showers deposit energy throughout the HF. Therefore, by reading out each type of fiber separately, the two types of showers can be distinguished. The fibers measure particle showers using Cherenkov light, which is read out by photomultiplier tubes (PMTs). They measure approximately 1 photoelectron for every 4 GeV of energy deposited. This alternative design was necessary to ensure the radiation hardness of the HF, parts of which can experience 100 Mrad/year.

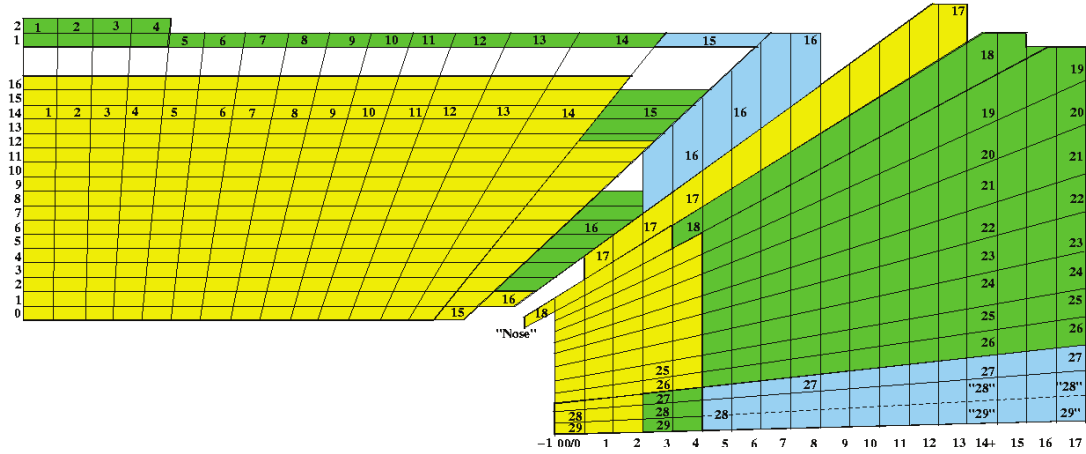


Figure 3.8: A diagram of the depth segmentation scheme in the HB and the HE.

3.5 Solenoid

The superconducting solenoid is the central feature of the CMS detector. The solenoid provides a magnetic field of 3.8 T within the volume formed by its diameter of 6 m and length of 12.5 m. This strong magnet field is necessary so that high energy charged particles bend sufficiently for the tracker to measure their momenta accurately. At full current, the solenoid has a stored energy of 2.35 GJ. The magnet is constructed from a 4-layer winding of reinforced NbTi conductor, cooled to 4.5 K. It is split into five rings of equal length. The cold mass of the magnet is 220 tons, and the high ratio between the stored energy and the cold mass, 11.6 KJ/kg, causes a significant mechanical deformation of 0.15% when the magnet is powered. Figure 3.9 shows an artistic rendering of the solenoid.

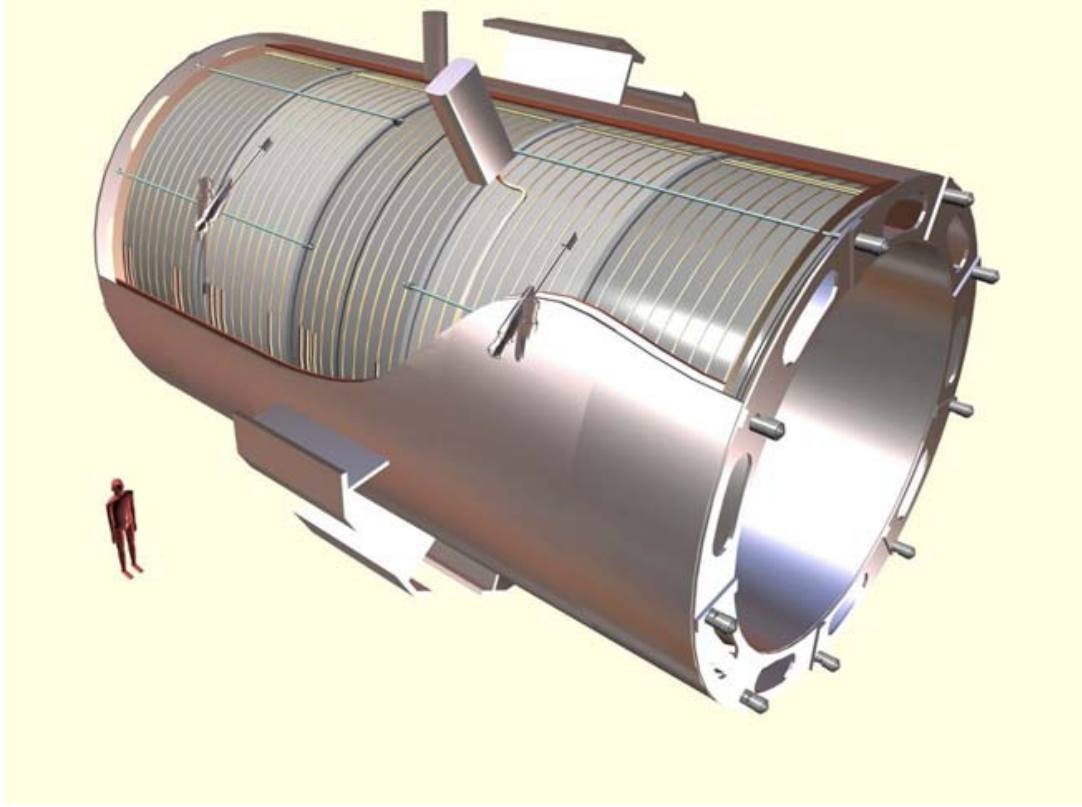


Figure 3.9: An artistic rendering of the CMS solenoid, showing the five rings placed inside the cryostat, along with the support structure.

3.6 Muon System

The identification and measurement of muons is a major focus of the CMS experiment. The CMS muon system comprises three subsystems, each utilizing different gaseous particle detection technologies. In the barrel region, drift tubes (DTs) are used. In the endcap region, cathode strip chambers (CSCs) are used. Resistive plate chambers (RPCs) are also used in both regions. The muon systems are built into the iron yoke, which consists of five barrel rings and six endcap disks weighing 10000 tons in total. The yoke confines the outer magnetic field from the return flux from the solenoid and absorbs stray hadrons. The layout of the muon

system is shown in Fig. 3.10. For 1 TeV muons, the resolution varies between 15% and 40%, depending on $|\eta|$. When the muon system measurements are combined with measurements from the tracker, the 1 TeV muon resolution is improved to 5%.

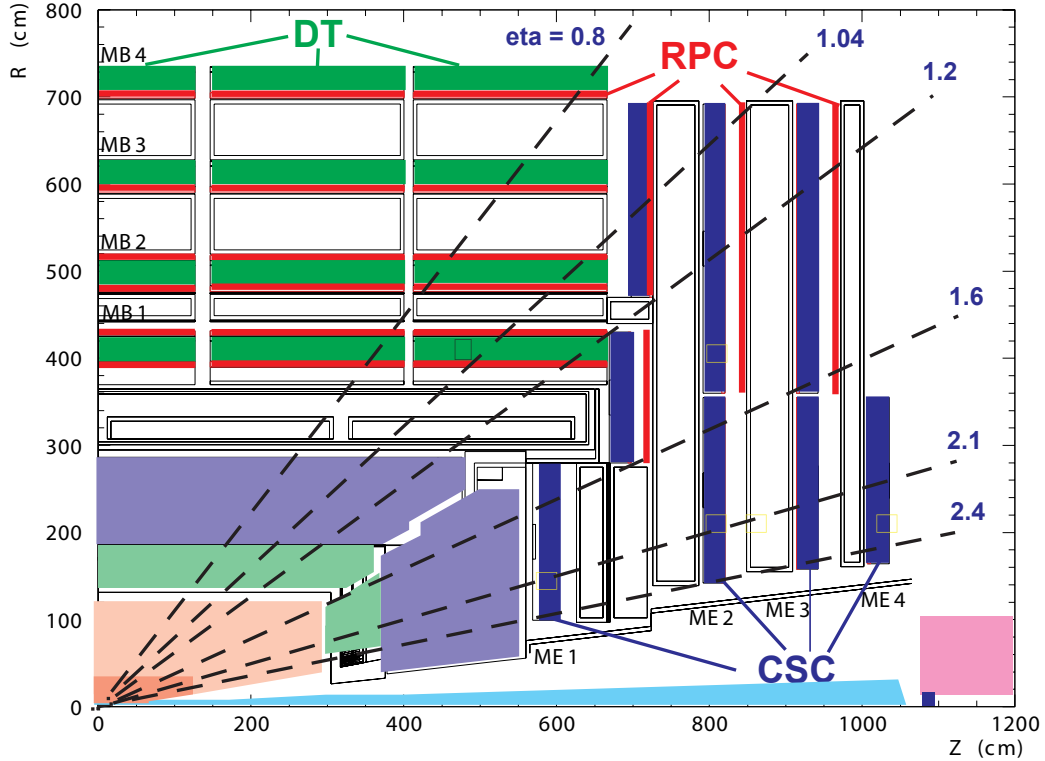


Figure 3.10: The layout of the muon system, with the three subsystems labeled.

The DTs are divided into four stations, which together cover the range $|\eta| < 1.2$ and are labeled MB1 through MB4 (Muon Barrel). The first three stations each contain twelve chambers divided into three groups of four. Two of the groups of four measure the r - ϕ coordinate of muons, while the third group of four measures the z coordinate. MB4 does not include a group of chambers that measures z . All four stations together contain 250 DTs with a total of 172000 sensitive wires. The gas used in the DTs is a mixture of 85% Ar and 15% CO₂, and the anode wires

are gold-plated stainless steel with a diameter of $50\ \mu\text{m}$. Within $|\eta| < 0.8$, the MB stations alone can reconstruct high- p_{T} muon tracks with an efficiency exceeding 95%. The global r - ϕ resolution is $100\ \mu\text{m}$.

The CSCs are also divided into four stations and cover the range $0.9 < |\eta| < 2.4$. The four stations are labeled ME1 through ME4 (Muon Endcap). Each station is divided into several groups as follows: ME1 has three groups of 72 CSCs; ME2 and ME3 each have one group of 36 CSCs and another group of 72 CSCs; ME4 has one group of 36 CSCs. The total number of CSCs is thus 468. The cathode strips are arranged in the radial direction to measure the r - ϕ coordinate, while the anode wires are perpendicular to the strips to measure η . There are approximately 220000 cathode strip readout channels and 180000 anode wire readout channels. The CSC gas mixture is set at 40% AR, 50% CO₂, and 10% CF₄. The cathode strips are formed from a fiberglass/epoxy material called FR4, coated with 36- μm -thick copper. The anode wires are made of gold-plated tungsten with a diameter of $50\ \mu\text{m}$. The first group of ME1 CSCs uses slightly thinner wire with $30\ \mu\text{m}$ diameter and has some other slightly modified properties.

To complement the DTs and CSCs, RPCs are installed in both the barrel and endcap regions, covering the range $|\eta| < 1.6$. The RPCs are primarily used to provide muon trigger information, due to fast tagging capabilities which allow them to precisely identify the bunch crossing time of muon candidate events. The time resolution of the RPCs is less than 3 ns, compared to maximum drift times of 400 ns for the DTs and 60 ns for the CSCs. MB1 and MB2 each have one internal and one external group of RPCs, relative to the DTs; MB3 and MB4 each have two internal

groups of RPCs. These groups together comprise 480 chambers. In the endcap, there are three RPC stations mounted in concentric circles on the iron yoke disks, with a total of 144 chambers. The RPCs are parallel plate detectors filled with a gas mixture of 96.2% $\text{C}_2\text{H}_2\text{F}_4$, 3.5% $i\text{C}_4\text{H}_{10}$, and 0.3% SF_6 .

3.7 Trigger

The LHC operates at a high instantaneous luminosity, up to $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$. With an expected proton-proton cross section of 100 mb at the LHC center-of-mass energies, the collision rate is approximately 1 MHz. The CMS trigger is necessary to reduce this high rate of collision events to a rate which can be stored and processed. The trigger system consists of two stages. The first stage uses hardware and is called the Level-1 (L1) Trigger. The L1 Trigger is designed to have a maximum output rate of 100 kHz. The second stage is the High Level Trigger (HLT), which uses software and reduces the output rate to $\mathcal{O}(100 \text{ Hz})$.

The L1 Trigger uses custom-built programmable electronics, including field-programmable gate arrays (FPGAs), memory lookup tables (LUTs), and application-specific integrated circuits (ASICs). All of the subdetectors send input to the L1 Trigger, which is organized into local, regional, and global components as shown in Fig. 3.11. The local components, Trigger Primitive Generators (TPGs), are constructed from energy deposits in the calorimeters and track segments or hit patterns from the muon system.

The TPGs from the ECAL, the HCAL, and the HF are combined into the

Regional Calorimeter Trigger (RCT). The RCT groups calorimeter trigger towers into regions, which are made up of four towers in the barrel and endcap, and one tower in the HF. These regions are used to determine electron and photon candidates, as well as transverse energy sums ($\sum E_T$) and tau-jet vetoes. The RCT also passes information to the muon triggers about minimum ionizing particle (MIP) energy deposits and surrounding energy deposits that indicate whether muon candidates are isolated from other particles. Using information from the RCT, the Global Calorimeter Trigger (GCT) determines jet candidates and counts, providing up to four jets and four tau-jets from the central HCAL and four jets from the HF. The GCT also determines total E_T , \cancel{E}_T , and H_T , which is calculated as $\sum E_T$ for all jets above a certain threshold.

In parallel, the muon DT, CSC, and RPC systems each produce their own local triggers. The Regional Muon Trigger (RMT) contains the DT and CSC Track Finders (DTTF, CSCTF) which make tracks using segments from their respective subdetectors. As mentioned in Sec. 3.6, the RPCs act as a dedicated trigger using their timing resolution of 1 ns to determine bunch crossing times. The Global Muon Trigger (GMT) combines the information from the RMT and RPCs to produce up to four muon candidates in each of the barrel and endcap regions. These candidates include the following information: p_T , charge, η , ϕ , a quality code, MIP, and isolation.

Finally, the Global Trigger (GT) combines the GCT and GMT candidates and quantities to decide whether or not to keep the event, based on a set of L1 triggers with different criteria. The GT also uses information from the Trigger Control

System (TCS) regarding the status of the subdetector readout and data acquisition systems. The Timing, Trigger, and Control (TTC) system is used to return the GT decision, called the Level-1 Accept (L1A), to the various subdetectors. This entire process is completed within $3.2 \mu\text{s}$. During this time, the high-resolution data for the event must be stored in memory, while $\mathcal{O}(100)$ subsequent bunch crossings occur. All of this incoming data must be pipelined in order to synchronize the results of the various steps of the trigger system for each event.

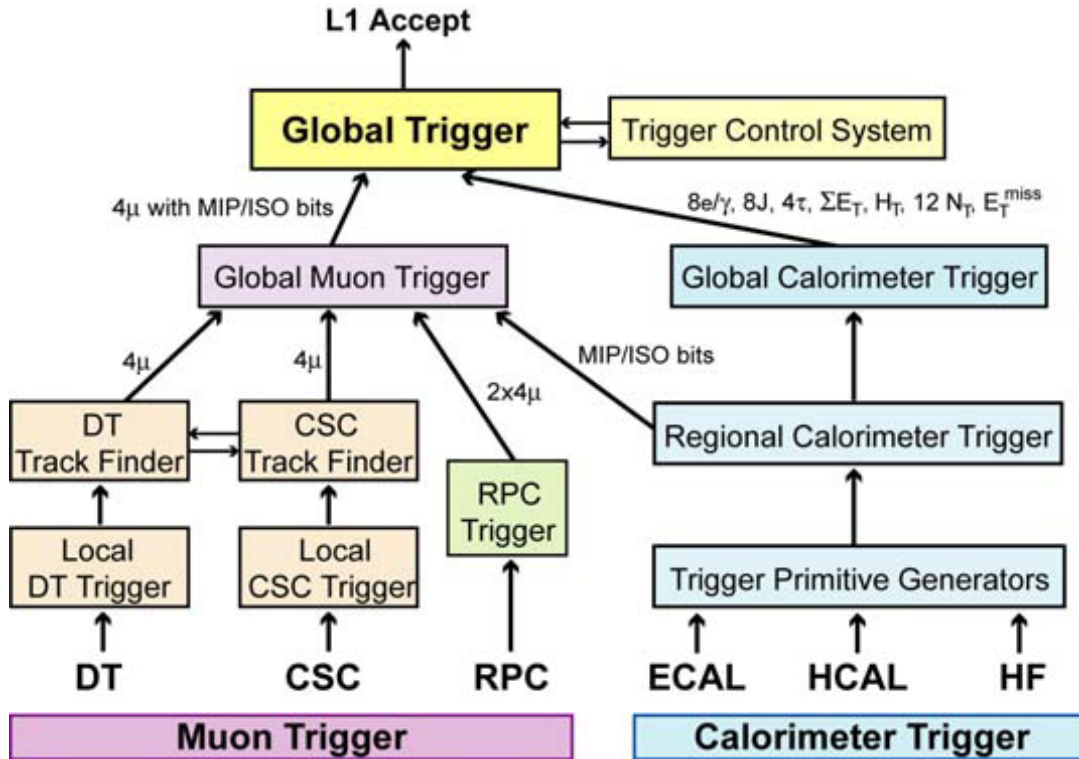


Figure 3.11: The architecture of the L1 Trigger.

The HLT further analyzes events which pass the L1A decision. Using a farm of roughly 1000 commercial processors comprised of over 13000 central processing units (CPUs), it emulates the full offline reconstruction algorithms described in Ch. 4. Like the L1 Trigger, the HLT uses a set of triggers with different criteria, called

the trigger menu. Different trigger menus are constructed for various conditions, including different instantaneous luminosity levels and different types of collisions or measurements. The selected menu can be changed during the operation of the detector in response to new conditions. Events which pass the HLT decision are sorted into primary datasets (PDs) with minimal overlap. The HLT output includes several streams, including monitoring and calibration streams in addition to the primary stream of physics events.

During the 2012 run, the L1 Trigger operated at rates up to 100 kHz with only 3% dead time [49]. The HLT operated at rates up to 1 kHz and took an average of 200 ms to process an event [50]. This processing speed is two orders of magnitude faster than the full offline reconstruction. The HLT achieves this fast processing time using several optimizations. The reconstruction algorithms in a given trigger path are arranged so that the fastest algorithms run first. If an algorithm's product does not pass a specified quality filter, the rest of the trigger path is skipped. In addition, the reconstruction algorithms only consider small regions of the detector output, based on the locations of L1 candidates.

3.8 Luminosity Measurement

The fine resolution of the CMS pixel detector (Sec. 3.2) implies that a given pixel will tend to be activated by one track at most per bunch crossing. Clusters are created from groups of nearby activated pixels in the tracker. A minimum bias interaction creates an average of 200 clusters, with each cluster containing an

average of 5 pixels [51]. Even with 100 pileup events per bunch crossing, the pixel detector will have an occupancy as low as 0.1%. The number of pixel hits should thus scale linearly with the number of interactions per bunch crossing for instantaneous luminosities up to and even beyond the LHC design performance. Equation (3.2) shows how the instantaneous luminosity \mathcal{L} is related to the average number of pixel clusters per event $\langle n \rangle$ [52]:

$$\mathcal{L} = \frac{\nu \langle n \rangle}{\sigma_{\text{vis}}}. \quad (3.2)$$

Here, $\nu = 11246$ Hz is the LHC revolution frequency and σ_{vis} is the the visible cross section, calibrated by a Van der Meer scan [53]. In 2012, CMS measured the total integrated luminosity with a systematic uncertainty of 2.6% using this method.

The HF is used as a second method of measuring the luminosity. This is possible because the HF can be run safely during unstable beams [52]. Information from the HF can be used to measure the luminosity in two different ways. The average fraction of empty towers can be related to the mean number of interactions per crossing, or the average transverse energy per tower can be linearly related to the luminosity. It can make an online determination of the average luminosity to a statistical uncertainty of 1% in under 1 s. However, the calibration of this measurement can drift over long time periods due to changes in the gain of the HF PMTs. In practice, the increase in pileup interactions observed during the 2012 run moves the HF response into a nonlinear regime, limiting the accuracy of this method. Because of these limitations, the HF method is utilized primarily as a cross-check for the pixel cluster counting method.

Chapter 4: Event Reconstruction

Particles created in proton-proton collisions pass through the CMS detector and leave signals in different subdetectors. Figure 4.1 shows examples of typical signals for different types of particles. Each type of particle has a different characteristic signature from which it can be identified using information from the various subdetectors. Muons, electrons, and charged hadrons create tracks in the tracker, while photons and neutral hadrons do not. Muons also create hits in the muon systems. Electrons and photons deposit energy in the ECAL, while charged and neutral hadrons deposit most of their energy in the HCAL. Neutrinos do not deposit any energy in the detector. Their presence must be inferred from missing transverse energy.

The raw output from each subdetector is processed in several steps in order to reconstruct the different types of particles [55]. The first step is local reconstruction, which involves the creation of reconstructed hits or “RecHits” for each subsystem of each subdetector. The tracker RecHits include information about the positions of clusters, which are combinations of contiguous strips or pixels that contain signals, and energy deposition information used for particle identification. The muon system RecHits also provide the positions of signals. In the DT and CSC subsystems,

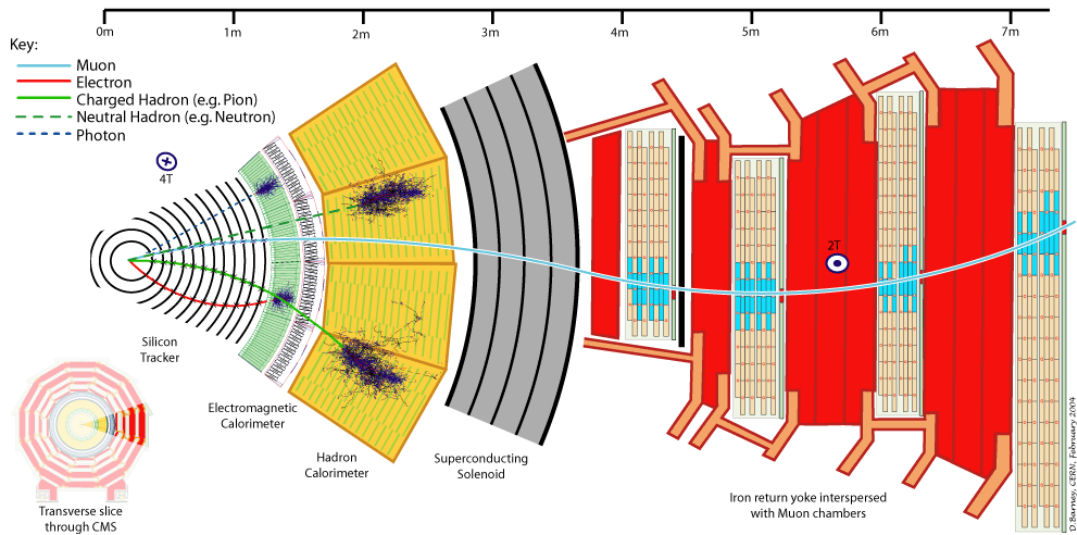


Figure 4.1: A cross-sectional view of the CMS detector with all subdetectors labeled and examples of signals left by muons, electrons, charged hadrons, neutral hadrons, and photons [54].

these RecHits can be combined into three-dimensional track segments, which provide information about the direction of the particle that created them. The ECAL and HCAL RecHits contain the energy, position, and time of energy deposits from traversing particles.

In the second step, global reconstruction, the RecHits from the different subsystems of a given subdetector are combined and further processed. In the tracker, pattern recognition algorithms are employed to reconstruct tracks for various cases, including displaced vertices, low p_T tracks, and high p_T tracks. The ECAL and HCAL RecHits are summed if they are in the same tower, forming calorimeter towers or “CaloTowers” using a projective η - ϕ geometry. “Standalone” muons are created by the muon system global reconstruction, which associates RecHits and track segments that have compatible radial trajectories, accounting for bending by

the residual magnetic field, and uses a vertex-constrained fit.

High-level reconstruction is the final step, in which information from all subdetectors is used to reconstruct various types of particles as precisely as possible. The particle types used in this search include electrons, muons, taus, jets, and b-jets. \cancel{E}_T is also used in the definitions of some samples used for background estimations. The reconstruction algorithms for these particles will be described in more detail in the following sections of this chapter. Many of these algorithms use a particle flow technique that will be described in Sec. 4.4.

4.1 Event Generation

Protons are primarily composed of two up quarks and one down quark. Because protons are QCD bound states, those three quarks should be treated as valence quarks, the most prominent features in a quark-gluon sea of virtual particles. At high energies such as those present in LHC collisions, the presence of additional quarks and gluons, collectively called partons, becomes significant. The fraction of the momentum of a hadron (such as a proton or neutron) A carried by a parton a is defined as x_a , and the PDF for that parton is $f_{a/A}(x_a, Q^2)$, where Q^2 is the momentum scale of the interaction, typically the square of the total four-momentum in a collision. PDFs are calculated using experimental data sets, with different groups taking different approaches to analyzing the data and modeling proton behavior. The CMS experiment primarily uses PDFs from Martin-Stirling-Thorne-Watt (MSTW) and the Coordinated Theoretical-Experimental Project on

QCD (CTEQ). Figure 4.2 shows an example set of PDFs calculated at NLO by MSTW [56]. At high momentum scales, even bottom quarks may be present in the quark-gluon sea of a proton. The dependence of the PDFs on Q^2 is given by the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations [57].

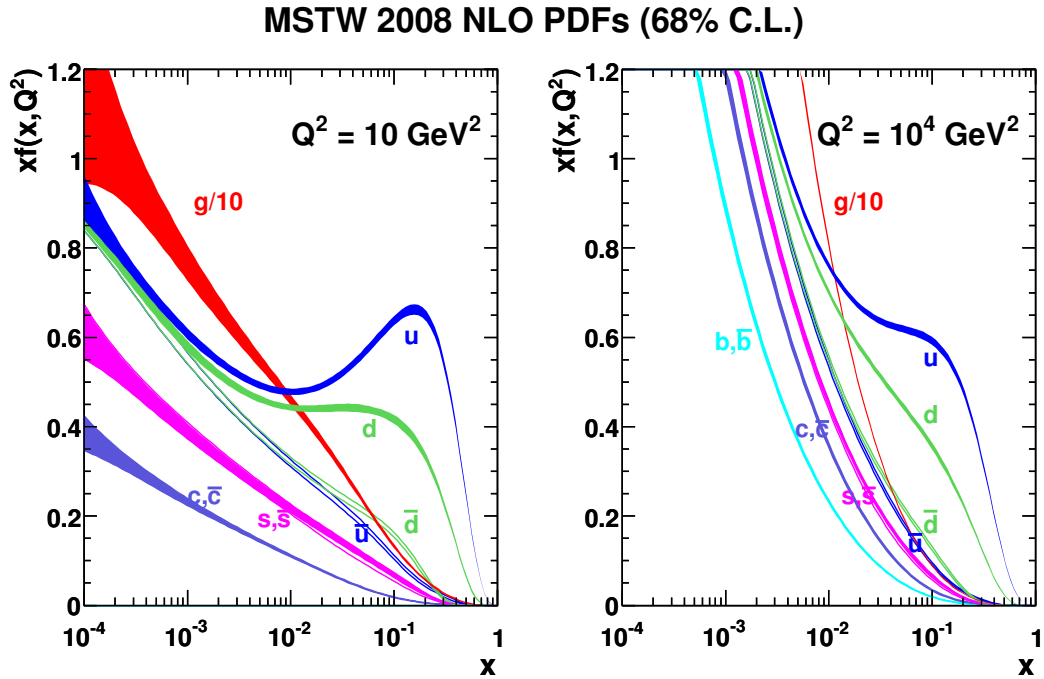


Figure 4.2: PDFs calculated at NLO by MSTW, plotted against momentum fraction x for two different values of the momentum scale Q^2 [56].

The QCD factorization theorem states that large logarithmic factors, for example from collinear emission of gluons, can be included in the definition of the PDFs for all hard-scattering processes. The DGLAP equations are used to account for the dependence of this factorization on Q^2 . With this theorem, the interaction of parton a in proton A with parton b in proton B , producing the final state X , can

be written simply:

$$\sigma_{AB \rightarrow X}(s) = \int dx_a dx_b f_{a/A}(x_a, \mu_F^2) f_{b/A}(x_b, \mu_F^2) \hat{\sigma}_{ab \rightarrow X}(\hat{s}, \mu_R^2) \quad (4.1)$$

For such an interaction at high energy, the center-of-mass energy is $\sqrt{\hat{s}} = \sqrt{x_a x_b s}$, where \sqrt{s} is the center-of-mass energy of the proton-proton system. There are two scales in Eq. 4.1: the factorization scale μ_F which separates long-distance physics from short-distance physics, and the renormalization scale μ_R of the QCD running coupling. Values for these scales are typically chosen to be characteristic of the hard scattering, with $\mu_F = \mu_R$. In addition to the primary hard-scattering process, the incoming protons and outgoing final state particles may radiate photons and gluons in processes called, respectively, initial state radiation (ISR) and final state radiation (FSR). The remnant partons which did not participate in the primary hard scatter may undergo soft interactions; these are collectively considered to be the “underlying event”. All of these possible interactions in a proton-proton collision are visualized in Fig. 4.3.

Various event generation programs exist to simulate proton-proton collisions. The simulation of the hard scattering relies on a $2 \rightarrow 2$ matrix element to calculate $\hat{\sigma}_{ab \rightarrow X}(\hat{s}, \mu_R^2)$. Usually, the LO matrix element and PDFs are used for the simulation, and the results are scaled to NLO or NNLO using a K -factor derived from the ratio of the relevant cross sections. The program PYTHIA [58] uses a parton showering approach to model ISR and FSR. Other generators such as MADGRAPH [59] use alternate approaches which more accurately simulate additional hard radiation outside of the primary hard scattering process, but must be combined with a pro-

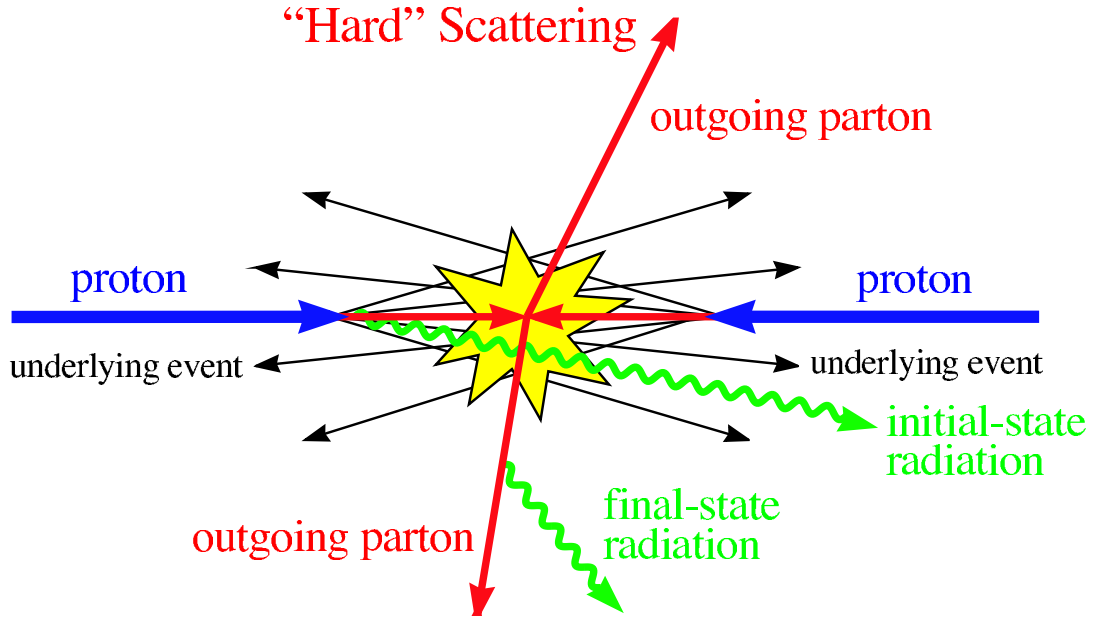


Figure 4.3: An illustration of a proton-proton collision, showing the hard scattering, ISR, FSR, and the underlying event [57].

gram like PYTHIA for showering from soft and collinear radiation. POWHEG [60, 61] is a generator which uses NLO matrix elements and PDFs by matching them with parton shower contributions to prevent double counting. After the initial event generation, MADGRAPH and POWHEG are interfaced with PYTHIA for hadronization. In addition, a specialized program called TAUOLA [62] can be applied for accurate handling of tau lepton decays.

4.2 Detector Simulation

Monte Carlo (MC) simulation is used to model the response of the CMS detector to a proton-proton collision event, using the final state particles output by event generators. The progression of each particle through the detector is tracked using the GEANT4 software [63, 64]. The geometry of every subdetector and subsystem is

carefully simulated to ensure the accuracy of the simulation. The conditions of the real detector, including alignment and calibration changes, are measured periodically and stored in a database which can be used to configure the simulation. GEANT4 includes customizable physics lists containing models of various physical processes to simulate the interactions of particles with matter. The models of electromagnetic processes are generally precise, while the models of hadronic interactions have larger uncertainties. The detector simulation creates simulated hits or “SimHits” for each subdetector from the deposition of energy or charge based on those interactions. The effects of photodetectors and readout electronics on these SimHits are then simulated, mimicing the real subdetectors’ data acquisition processes. The ROOT software library [65] is used to store and analyze both the simulated and observed data.

4.3 Tracks and Vertices

Hits from the different tracker subsystems are reconstructed into charged-particle tracks using the Combinatorial Track Finder (CTF) algorithm [66]. CTF is an iterative algorithm which first identifies the tracks that are the easiest to find, in order to remove the associated hits from consideration. This reduces the complexity of identifying the tracks that are more difficult to find, which is done in subsequent iterations. Each iteration follows the same four-step procedure, varying the type of seed used and the selection criteria applied.

1. A seed is generated using only a few hits.

2. Additional hits are added to the track based on the extrapolated trajectory of the seed.
3. The parameters of the track are estimated using a fit which considers all hits in the trajectory.
4. Selection criteria are applied to determine the quality of the track, and tracks which do not pass the selection are excluded.

The types of seeds are categorized based on the number of hits included and the source of those hits. Initial iterations use pixel triplet and pair seeds, created from three and two pixel hits, respectively. These are the highest-quality seeds and are used to reconstruct prompt tracks (those emitting from primary vertices near the IP). A subsequent iteration uses a mixed triplet seed, containing 1–3 pixel hits and <3 strip hits. This iteration typically finds displaced tracks from heavy flavor decays, nuclear interactions, and photons which convert to e^+e^- pairs in the tracker. The final iterations use strip pair seeds, consisting of two matched hits from the strip detectors, usually generated by charged particles which did not enter the pixel detector.

The iterations that use seeds with strip hits may also find prompt tracks which lack pixel hits. The specific sequence of iterations has been modified several times to improve the computing and physics performance of CMS tracking [67]. Table 4.1 lists the sequence used during the 2012 run. The selection criteria for each iteration are given, including cuts on p_T , transverse impact parameter d_0 , and longitudinal impact parameter z_0 . Some of the z_0 cuts are in terms of σ , the length of the beam

spot in the z -direction as determined by a Gaussian fit.

step	seed type	seed subdetectors	p_T [GeV/ c]	d_0 [cm]	$ z_0 $
0	triplet	pixel	>0.6	<0.02	$<4.0\sigma$
1	triplet	pixel	>0.2	<0.02	$<4.0\sigma$
2	pair	pixel	>0.6	<0.015	<0.09 cm
3	triplet	pixel	>0.3	<1.5	$<2.5\sigma$
4	triplet	pixel/TIB/TID/TEC	$>0.5-0.6$	<1.5	<10.0 cm
5	pair	TIB/TID/TEC	>0.6	<2.0	<10.0 cm
6	pair	TOB/TEC	>0.6	<2.0	<30.0 cm

Table 4.1: The sequence of tracking iterations used during the 2012 run, including information on the seeds and selection criteria used in each step [67].

The reconstructed tracks are used to reconstruct the primary vertices from the event [66]. This includes both the main hard scatter vertex and additional vertices from pileup collisions. First, selection requirements are imposed on the tracks, in order to consider only prompt tracks near the IP. The selection requirements include cuts on the significance of d_0 , the number of pixel and strip hits in the track, and the normalized χ^2 from the fit of the track trajectory. The tracks which pass the selection requirements are clustered together using their z -coordinates, determined at each track’s closest approach to the beam spot. A deterministic annealing algorithm is used to perform this clustering, in which each track may have a different probability to be associated with each vertex. The algorithm uses analogues of statistical mechanics quantities, slowly reducing the “temperature” and minimizing the “free energy” during each temperature iteration.

Once the deterministic annealing algorithm produces a list of vertex candidates, an adaptive vertex fitter is applied to each vertex candidate. This fitter weights each track in the vertex based on the agreement between the track and

vertex positions. Using those weights, it fits the parameters of the vertex, including the (x, y, z) position, the covariance matrix, and the number of degrees of freedom, $n_{\text{dof}} = -3 + 2 \sum w_i$, where w_i is the weight for the i th track. The variable n_{dof} can be used to select vertices which correspond to actual proton-proton interactions, as it is closely related to the number of tracks compatible with the vertex.

Several requirements are applied to the reconstructed primary vertices to ensure high quality [68]. The number of degrees of freedom n_{dof} must be greater than 4. The longitudinal position z of the vertex must obey $|z| < 24$ mm, and the transverse position ρ must obey $|\rho| < 2$ mm. The other reconstructed objects described below can be assigned to a reconstructed vertex. The closest vertex to the track associated with the object is chosen as its vertex. For electrons, the GSF track is used; for muons, the best track is used; and for hadronically decaying tau leptons, the track of the leading charged hadron is used.

4.4 Particle Flow

The CMS experiment uses a technique called particle flow (PF) to combine information from all subdetectors in order to identify all stable particles in each event [69]. As described at the beginning of this chapter and shown in Fig. 4.1, each type of stable particle is expected to create signals in a certain subset of the subdetectors. The performance of the PF algorithm was validated using early CMS data [70, 71], along with newer CMS data for more recent versions of the algorithm [72], demonstrating significant improvement over simpler approaches. The reconstructed

particles are known as PF candidates, which can be treated as input particles by the various high-level reconstruction algorithms.

The RecHits from the local reconstruction process are used to create the basic elements for this technique: tracks and clusters. Charged-particle tracks are created from tracker RecHits using an iterative algorithm as described in Sec. 4.3, and muon tracks are created from muon system RecHits. Calorimeter energy deposits are grouped into clusters by identifying seed hits as local energy maxima exceeding a certain threshold, and then adding neighboring hits with energy above subsystem-specific thresholds meant to eliminate photodetector noise. Further removal of noise from the calorimeters is performed by rejecting clusters with characteristics matching those expected from leading sources of noise. Tracks and clusters are associated together using a linking algorithm that determines if they were likely produced by the same particle. The algorithm considers a possible link between each element based on the η - ϕ distance between a charged-particle track and a cluster, accounting for propagation in the magnetic field, or between two clusters. For links between a charged-particle track and a muon track, the χ^2 value from a global fit is used as the link distance. Groups of elements are associated based on minimizing the link distance and are called “blocks”.

PF reconstruction algorithms classify the blocks as different types of particles. When a block is classified as a certain type of particle, it is removed from the list of unclassified blocks. A global muon block is accepted as a PF muon if the momentum of the combined charged-particle and muon tracks agrees with the momentum of the charged-particle track alone. The energy expected to have been deposited by the

PF muons in the calorimeters due to minimum ionization is subtracted from the clusters. The remaining charged-particle tracks are checked for compatibility with electrons, which tend to radiate energy via bremsstrahlung, causing the curvature of their tracks to increase with radial distance in the tracker. A Gaussian Sum Filter (GSF) is used to fit the compatible tracks in order to match their trajectories with ECAL clusters, and the combination of a track and one or more clusters is classified as a PF electron. Using a mixture of Gaussians to select the electron track better accounts for the energy loss from bremsstrahlung, as compared to the standard CMS track finding procedure [73].

The remaining tracks are linked to clusters to form PF charged hadrons, when the total cluster energy is similar to but smaller than the total track momentum. More than one track can link to a given cluster, but for a given track, only the link to the closest cluster is kept. This reflects the coarser segmentation of the calorimeter system as compared to the tracker. In cases where the total cluster energy is significantly smaller than the total track momentum, additional PF muons may be found using tracks from the block, and some tracks may be classified as fake and removed from consideration. Finally, an excess of energy in the clusters, above the total track momentum, is assumed to come from neutral particles. Any excess energy in the ECAL is typically classified as a PF photon. If additional excess energy remains after this, a PF neutral hadron is created. The remaining clusters not linked to any tracks are used to create PF photons in ECAL and PF neutral hadrons in HCAL.

4.5 Electrons

The electron candidates reconstructed by the PF algorithm are considered to be “tracker-driven” [74]. This approach is suitable for low- p_T electrons and electrons produced by jets. For higher- p_T electrons, an alternative “ECAL-driven” approach is used, distinct from the PF framework. ECAL clusters are grouped into superclusters to account for bremsstrahlung photons radiated by electrons as they traverse the tracker, as well as the spread of energy in the ϕ -direction due to the magnetic field [75]. Similarly to the PF algorithm, these superclusters are matched with track seeds and a GSF is used to reconstruct the trajectory of the electron track. The lists of tracker-driven and ECAL-driven electron candidates are compared to avoid double counting.

Quality cuts on various kinematic variables are applied to the GSF electron candidates to identify whether or not they are genuine electrons [76]. A set of cut values is called a working point, and multiple working points are defined based on the strictness of the cut values. The η width of the supercluster, $\sigma_{i\eta\eta}$, is taken from the covariance matrix of a weighted difference between the η positions of the included crystals and the seed cluster. A modified η variable which accounts for the crystal spacing is used, and each crystal’s contribution is weighted using the logarithm of the ratio of its energy to the seed cluster energy [77]. The differences in position of the supercluster, (η_{sc}, ϕ_{sc}) , and the extrapolated track, $(\eta_{in}^{\text{extrap}}, \phi_{in}^{\text{extrap}})$, are defined as $|\Delta\eta_{in}| = |\eta_{sc} - \eta_{in}^{\text{extrap}}|$ and $|\Delta\phi_{in}| = |\phi_{sc} - \phi_{in}^{\text{extrap}}|$. The leakage energy H in the HCAL tower located behind the ECAL seed cluster is compared to the energy of

the seed cluster E in the variable H/E . The transverse and longitudinal impact parameters of the electron track compared to its associated vertex, d_0^{vtx} and z_0^{vtx} , are used. The variable $|1/E - 1/p|$, comparing the electron energy and momentum, is also considered. Finally, isolation is computed by summing the p_{T} of charged hadron (CH), neutral hadron (NH), and photon (γ) PF candidates within a cone of $\Delta R < 0.3$ of the electron candidate, including a correction for contributions from pileup based on the median energy per area ρ and the effective area of the electron A_{eff} :

$$I_{\text{e}}^{\text{PF}} = \sum_{\Delta R < 0.3} p_{\text{T}}^{(\text{CH})} + \max \left(\sum_{\Delta R < 0.3} p_{\text{T}}^{(\text{NH})} + \sum_{\Delta R < 0.3} p_{\text{T}}^{(\gamma)} - \rho A_{\text{eff}}, 0 \right). \quad (4.2)$$

The relative isolation $I_{\text{e}}^{\text{PF}}/p_{\text{T}}^{(\text{e})}$, scaled by the p_{T} of the electron, is used for the quality cuts.

Two working points for the quality cuts described in Sec. 4.5 are used to identify reconstructed electrons in the analysis performed in this dissertation. Table 4.2 lists the various requirements for each working point. Electrons are required to be in the barrel with $|\eta| < 1.444$ or in the endcap with $1.56 < |\eta| < 2.5$. The leading electron in the $e\tau_{\text{h}}$ channel is selected with the medium working point. In addition to these quality cuts, it is required to have $p_{\text{T}} > 30 \text{ GeV}$ to match the specifications of the HLT criterion used to collect the PD, as described in Sec. 5.1.1. The loose working point is used to veto additional electrons and for selections in some samples used for background estimations. The less restrictive kinematic cut $p_{\text{T}} > 20 \text{ GeV}$ is applied to electrons identified with the loose working point.

Cut Variable	Cut Value			
	Medium		Loose	
	Barrel	Endcap	Barrel	Endcap
$I_e^{\text{PF}}/p_T <$	0.15	0.15	0.15	0.15
$\sigma_{i\eta i\eta} <$	0.01	0.03	0.01	0.03
$ \Delta\phi_{\text{in}} <$	0.06	0.03	0.15	0.10
$ \Delta\eta_{\text{in}} <$	0.004	0.007	0.007	0.009
$H/E <$	0.12	0.10	0.12	0.10
$ d_0^{\text{vtx}} <$	0.02	0.02	0.02	0.02
$ d_z^{\text{vtx}} <$	0.1	0.1	0.2	0.2
$ 1/E - 1/p <$	0.05	0.05	0.05	0.05

Table 4.2: The quality cuts for the medium and loose working points of the electron identification.

4.6 Muons

The PF muon candidates are used for muon reconstruction. In addition, two supplementary methods are used to reconstruct muons [78]. The first method considers all charged-particle tracks from the tracker, above minimal p_T and p cuts, and creates a “tracker muon” from any track whose extrapolated position matches a track segment in the muon system. This method is efficient for low-momentum muons. The second method produces “global muons”. This method starts with a standalone muon from a track segment in the muon system and finds a matching track from the tracker. A global muon fit is then performed using both the muon system and tracker tracks, which can provide better momentum resolution for high- p_T muons. The global and track muon candidates, along with any remaining standalone muons which were not matched to a charged-particle track, are combined into one collection in order to prevent double counting.

As with electrons, working points are defined based on sets of quality cuts with

different strictness. These cuts can include minimum numbers of muon system hits and segments, as well as minimum numbers of pixel hits and overall tracker hits. The reduced χ^2 from the global muon fit is considered. The distances between the primary vertex and the transverse and longitudinal impact parameters d_0 and d_z of the tracker track are also used. The isolation is calculated using PF candidates within a cone of $\Delta R < 0.4$ of the muon:

$$I_{\mu}^{\text{PF}} = \sum_{\Delta R < 0.4} p_{\text{T}}^{(\text{CH})} + \max \left(\sum_{\Delta R < 0.4} p_{\text{T}}^{(\text{NH})} + \sum_{\Delta R < 0.4} p_{\text{T}}^{(\gamma)} - \Delta\beta \sum_{\Delta R < 0.4} p_{\text{T}}^{(\text{PU})}, 0 \right). \quad (4.3)$$

A $\Delta\beta$ pileup (PU) correction is applied using the p_{T} of PU particles, which are identified as charged PF candidates from a different vertex than the muon. The $\Delta\beta$ factor is assigned a value of 0.5 based on the expected ratio of charged to neutral particles in pileup [70]. Again, cuts are made on the relative isolation $I_{\mu}^{\text{PF}}/p_{\text{T}}^{(\mu)}$. The selections on these quantities are intended to minimize the contribution from cosmic ray muons, muons from heavy flavor decays, and leakage from hadronic showers. They also ensure precise measurement of the muon p_{T} .

Two working points for these quality cuts are used to identify reconstructed muons in the analysis performed in this dissertation. Table 4.3 lists the various requirements for each working point. The leading muon in the $\mu\tau_{\text{h}}$ channel is selected with the tight working point. In addition to these quality cuts, it is required to have $p_{\text{T}} > 30 \text{ GeV}$ and $|\eta| < 2.1$ to match the specifications of the HLT criterion used to collect the PD, as described in Sec. 5.1.1. The loose working point is used to veto additional muons and for selections in some samples used for background estimations. Less restrictive kinematic cuts $p_{\text{T}} > 20 \text{ GeV}$ and $|\eta| < 2.4$ are applied

to muons identified with the loose working point.

Working Point	
Tight	Loose
PF muon Global muon $I_{\mu}^{\text{PF}}/p_{\text{T}} < 0.12$ $d_0 < 0.2 \text{ cm}$ $d_z < 0.5 \text{ cm}$ Global track fit $\chi^2/n_{\text{dof}} < 10$ Global track fit $n_{\text{muon segment}} > 0$ $n_{\text{hits}}(\text{pixel}) > 0$ $n_{\text{layers}}(\text{tracker}) > 5$ $n_{\text{stations}}(\text{muon}) > 1$	PF muon Global muon OR tracker muon $I_{\mu}^{\text{PF}}/p_{\text{T}} < 0.3$

Table 4.3: The quality cuts for the tight and loose working points of the muon identification.

4.7 Jets

Due to QCD confinement, strongly interacting particles (quarks and gluons) cannot exist in a bare state. They immediately form multiple color singlet bound states, hadrons, using the energy from the gluon field in a process called hadronization. These hadrons tend to be produced in a narrow spray, which is called a jet [79]. To reconstruct a jet, the component particles must be clustered together. CMS uses the anti- k_{T} algorithm [80] to perform this clustering. This algorithm is a specific

case of a generalized iterative cone algorithm, using the following equations:

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.4)$$

$$\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2, \quad (4.5)$$

$$d_{iB} = p_{Ti}^{2p}. \quad (4.6)$$

Two distance variables are defined: the distance between particles i and j , d_{ij} , and the distance between particle i and the beam, d_{iB} . These distances are weighted by the p_T of the particles as indicated, and R is a size parameter. At each iteration, both distance variables are calculated for all particles. If the minimum distance is d_{iB} , particle i is considered to be a fully-clustered jet and removed from the list of particles. Otherwise, particles i and j from the minimum d_{ij} are grouped together in the particle list. The algorithm continues to iterate until the particle list is empty. The anti- k_T algorithm is a special case of these equations with parameter $p = -1$. It has many desirable properties, including infrared and collinear safety, and the creation of circular jets with radius R as shown in Fig. 4.4, which aids in the jet calibration.

CMS chooses the value $R = 0.5$ for the size parameter and uses PF candidates for clustering, producing PF jets [69, 70]. Alternative jet methods use CaloTowers and charged-particle tracks separately or together, but these do not perform as well as PF jets. Typically, 65% of jet energy goes into charged particles, 25% goes into photons, and 10% goes into neutral hadrons. Using PF candidates takes advantage of the excellent energy and position resolution for charged particles and photons provided by the combination of the tracker and the calorimeters in the PF algorithm.

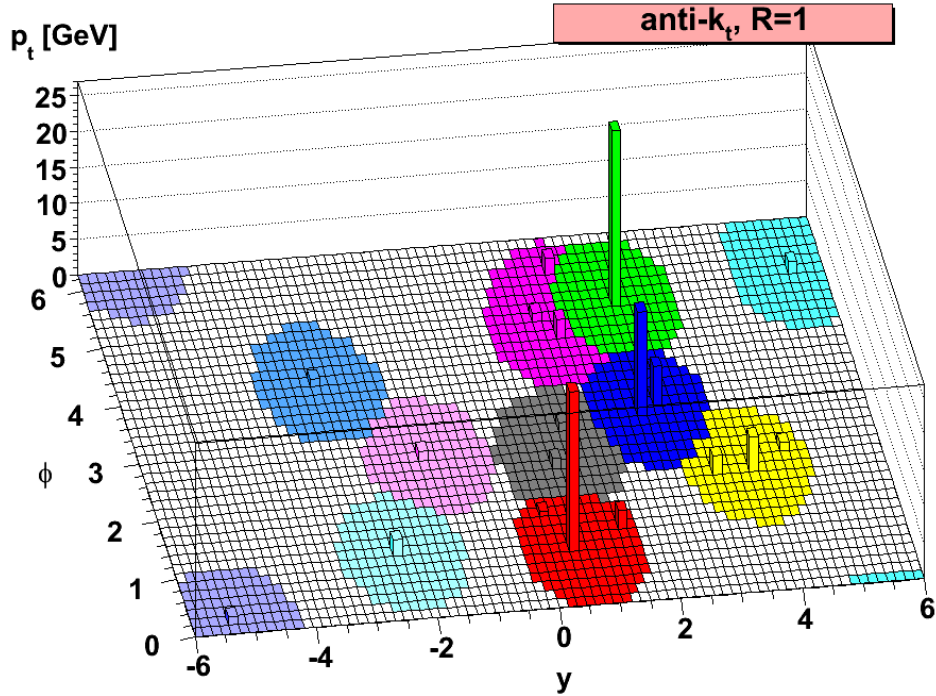


Figure 4.4: An example of jets reconstructed with the anti- k_T algorithm, demonstrating the characteristic circular shape [80].

Once a jet is reconstructed, several important types of corrections are applied to its energy response [81]. First, minimum bias events are used to estimate the contributions from electronic noise and pileup, and those are subtracted in the offset correction. Next, the energy response is made uniform in η using the multiplicative relative correction, derived from simulated dijet events. Finally, the multiplicative absolute correction is derived from observed γ/Z + jets events, exploiting the precise energy resolution of the ECAL and the tracker, and applied to make the energy response uniform in p_T .

To ensure the quality of reconstructed jets used in data analysis, a set of variables is used for PF jet identification [82]. These variables include: the fraction

of neutral hadrons in the jet f_{NH} , the fraction of neutral EM particles f_{γ} , the fraction of charged hadrons f_{CH} , the fraction of charged EM particles f_{EM} , the number of constituents $n_{\text{constituents}}$, and the multiplicity of charged particles n_{charged} . Multiple sets of cuts on these variables are defined as working points [83]. The loose working point of the PF jet identification algorithm is used to identify reconstructed jets in the analysis performed in this dissertation. The requirements for the loose working point are summarized in Table 4.4. To eliminate many jets from pileup interactions, each selected jet is required to have $p_{\text{T}} > 30 \text{ GeV}$. The cut $|\eta| < 2.4$ is also applied, in order to include only jets measured in the best-performing regions of the detector.

Cut Variable	Cut Value
	Loose
$f_{\text{CH}} >$	0.0
$f_{\text{NH}} <$	0.99
$f_{\gamma} <$	0.99
$f_{\text{EM}} <$	0.99
$n_{\text{charged}} >$	0
$n_{\text{constituents}} >$	1

Table 4.4: The quality cuts for the loose working point of the jet identification.

4.8 Taus

Tau leptons decay into hadrons approximately 64.76% of the time [8]. These hadronic decays produce objects similar to jets, but typically narrower and more isolated. For this reason, PF jets are used as the basis for reconstructing hadronically decaying tau leptons (hadronic taus or τ_{h} s). The Hadron Plus Strips (HPS) algorithm is used to identify tau leptons [84, 85]. The vast majority of τ_{h} decays

consist of a tau neutrino ν_τ , one or three charged hadrons h^- that are either π^- or K^- , and zero or more neutral hadrons π^0 that almost immediately decay to two photons. The HPS algorithm only considers the visible decay products, so the ν_τ is ignored. Table 4.5 lists the leading hadronic decays, which account for 61.56% of all tau lepton decays. Some of these decays include intermediate hadronic resonances, as noted in the table. In the decays with indicated resonances, the non-resonant contribution is negligible.

Decay	Resonance	Mass (MeV/c ²)	Branching fraction (%)
$\tau^- \rightarrow h^- \nu_\tau$			11.53%
$\tau^- \rightarrow h^- \pi^0 \nu_\tau$	ρ^-	775	25.95%
$\tau^- \rightarrow h^- \pi^0 \pi^0 \nu_\tau$	a_1^-	1230	9.52%
$\tau^- \rightarrow h^- h^+ h^- \nu_\tau$	a_1^-	1230	9.80%
$\tau^- \rightarrow h^- h^+ h^- \pi^0 \nu_\tau$			4.76%

Table 4.5: The leading hadronic decays of tau leptons, including branching fractions and intermediate hadronic resonances [8]. The symbol h^- can be either π^- or K^- .

The strips in the HPS algorithm consist of PF photon candidates. Starting with the most energetic photon in the PF jet, the strip is built using an iterative search for other photons within a range 0.20×0.05 in η - ϕ around the center of the strip. Each iteration accepts the most energetic photon found and then recalculates the four-momentum of the updated strip. The strip is complete once no remaining photons are found in the given window, and it is kept if it passes a minimum p_T cut. The procedure is repeated with any remaining ungrouped photons in the jet. The formation of strips from photons accounts for spreading of their energy due to the effect of the magnetic field on conversions in the tracker.

Using the constituents of the PF jet, all combinations of strips with one

or three charged hadrons are tested, with the charged hadrons assumed to be pions. All strips and charged hadrons must be found within a cone of $\Delta R = \max(\min(0.10, 3.0/p_T^{\tau_h}), 0.05)$, where $p_T^{\tau_h}$ is the p_T of the τ_h candidate in GeV. The invariant mass of each combination, calculated from all constituent strips and charged hadrons, is required to be within a certain range based on the tau lepton mass and any hadronic resonance in the decay [86]. Several decay mode topologies are included in the algorithm, as depicted in Fig. 4.5 and listed below:

1. Single hadron, when no strips are found.
2. One hadron + one strip, when the π^0 decay creates one strip from two narrowly separated photons. The invariant mass of the τ_h candidate, M_{τ_h} , must be in the range $0.3 < M_{\tau_h} < \max(1.3, \min(1.3\sqrt{p_T^{\tau_h}/200}, 2.1))$ GeV.
3. One hadron + two strips, when the π^0 decay creates one strip from two narrowly separated photons. In this case, the invariant mass of the two strips combined, M_{strips} , must be in the range $50 < M_{\text{strips}} < 200$ MeV, and M_{τ_h} must be in the range $0.4 < M_{\tau_h} < \max(1.2, \min(1.2\sqrt{p_T^{\tau_h}/200}, 2.0))$ GeV.
4. Three hadrons, which requires all three charged hadron candidates to originate from the same primary vertex and to have the appropriate electric charges. M_{τ_h} must be in the range $0.8 < M_{\tau_h} < 1.5$ GeV.

If more than one combination of PF constituents passes these decay mode finding requirements, the combination with the highest $p_T^{\tau_h}$ is selected.

Isolation is an important tool in discriminating between τ_h s and jets. The

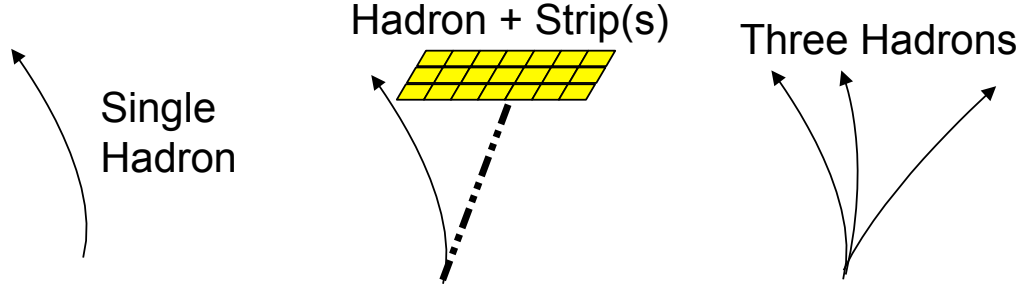


Figure 4.5: A simple diagram of the different hadronic tau decay mode topologies reconstructed by the HPS algorithm [87].

isolation variable is computed using charged hadron and photon PF candidates within a cone of $\Delta R < 0.5$ around the τ_h candidate. A $\Delta\beta$ PU correction is applied using PU particles within a cone of $\Delta R < 0.8$ which originate from a different vertex than the τ_h candidate, with the factor $\Delta\beta = 0.4576$ [87].

$$I_{\tau_h}^{\text{PF}} = \sum_{\Delta R < 0.5} p_T^{(\text{CH})} + \max\left(\sum_{\Delta R < 0.5} p_T^{(\gamma)} - \Delta\beta \sum_{\Delta R < 0.8} p_T^{(\text{PU})}, 0\right). \quad (4.7)$$

The HPS algorithm uses the absolute isolation $I_{\tau_h}^{\text{PF}}$ for the different working point quality cuts. The isolation discrimination also requires that each track associated with the τ_h contain at least three hits in the tracker. In addition to isolation, it is necessary to discriminate against electrons and muons which are misidentified as τ_h s. The electron-tau discriminator uses a multivariate (MVA) approach, training boosted decision trees (BDTs) using numerous variables depending on different cases of τ_h . These cases include: the possible association of the primary charged hadron in the τ_h with a GSF track; the possible association of the τ_h with a GSF electron within a cone of $\Delta R < 0.3$; whether or not the τ_h includes strips; and whether the τ_h η coordinate lies in the EB or EE range. Multiple trainings for the anti-electron MVA discriminator were performed using simulated events and multiple working

points are defined. The muon-tau discriminator uses a cut-based approach, with multiple sets of cuts and working points defined. The cuts include requirements to minimize the activity in the muon system in the direction of the τ_h and to veto MIP muons based on energy and momentum.

The working points for each discriminator are listed in Table 4.6 for each channel of the analysis performed in this dissertation. The loose working point for the combined isolation requires $I_{\tau_h}^{\text{PF}} < 2 \text{ GeV}$. A tighter working point for the muon-tau discriminator is used in the $\mu\tau_h$ channel to increase the rejection of muons misidentified as tau leptons from, e.g., the $Z \rightarrow \mu\mu$ process. The loose working point for the electron-tau discriminator was found to be optimal for both channels. In addition to these discriminators, each reconstructed tau lepton is required to have $p_T > 30 \text{ GeV}$ and $|\eta| < 2.3$.

Discriminator	Working Point	
	$e\tau_h$	$\mu\tau_h$
Isolation	Loose (3 hits)	Loose (3 hits)
$e\text{-}\tau$	Loose MVA (v3)	Loose MVA (v3)
$\mu\text{-}\tau$	Loose (v2)	Tight (v2)

Table 4.6: The working points for the different tau discriminators used in the tau lepton identification.

4.9 b-tagging

Bottom quarks are associated with many interesting physical signatures, including decays of top quarks and Higgs bosons, as well as many BSM theories, including supersymmetry, that privilege or otherwise relate to third-generation fermions.

Jets generated by bottom quark hadronization (b-jets) can be identified using certain properties which set them apart from jets from lighter quarks or gluons [88]. Bottom quarks form hadrons with relatively long lifetimes and decay products that tend to have high p_T . Numerous algorithms have been created to identify b-jets based on information from tracks or reconstructed secondary vertices from b-hadron decays. The most successful of these algorithms is the Combined Secondary Vertex (CSV) algorithm, which uses secondary vertices and adds information from the tracks. The CSV algorithm calculates a likelihood-based discriminator to separate b-jets and other jets. Loose, medium, and tight working points are defined for this discriminator, based on setting the probability of misidentifying a light quark or gluon jet as a b-jet to be 10%, 1%, and 0.1%, respectively [89]. In this dissertation, the loose working point is used to select b-tagged jets, which requires the discriminator calculated by the CSV algorithm to have a value greater than 0.244. The loose working point has an efficiency of $\sim 85\%$.

For a given jet, the CSV algorithm subjects the tracks within that jet to additional purity requirements, beyond those specified in Sec. 4.3. The track must fall within a cone of $\Delta R < 0.3$ relative to the direction of the jet and the distance between the track and the jet at closest approach must be less than 0.07 cm. The track must include at least two pixel hits and at least eight total tracker hits, with $\chi^2/n_{\text{dof}} < 5$. The impact parameters between the track and the primary vertex must satisfy $d_0 < 0.2$ cm and $d_z < 17$ cm. Finally, the track must have $p_T > 1$ GeV/c and decay length less than 5 cm, where decay length is the distance between the primary vertex and the closest approach of the track and the jet. The significance of the

track's impact parameter S_{ip} , defined as the value of the impact parameter divided by its uncertainty, is a powerful observable. The number of high-quality tracks in the jet is also used.

Secondary vertices are reconstructed using the adaptive vertex fitter, which was described in Sec. 4.3. Candidates are rejected if $\geq 65\%$ of their tracks are also associated with the primary vertex; if they are outside a cone $\Delta R < 0.5$ with respect to the jet direction; or if they are radially separated from the primary vertex by more than 2.5 cm with an invariant mass close to the K^0 mass. The CSV algorithm assigns jets into one of three categories based on the presence of a secondary vertex: real vertex, pseudo-vertex, and no vertex. In the pseudo-vertex case, the algorithm uses tracks with $S_{ip} > 2$ to create an effective secondary vertex when the adaptive vertex fitter fails. In the no-vertex case, the algorithm defaults to the track-based variables described previously. Otherwise, the vertex variables used include:

- The significance of the flight distance between the secondary and primary vertices in the transverse plane.
- The mass of the secondary vertex.
- The number of tracks in the secondary vertex.
- The ratio between the energy in the secondary vertex tracks and the energy in all the tracks in the jet.
- The $\Delta\eta$ values between each the secondary vertex track and the jet.
- The transverse impact parameter significance for the track that pushes the

invariant mass of the secondary vertex above $1.5 \text{ GeV}/c^2$, the threshold for charm quarks. This is calculated by sorting the tracks by S_{ip} and combining them one by one.

4.10 Missing Transverse Energy

Missing transverse momentum is reconstructed as the negative vector sum of \vec{p}_{T} for all PF candidate particles in the event: $\vec{\cancel{E}}_{\text{T}} = -\sum_i \vec{p}_{\text{T}}^{(i)}$. The magnitude of this quantity is called the missing transverse energy, denoted as \cancel{E}_{T} . Because the incident proton beams have momentum only in the z direction, any imbalance of measured momentum in the transverse plane indicates particles which were not detected. Particles which produce \cancel{E}_{T} include neutrinos and hypothetical particles such as the LSP in RPC SUSY. The use of PF candidates allows the contributions to \cancel{E}_{T} from photon, electron, and muon candidates, jets, and unclustered energies to be considered separately. This is important for several types of corrections, which improve the measurement of genuine \cancel{E}_{T} from invisible particles [90].

The jet energy scale corrections are applied to the jet contributions in the \cancel{E}_{T} calculation, significantly reducing the effects of calorimeter thresholds and nonlinearities on the estimation of the \cancel{E}_{T} magnitude. These corrections are applied for jets, with the requirements $f_{\text{EM}} < 0.9$ and $p_{\text{T}} > 10 \text{ GeV}$ used to exclude electrons from the jet collection. Pileup interactions tend to induce non-genuine \cancel{E}_{T} , also due to the effects of calorimeter thresholds and response nonlinearities. The correction to reduce the contribution from pileup is parameterized as a function of the vector

sum of \vec{p}_T from the charged hadron candidates that are identified as pileup based on their association with vertices other than the primary vertex. A ϕ asymmetry is also induced by imperfect detector calibrations and alignment, scaling roughly linearly with the number of reconstructed vertices. This scaling is used to derive corrections which are applied separately to the two components of \cancel{E}_T , \cancel{E}_x and \cancel{E}_y , event-by-event. Various filters are applied to reduce contributions from calorimeter noise and the beam halo, based on the physical characteristics of these contributions [91].

Chapter 5: Data Analysis

5.1 Data Samples

5.1.1 Data

The data analyzed in this dissertation consist of 19.712 fb^{-1} of integrated luminosity, collected by the CMS detector during the 2012 run of the LHC. Figure 5.1 shows the total delivered, recorded, and validated integrated luminosity versus time throughout the 2012 run. Events are included in the validated data only if the LHC and the CMS detector were fully operational when the event was recorded. Events are selected from the single electron and single muon PDs, listed in Table 5.1, for the $e\tau_h$ and $\mu\tau_h$ channels if they pass the HLT criteria `HLT_Ele27_WP80` and `HLT_IsoMu24`, respectively. The criterion `HLT_Ele27_WP80` requires a reconstructed electron with $p_T > 27 \text{ GeV}$ using the identification working point with 80% efficiency. The criterion `HLT_IsoMu24` requires an isolated reconstructed muon with $p_T > 24 \text{ GeV}$. The average number of interactions per bunch crossing was 21 [44].

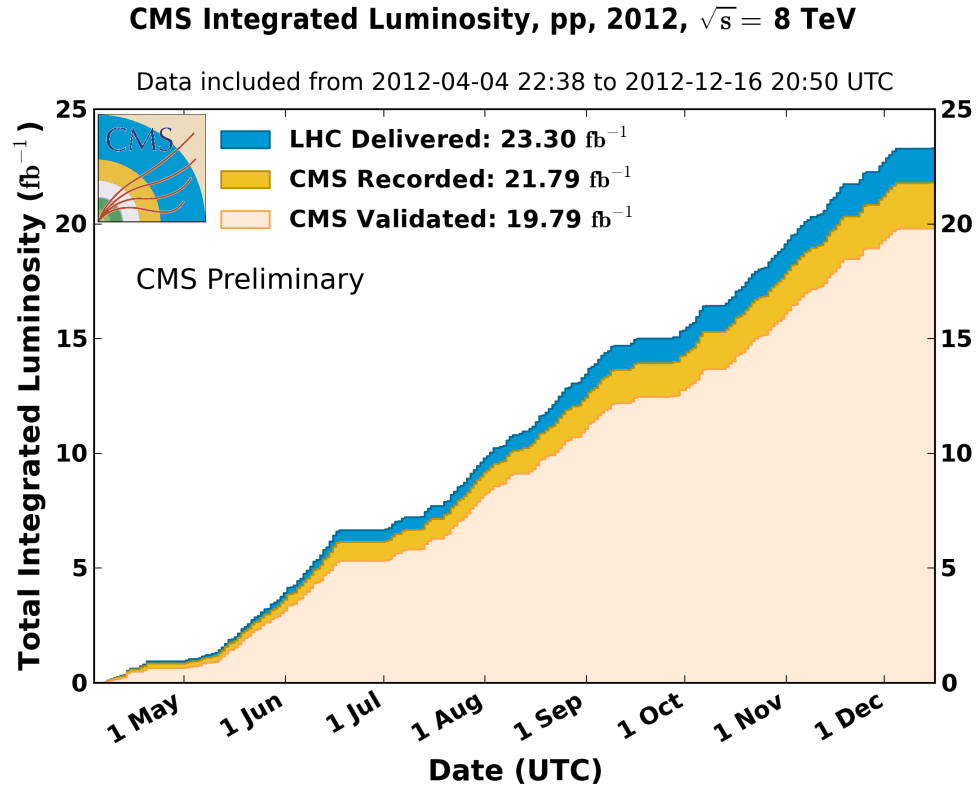


Figure 5.1: A plot of the total integrated luminosity over time for the 2012 run of the LHC, showing the luminosity delivered by the LHC (blue), recorded by the CMS detector (dark orange), and validated as good for physics analysis (light orange) [92].

5.1.2 Monte Carlo

MC simulation is used to study the SM background processes which can mimic the signatures of the signal processes. For some processes, the MC simulation is used for the estimation of the final yields and kinematic distributions. For other processes, the final yields and/or kinematic distributions may be estimated from control regions in the observed data. More details on the final background estimations are provided in Sec. 5.3. The leptoquark and top squark signal processes are also modeled using MC. A full list of the MC samples used in the analysis described in this dissertation

$e\tau_h$ channel
Dataset Name
/SingleElectron/Run2012A-22Jan2013-v1/AOD
/SingleElectron/Run2012B-22Jan2013-v1/AOD
/SingleElectron/Run2012C-22Jan2013-v1/AOD
/SingleElectron/Run2012D-22Jan2013-v1/AOD
$\mu\tau_h$ channel
Dataset Name
/SingleMu/Run2012A-22Jan2013-v1/AOD
/SingleMu/Run2012B-22Jan2013-v1/AOD
/SingleMu/Run2012C-22Jan2013-v1/AOD
/SingleMu/Run2012D-22Jan2013-v1/AOD

Table 5.1: The list of data samples used in each channel.

can be found in App. C.

The MADGRAPH v5.1.3.30 generator [59] is used to model the $t\bar{t}$, $W + \text{jets}$, and $Z + \text{jets}$ processes. For the $t\bar{t}$ process, an inclusive sample is generated including all decay modes of the two W bosons from the top quark decays. Exclusive samples with larger numbers of events are also generated, each including only one of three categories for the W decays: fully leptonic, semileptonic, and fully hadronic. For the $W + \text{jets}$ and $Z + \text{jets}$ processes, inclusive samples including any number of jets are generated. Exclusive samples with larger numbers of events are also generated separately for events with 1 jet, 2 jets, 3 jets, and 4 jets. Single top quark production is modeled with the POWHEG 1.0 r138 [93, 94, 95] generator. The PYTHIA v6.4.24 generator [58] is used to model the diboson processes, including the production of W^+W^- , $W^\pm Z$, and ZZ , collectively denoted as VV . The cross sections for the $t\bar{t}$ and single top quark processes are calculated at next-to-next-to-leading logarithmic (NNLL) accuracy [96]. The cross sections for the inclusive $W + \text{jets}$ and $Z + \text{jets}$ processes are calculated at NNLO accuracy, while the cross sections for the exclusive

production with a specific number of jets are calculated at LO accuracy [97]. For the VV processes, the cross sections are calculated at NLO accuracy [98].

The leptoquark and top squark signal processes are generated using PYTHIA v6.4.24 [58]. The leptoquark pair production process is generated for $M_{\text{LQ}} = 200 \text{ GeV}$ to $M_{\text{LQ}} = 1000 \text{ GeV}$ in steps of 50 GeV, with the decay $\text{LQ} \rightarrow \tau b$ required. The cross section for leptoquark pair production is calculated at NLO accuracy [27]. The theoretical uncertainties on these cross sections have been discussed in Sec. 2.3 and are listed in Table 2.3. As mentioned in Sec. 2.4, the deviation between these LQ production cross sections and the corresponding top squark production cross sections is less than 2%, so the same cross sections can be used to normalize the samples for both signals. The simulation of the chargino-mediated λ'_{3jk} decay of the top squark is summarized in Table 5.2. Samples are generated for $M_{\tilde{t}} = 200 \text{ GeV}$ to $M_{\tilde{t}} = 900 \text{ GeV}$ in steps of 100 GeV, where $M_{\tilde{t}}$ is the top squark mass. The MSSM parameters for the event generation are $M_1 = 1000 \text{ GeV}$, $M_2 = 1000 \text{ GeV}$, $M_3 = 1500 \text{ GeV}$, $m_{\tilde{e}} = 1500 \text{ GeV}$, $m_{\tilde{q}} = 1500 \text{ GeV}$, $m_{\tilde{\nu}} = 2000 \text{ GeV}$ and $\tan(\beta) = 40$. Here, M_1 is the bino mass term, M_2 is the wino mass term, and M_3 is the gluino mass term. The λ'_{321} coupling is selected for the generation; because reconstructed light-quark jets are generally indistinguishable regardless of the flavor of the initial quark, these samples can be used to model signals for all λ'_{3jk} couplings with $j, k = 1, 2$.

To account for known discrepancies between the MC simulation and the observed data, various correction factors are applied to the simulated samples. Corrections for the efficiency of muon trigger criteria, identification, and isolation are

$M_{\tilde{t}}$ [GeV]	$M_{\tilde{\chi}^{\pm}}$ [GeV]	μ [GeV]	λ'_{321}	N_{events}
200	100	100.98	1	539,000
300	200	201.69	1	50,000
400	300	302.49	1	50,000
500	400	403.46	1	50,000
600	500	504.73	1	50,000
700	600	606.55	1	50,000
800	700	709.47	1	50,000
900	800	815.16	1	50,000

Table 5.2: A summary of the generation of signal samples for the chargino-mediated λ'_{3jk} decay of the top squark. The chargino mass is determined by the value of μ .

calculated using tag and probe methods [99]. Similar correction factors are calculated for electrons [100, 101]. These correction factors depend on the lepton p_T and η , and they are usually consistent with 1 within the uncertainties. The chosen working points for the hadronic tau lepton discriminators were found not to need any corrections in the simulation [102]. Correction factors for b-tagging and mistagging efficiencies, typically in the range 5–10%, are applied using the method called “event reweighting using scale factors and MC b-tagging efficiencies” with the EPS13 prescription [103]. When \cancel{E}_T is used to define control regions, the energies of the PF jet candidates used to calculate the \cancel{E}_T in the simulation are smeared to improve the agreement with the observed jet energy resolution [81]. This improves the modeling of \cancel{E}_T in the simulation [104]. Each simulated event is weighted based on the true number of proton-proton interactions in the simulated collision in order to ensure that the overall distribution of interactions in the simulated sample matches the observed data.

5.2 Selection and Optimization

The events analyzed in this dissertation are selected based on requirements for the numbers and types of different reconstructed objects and the relationships between those objects within each event. The selection criteria are applied in three stages. The most basic criteria are applied first in the preselection. This stage is used to validate the Monte Carlo background simulations by examining various relevant kinematic distributions. Second, the main selection applies additional criteria, which more closely address the expected properties of signal events. Kinematic distributions in the main selection are used to optimize the cuts for the final selection, in order to reduce the contributions from background while enhancing the signal. These optimizations maximize the expected limit on the mass of the new particle assuming that only background processes occur, and they use the MC background simulations. Separate final selections are defined for the leptoquark search and the top squark search.

5.2.1 Preselection

At the preselection stage, a single primary well-identified light lepton ℓ is required. In the $e\tau_h$ channel, this is an electron satisfying the medium working point and kinematic criteria described in Sec. 4.5. In the $\mu\tau_h$ channel, this is a muon satisfying the tight working point and kinematic criteria described in Sec. 4.6. In addition to the light lepton, the event must contain a hadronic tau identified by the HPS algorithm and associated discriminators as defined in Sec. 4.8. The

two objects, light lepton and hadronic tau, must have opposite charges, originate from the same vertex, and be separated by $\Delta R > 0.5$. Two jets are also required, satisfying the loose working point and kinematic criteria described in Sec. 4.7. The jets must be separated from the light lepton and the hadronic tau by $\Delta R > 0.5$.

Several cases of additional light leptons are vetoed. To suppress background from the $Z + \text{jets}$ process, events are rejected if they contain additional light leptons satisfying the loose working point and having the same flavor and opposite charge with respect to the primary lepton. This creates control regions, separate from the signal region, containing $Z \rightarrow \ell\ell + \text{jets}$ events, which will be used for background estimation. To avoid overlap between the two channels, events are rejected if they contain additional light leptons satisfying the primary working point and kinematic criteria for leptons (medium for electrons and tight for muons) and having the opposite flavor and opposite charge with respect to the primary lepton. This also defines an $e\mu$ control region which will be used for background estimation.

Figures 5.2 and 5.3 compare the observed data to the MC background simulation after the preselection in the $e\tau_h$ and $\mu\tau_h$ channels, respectively. All histograms in this dissertation show the overflow, if any, in the last bin. The event yields for the observed data and the expected SM backgrounds after the preselection are summarized in Table 5.3. The discrepancy between the observed data and simulated background yields in the $e\tau_h$ channel is due to the presence of QCD multijet events in the data. This discrepancy is primarily observed at low p_T and high η , as expected from QCD. This process is not modeled in the MC simulation because of the difficulty in simulating enough events to be comparable to the amount of data

collected by the CMS experiment in 2012. Instead, the yield from QCD is estimated using a same-sign/opposite-sign method which is described in detail in Sec. 5.3.2.1.

	$e\tau_h$ channel	$\mu\tau_h$ channel
W + jets	4221.6 ± 188.1	4846.3 ± 233.6
Z + jets	4766.7 ± 85.1	2369.1 ± 80.7
$t\bar{t}$	6272.2 ± 65.5	6430.5 ± 69.8
Single t	462.9 ± 14.4	512.3 ± 16.0
VV	223.4 ± 4.4	212.5 ± 4.6
QCD multijets	(2452.6 ± 512.1)	—
Total Bkg. (no QCD)	15946.8 ± 232.9	14370.8 ± 257.3
Data	18177	14351

Table 5.3: The simulated background and observed event yields after the preselection in the $e\tau_h$ and $\mu\tau_h$ channels. The statistical uncertainties are given for each simulated background. The contribution from the QCD multijets process is not taken into account in the total background yield.

5.2.2 Main Selection

A preselected event satisfies the main selection if at least one of the selected jets in the event is b-tagged. Though the expected final state from the signal includes two b-jets, it was found that requiring b-tagging for only one jet improved the signal yield S in comparison to the SM backgrounds yield B , compared using the quantity S/\sqrt{B} . This is expected due to the reduced efficiency of applying the CSV algorithm for b-tagging multiple times per event, and because the major $t\bar{t}$ background also contains two b-jets. The second jet in the event may or may not be b-tagged. Plots comparing observed data to the simulated backgrounds after the main selection are shown in Figs. 5.4 and 5.5, and event yields are given in Table 5.4.

At this stage of the selection, several new observables are introduced. The first

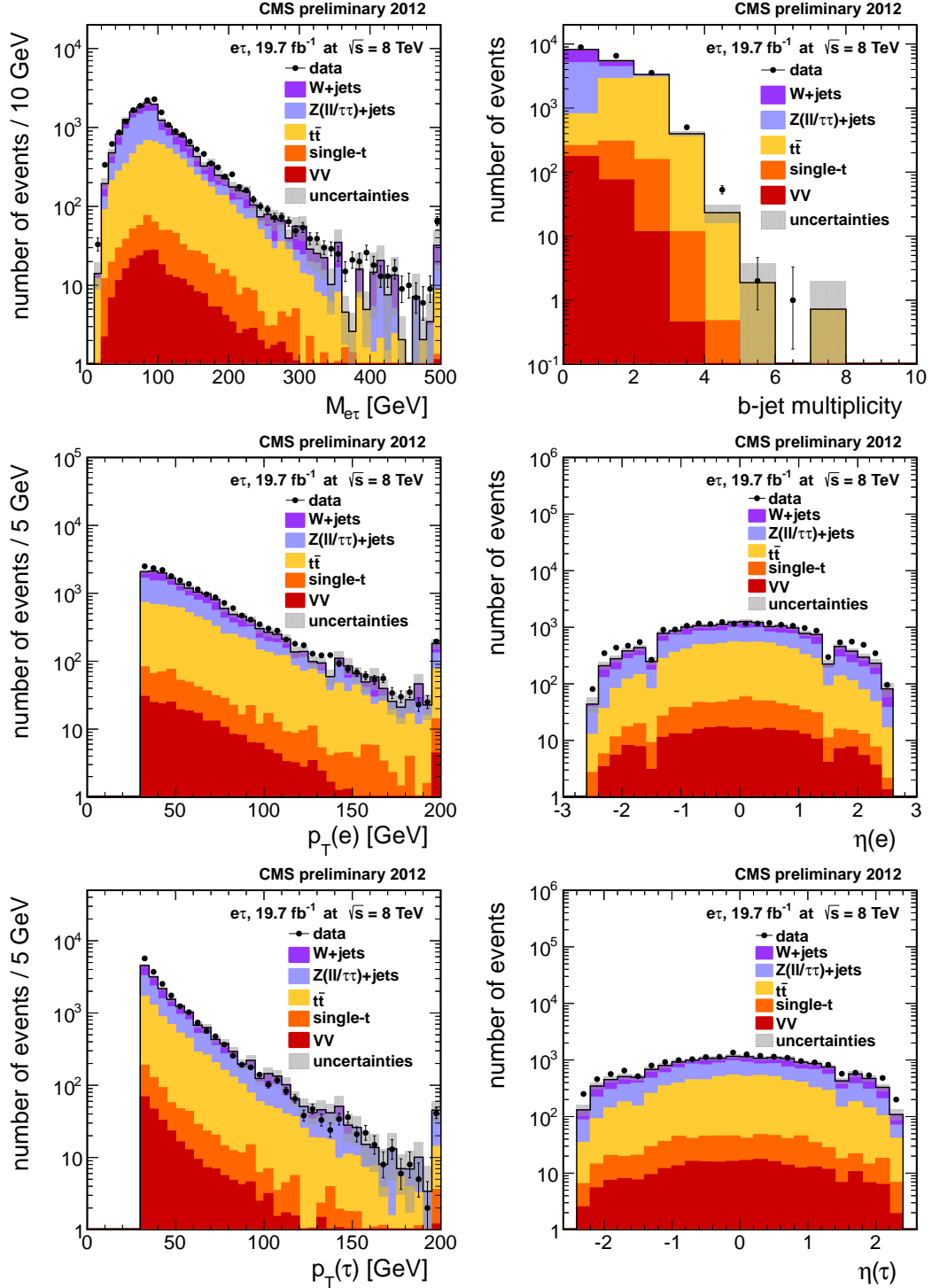


Figure 5.2: Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $e\tau_h$ channel after the preselection: the visible mass of the electron and hadronic tau (top left), the number of b-tagged jets (top right), the p_T (left) and η (right) of the electron (middle) and hadronic tau (bottom). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.

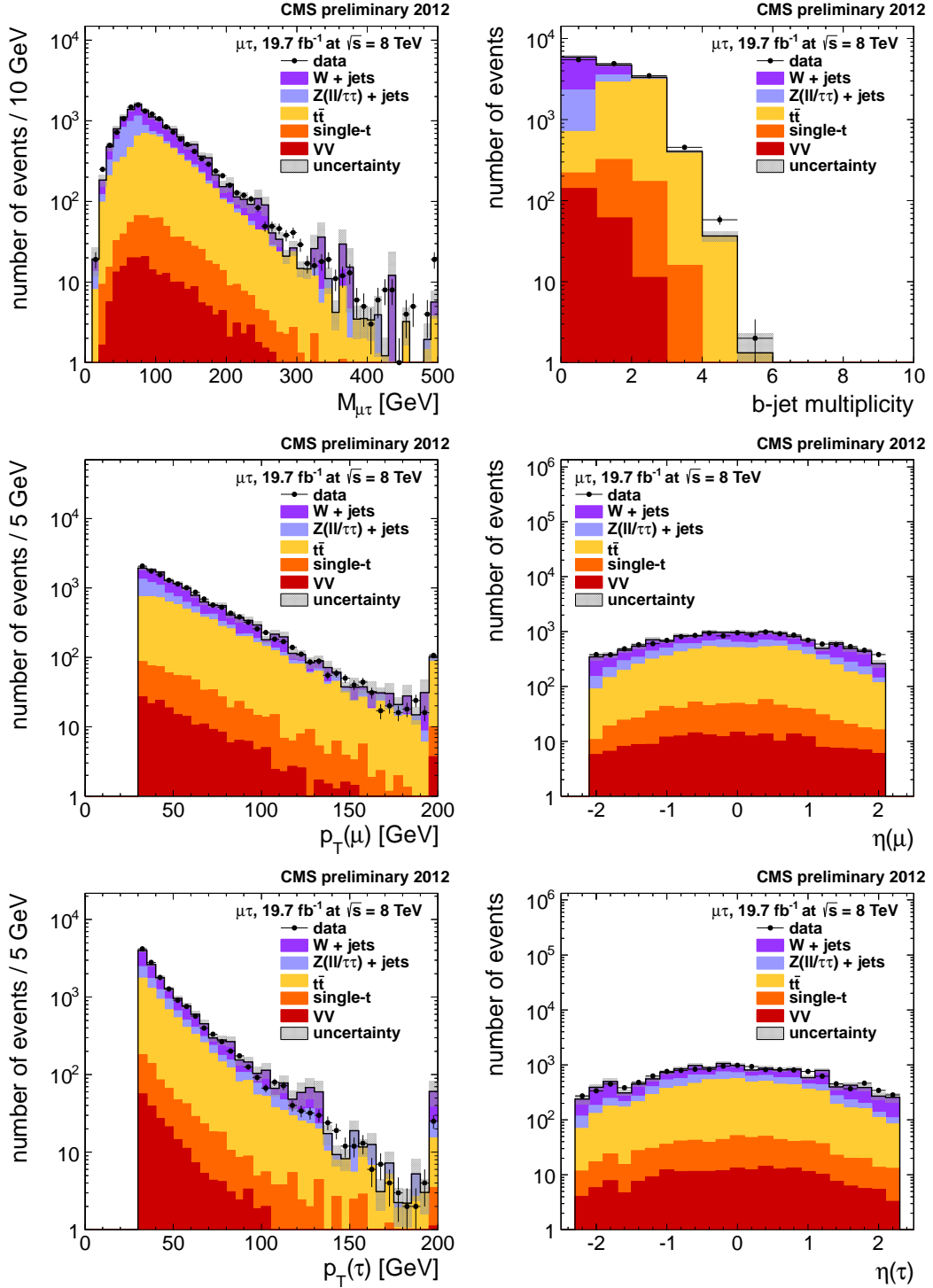


Figure 5.3: Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $\mu\tau_h$ channel after the preselection: the visible mass of the muon and hadronic tau (top left), the number of b-tagged jets (top right), the p_T (left) and η (right) of the muon (middle) and hadronic tau (bottom). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.

is the visible mass of the hadronic tau and a jet, $M(\tau_h, \text{jet})$. There are two possible pairings of the light lepton and the hadronic tau with the two selected jets. The pairing is chosen that minimizes the difference between $M(\tau_h, \text{jet})$ and $M(\ell, \text{jet})$. In signal events, these two values are expected to be similar, as both sets of correctly-paired particles will originate from the same mother particles, leptoquarks or top squarks. However, the visible mass variables will not directly measure the true mass of the mother particles; some of the energy will be lost in the form of neutrinos produced in tau lepton decays. According to the simulation of the signal process, this minimization selects the correct pairing in approximately 70% of events. This observable will be used in the final selection for the leptoquark search.

The second new variable is S_T , which is defined as the scalar sum of transverse momenta for all final-state objects:

$$S_T(\text{LQ}) = p_T(\ell) + p_T(\tau_h) + p_T(\text{b-jet}) + p_T(\text{jet}). \quad (5.1)$$

As indicated in Eq. (5.1), this definition of S_T is appropriate for the leptoquark search, which requires four objects in the final state: a light lepton, a hadronic tau, a b-tagged jet, and another jet. Figures 5.4 and 5.5 show this version of S_T . An alternate definition appropriate for the top squark search, which has a slightly different final state, is given in Eq. (5.2).

5.2.3 Final Selections

In the final stage of the selection for the leptoquark search, two additional requirements are applied. The selected hadronic tau must have $p_T > 50 \text{ GeV}$, and

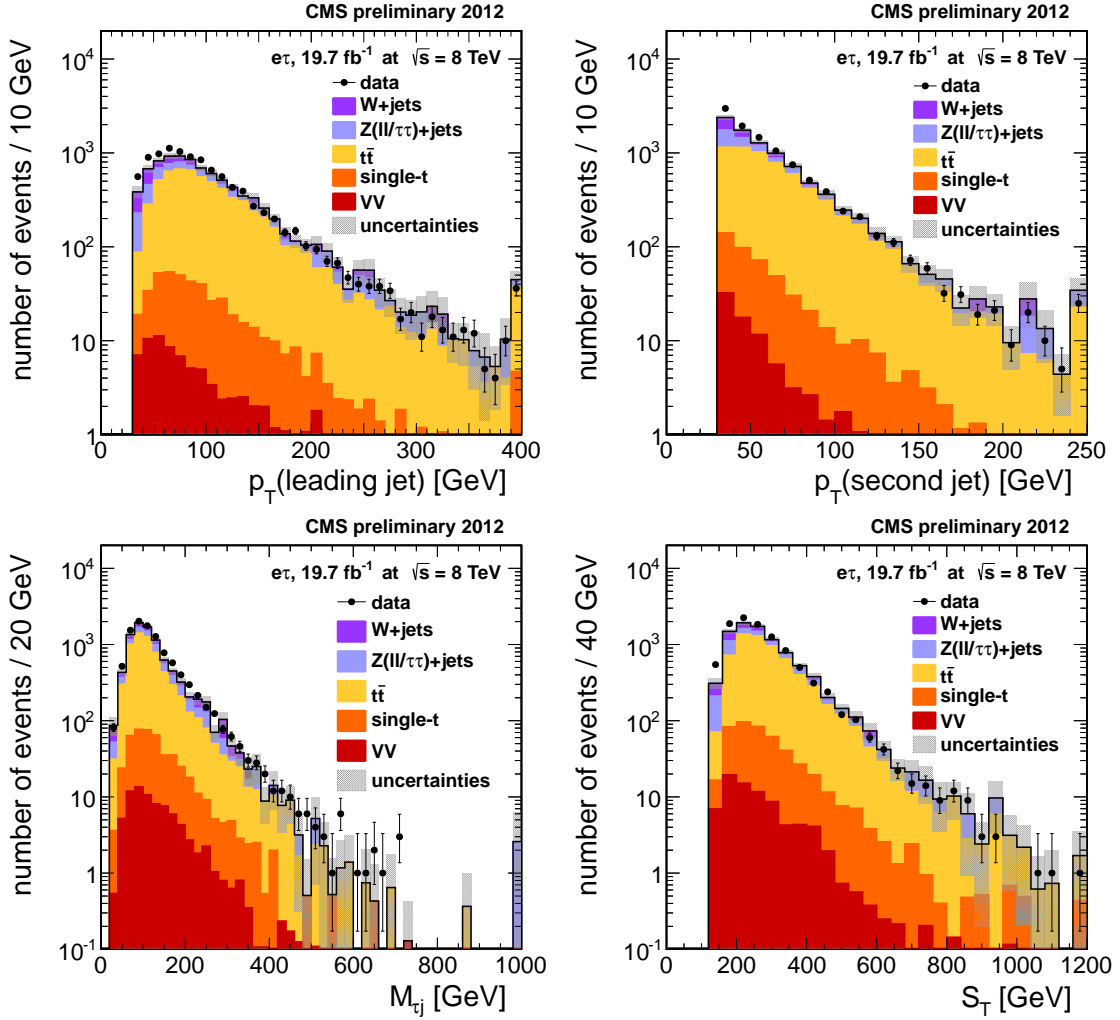


Figure 5.4: Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $e\tau_h$ channel after the main selection: the p_T spectra (top) of the first (left) and second (right) selected jets, $M(\tau_h, \text{jet})$ (bottom left), and S_T (bottom right). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.

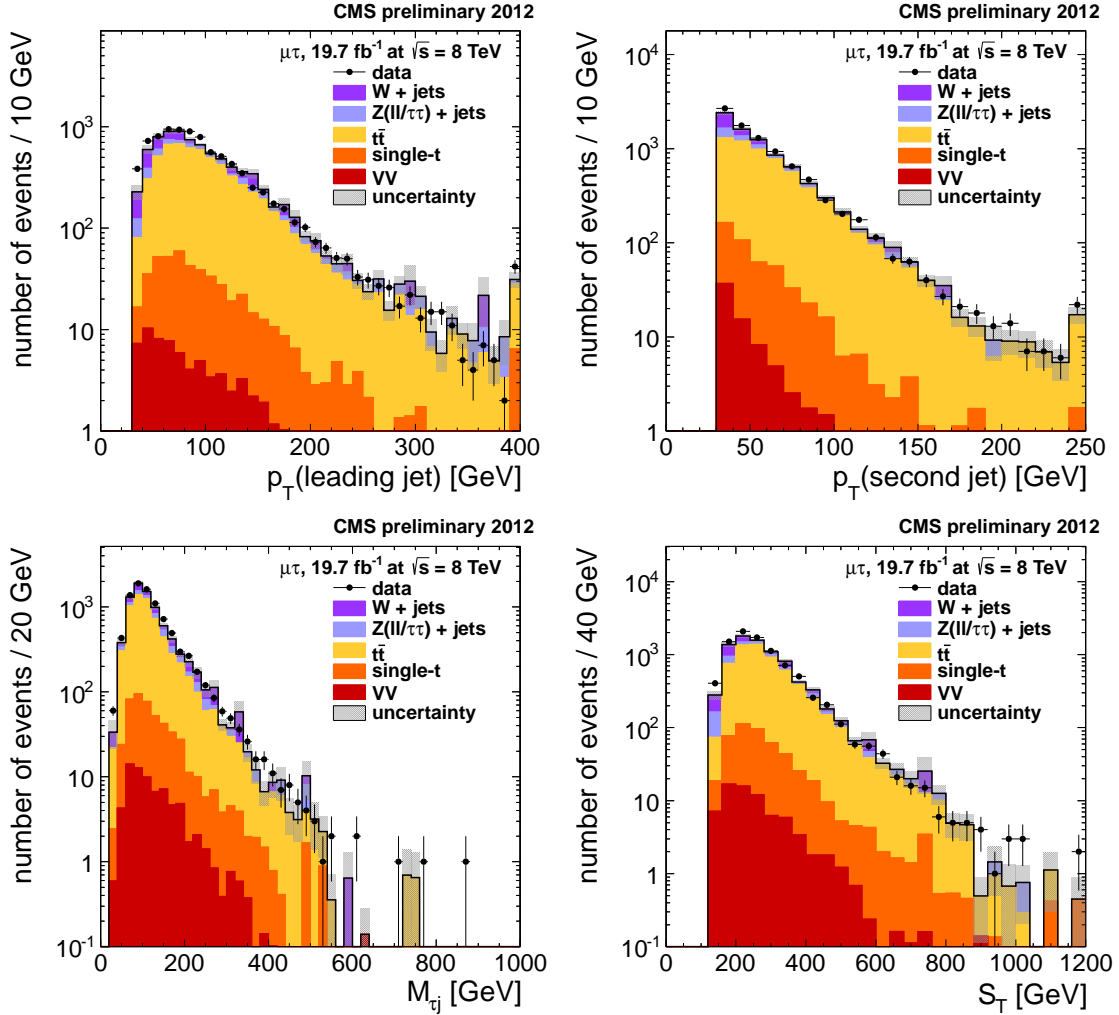


Figure 5.5: Plots of various kinematic quantities comparing observed data and simulated backgrounds in the $\mu\tau_h$ channel after the main selection: the p_T spectra (top) of the first (left) and second (right) selected jets, $M(\tau_h, \text{jet})$ (bottom left), and S_T (bottom right). The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.

	$e\tau_h$ channel	$\mu\tau_h$ channel
W + jets	1201.9 ± 104.0	1285.8 ± 122.7
Z + jets	1534.0 ± 49.5	774.4 ± 46.8
$t\bar{t}$	5783.6 ± 63.5	5699.5 ± 64.4
Single t	390.2 ± 13.5	418.6 ± 14.2
VV	82.9 ± 2.7	74.1 ± 2.7
QCD multijets	(1226.0 ± 131.0)	—
Total Bkg. (no QCD)	8992.6 ± 141.2	8252.3 ± 147.0
Data	10113	8866

Table 5.4: The simulated background and observed event yields after the main selection in the $e\tau_h$ and $\mu\tau_h$ channels. The statistical uncertainties are given for each simulated background. The contribution from the QCD multijets process is not taken into account in the total background yield.

the selected events must have $M(\tau_h, \text{jet}) > 250 \text{ GeV}$. This cut on the $M(\tau_h, \text{jet})$ value gives the optimum expected limit (calculated using simulated samples) for the entire range of leptoquark masses under consideration. The change in the kinematic properties of the signal for different mass values is taken into account when the S_T distribution is used to set CL_s limits. The procedure for setting limits with a distribution is described in App. A.

Slightly different final selection criteria are applied for the top squark search. Again, the selected hadronic tau must have $p_T > 50 \text{ GeV}$. Instead of a cut on $M(\tau_h, \text{jet})$, which is not a meaningful variable for the top squark λ'_{3jk} decay chain, the selected events are required to have $N_{\text{jets}} \geq 5$. The expected final state of the top squark signal has at least $N_{\text{jets}} = 6$, and though it was found that requiring $N_{\text{jets}} \geq 6$ would maximize the expected limit based on the simulation, the number of both signal and background events passing that cut was greatly reduced. Because this would also reduce the number of events in the control regions used to estimate

the major backgrounds, the corresponding increase in systematic uncertainty would render the limit indistinguishable from the $N_{\text{jets}} \geq 5$ case. Thus, the slightly looser N_{jets} cut was chosen. The S_T variable for the top squark search is defined as:

$$S_T(\tilde{t}) = p_T(\ell) + p_T(\tau_h) + p_T(\text{b-jet}) + \sum_{i=1}^4 p_T(\text{jet } i). \quad (5.2)$$

Figure 5.6 shows the final selection variables for each search with both channels combined. Simulated signal distributions are added on top of the background to demonstrate the effect of the final selection cuts in removing background while preserving signal events. The common requirement $p_T(\tau_h) > 50 \text{ GeV}$ is applied in these plots, and the major background yields are estimated from observed data. The background estimation techniques will be described in Sec. 5.3.

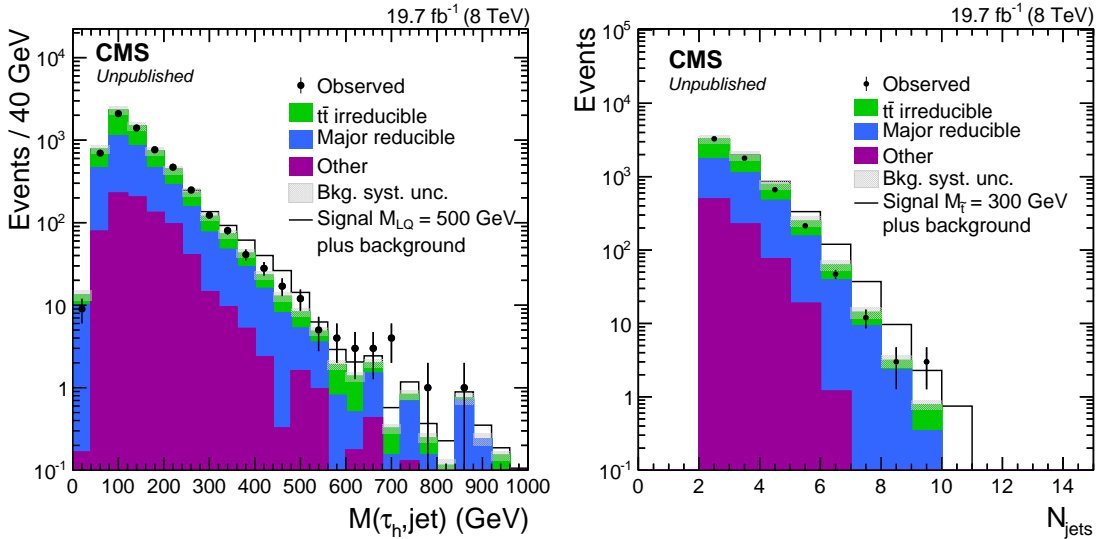


Figure 5.6: $M(\tau_h, \text{jet})$ (left) and N_{jets} (right) before the respective final selection cuts on these variables, with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for leptoquarks with a mass of 500 GeV (left) or top squarks with a mass of 300 GeV (right) is added on top of the background prediction. Both plots include the cut $p_T(\tau_h) > 50 \text{ GeV}$ and the major backgrounds are estimated from observed data.

5.3 Background Estimations

Several SM processes can mimic the final-state signatures expected from leptoquark or top squark pair production and decay. In this dissertation, the backgrounds are divided into three groups, which are denoted as $t\bar{t}$ irreducible, major reducible, and other. The $t\bar{t}$ irreducible background comes from the pair production of top quarks when both the light lepton and τ_h are genuine, each produced from the decay of a W boson. In this case, the light lepton can originate either directly from the W boson decay or from a decay chain $W \rightarrow \tau\nu_\tau \rightarrow \ell\nu_\ell\nu_\tau\nu_\tau$. The major reducible background consists of events in which a quark or gluon jet is misidentified as a τ_h . The processes contributing to the major reducible background are W + jets, Z + jets, and $t\bar{t}$. Additionally, a small contribution from the QCD multijet process is included, in which both the light lepton and the τ_h are misidentified jets. The contribution from QCD is only significant in the $e\tau_h$ channel of the leptoquark search. The third group, other backgrounds, consists of processes that make small contributions and may contain either genuine or misidentified tau leptons. The other backgrounds are estimated using the MC simulation, while the $t\bar{t}$ irreducible and major reducible backgrounds are estimated using observed data. Both the major reducible and other backgrounds include events with both genuine and misidentified light leptons.

5.3.1 $t\bar{t}$ Irreducible Background

A large fraction of the background from the $t\bar{t}$ process is irreducible, due to its similar signature to the signal when both W bosons produce one muon or electron and one hadronic tau. Thus, the $t\bar{t}$ irreducible background is one of the dominant backgrounds and has to be estimated precisely. The use of single muon and single electron HLT criteria, along with extra lepton vetoes, enables the definition of a control region of observed events containing one electron and one muon. This $e\mu$ control region is used to estimate the yield of the $t\bar{t}$ background in the signal region, using the simulation to correct for the differences between the two regions. It is defined using the selection from the $\mu\tau_h$ channel, except requiring an electron instead of a hadronic tau. For example, the final cut on $p_T(\tau_h)$ is applied to the electron in this selection. Figure 5.7 shows the S_T distributions after the leptoquark final selection in the $e\mu$ channel, with good agreement between the observed data and the simulation. The $e\mu$ sample consists of approximately 87% $t\bar{t}$ events; the residual background from other processes is simulated and subtracted from the observed data. The residual background consists of 10% single top quark events, 2% diboson events, and 1% W/Z + jets events. The signal contamination in this sample has been found to be negligible for any signal mass hypothesis, as expected due to the small branching fraction (only 6%) for both tau leptons in the signal final state to decay leptonically.

The $t\bar{t}$ yields in the $e\mu$ control region, $N_{e\mu}$, and in the signal regions, $N_{\ell\tau_h}$, are defined in Eqs. (5.3) and (5.4). Throughout this section, the convention $\ell = e, \mu$

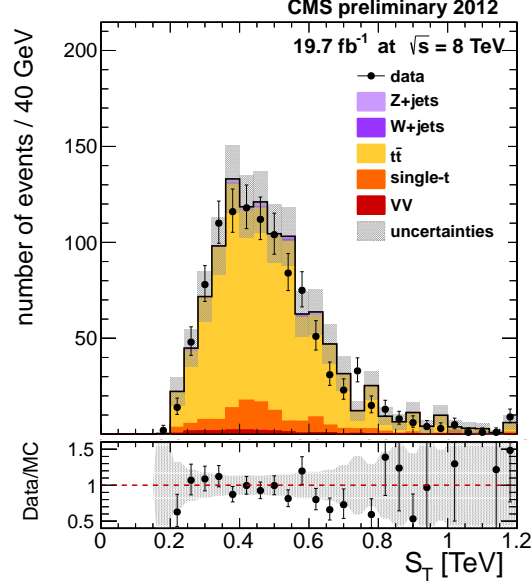


Figure 5.7: The S_T distribution in the $e\mu$ control region after the leptoquark final selection. The uncertainty band reflects the statistical uncertainty in the simulated backgrounds.

will be used.

$$\begin{aligned}
N_{e\mu} &= \sigma \mathcal{L} \times (\varepsilon_e^{\text{ID}} \varepsilon_\mu^{\text{ID}} \varepsilon_{e\mu}^{\text{sel}} \rho_{e\mu}^{\text{sel}}) \\
&\times 2 \left[\mathcal{A}_{e\mu} \mathcal{B}_{W_e} \mathcal{B}_{W_\mu} + \mathcal{A}_{\mu\tau_e} \mathcal{B}_{W_\mu} \mathcal{B}_{W_{\tau_e}} \right. \\
&\quad \left. + \mathcal{A}_{e\tau_\mu} \mathcal{B}_{W_e} \mathcal{B}_{W_{\tau_\mu}} + \mathcal{A}_{\tau_e\tau_\mu} \mathcal{B}_{W_{\tau_e}} \mathcal{B}_{W_{\tau_\mu}} \right],
\end{aligned} \tag{5.3}$$

$$\begin{aligned}
N_{\ell\tau_h} &= \sigma \mathcal{L} \times (\varepsilon_\ell^{\text{ID}} \varepsilon_{\tau_h}^{\text{ID}} \varepsilon_{\ell\tau_h}^{\text{sel}} \rho_{\ell\tau_h}^{\text{sel}}) \\
&\times 2 \left[\mathcal{A}_{\ell\tau_h} \mathcal{B}_{W_\ell} \mathcal{B}_{W_{\tau_h}} + \mathcal{A}_{\tau_\ell\tau_h} \mathcal{B}_{W_{\tau_\ell}} \mathcal{B}_{W_{\tau_h}} \right].
\end{aligned} \tag{5.4}$$

Here, σ is the $t\bar{t}$ production cross section; \mathcal{L} is the integrated luminosity; $\varepsilon_x^{\text{ID}}$ is the selection efficiency for the identification and isolation of a lepton x in the simulation; $\varepsilon_i^{\text{sel}}$ is the analysis selection efficiency in the simulation, including the jet multiplicity cut, the b-tagging requirement, and so forth, for the channel i ; ρ_i^{sel} is the simulation to data scale factor associated with the selection efficiency for the channel i ; $\mathcal{A}_{\ell\ell'}$

channel	\mathcal{A} (%)	channel	\mathcal{A} (%)
$e\tau_h$	23.6 ± 0.3	$\mu\tau_h$	23.7 ± 0.3
$\tau_e\tau_h$	10.4 ± 0.5	$\tau_\mu\tau_h$	10.5 ± 0.4
$e\tau_\mu$	13.9 ± 0.4	$\mu\tau_e$	12.3 ± 0.4
$e\mu$	41.6 ± 0.3	$\tau_e\tau_\mu$	3.4 ± 0.5

Table 5.5: Acceptances (in %) for the $t\bar{t}$ process in each leptonic decay channel of the W boson produced by the top quark decays. Variations in the acceptances are mostly due to differences in the p_T distributions for each type of lepton.

is the kinematic acceptance for a process that produces one lepton ℓ and another lepton ℓ' ; and \mathcal{B} is the branching fraction for a given leptonic decay channel.

The branching fractions \mathcal{B} of the W boson decay for each leptonic decay channel are labeled as follows:

$$\mathcal{B}(W \rightarrow \ell\nu_\ell) = \mathcal{B}_{W\ell}, \quad (5.5)$$

$$\mathcal{B}(W \rightarrow \tau\nu_\tau \rightarrow \ell\nu_\ell\nu_\tau) = \mathcal{B}_{W\tau\ell}, \quad (5.6)$$

$$\mathcal{B}(W \rightarrow \tau\nu_\tau \rightarrow \text{hadrons} + \nu_\tau) = \mathcal{B}_{W\tau_h}. \quad (5.7)$$

The kinematic acceptances \mathcal{A} are computed using the MC simulation of the $t\bar{t}$ process and are defined for each leptonic decay channel:

$$\mathcal{A} = \frac{N_{\text{sel}}(\ell\ell')}{N_{\text{gen}}(\ell\ell')}. \quad (5.8)$$

$N_{\text{gen}}(\ell\ell')$ is the number of generated events containing two leptons ℓ and ℓ' , and $N_{\text{sel}}(\ell\ell')$ is the number of those generated events that pass the p_T and η cuts used for the selection of the primary leptons (electrons, muons, or hadronic taus) in the signal region. The acceptances for each decay channel are listed in Table 5.5.

The efficiencies ε and the data/MC scale factors ρ are derived from several different sources. The standard efficiencies for the lepton identification algorithms

lepton	ε^{ID} from $t\bar{t}$ events (%)	ρ from Z events
e	77.3 ± 0.1	1.011 ± 0.001
μ	81.3 ± 0.1	0.995 ± 0.001
τ_{h}	35.5 ± 4.9	1.000

Table 5.6: Lepton identification efficiencies ε^{ID} measured in simulated $t\bar{t}$ events and associated scale factors ρ measured in Z events.

are estimated using Z events. However, differences in topology between $t\bar{t}$ and Z events make these values inappropriate for the $t\bar{t}$ background estimation. The identification efficiencies are therefore calculated using simulated $t\bar{t}$ events. The reconstructed leptons in these events are required to be matched with genuine leptons at the generator level, and the results are listed in Table 5.6. The standard efficiencies for the HLT criteria are used. The standard data/MC scale factors for the lepton identification, HLT efficiency, and b-tagging are also used. The difference in performance between the observed data and the MC simulation is assumed to be due to inherent properties of the simulation of the individual objects, without any dependence on the type of event.

The analysis selection efficiencies are calculated using the $t\bar{t}$ simulation. The efficiencies are computed for each selection requirement, including: vertex matching between the two leptons, opposite charge between the two leptons, additional lepton veto, $N_{\text{jets}} \geq 2$, $N_{\text{b-jets}} \geq 1$, and $p_{\text{T}}(\tau_{\text{h}}) > 50 \text{ GeV}$. Efficiencies are also computed for the final selection cuts ($M(\tau_{\text{h}}, \text{jet}) > 250 \text{ GeV}$ for the leptoquark search and $N_{\text{jets}} \geq 5$ for the top squark search). To create a $t\bar{t}$ enriched sample in the observed data for the computation of the data/MC scale factors, the order of the selection cuts listed in Sec. 5.2 must be altered. In the altered order, the jet multiplicity requirement

is applied first: $N_{\text{jets}} \geq 2$ for the leptoquark search or $N_{\text{jets}} \geq 5$ for the top squark search. Next, the b-tagging requirement is applied, and then the $M(\tau_{\text{h}}, \text{jet})$ cut is applied for the leptoquark search. The other selection requirements are applied subsequently.

The overall data/MC scale factors associated with the vertex matching, opposite charge, lepton veto, and final hadronic tau p_{T} cut are computed in the control region defined by requiring events with two b-jets and $M(\tau_{\text{h}}, \text{jet}) < 120 \text{ GeV}$. The separate control region with $120 \text{ GeV} < M(\tau_{\text{h}}, \text{jet}) < 250 \text{ GeV}$ is used to check that the scale factors do not depend on the value of the $M(\tau_{\text{h}}, \text{jet})$ cuts. Due to the lack of a pure $t\bar{t}$ control region to estimate the data/MC scale factor at the level of the jet multiplicity requirement, no scale factor is applied and an additional systematic uncertainty is assigned. For similar reasons, no scale factor is applied for the $M(\tau_{\text{h}}, \text{jet})$ cut. The results of these selection efficiency and scale factor calculations are listed in Table 5.7. The primary source of disagreement between the data and the simulation is mismodeling of the hadronic tau energy scale and momentum in the $t\bar{t}$ MC samples.

The systematic uncertainty assigned to the jet multiplicity requirement is defined as the difference between data and simulation in the efficiency of the requirement. This difference is calculated in a control region consisting of 95% $t\bar{t}$ events, which is selected by requiring both W bosons in each event to decay to muons. Additional requirements are applied to reduce the contamination from non- $t\bar{t}$ processes: $\cancel{E}_{\text{T}} > 50 \text{ GeV}$, $20 \text{ GeV} < M_{\mu\mu} < 70 \text{ GeV}$ or $M_{\mu\mu} > 120 \text{ GeV}$. This control region contains enough events for an accurate comparison between data and MC.

Selection	$e\tau_h$ channel		$\mu\tau_h$ channel		$e\mu$ channel	
	ε (%)	ρ	ε (%)	ρ	ε (%)	ρ
$\varepsilon_\ell\varepsilon_{\ell'}$	27.4 ± 0.1	1.011 ± 0.001	28.9 ± 0.1	0.995 ± 0.001	62.8 ± 0.1	1.006 ± 0.001
ε_{HLT}	91.4 ± 0.1	–	89.2 ± 0.1	–	89.2 ± 0.1	–
	Leptoquark search					
ε_{jet}	76.1 ± 0.5	–	74.1 ± 0.5	–	79.5 ± 0.2	–
$\varepsilon_{\text{b-tag}}$	94.2 ± 0.3	–	92.7 ± 0.4	–	94.0 ± 0.1	–
$\varepsilon_{\text{vtx, OS, } \ell \text{ veto, } p_T(\tau_h)}$	38.2 ± 1.5	0.987 ± 0.042	38.2 ± 1.6	0.947 ± 0.038	60.5 ± 1.0	1.002 ± 0.020
$\varepsilon_{\text{mass}}$	6.8 ± 0.6	–	5.3 ± 0.6	–	7.1 ± 0.1	–
	Top squark search					
ε_{jet}	3.7 ± 0.2	–	2.8 ± 0.2	–	4.2 ± 0.2	–
$\varepsilon_{\text{b-tag}}$	98.3 ± 0.3	–	99.0 ± 0.3	–	98.0 ± 0.1	–
$\varepsilon_{\text{vtx, OS, } \ell \text{ veto, } p_T(\tau_h)}$	51.7 ± 2.0	0.987 ± 0.042	45.9 ± 2.1	0.947 ± 0.038	66.7 ± 2.0	1.002 ± 0.020

Table 5.7: The relative selection efficiencies (in %) and data/MC scale factors ρ for the $t\bar{t}$ process in each channel $e\tau_h$, $\mu\tau_h$, and $e\mu$. The total selection efficiency is the product of all the relative selection efficiencies for a given search.

It is used to derive a systematic uncertainty because it is not kinematically similar enough to the signal region to provide an appropriate scale factor. The data/MC differences in the efficiency of the jet multiplicity requirements are found to be 1% for $N_{\text{jets}} \geq 2$ in the leptoquark search and 3% for $N_{\text{jets}} \geq 5$ in the top squark search. These differences are propagated through the $t\bar{t}$ estimation method as systematic uncertainties.

The systematic uncertainty assigned to the $M(\tau_h, \text{jet})$ requirement is calculated from two sources. The first source is the relative difference between the efficiencies in data and simulation in the $t\bar{t} \rightarrow \mu\mu$ control sample, which is found to be 6%. The second source is the difference between the efficiencies obtained in data in the $e\mu$ channel with and without the residual non- $t\bar{t}$ background. This difference is also found to be 6%. The two sources are taken to be independent, so they are added in quadrature to produce a total uncertainty of 8.5%.

Equations (5.4) and (5.3) can be combined to eliminate $\sigma\mathcal{L}$, which produces a formula to estimate the $t\bar{t}$ yield in the $\ell\tau_h$ signal region from the $e\mu$ control region:

$$N_{\ell\tau_h} = N_{e\mu} \times \frac{\varepsilon_{\ell}^{\text{ID}} \varepsilon_{\tau_h}^{\text{ID}}}{\varepsilon_e^{\text{ID}} \varepsilon_{\mu}^{\text{ID}}} \times \frac{\varepsilon_{\ell\tau_h}^{\text{sel}} \rho_{\ell\tau_h}^{\text{sel}}}{\varepsilon_{e\mu}^{\text{sel}} \rho_{e\mu}^{\text{sel}}} \times \frac{\mathcal{A}_{\ell\tau_h} \mathcal{B}_{W\ell} \mathcal{B}_{W\tau_h} + \mathcal{A}_{\tau\ell\tau_h} \mathcal{B}_{W\tau\ell} \mathcal{B}_{W\tau_h}}{\mathcal{A}_{e\mu} \mathcal{B}_{We} \mathcal{B}_{W\mu} + \mathcal{A}_{\mu\tau_e} \mathcal{B}_{W\mu} \mathcal{B}_{W\tau_e} + \mathcal{A}_{e\tau\mu} \mathcal{B}_{We} \mathcal{B}_{W\tau\mu} + \mathcal{A}_{\tau_e\tau\mu} \mathcal{B}_{W\tau_e} \mathcal{B}_{W\tau\mu}}. \quad (5.9)$$

$N_{e\mu}$ is the yield from the $e\mu$ control region in the observed data, with the residual non- $t\bar{t}$ background subtracted. The yield in the signal region $N_{\ell\tau_h}$ is calculated from $N_{e\mu}$ via multiplication by selection and identification efficiencies, data/MC scale factors, acceptances, and branching fractions.

To describe the calculation of the total systematic uncertainty from the prop-

agation of the uncertainty in the acceptances, efficiencies, and scale factors, a shorthand notation is used to simplify Eq. (5.9):

$$N_{\ell\tau_h} = N_{e\mu} \times R_{\text{ID}} \times R_{\text{sel}} \times R_{\text{AB}}. \quad (5.10)$$

As indicated by comparing Eq. (5.9) and Eq. (5.10), R_{ID} is the ratio of the identification efficiencies, R_{sel} is the ratio of the selection efficiencies and scale factors, and R_{AB} is the ratio of the acceptances and branching fractions. Equation (5.11) shows a simplified version of the propagation of the uncertainties from these three ratio factors:

$$\delta_{\text{syst}} = \sqrt{(\delta_{\text{ID}} \times R_{\text{sel}} \times R_{\text{AB}})^2 + (R_{\text{ID}} \times \delta_{\text{sel}} \times R_{\text{AB}})^2 + (R_{\text{ID}} \times R_{\text{sel}} \times \delta_{\text{AB}})^2}. \quad (5.11)$$

The uncertainty in the ratio R_x is written as δ_x . In the particular case of δ_{sel} , the uncertainty value is modified by the additional uncertainties assigned to the N_{jets} and $M(\tau_h, \text{jet})$ cuts, as discussed above:

$$\delta_{\text{sel}} \rightarrow \sqrt{\delta_{\text{sel}}^2 + (R_{\text{sel}} \times U_{N_{\text{jets}}})^2 + (R_{\text{sel}} \times U_{M(\tau_h, \text{jet})})^2}. \quad (5.12)$$

U_x is the additional uncertainty associated with the cut x . In the top squark search, the uncertainty associated with the $M(\tau_h, \text{jet})$ cut is not included, as that cut is not applied.

Using Eq. 5.9, the estimated yield of the $t\bar{t}$ irreducible background is computed for each $\ell\tau_h$ channel. The results are summarized in Table 5.8 for the leptoquark search and in Table 5.9 for the top squark search. This estimation method is applied to a simulated $e\mu$ sample as a cross-check, and the results agree within one standard deviation when compared to the expectation from the direct simulation of the irre-

ducible $t\bar{t}$ contributions to the $\ell\tau_h$ channels. The calculation from observed data is also in agreement with both the expectation from the direct $t\bar{t}$ $\ell\tau_h$ simulation and the closure test using the $e\mu$ simulation. For each $\ell\tau_h$ channel, the $t\bar{t}$ S_T distribution is estimated from the exclusive $t\bar{t} \rightarrow b\ell\nu b\ell\nu$ MC simulation.

Leptoquark search			
$e\mu$ channel			
	$t\bar{t}$ MC	data	data – residual MC
	966.71 ± 27.3	$1065 (\pm 32.6)$	$920.0 \pm 8.3 \pm 32.6$
$\ell\tau_h$ channel			
channel	$t\bar{t}$ MC (genuine τ_h)	$t\bar{t}$ MC closure test	data result
$e\tau_h$	94.4 ± 8.3	$94.0 \pm 2.7 \pm 14.9$	$98.7 \pm 3.6 \pm 17.7$
$\mu\tau_h$	72.8 ± 8.5	$72.2 \pm 2.0 \pm 12.6$	$64.2 \pm 2.3 \pm 12.4$

Table 5.8: The estimated $t\bar{t}$ irreducible yields after the leptoquark final selection, from: direct simulation, closure test from the $e\mu$ simulation, and calculation from the observed $e\mu$ data. In the direct simulation, only the statistical uncertainty is given. In the closure test and the data result, the first uncertainty value corresponds to the statistical uncertainty in the simulation and observed data, while the second uncertainty value corresponds to the propagation of the uncertainties in the acceptances, efficiencies and scale factors.

This estimation of the $t\bar{t}$ irreducible background estimation does not account for events with jets or leptons misidentified as hadronic taus, but does include events with jets or photons misidentified as electrons or muons. The yield from events in which an electron or a muon is identified as a hadronic tau is estimated from the simulation and added to the yield of the $t\bar{t}$ irreducible background. This contribution is estimated to be 6.9 ± 2.1 (2.5 ± 1.3) for the $e\tau_h$ ($\mu\tau_h$) channel in the leptoquark search and 11.7 ± 2.7 (2.8 ± 1.3) for the $e\tau_h$ ($\mu\tau_h$) channel in the top squark search. The small statistical uncertainty from this contribution is neglected when adding it to the $t\bar{t}$ irreducible yield. The estimation of the background from events with jets

Top squark search

Top squark search			
$e\mu$ channel			
	$t\bar{t}$ MC	data	data – residual MC
	823.9 ± 26.3	$733 (\pm 27.1)$	$700.0 \pm 7.2 \pm 27.1$
$\ell\tau_h$ channel			
channel	$t\bar{t}$ MC (genuine τ_h)	$t\bar{t}$ MC closure test	data result
$e\tau_h$	80.4 ± 7.5	$81.7 \pm 2.6 \pm 12.5$	$76.6 \pm 3.1 \pm 13.3$
$\mu\tau_h$	57.9 ± 6.4	$65.8 \pm 2.1 \pm 10.3$	$52.2 \pm 2.1 \pm 9.3$

Table 5.9: The estimated $t\bar{t}$ irreducible yields after the top squark final selection, from: direct simulation, closure test from the $e\mu$ simulation, and calculation from the observed $e\mu$ data. In the direct simulation, only the statistical uncertainty is given. In the closure test and the data result, the first uncertainty value corresponds to the statistical uncertainty in the simulation and observed data, while the second uncertainty value corresponds to the propagation of the uncertainties in the acceptances, efficiencies and scale factors.

misidentified as hadronic taus will be discussed in the next section.

5.3.2 Major Reducible Background

The major reducible background comes from events with jets misidentified as hadronic taus. This background has contributions from the $W + \text{jets}$, $Z + \text{jets}$, $t\bar{t}$, and QCD multijet processes. The contribution from $W + \text{jets}$, $Z + \text{jets}$, and $t\bar{t}$ is estimated from observed data by measuring the misidentification probability or “fake rate”. The QCD events contain mostly gluon jets, while the events from the other processes contain mostly quark jets. Gluon jets tend to be wider than quark jets and to have higher multiplicity, with a correspondingly lower energy per particle. Therefore, gluon jets are less likely to pass isolation and therefore have a lower fake rate. The fake rate used in this estimation is measured in a control region that is appropriate for the quark jet processes. Therefore, this fake rate cannot

produce an accurate estimate of the contribution from QCD. This contribution is estimated using a same-sign/opposite-sign (SS/OS) method, which is described in Sec. 5.3.2.1. The contribution from QCD is non-negligible only in the $e\tau_h$ channel of the leptoquark search. In this case, the reconstructed electron is also a misidentified jet. A jet is much less likely to be misidentified as a muon, so QCD does not contribute in the $\mu\tau_h$ channel. In the top squark search, the high jet multiplicity requirement eliminates any significant contribution from QCD.

The probability for a jet to be misidentified as a hadronic tau is measured from the observed data in a $Z \rightarrow \mu\mu + \text{jets}$ control region. The selection of the primary muon uses the same criteria as the $\mu\tau_h$ channel in the signal region. The second muon is selected using the loose working point for identification and isolation, along with looser kinematic cuts. The two muons are required to have the same vertex, opposite charges, and a separation of $\Delta R > 0.5$. This ensures the orthogonality of the control region because of the opposite-sign additional lepton veto in the signal region. The selection of hadronic tau candidates, which are misidentified jets in the $Z \rightarrow \mu\mu + \text{jets}$ control region, uses the same identification, discriminator working points, and kinematic cuts as the signal selection. However, the hadronic tau candidates are not required to be isolated, as the isolation variable is used to compute the fake rate. All selected hadronic taus must originate from the same vertex as the two muons and must be separated from each muon with $\Delta R > 0.5$. A cut on the invariant mass of the two muons is also applied, $M_{\mu\mu} > 50 \text{ GeV}$, in order to match the generation of the simulated $Z + \text{jets}$ samples.

This selection produces a control region which consists of 95% $Z + \text{jets}$ events.

The simulated $Z + \text{jets}$ sample is normalized to the yield from the observed data, with the yield from the other simulated backgrounds subtracted. The normalization factor is $N[\text{data} - \text{other MC}]/N[Z + \text{jets}] = 0.924 \pm 0.003$, calculated using the range $70 < M_{\mu\mu} < 110 \text{ GeV}$. Figure 5.8 shows the dimuon mass $M_{\mu\mu}$ and the multiplicity of extra jets, which are not misidentified as hadronic taus, in the $Z \rightarrow \mu\mu + \text{jets}$ control region. Good agreement is seen between data and simulation, demonstrating the validity of the control region. The majority of events contain no extra jets, but an appreciable percentage, around 40%, contain one or more extra jets.

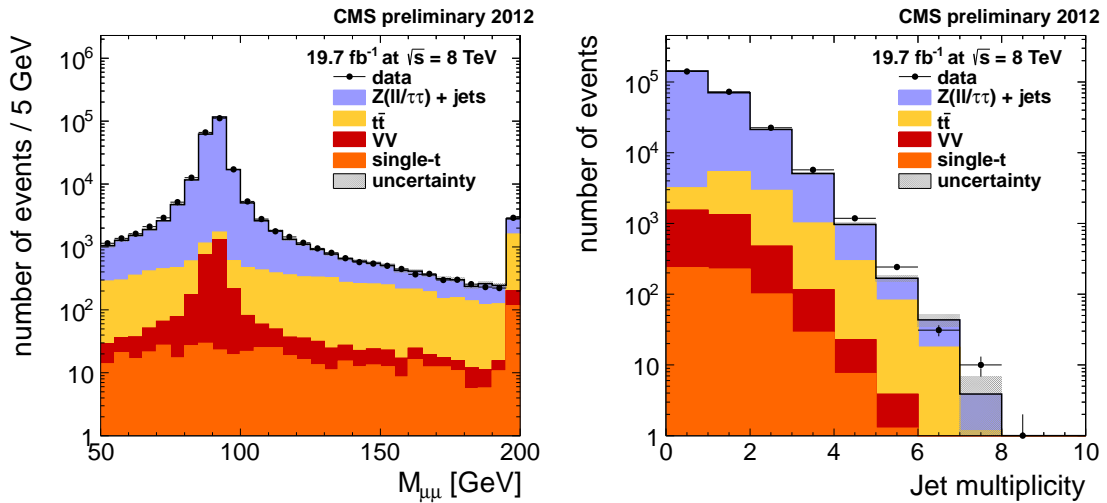


Figure 5.8: Comparison of observed data and simulation in the $Z \rightarrow \mu\mu + \text{jets}$ control region for $M_{\mu\mu}$ (left) and multiplicity of extra jets, which are not misidentified as hadronic taus (right).

All hadronic tau candidates in the $Z \rightarrow \mu\mu + \text{jets}$ control region are considered when calculating the fake rate. These hadronic tau candidates are misidentified jets which pass all selection requirements for hadronic taus, except that isolation is not required. The subset of these candidates which also pass the isolation requirement

would be fully misidentified as hadronic taus. The fake rate is therefore the rate at which the hadronic tau candidates pass the isolation requirement. It is parameterized as a function of the transverse momentum of the tau candidates, as shown in Eq. (5.13):

$$f(p_T) = \frac{N_{\text{iso } \tau}^{(Z \rightarrow \mu\mu)}(p_T)}{N_{\text{all } \tau}^{(Z \rightarrow \mu\mu)}(p_T)} \quad (5.13)$$

When calculating the fake rate from the observed data, the contribution from the residual background is subtracted from both the numerator and the denominator of Eq. (5.13). The residual background consists of 4% $t\bar{t}$ events, 1% diboson events, and 0.2% single top quark events. The observed data and simulated fake rates calculated from all $Z \rightarrow \mu\mu + \text{jets}$ events, from events with one or more extra jets, and from events with two or more extra jets are shown in Fig. 5.9. The fake rate calculated from all $Z \rightarrow \mu\mu + \text{jets}$ events is called the inclusive fake rate, since it includes all events regardless of the number of extra jets.

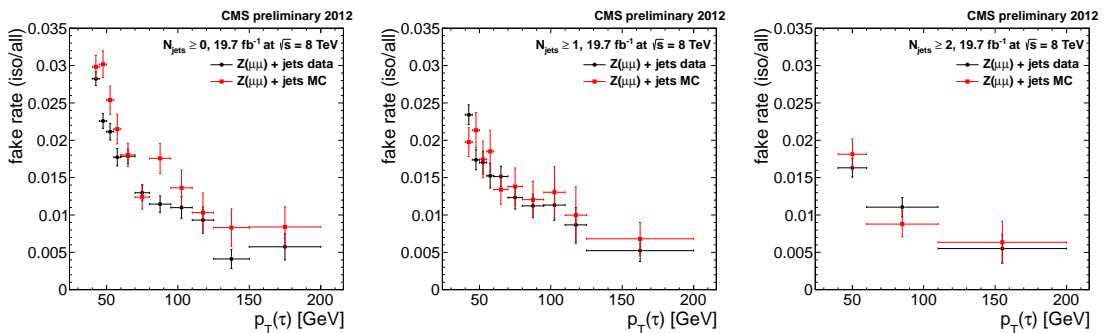


Figure 5.9: The fake rates measured in the $Z \rightarrow \mu\mu + \text{jets}$ control region, using all events (left), only events with one or more extra jets (middle), only events with two or more extra jets (right).

Events from semileptonic $t\bar{t}$ decays make up a significant portion of the major reducible background. These events resemble events from the $W + \text{jets}$ process,

but with higher jet multiplicities. The relative difference in the fake rate between $Z \rightarrow \mu\mu + \text{jets}$ events with $N_{\text{jet}} \geq 1$ and semileptonic $t\bar{t}$ events is estimated from the simulation and found to be 18%. Figure 5.10 shows the simulated fake rates versus hadronic tau p_T and the calculation of the relative difference, which is independent of p_T . This relative difference is multiplied by the percentage of the major reducible background consisting of semileptonic $t\bar{t}$ events, 35% in the leptoquark search and 95% in the top squark search according to the simulation, and applied as a systematic uncertainty to account for any differences between the types of processes.

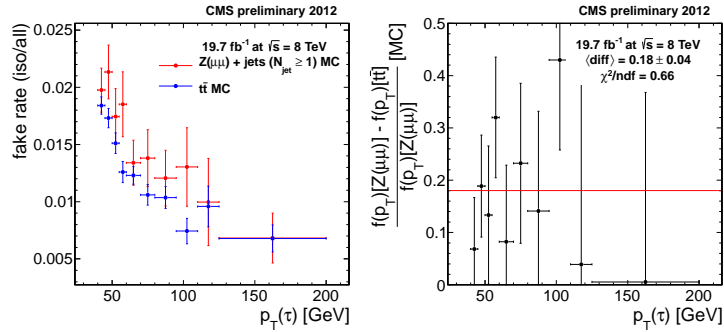


Figure 5.10: Comparison of the fake rates between simulated $Z \rightarrow \mu\mu + \text{jets}$ events with at least one extra jet and simulated semileptonic $t\bar{t}$ events (left). The difference in the two fake rates is found to be 18%, independent of the hadronic tau p_T (right).

To estimate the total yield of the major reducible background in the signal region, a second control region is needed. This control region is defined identically to the signal region, except that events are rejected if any selected hadronic tau passes isolation. This creates an orthogonal region of events in which all selected hadronic taus fail isolation, called the “anti-isolated” region. Figures 5.11 and 5.12 shows the hadronic tau p_T spectrum and multiplicity in the anti-isolated region for both channels after the leptoquark and top squark final selections, respectively.

There is good agreement between observed data and simulated backgrounds in the $\mu\tau_h$ channel. In the $e\tau_h$ channel in the leptoquark search, there is an expected disagreement due to a contribution from QCD multijet events in the observed data, which will be addressed. The simulated contributions labeled W + jets, Z + jets, and $t\bar{t}$ contain only events where the hadronic tau candidates are misidentified jets based on matching between the generated and reconstructed particles. The residual background is defined to include events from two categories. The first category consists of events from the diboson and single-top-quark processes, which are not included in the major reducible background. The second category consists of events from the W + jets, Z + jets, and $t\bar{t}$ processes with genuine hadronic taus that fail isolation. The single top quark process contributes 2–5% depending on the channel and the search, while the other processes contribute less than 1% each.

The fake rate is used to weight the anti-isolated events in order to estimate the yield in the signal region. Because $f(p_T)$ is the probability for a fake hadronic tau candidate to be isolated, $1 - f(p_T)$ is the probability for a fake hadronic tau candidate to fail isolation:

$$P(\tau \text{ w/ given } p_T \text{ is isolated}) = f(p_T(\tau)), \quad (5.14)$$

$$P(\tau \text{ w/ given } p_T \text{ is anti-isolated}) = 1 - f(p_T(\tau)). \quad (5.15)$$

The signal region and the anti-isolated region are two complementary parts of a total region containing all events with hadronic tau candidates, passing or failing isolation. The probability that any given event in this total region is part of the anti-isolated region is equivalent to the probability that all hadronic taus in that

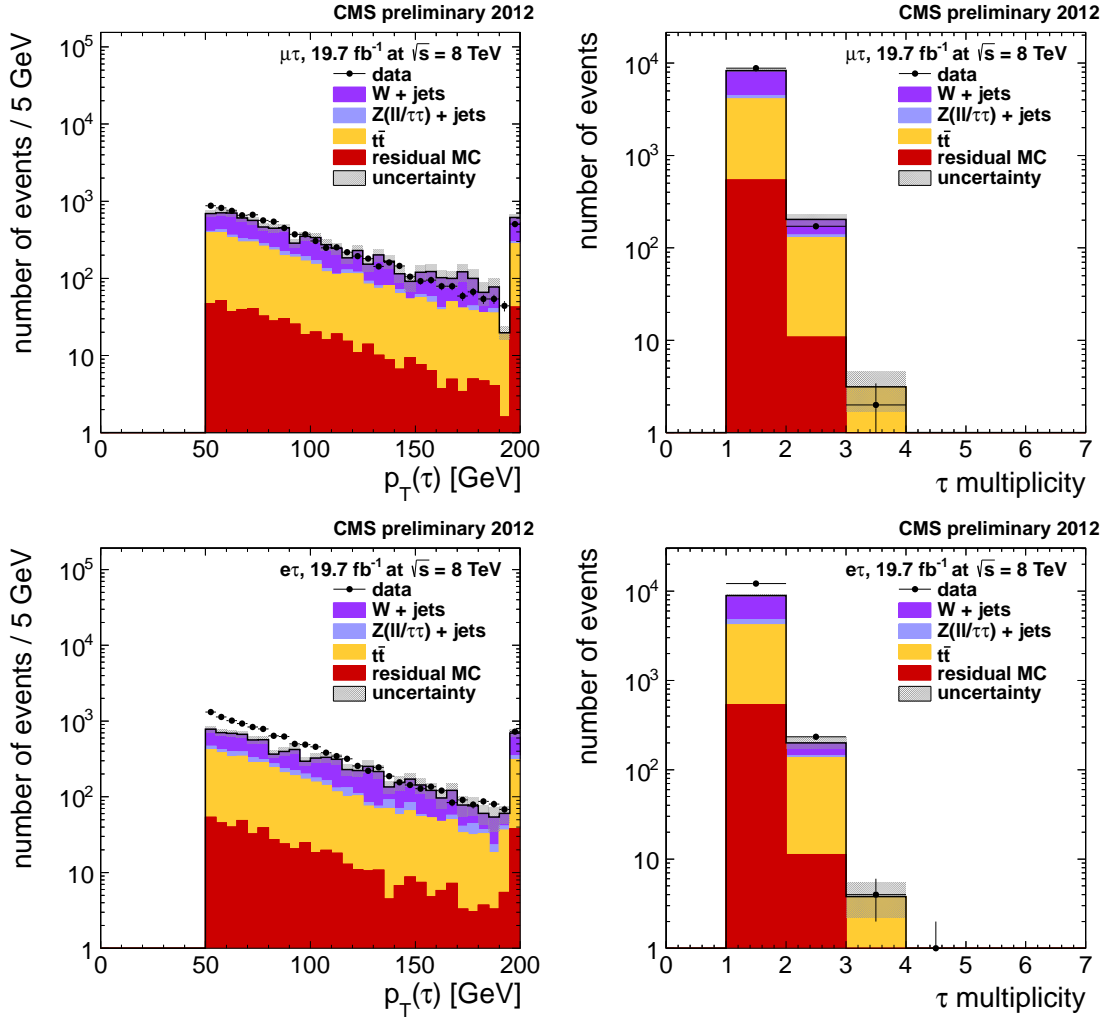


Figure 5.11: Plots of the anti-isolated control region after the leptoquark final selection, showing the hadronic tau p_T spectrum (left) and multiplicity (right) in the $\mu\tau_h$ channel (top) and the $e\tau_h$ channel (bottom).

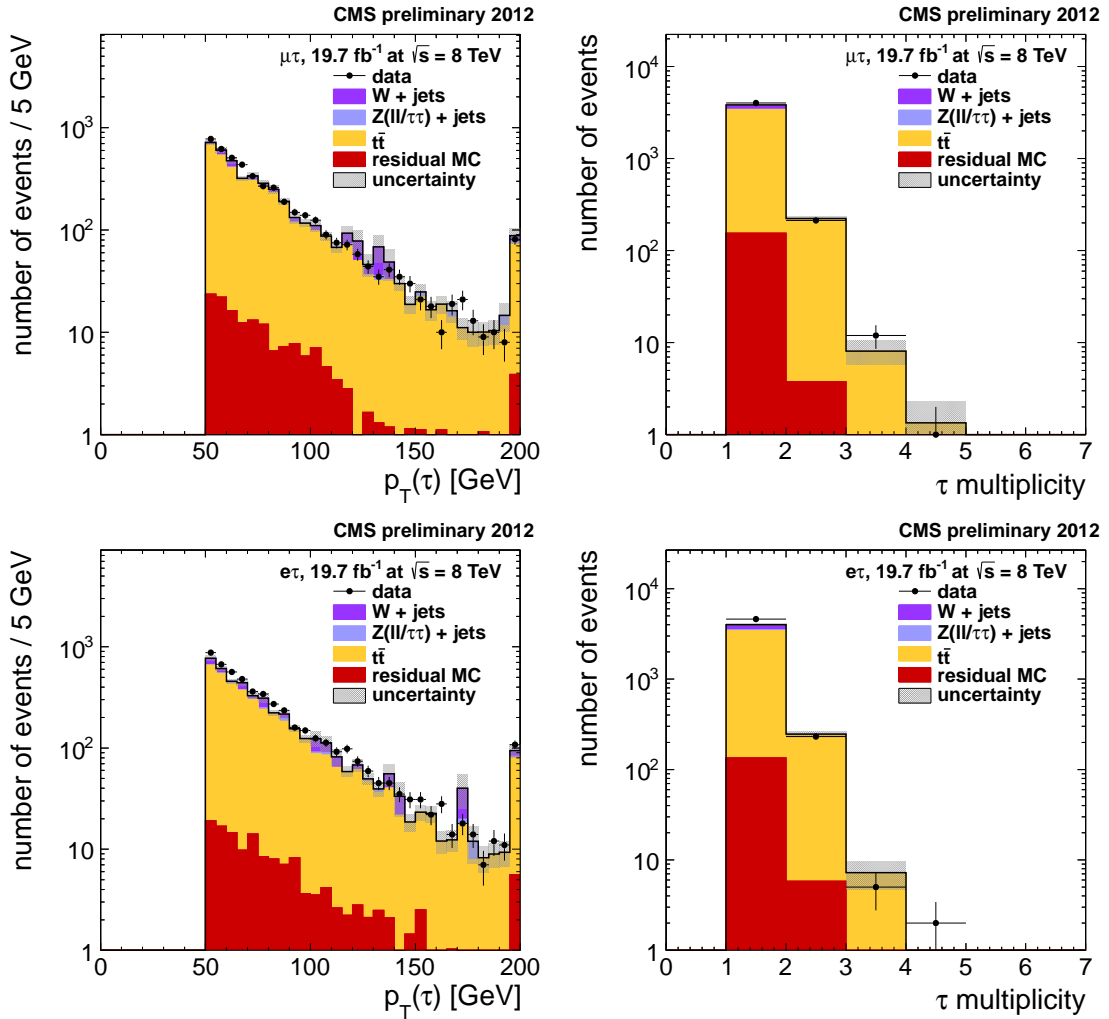


Figure 5.12: Plots of the anti-isolated control region after the top squark final selection, showing the hadronic tau p_T spectrum (left) and multiplicity (right) in the $\mu\tau_h$ channel (top) and the $e\tau_h$ channel (bottom).

event fail isolation. The isolation status of each tau in an event is independent, so this probability is given by the product of Eq. (5.15) for each tau, shown in Eq. (5.16). The signal region contains all events with at least one isolated hadronic tau. Therefore, the probability that an event in the total region is part of the signal region is given by the complement of the anti-isolated region probability, shown in Eq. (5.17).

$$P(\text{all } \tau\text{s in event are anti-isolated}) = \prod_{\tau}^{(\text{event})} [1 - f(p_{\text{T}}(\tau))] \quad (5.16)$$

$$P(\text{at least 1 } \tau \text{ in event is isolated}) = 1 - \prod_{\tau}^{(\text{event})} [1 - f(p_{\text{T}}(\tau))] \quad (5.17)$$

Equation (5.18) sums over all events in the anti-isolated region, using these complementary probabilities to produce a data-driven estimate of the major reducible background in the signal region.

$$N_{\text{misID } \tau} = \sum_{\text{events}}^{(\text{anti-iso})} \frac{1 - \prod_{\tau} [1 - f(p_{\text{T}}(\tau))]}{\prod_{\tau} [1 - f(p_{\text{T}}(\tau))]} \quad (5.18)$$

Events in the signal region from the exclusive samples simulated for the $W + \text{jets}$, $Z + \text{jets}$, and $t\bar{t}$ processes are used to predict the S_{T} distributions for the major reducible background. Because the dependency of the fake rate on p_{T} is similar in data and simulation, as shown in Fig. 5.9, the simulated distributions are expected not to be biased with respect to the observed data. Only the events in which the leading selected hadronic tau is a misidentified jet, based on matching between reconstructed and generated particles, are considered. The simulated distributions are normalized to the data-driven estimation of the major reducible background yield.

In the leptoquark search, the central value for the yield estimation is calculated using the inclusive fake rate from the observed data and the observed events in the anti-isolated region. Systematic errors on this estimation are assigned based on variations in the fake rate and in the anti-isolated region. The inclusive fake rate is varied by the statistical uncertainty σ_{stat} in both directions $f + 1\sigma_{\text{stat}}$ and $f - 1\sigma_{\text{stat}}$ separately for each bin. Each of the two variations is applied to the observed events in the anti-isolated region. The larger difference between the estimations from the two statistical variations and the central value is taken as a systematic uncertainty. Similarly, the $N_{\text{jet}} \geq 1$ fake rate from the observed data is applied to the observed events in the anti-isolated region, and the difference is taken as a systematic uncertainty. The contribution from residual backgrounds in the anti-isolated region is estimated using the inclusive fake rate from the observed data and included as an additional uncertainty. As mentioned previously, an uncertainty is also assessed due to the systematic difference in fake rates between $Z \rightarrow \mu\mu + \text{jets}$ events and semileptonic $t\bar{t}$ events.

In the top squark search, the systematic uncertainties are computed in the same way. However, the central value for the yield estimation is calculated using the $N_{\text{jet}} \geq 1$ fake rate from the observed data, in order to account for the higher jet multiplicity. Correspondingly, the $N_{\text{jet}} \geq 2$ fake rate from the observed data is used as a variation. The overall systematic uncertainties are larger for both channels in the top squark search, compared to the leptoquark search. This is caused by the higher jet multiplicities required when calculating the fake rates in the $Z \rightarrow \mu\mu + \text{jets}$ control region, which reduces the number of events. In addition, the background

in the top squark search is composed of a higher percentage of $t\bar{t}$ events, which increases the contribution from the systematic deviation between the $Z \rightarrow \mu\mu + \text{jets}$ fake rate and the $t\bar{t}$ fake rate.

Tables 5.10 and 5.11 show the results of the major reducible background estimation in the $\mu\tau_h$ channel for the leptoquark search and the top squark search, respectively. The predictions from the simulation are in agreement with the data-driven values and with the closure test performed using the simulated fake rate and simulated anti-isolated events. These predictions are taken from $W + \text{jets}$, $Z + \text{jets}$, and $t\bar{t}$ events in the signal region with matching between generated jets and reconstructed hadronic taus, normalized to $\mathcal{L} = 19.7 \text{ fb}^{-1}$. When contributions from all sources of systematic uncertainty are added in quadrature, the overall uncertainties are 16% for the leptoquark search and 23% for the top squark search. The statistical uncertainty from the anti-isolated region is approximately 1 – 2% in all cases, which is negligible in comparison.

In the $e\tau_h$ channel, there is a significant contribution from QCD events in the anti-isolated region. As described previously at the beginning of Sec. 5.3.2, the fake rate measured in the $Z \rightarrow \mu\mu + \text{jets}$ control region is not appropriate for QCD events. The QCD contribution must be subtracted from the major reducible background estimation calculated using the fake rate. Tables 5.12 and 5.13 show the results of the estimation in the $e\tau_h$ channel for the leptoquark search and the top squark search, respectively. The QCD contribution is already subtracted in these tables; the details of the subtraction will be discussed in Sec. 5.3.2.1. The sources of systematic uncertainty are the same as those considered in the $\mu\tau_h$ channel, with

$\mu\tau_h$ channel

Sources		LQ Results	
anti-iso	fake rate	yield	variation
data	data (incl.)	117.3	(central)
data	data (incl.) $\pm 1\sigma_{\text{stat}}$	128.3	11.0
data	data ($N_{\text{jet}} \geq 1$)	106.5	10.7
residual bkg.	data (incl.)	—	7.0
$Z \rightarrow \mu\mu$ vs. $t\bar{t}$ fake rates (18% \times 35%)		—	7.4
Final Result		117.3 ± 18.5	
Simulated Prediction		117.7 ± 28.7	
Closure Test		121.7 ± 3.5	

Table 5.10: The results of the data-driven major reducible background estimation for the leptoquark search in the $\mu\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.

an additional systematic uncertainty from the uncertainty on the subtracted QCD contribution. The results from the data-drive estimations are consistent with the simulated predictions and closure tests. The total uncertainty on the $e\tau_h$ channel yield is 17% for the leptoquark search and 24% for the top squark search.

5.3.2.1 QCD Multijet Reducible Background

The estimation of the contribution from the QCD multijet process in the $e\tau_h$ channel requires two steps. The presence of QCD events in the observed data in the $e\tau_h$ anti-isolated control region biases the previous estimation, as these events are weighted by the inappropriate quark-jet fake rate. Therefore, the first step must be to estimate the yield and p_T distribution of the QCD events in the anti-isolated observed data, in order to correct for that bias by subtracting the contribution from QCD. The second step is to perform an independent estimation of the QCD

$\mu\tau_h$ channel

Sources		\tilde{t} Results	
anti-iso	fake rate	yield	variation
data	data ($N_{\text{jet}} \geq 1$)	59.8	(central)
data	data ($N_{\text{jet}} \geq 1$) $\pm 1\sigma_{\text{stat}}$	67.0	7.2
data	data ($N_{\text{jet}} \geq 2$)	54.4	5.4
residual MC	data ($N_{\text{jet}} \geq 1$)	—	2.1
$Z \rightarrow \mu\mu$ vs. $t\bar{t}$ fake rates (18% \times 95%)		—	10.2
Final Result		59.8 ± 13.8	
Simulated Prediction		57.5 ± 6.5	
Closure Test		58.0 ± 1.3	

Table 5.11: The results of the data-driven major reducible background estimation for the top squark search in the $\mu\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.

contribution in the signal region using the observed data.

In both steps, a same-sign/opposite-sign (SS/OS) method is used to estimate the number of QCD events in a given region of the data. A same-sign control region is defined by requiring that the light lepton and hadronic tau have the same electric charge, instead of the opposite charge. Because both objects are misidentified jets in QCD events, their charge assignments are expected to be random, so the number of events in which both objects have the same charge should be the same as for the opposite charge. In practice, a slight deviation between the numbers of same-sign and opposite-sign events is possible, for example due to charge asymmetries in proton-proton collisions. The scale factor relating the numbers of same-sign and opposite-sign events was measured to be 1.06 in Ref. [105].

The non-QCD background in the same-sign control region is estimated using the simulation and subtracted from the observed data; any remaining events are

$e\tau_h$ channel

Sources		LQ Results	
anti-iso	fake rate	yield	variation
data	data (incl.)	124.2	(central)
data	data (incl.) $\pm 1\sigma_{\text{stat}}$	138.5	14.3
data	data ($N_{\text{jet}} \geq 1$)	113.9	10.3
residual MC	data (incl.)	—	7.1
QCD	data (incl.)	—	2.7
$Z \rightarrow \mu\mu$ vs. $t\bar{t}$ fake rates (18% \times 35%)		—	7.8
Final Result		124.2 ± 20.7	
Simulated Prediction		106.8 ± 27.3	
Closure Test		129.5 ± 3.6	

Table 5.12: The results of the data-driven major reducible background estimation for the leptoquark search in the $e\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.

assumed to originate from the QCD process. The scale factor of 1.06 is applied to extrapolate from this yield $N_{\text{QCD}}^{\text{SS}}$ to the opposite-sign region $N_{\text{QCD}}^{\text{OS}}$:

$$N_{\text{QCD}}^{\text{SS}} = N_{\text{data}}^{\text{SS}} - N_{\text{MC}}^{\text{SS}}, \quad (5.19)$$

$$N_{\text{QCD}}^{\text{OS}} = 1.06 N_{\text{QCD}}^{\text{SS}}. \quad (5.20)$$

The subtraction of non-QCD backgrounds can potentially involve a large number of events. In order to ensure the validity of the subtraction, the normalization of the simulated samples must be carefully checked.

To check the normalization of the simulated $W + \text{jets}$, $Z \rightarrow \ell^+\ell^-$, and $Z \rightarrow \tau^+\tau^-$ samples, a region of data is defined using the preselection criteria listed in Sec. 5.2.1 without the cut $N_{\text{jets}} \geq 2$. Most of the events at the preselection level are eventually eliminated from the signal region by the stricter main and final selections, so the contents of the signal region make up a relatively small portion of this larger

$e\tau_h$ channel

Sources		\tilde{t} Results	
anti-iso	fake rate	yield	variation
data	data ($N_{\text{jet}} \geq 1$)	65.7	(central)
data	data ($N_{\text{jet}} \geq 1$) $\pm 1\sigma_{\text{stat}}$	74.6	8.9
data	data ($N_{\text{jet}} \geq 2$)	59.8	5.9
residual MC	data ($N_{\text{jet}} \geq 1$)	—	1.8
QCD	data ($N_{\text{jet}} \geq 1$)	—	0.1
$Z \rightarrow \mu\mu$ vs. $t\bar{t}$ fake rates (18% \times 95%)		—	11.2
Final Result		65.7 ± 15.6	
Simulated Prediction		66.7 ± 7.6	
Closure Test		61.6 ± 1.4	

Table 5.13: The results of the data-driven major reducible background estimation for the top squark search in the $e\tau_h$ channel. This table shows all sources of systematic uncertainty and a comparison to the simulated prediction. Only statistical uncertainty is given for the simulated prediction and the closure test. From the two variations of the fake rate $\pm 1\sigma_{\text{stat}}$, the estimation with the larger difference from the central value is used.

region. The transverse mass of the electron and \cancel{E}_T system, $M_T(e, \cancel{E}_T)$, is defined in Eq. (5.21):

$$M_T(e, \cancel{E}_T) = \sqrt{2p_T^{(e)} \cancel{E}_T [1 - \cos(\Delta\phi(e, \cancel{E}_T))]} \quad (5.21)$$

This definition assumes both particles in the system are massless, which is a good approximation for highly relativistic electrons and neutrinos. The M_T distribution in the region defined above is used to determine normalization parameters for the simulated samples by comparing them to the observed data:

$$M_T^{(\text{data})} = r_W M_T^{(W + \text{jets})} + r_\tau M_T^{(Z \rightarrow \tau^+ \tau^-)} + r_\ell M_T^{(Z \rightarrow \ell^+ \ell^-)} + M_T^{(t\bar{t}, t, VV)}. \quad (5.22)$$

The difference between the sum of the simulated M_T distributions and the observed M_T distribution is minimized by varying the three normalization parameters r_W , r_τ , and r_ℓ . The normalization of the other processes, $t\bar{t}$, single top, and diboson, is not

addressed in this minimization. The observed distribution contains a contribution from QCD which is not modeled in the simulation. Therefore, at least one of the normalization parameters will be inflated during the minimization in order to include this contribution. Measurements of the inclusive W and Z production cross sections have demonstrated a similarity between the M_T distributions for $Z \rightarrow \ell^+\ell^-$ and QCD events when using a selection tuned for $W \rightarrow \ell\nu$ events [106, 107]. Hypothesizing that this similarity holds for the selection considered here, the parameter r_ℓ should scale the simulated $Z \rightarrow \ell^+\ell^-$ yield to include the QCD yield.

The minimization in Eq. (5.22) provides the following values for the r parameters: $r_W = 0.86$, $r_\tau = 1.21$, $r_\ell = 2.02$. Using these parameters, corrected yields N can be defined for the simulated samples, based on the uncorrected yields N_{MC} :

$$N(\text{W} + \text{jets}) = r_W N_{\text{MC}}(\text{W} + \text{jets}), \quad (5.23)$$

$$N(Z \rightarrow \tau^+\tau^-) = r_\tau N_{\text{MC}}(Z \rightarrow \tau^+\tau^-), \quad (5.24)$$

$$N(Z \rightarrow \ell^+\ell^-) + N(\text{QCD}) = r_\ell N_{\text{MC}}(Z \rightarrow \ell^+\ell^-). \quad (5.25)$$

Equation (5.25) arises from the hypothesis that the $Z \rightarrow \ell^+\ell^-$ yield will be scaled to include the QCD yield. This can be used to check the validity of the M_T minimization method for correcting the simulated sample normalizations. The parameter r_τ is assigned to be the normalization correction for both the $Z \rightarrow \ell^+\ell^-$ and $Z \rightarrow \tau^+\tau^-$ samples, in order to solve for $N(\text{QCD})$ in terms of known quantities:

$$N(Z \rightarrow \ell^+\ell^-) = r_\tau N_{\text{MC}}(Z \rightarrow \ell^+\ell^-), \quad (5.26)$$

$$N(\text{QCD}) = r_\ell N_{\text{MC}}(Z \rightarrow \ell^+\ell^-) - r_\tau N_{\text{MC}}(Z \rightarrow \ell^+\ell^-). \quad (5.27)$$

Equation (5.27) follows from the combination of Eqs. (5.25) and (5.26). This equation produces the value $N(\text{QCD}) = 24765 \pm 1411$ for the region defined by the preselection criteria without the cut $N_{\text{jets}} \geq 2$. In comparison, the SS/OS method applied directly to this region predicts the value $N(\text{QCD}) = 26424 \pm 1250$. These two values agree within uncertainties, validating the assumptions made in the M_T minimization method. The normalization parameters r_W and r_τ will be used for the $W + \text{jets}$ and $Z + \text{jets}$ normalizations in the various same-sign and anti-isolated control regions necessary to conduct the two steps of the QCD estimation in the signal region. A separate control region requiring $N_{\text{b-jet}} \geq 2$ and $M_T > 70 \text{ GeV}$ is used to check the normalization of the $t\bar{t}$ simulation, and no correction is found to be necessary.

To measure the QCD portion of the data in the anti-isolated region, the SS/OS method described by Eqs. (5.19) and (5.20) is applied, using the normalization parameters r_W and r_τ . Figure 5.13 shows the $M_T (e, \cancel{E}_T)$ distribution in the same-sign anti-isolated control region for the leptoquark search, with an overall excess in the observed data indicating the presence of QCD. This method predicts 3026 ± 210 QCD events after the leptoquark final selection and 152 ± 15 events after the top squark final selection. In order to subtract the contribution of these events from the reducible background estimation, the p_T distribution of the hadronic tau candidates in the QCD events is needed. Figure 5.11 shows that the QCD events in the $e\tau_h$ anti-isolated region are found primarily in the single tau multiplicity bin. Therefore, it is sufficient to subtract the simulated tau p_T distribution from the observed tau p_T distribution to obtain the QCD tau p_T distribution. A simplified version of Eq.

(5.18) can be used to weight these single-tau events:

$$N_{\text{misID } \tau}^{(\text{QCD})} = N_{\text{anti-iso}}^{(\text{QCD})} \sum_{p_T} \frac{f(p_T)}{1 - f(p_T)}. \quad (5.28)$$

Using Eq. (5.28), the QCD yield to subtract is found to be 38.5 ± 2.7 events in the leptoquark search and 1.5 ± 0.1 events in the top squark search. The subtraction of these QCD yields is already included in the results shown in Tables 5.12 and 5.12, with the statistical uncertainties on the QCD yields included as additional systematic uncertainties on the final major reducible background estimation.

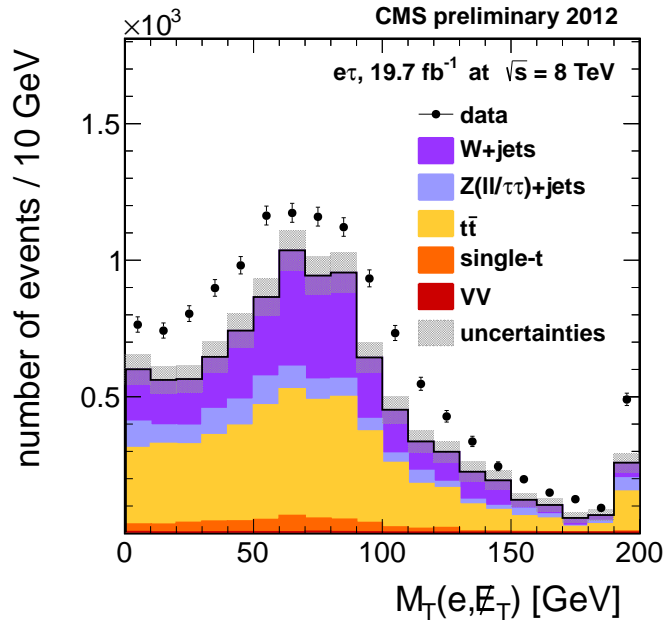


Figure 5.13: The transverse mass of the electron and missing transverse energy system for the same-sign anti-isolated control region, after the leptoquark final selection. The overall excess in the observed data, not localized in a specific range of M_T values, indicates the presence of QCD events.

After applying the correction to account for the presence of QCD in the anti-isolated region, the actual contribution from QCD, if any, must be obtained. Again, the SS/OS method is used, now with the signal region. To decrease the statistical

uncertainty in this estimation, the same-sign control region is defined before the final selection. For the leptoquark search, this means that the cut $M(\tau_h, \text{jet}) > 250 \text{ GeV}$ is not applied. Figure 5.14 shows the \cancel{E}_T distribution in this control region, with an excess in the observed data at low \cancel{E}_T indicating the presence of QCD. Using the normalization corrections derived previously, the simulated yield in this control region is found to be 474 ± 18 events, while the observed yield is found to be 736 ± 27 , which gives $N_{\text{QCD}}^{\text{SS}} = 262 \pm 32$ and correspondingly $N_{\text{QCD}}^{\text{OS}} = 277 \pm 34$. To extrapolate this QCD yield to the region defined by the leptoquark final selection, the efficiency of the cut $M(\tau_h, \text{jet}) > 250 \text{ GeV}$ is measured in a same-sign control region which vetoes events containing one or more b-tagged jets, enhancing the contribution from QCD. The QCD yields before and after the mass cut are defined as the subtraction of the simulated yields from the observed yields:

$$N_{\text{QCD}}^{\text{before}} = N_{\text{data}}^{\text{before}} - N_{\text{MC}}^{\text{before}} = (793 \pm 28) - (469 \pm 19) = 324 \pm 34, \quad (5.29)$$

$$N_{\text{QCD}}^{\text{after}} = N_{\text{data}}^{\text{after}} - N_{\text{MC}}^{\text{after}} = (93 \pm 10) - (66 \pm 7) = 27 \pm 12. \quad (5.30)$$

The ratio of these two yields is the efficiency of the cut, $\varepsilon_{M(\tau_h, \text{jet})} = 8.5\% \pm 4.0\%$.

Thus, Eq. (5.31) calculates the final yield from QCD in the leptoquark search:

$$N_{\text{QCD}}^{\text{final}} = N_{\text{QCD}}^{\text{OS}} \times \varepsilon_{M(\tau_h, \text{jet})} = (277 \pm 34) \times (8.5\% \pm 4\%) = 23.6 \pm 12. \quad (5.31)$$

To check this estimation, the SS/OS method is applied to the signal region after the cut $M(\tau_h, \text{jet}) > 250 \text{ GeV}$. This check gives a less precise result $N_{\text{QCD}}^{\text{SS/OS}} = 31.8 \pm 21.2$, which is fully compatible with the above result within uncertainties.

In addition to the yield, the S_T distribution for the QCD process has to be estimated from the observed data, as no simulated sample is available to produce

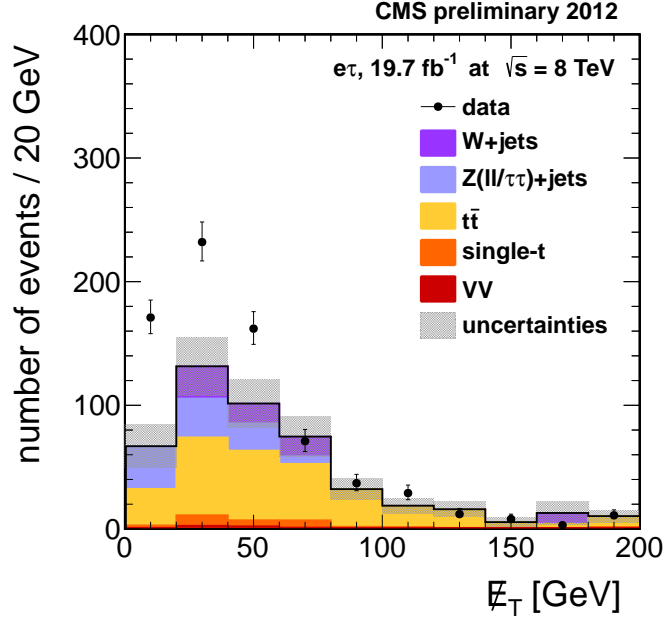


Figure 5.14: The missing transverse energy spectrum for the same-sign control region selected before the $M(\tau_h, \text{jet}) > 250$ GeV requirement for the leptoquark search. The excess of observed events at low \cancel{E}_T is an indication of the presence of QCD events.

it. The distribution is obtained by subtracting the simulated S_T distribution of the non-QCD backgrounds from the observed S_T distribution in the same-sign control region after the full leptoquark final selection, as shown in Fig. 5.15. The bins with negative values, all of which are equivalent to zero within their statistical uncertainties, are set to be zero to avoid unphysical values in the distribution. The QCD S_T distribution obtained from this method is then added to the major reducible S_T distribution for the $e\tau_h$ channel. The propagated statistical uncertainty on the QCD yield, ± 12 events, is added in quadrature with the systematic uncertainty from the major reducible background estimation.

In the top squark search, the QCD background is found to be 0 ± 13 using the SS/OS method with an extrapolation of the efficiency of the cut $N_{\text{jets}} \geq 5$.

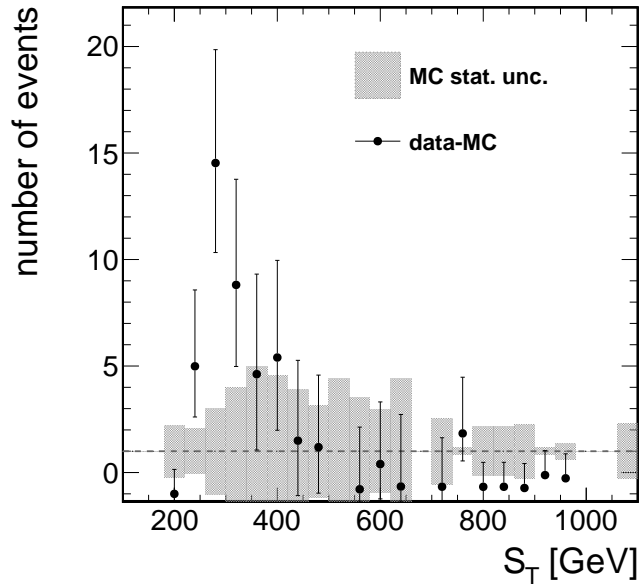


Figure 5.15: The QCD S_T distribution estimated by subtracting the simulated distribution from the observed distribution in the same-sign control region after the leptoquark final selection. The negative values are set to zero when using the distribution.

Applying the SS/OS method after the final selection to check that prediction finds 0 ± 18 . These predictions agree and are both equivalent to zero events, so no QCD background is added to the $e\tau_h$ channel in the top squark search. This lack of QCD background is expected due to the high jet multiplicity requirement in this selection.

5.3.3 Other Backgrounds

The $t\bar{t}$ irreducible and major reducible backgrounds described above are the major backgrounds for the signal processes considered in this dissertation. Some background categories are not included in these data-driven estimations. These backgrounds contribute only a small number of events after the final selections. They

are estimated from the simulation, normalized using the cross sections listed in Table C.2. The contributions from the diboson and single-top-quark processes are entirely estimated using the simulation. Matching between the generator-level particles and the reconstructed tau object is used to identify the irreducible contribution from $Z \rightarrow \tau^+\tau^- + \text{jets}$. The reducible contributions from $t\bar{t}$ with two light leptons and $Z \rightarrow \ell^+\ell^- + \text{jets}$, when a light lepton is misidentified as a hadronic tau, are also estimated from the simulation.

5.4 Systematic Uncertainties

There are a number of systematic uncertainties associated with both the background estimations and the simulation of the signals. For the simulated backgrounds and signals, these uncertainties arise from several sources, including discrepancies between the observed data and the simulation in the performance of reconstruction algorithms. Some sources of uncertainty affect both the S_T distributions and the simulated yields. Table 5.14 summarizes all of the systematic uncertainties in the simulated samples. For the data-driven background estimations, variations and uncertainties in the components of the methods used for the estimations are combined to compute the overall systematic uncertainties, as discussed in Secs. 5.3.1 and 5.3.2. The overall systematic uncertainties in the data-driven background estimations are listed in Table 5.15.

The identification of light leptons can have systematic uncertainties associated with the efficiency of the HLT criteria, the identification algorithms, and the

Source	Uncertainty	Effect on:			
		Signal	Z + jets	Single t	VV
(e, μ) ID, iso, HLT	2%	2%	2%	2%	2%
τ_h ID, iso	6%	6%	6%	6%	6%
b-tagging	$\sim 4\%$	3%	1%	3%	1%
mistagging	$\sim 10\%$	1%	4%	1%	2%
pileup	6%	3%	3%	3%	3%
luminosity	2.6%	2.6%	2.6%	2.6%	2.6%
cross section	–	–	2%	14%	5–15%
statistical	–	–	20–40%	20–40%	20–40%
ISR/FSR	–	4%	–	–	–
τ_h energy scale	3%	0–5%	5–19%	5–19%	5–19%
τ_h energy resolution	10%	1–9%	20%	20%	20%
jet energy scale	$\sim 4\%$	1%	0–7%	0–7%	0–7%
jet energy resolution	5–10%	1%	0–5%	0–5%	0–5%

Table 5.14: The relative systematic uncertainties on the yields of the simulated signal and the simulated backgrounds. The magnitude of the effects may vary within indicated ranges for different signal masses, for different background processes within a given category, or for the different searches.

Channel		$t\bar{t}$ irreducible	Major reducible
LQ	$e\tau_h$	17%	16%
	$\mu\tau_h$	19%	16%
\tilde{t}	$e\tau_h$	16%	24%
	$\mu\tau_h$	17%	23%

Table 5.15: The relative systematic uncertainties on the yields of the major backgrounds estimated from the observed data for each channel of each search.

computation of the isolation. The contributions from these sources are measured together using tag and probe methods to compare $Z \rightarrow \ell^+\ell^-$ events in the observed data and the simulation for electrons [108] and muons [109]. For both types of light leptons, the systematic uncertainty is found to be 2%. The uncertainty in the identification and isolation of hadronic taus using the HPS algorithm is measured to be 6%, using tag and probe methods with $Z \rightarrow \tau^+\tau^-$ events [86]. The uncertainty in the performance of the lepton-tau discriminators is negligible.

The correction factors applied to the simulation for b-tagging and mistagging efficiencies have associated uncertainties of $\sim 4\%$ and $\sim 10\%$, respectively, with dependence on p_T and η . The effects of these uncertainties are propagated to the yields estimated from the simulation by varying the correction factors. The relative systematic uncertainties in the simulated yields are 1–3% from the b-tagging efficiency corrections and 1–4% from the mistagging efficiency corrections.

Pileup interactions at the LHC contribute additional energy to events beyond the energy from the primary hard-scattering interaction. The uncertainty in the modeling of pileup interactions in the simulation is estimated to be 6% [110]. This can affect lepton isolation and the jet energy scale. However, the CMS reconstruction algorithms for leptons and jets have approximately pileup-independent performance after subtracting pileup contributions when calculating isolation and energy scales. Therefore, the effect of the pileup uncertainty on the final event yields is expected to be small. A conservative relative uncertainty of 3% is assigned to the simulated signal and background yields.

The normalization of the simulated samples involves the calculated cross sections and the measured integrated luminosity for the observed data. As discussed in Sec. 3.8, the uncertainty in the measured luminosity is 2.6%. The uncertainties in the calculated cross sections of the simulated background processes are assessed by comparison to measurements in observed data. For the Z + jets process, the uncertainty is found to be 2% [111]. For the diboson processes, the uncertainty varies from 5–14% depending on the process [112]. For the single top quark process, the uncertainty is 14% [113]. An additional uncertainty of 20–40% is assigned to

the simulated backgrounds, based on the statistical uncertainty due to the limited number of events in the simulation. The uncertainty in the modeling of initial- and final-state radiation in the simulation affects the signal yield at the level of 4% and has a negligible effect on the background yields. The theoretical cross section for the signal has uncertainties of 7–32% from the measurement of PDFs and 14–80% from the variation of QCD renormalization and factorization scales, as shown in Sec. 2.3.

The uncertainty in the modeling of energy scales and energy resolutions for reconstructed objects in the simulation can affect both the yields and the S_T distributions. To account for this, the energy scale or energy resolution is varied independently for each type of object, and then the whole analysis is repeated with the varied quantity. This produces a varied S_T distribution whose difference from the nominal S_T distribution is considered to be the uncertainty in the distribution. The effects of these uncertainties on all of the S_T distributions, which are estimated from the simulation for both the simulation-based and data-driven backgrounds as well as the signal, are considered. The effect of the light lepton energy scales is negligible, as the disagreement between the simulation and the observed data is only 1%. The uncertainties in the tau energy scale and energy resolution are 3% and 10% [102]. These lead to uncertainties in the signal yield of 0–5% and 1–9%, respectively, and uncertainties in the simulated background yields of 5–19% and 20%, respectively. The jet energy scale uncertainty varies with p_T and η , with a typical level of $\sim 4\%$; the jet energy resolution uncertainty varies with η between 5–10% [81, 114]. These both cause uncertainties in the signal yield of 1%, and in the simulated background yields 0–7% and 0–5%, respectively.

5.5 Results

The results of these searches are published in Ref. [115]. The numbers of observed events and expected signal and background events after the final selections for the leptoquark and top squark searches are listed in Tables 5.16 and 5.17, respectively, and the selection efficiencies for the two signals are listed in Tables 5.18 and 5.19. The S_T distributions of the selected events from the observed data and from the background predictions, combining the $e\tau_h$ and $\mu\tau_h$ channels, are shown in Fig. 5.16 for the leptoquark search and Fig. 5.17 for the top squark search. The distribution from the 500 GeV (300 GeV) signal hypothesis is added to the background in Fig. 5.16 (Fig. 5.17) to illustrate how a hypothetical signal would appear above the background prediction. There is good agreement between the observed data and the SM background prediction.

An upper bound at the 95% confidence level (CL) is set on $\sigma\mathcal{B}^2$. In the leptoquark search, σ is the cross section for pair production of third-generation LQs and \mathcal{B} is the branching fraction for the decay $LQ \rightarrow \tau b$. In the top squark search, σ is the cross section for pair production of top squarks and \mathcal{B} is the branching fraction for the decay $\tilde{t} \rightarrow \tilde{\chi}^\pm b, \tilde{\chi}^\pm \rightarrow \tilde{\nu}\tau^\pm \rightarrow qq\tau^\pm$. The modified-frequentist construction CL_s [116, 117, 118] is used for the limit calculation. A maximum likelihood fit is performed to the S_T distribution simultaneously for the $e\tau_h$ and $\mu\tau_h$ channels, taking into account correlations between the systematic uncertainties. Appendix A contains more details on the computation of CL_s limits using a distribution. The expected and observed upper limits on $\sigma\mathcal{B}^2$ as a function of the signal mass are shown in

Fig. 5.18 for the leptoquark search and Fig. 5.19 for the top squark search.

We extend the current limits and exclude scalar leptoquarks and top squarks decaying through the coupling λ'_{333} with masses below 740 GeV, in agreement with the expected limit of 750 GeV. We exclude top squarks undergoing a chargino-mediated decay involving the coupling λ'_{3jk} with masses in the range 200–580 GeV, in agreement with the expected limit in the range 200–590 GeV. These upper limits assume $\mathcal{B} = 100\%$. Similar results are obtained when calculating upper bounds using a Bayesian method with a uniform positive prior for the cross section.

The upper bounds for the leptoquark search as a function of the leptoquark branching fraction and mass are shown in Fig. 5.20. Small \mathcal{B} values are not constrained by this search. Results from the CMS experiment on a search for top squarks decaying to a top quark and a neutralino [119] are used to improve the constraints on \mathcal{B} . If the neutralino is massless, the final state kinematic distributions for such a signal are the same as those for the pair production of leptoquarks decaying to a tau neutrino and a top quark. Limits can therefore be placed on this signal, which must have a branching fraction of $1 - \mathcal{B}$ if the leptoquark only decays to third-generation fermions. This reinterpretation is included in Fig. 5.20. The unexcluded region at $M_{LQ} = 200\text{--}230$ GeV corresponds to a portion of phase space where it is topologically very difficult to distinguish between the top squark signal and the $t\bar{t}$ process, due to small \cancel{E}_T . A top squark excess in this region would imply an excess in the measured $t\bar{t}$ cross section of $\sim 10\%$.

	$e\tau_h$	$\mu\tau_h$
$t\bar{t}$ irreducible	105.6 \pm 18.1	66.7 \pm 12.6
Major reducible	147.8 \pm 33.0	117.3 \pm 18.9
$Z(\ell\ell/\tau\tau) + \text{jets}$	21.4 \pm 7.4 \pm 4.9	7.5 \pm 4.6 \pm 0.2
Single t	16.0 \pm 2.8 \pm 4.4	17.3 \pm 2.8 \pm 4.7
VV	4.1 \pm 0.6 \pm 1.3	2.6 \pm 0.5 \pm 0.8
Total exp. bkg.	294.9 \pm 7.9 \pm 39.1	211.4 \pm 5.4 \pm 23.4
Observed	289	216
$M_{LQ} = 500 \text{ GeV}$	57.7 \pm 1.4 \pm 5.9	51.6 \pm 1.3 \pm 5.3
$M_{LQ} = 600 \text{ GeV}$	20.1 \pm 0.5 \pm 1.9	17.7 \pm 0.4 \pm 1.6
$M_{LQ} = 700 \text{ GeV}$	7.1 \pm 0.2 \pm 6.3	6.2 \pm 0.1 \pm 5.5
$M_{LQ} = 800 \text{ GeV}$	2.7 \pm 0.1 \pm 0.2	2.3 \pm 0.1 \pm 0.2

Table 5.16: The estimated backgrounds, observed event yields, and expected number of signal events for the leptoquark search. For the simulation-based entries, the statistical and systematic uncertainties are shown separately, in that order. Only the systematic uncertainties are shown for the data-driven background entries.

	$e\tau_h$	$\mu\tau_h$
$t\bar{t}$ irreducible	88.3 \pm 13.7	55.0 \pm 9.5
Major reducible	65.7 \pm 16.4	59.8 \pm 13.8
$Z(\ell\ell/\tau\tau) + \text{jets}$	4.9 \pm 2.5 \pm 1.1	11.6 \pm 5.5 \pm 2.7
Single t	3.9 \pm 1.5 \pm 1.1	3.5 \pm 1.3 \pm 0.9
VV	0.6 \pm 0.2 \pm 0.2	0.4 \pm 0.2 \pm 0.1
Total exp. bkg.	163.4 \pm 2.9 \pm 21.5	130.3 \pm 5.6 \pm 17.1
Observed	156	123
$M_{\tilde{t}} = 300 \text{ GeV}$	94.3 \pm 8.5 \pm 13.2	82.8 \pm 8.0 \pm 11.7
$M_{\tilde{t}} = 400 \text{ GeV}$	43.9 \pm 2.6 \pm 4.3	38.3 \pm 2.3 \pm 3.8
$M_{\tilde{t}} = 500 \text{ GeV}$	19.4 \pm 0.8 \pm 1.8	15.4 \pm 0.7 \pm 1.5
$M_{\tilde{t}} = 600 \text{ GeV}$	6.9 \pm 0.9 \pm 0.7	5.7 \pm 0.3 \pm 0.5

Table 5.17: The estimated backgrounds, observed event yields, and expected number of signal events for the top squark search. For the simulation-based entries, the statistical and systematic uncertainties are shown separately, in that order. Only the systematic uncertainties are shown for the data-driven background entries.

$M_{LQ} \text{ (GeV)}$	200	250	300	350	400	450	500	550	600
$e\tau_h$	0.1	0.3	1.0	1.9	2.4	3.0	3.6	4.0	4.4
$\mu\tau_h$	0.1	0.2	0.8	1.5	2.3	2.9	3.2	3.3	3.8
$M_{LQ} \text{ (GeV)}$	650	700	750	800	850	900	950	1000	
$e\tau_h$	4.5	4.7	4.9	5.1	5.4	5.1	5.4	5.5	
$\mu\tau_h$	4.0	4.1	4.2	4.3	4.4	4.4	4.3	4.4	

Table 5.18: Selection efficiencies in % for the signal in the leptoquark search, estimated from the simulation.

$M_{\tilde{t}}$ (GeV)	200	300	400	500	600	700	800	900
$e\tau_h$	0.02	0.3	0.7	1.2	1.5	1.8	1.8	1.5
$\mu\tau_h$	0.02	0.2	0.6	1.0	1.2	1.4	1.3	1.1

Table 5.19: Selection efficiencies in % for the signal in the top squark search, estimated from the simulation.

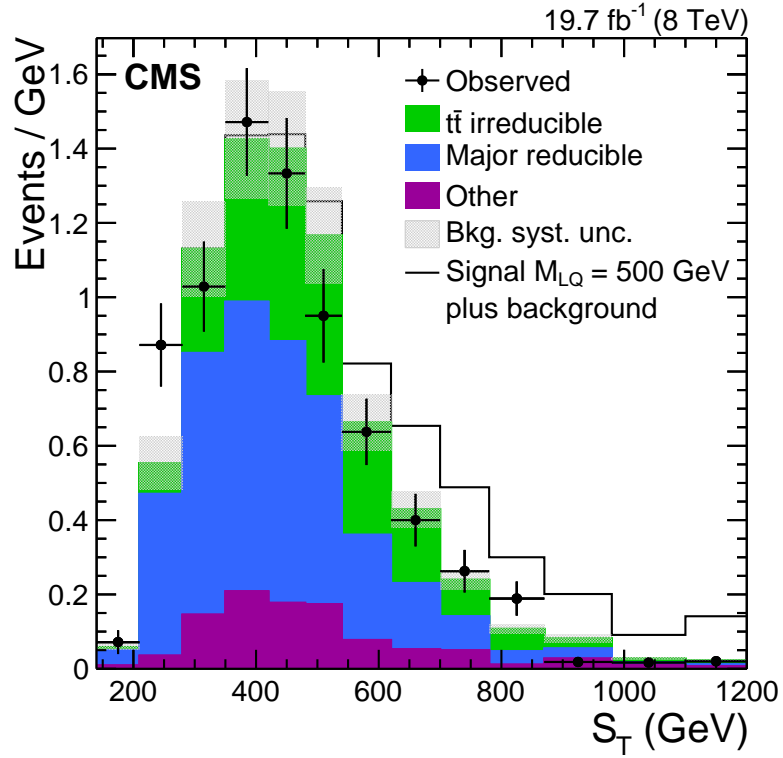


Figure 5.16: The final S_T distribution for the leptoquark search with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for leptoquarks with $M_{LQ} = 500$ GeV is added on top of the background prediction. The last bin contains the overflow events. The horizontal bar on each observed data point indicates the width of the bin in S_T .

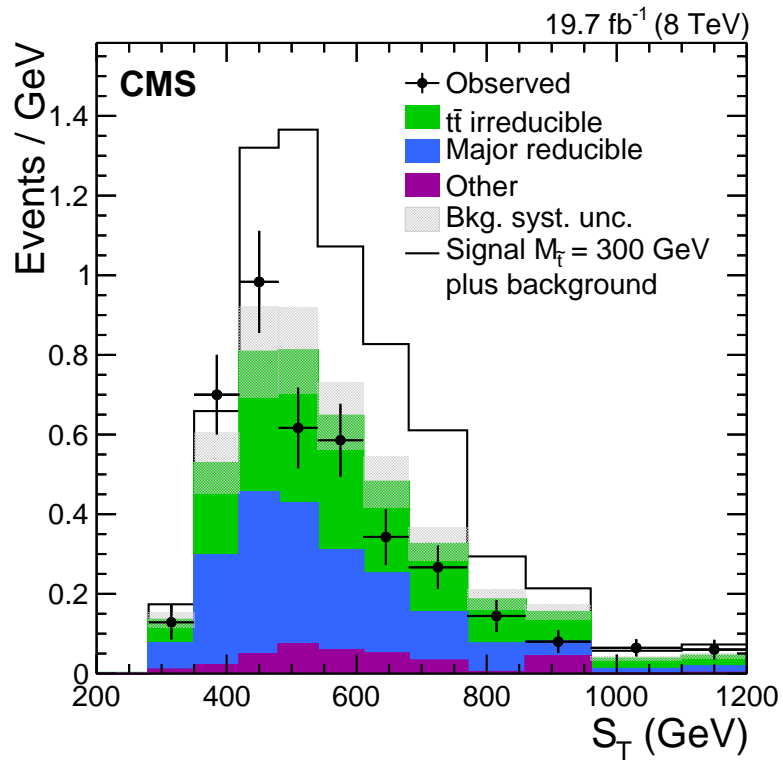


Figure 5.17: The final S_T distribution for the top squark search with the $e\tau_h$ and $\mu\tau_h$ channels combined. A signal sample for top squarks with $M_{\tilde{t}} = 300$ GeV is added on top of the background prediction. The last bin contains the overflow events. The horizontal bar on each observed data point indicates the width of the bin in S_T .

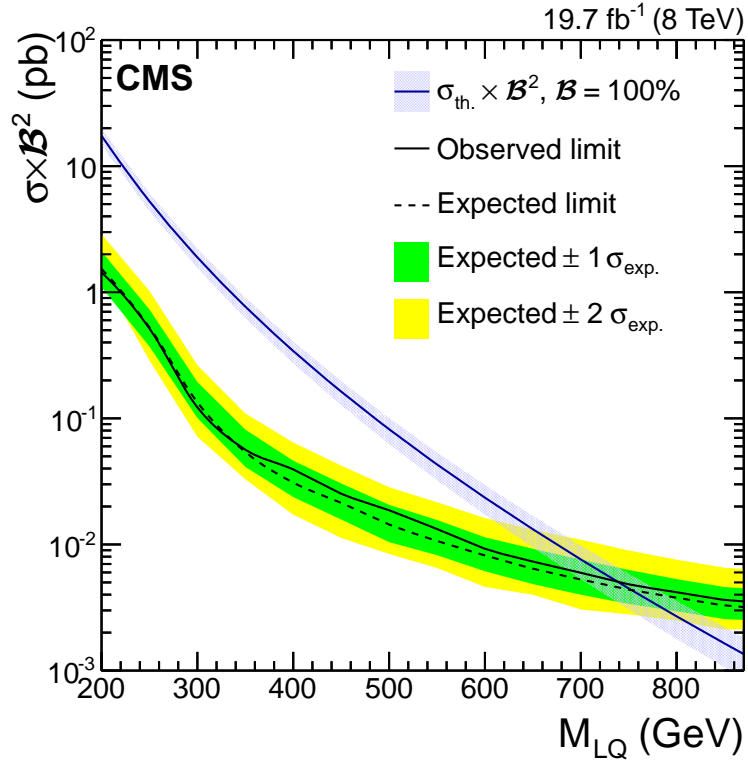


Figure 5.18: The expected and observed combined upper limits on the third-generation LQ pair production cross section σ times the square of the branching fraction, \mathcal{B}^2 , at the 95% CL, as a function of the LQ mass. These limits also apply to top squarks decaying directly via the coupling λ'_{333} . The green (darker) and yellow (lighter) uncertainty bands represent the 68% and 95% CL intervals on the expected limit. The dark blue curve and the hatched light blue band represent the theoretical LQ pair production cross section, assuming $\mathcal{B} = 100\%$, and the uncertainties due to the choice of PDF and renormalization/factorization scales.

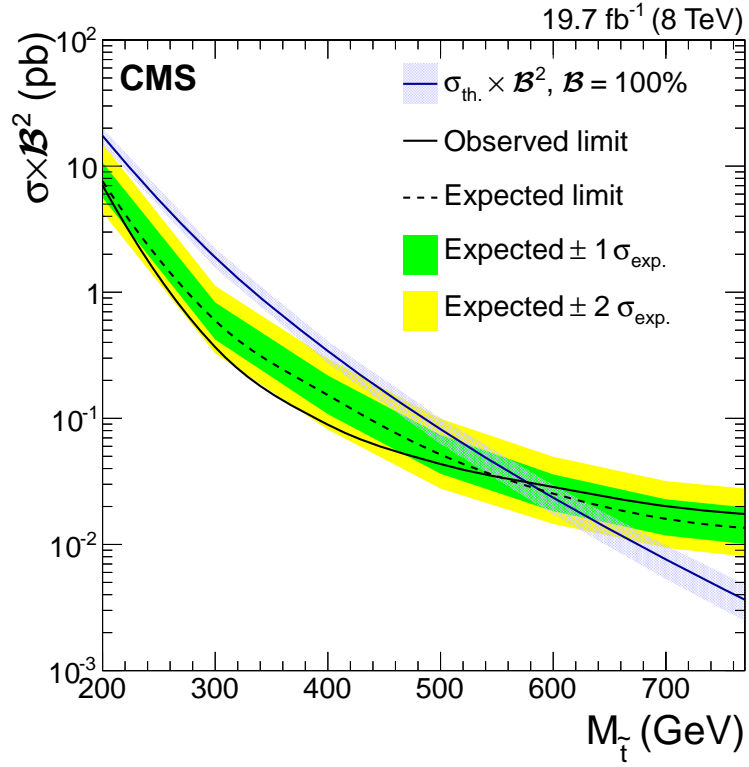


Figure 5.19: The expected and observed combined upper limits on the top squark pair production cross section σ times the square of the branching fraction, \mathcal{B}^2 , at the 95% CL, as a function of the top squark mass. These limits apply to top squarks with a chargino-mediated decay through the coupling λ_{3kj} . The green (darker) and yellow (lighter) uncertainty bands represent the 68% and 95% CL intervals on the expected limit. The dark blue curve and the hatched light blue band represent the theoretical top squark pair production cross section, assuming $\mathcal{B} = 100\%$, and the uncertainties due to the choice of PDF and renormalization/factorization scales.

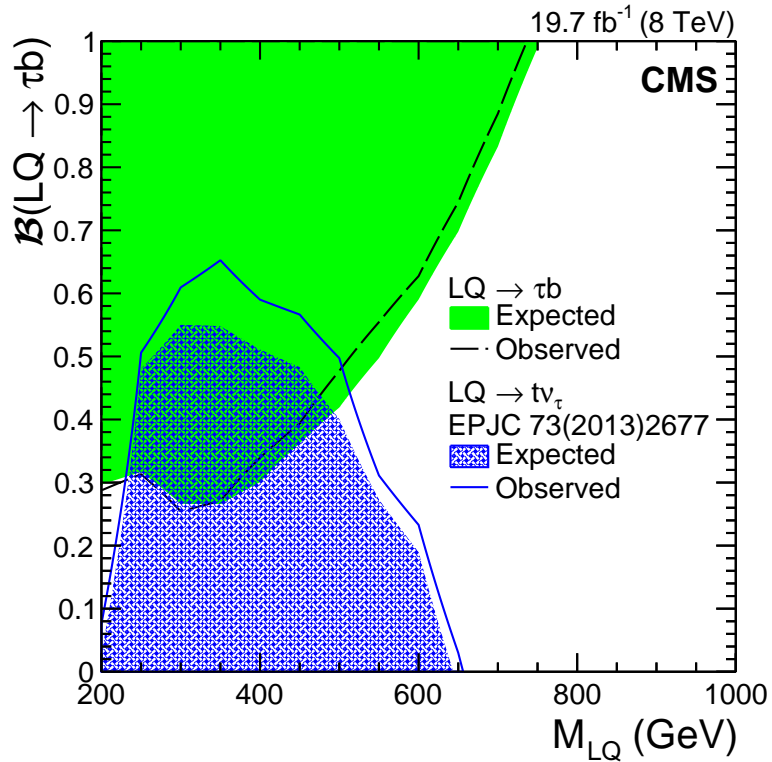


Figure 5.20: The expected (dashed black) and observed (green solid) 95% CL upper limits on the branching fraction for the leptoquark decay to a tau lepton and a bottom quark, as a function of the leptoquark mass. A search for top squark pair production [119] has the same kinematic signature as the leptoquark decay to a tau neutrino and a top quark. This search is reinterpreted to provide the expected (blue hatched) and observed (blue open) 95% CL upper limits for low values of \mathcal{B} , assuming the leptoquark only decays to third-generation fermions.

Chapter 6: Conclusions

This dissertation has presented a search for pair production of third-generation scalar leptoquarks with each leptoquark decaying to a tau lepton and a bottom quark. The search used 19.7 fb^{-1} of proton-proton collision data collected with the Compact Muon Solenoid experiment during the 2012 run of the Large Hadron Collider at a center-of-mass energy of $\sqrt{s} = 8 \text{ TeV}$. The existence of these leptoquarks is excluded at the 95% confidence level for masses up to 740 GeV. This mass limit applies directly to pair production of top squarks decaying through the R-parity violating coupling λ'_{333} , which has the same final-state signature and kinematic distributions as the third-generation scalar leptoquarks. This limit is a significant improvement over the previous limit of 530 GeV obtained using 7 TeV data [28, 29]. Limits are also set for varying leptoquark branching fraction, with the area of low branching fraction constrained by a reinterpretation of a search for top squarks decaying to a top quark and a neutralino [119].

The search is extended to cover top squarks undergoing a chargino-mediated decay involving the R-parity violating coupling λ'_{3jk} , in which each top squark decays to a final state including a tau lepton, a bottom quark, and two light quarks. Top squarks undergoing this decay are excluded at the 95% confidence level in the mass

range 200–580 GeV. This is the first direct search for the top squark decay involving the coupling λ'_{3jk} .

In 2015, Run 2 of the LHC will begin at approximately the design center-of-mass energy $\sqrt{s} = 13\text{--}14\text{ TeV}$. This increase in energy corresponds to an order-of-magnitude increase in the pair production cross section for leptoquarks at high masses. The cross section for $M_{LQ} = 1000\text{ GeV}$ will increase from $4.01 \times 10^{-4}\text{ pb}$ at $\sqrt{s} = 8\text{ TeV}$ to $8.36 \times 10^{-3}\text{ pb}$ [27]. With this significant increase in the cross section, the exclusion of leptoquarks at the TeV scale will be in reach with only a moderate amount of data [24]. Additionally, searches for single production of leptoquarks will become feasible, as the limits on the leptoquark Yukawa coupling only extend to the TeV scale [21, 22, 23].

The searches for R-parity violating supersymmetry were motivated by the existing limits on R-parity conserving supersymmetry from searches requiring large missing transverse energy. The limits set in these searches, which are the most stringent to date for the selected couplings, similarly approach the high edge of the conditions for naturalness [32]. However, supersymmetry is not fully excluded yet; significant regions of the parameter space remain unexamined. Run 2 of the LHC will have a high potential for either the discovery or more complete exclusion of supersymmetry [120].

Appendix A: Full CLs Shape-Based Limits

To set limits using the modified frequentist CL_s procedure [121], two hypotheses are defined. The first is the null or background-only hypothesis H_0 or b , and the second is the alternate or signal plus background hypothesis H_1 or $s + b$.

$\mathcal{P}(\theta; N_{H_i})$ is defined as the Poisson probability to observe θ events in data given the hypothesis H_i which predicts N_{H_i} events. This probability can be defined generally for the whole sample, but also per bin for a histogram of some quantity, e.g. S_T , and/or per channel.

To obtain this probability, it is necessary to integrate over all of the nuisance parameters:

$$\mathcal{P}(\theta; N_{H_i}) = \int \text{Poisson}(\theta; N_{H_i}, \eta) f(\eta) d\eta \quad (\text{A.1})$$

where f is the probability density function for the nuisance parameter η .

With those definitions, the test statistic \mathcal{Q} is written as a ratio of likelihoods for a basic counting experiment:

$$\mathcal{Q} = \frac{\mathcal{P}(\theta; N_{H_1})}{\mathcal{P}(\theta; N_{H_0})} \quad (\text{A.2})$$

Splitting into S_T bins and two channels ($e\tau_h, \mu\tau_h$) gives:

$$\mathcal{Q} = \prod_{i=e\tau_h, \mu\tau_h} \prod_{j=0}^{n_{\text{bin}}} \frac{\mathcal{P}_{i,j}(\theta; N_{H_1})}{\mathcal{P}_{i,j}(\theta; N_{H_0})} \quad (\text{A.3})$$

For simplicity of computation, another form of the test statistic can be defined using the log likelihood ratio:

$$q = -2 \ln \mathcal{Q} \tag{A.4}$$

To evaluate the test statistic as a function of the number of observed events θ , many simulated pseudo-experiments are performed. For each hypothesis, θ is varied according to the probability distribution of that hypothesis, and the value of \mathcal{Q} (or q) is kept for each θ value. To get \mathcal{Q} for the actual number of observed events, \mathcal{Q}_{obs} , the same procedure is followed using $\theta = N_{\text{obs}}$. The CL_{s+b} and CL_b variables correspond to the probability for \mathcal{Q}_{obs} to be greater than the \mathcal{Q} values obtained for the hypotheses H_1 and H_0 , respectively. When using q as the test statistic, the observed value should be smaller than the value for the hypothesis. A visual example of these variables is shown in Fig. A.1.

$$\text{CL}_{s+b} = \mathcal{P}(\mathcal{Q}_{H_1} \leq \mathcal{Q}_{\text{obs}}) = \mathcal{P}(q_{H_1} \geq q_{\text{obs}}) \tag{A.5}$$

$$\text{CL}_b = \mathcal{P}(\mathcal{Q}_{H_0} \leq \mathcal{Q}_{\text{obs}}) = \mathcal{P}(q_{H_0} \geq q_{\text{obs}}) \tag{A.6}$$

$$\text{CL}_s = \text{CL}_{s+b}/\text{CL}_b \tag{A.7}$$

To set a mass limit on the signal hypothesis, the calculation of CL_s is repeated for different signal masses. Masses with $\text{CL}_s < 1 - \alpha$ are excluded at the α confidence level, typically 95%.

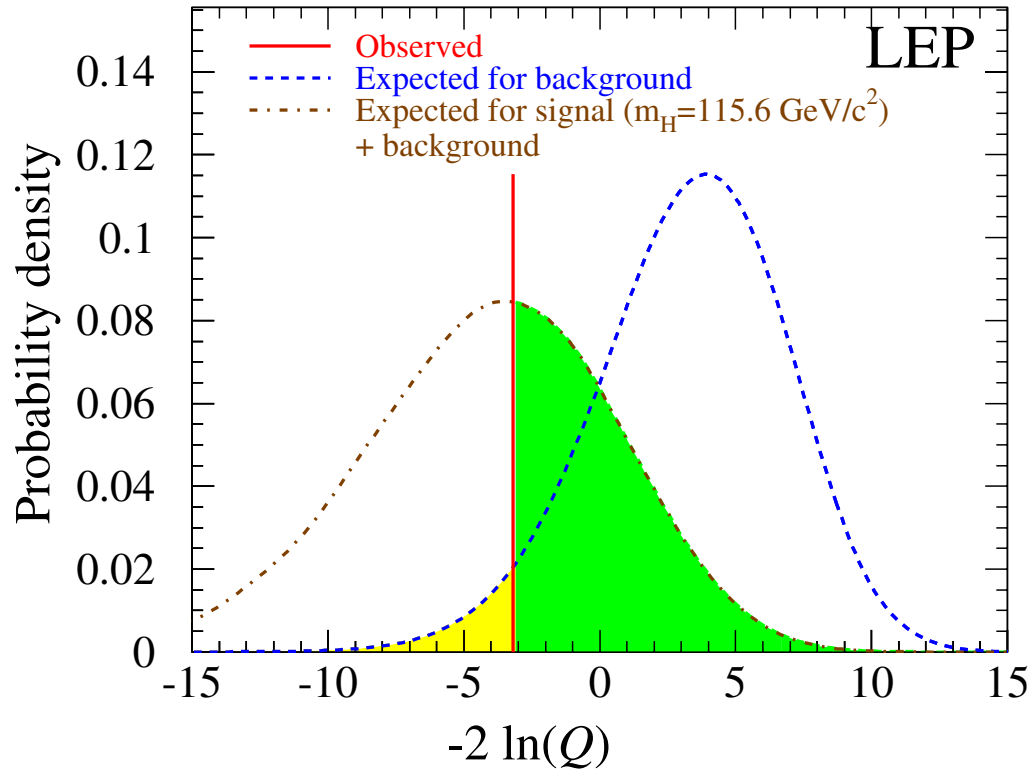


Figure A.1: Comparison of the observed value (red line) to the probability densities for H_0 (background only, blue line) and H_1 (signal + background, brown line) as a function of the log likelihood ratio. Green area: CL_{s+b} , yellow area: $1 - CL_b$. From [116].

Appendix B: Event Displays

B.1 Leptoquark Search

Figure B.1 shows the two-dimensional display in the transverse (r - ϕ) plane and the three-dimensional display for the highest S_T observed event in the $\mu\tau_h$ channel of the leptoquark search. Figure B.2 shows the same displays for the second-highest S_T observed event. The kinematic properties of the selected particles in those events are listed in Table B.1.

S_T [GeV]	Particle	p_T [GeV]	η	ϕ
1444.6	μ	92.0	-0.84	-0.16
	τ_h	87.8	0.43	1.76
	b-jet	125.2	1.63	1.57
	jet	1139.6	-0.60	-2.72
1012.1	μ	293.6	-0.49	-0.13
	τ_h	57.0	0.37	-1.03
	b-jet	77.4	1.92	-1.98
	jet	584.1	-0.06	3.08

Table B.1: The kinematic properties of the selected particles for the two highest S_T observed events in the $\mu\tau_h$ channel of the leptoquark search.

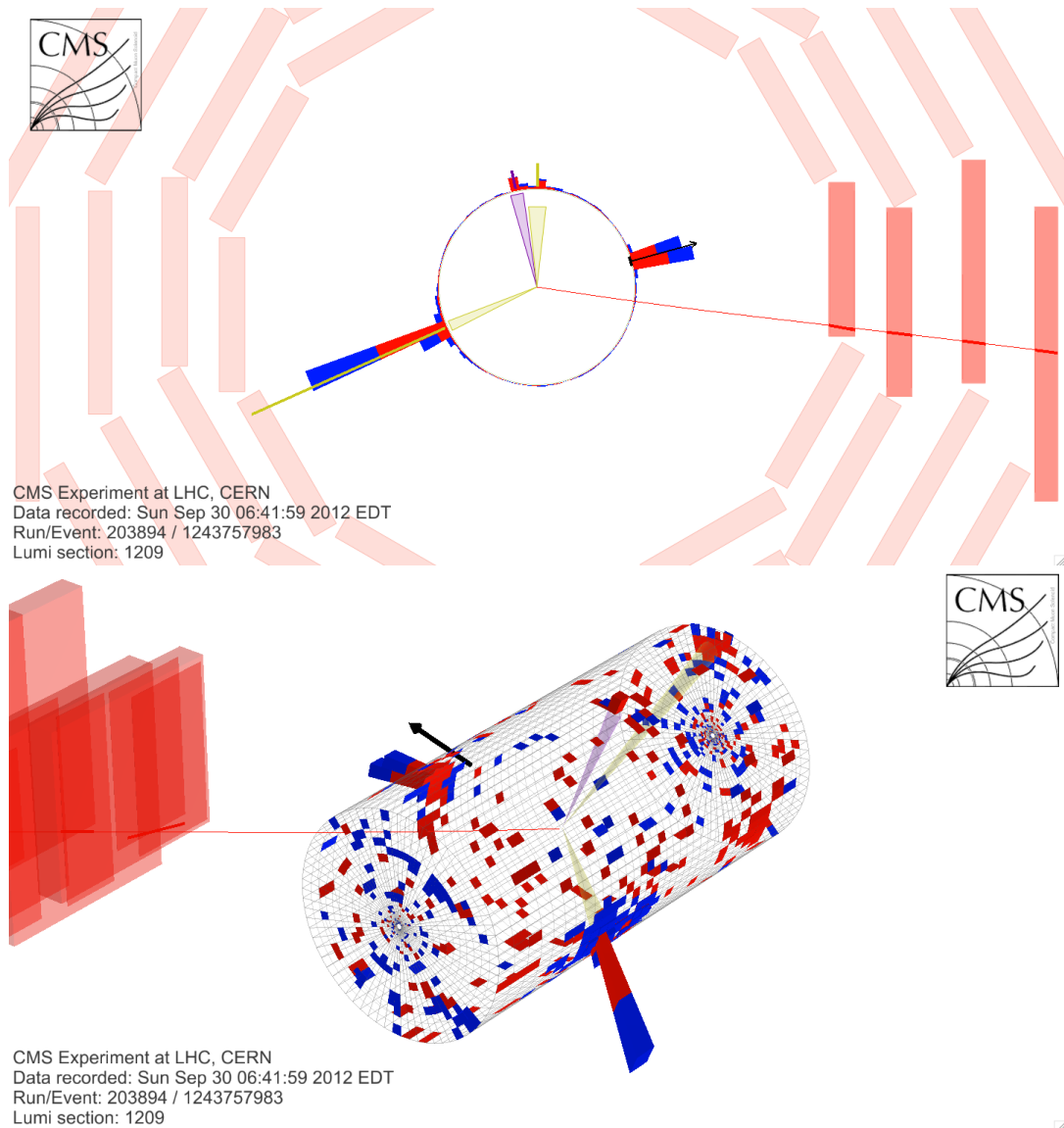


Figure B.1: A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the highest S_T observed event in the $\mu\tau_h$ channel of the leptoquark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.

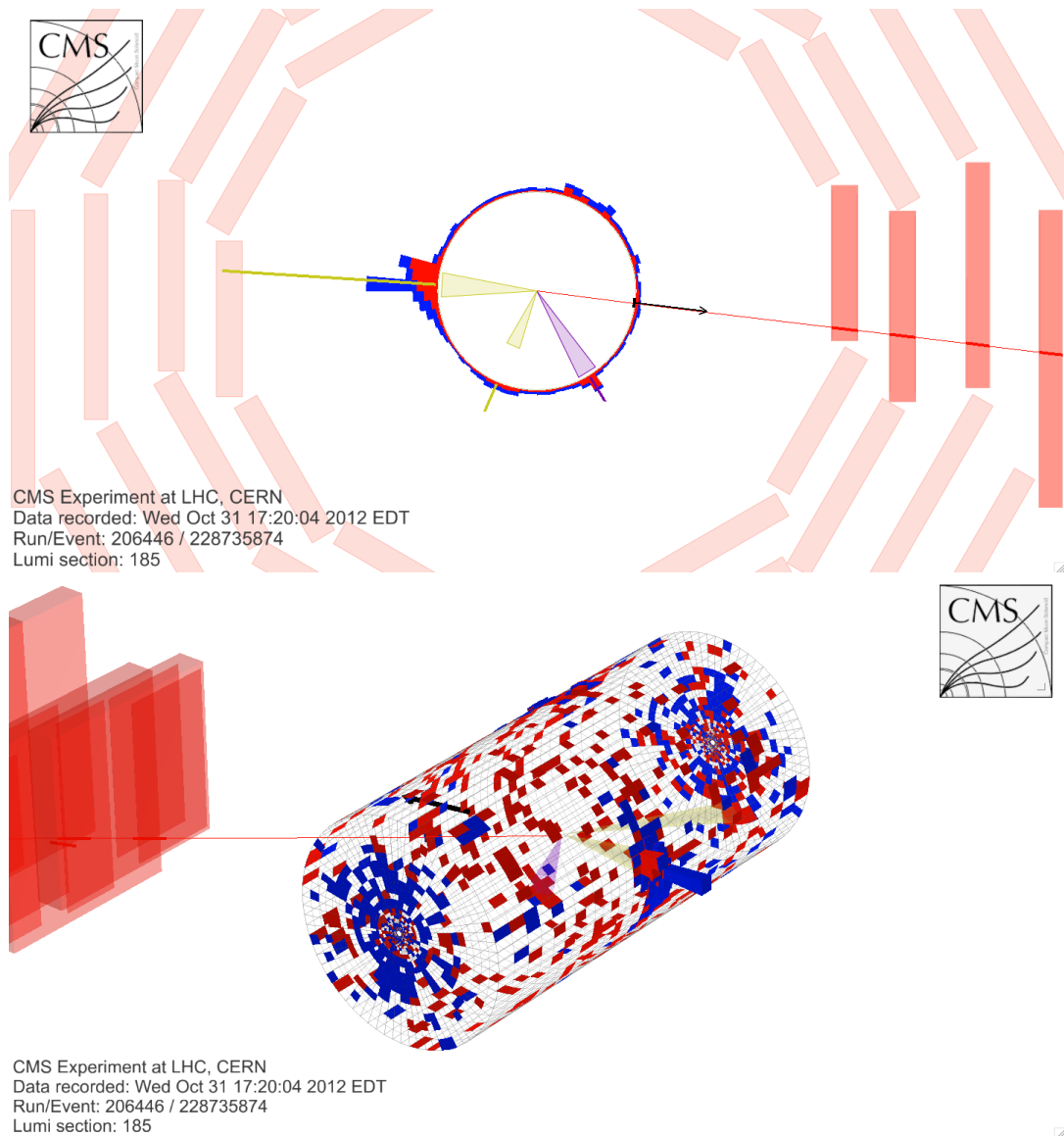


Figure B.2: A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the second-highest S_T observed event in the $\mu\tau_h$ channel of the leptoquark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.

B.2 Top Squark Search

Figure B.3 shows the two-dimensional display in the transverse (r - ϕ) plane and the three-dimensional display for the highest S_T observed event in the $\mu\tau_h$ channel of the top squark search. Figure B.4 shows the same displays for the second-highest S_T observed event. The kinematic properties of the selected particles in those events are listed in Table B.2.

S_T [GeV]	Particle	p_T [GeV]	η	ϕ
1586.2	μ	92.0	-0.84	-0.16
	τ_h	87.8	0.43	1.76
	b-jet	125.2	1.63	1.57
	jet 1	1139.6	-0.60	-2.72
	jet 2	63.5	-0.05	-2.77
	jet 3	42.9	1.10	-2.87
	jet 4	35.2	-1.69	-2.19
1136.3	μ	313.1	-0.09	-1.18
	τ_h	53.7	1.83	0.89
	b-jet	156.0	-0.09	1.52
	jet 1	325.3	0.00	2.49
	jet 2	123.1	-0.67	-1.57
	jet 3	103.0	-0.59	1.93
	jet 4	62.2	-0.62	-2.32

Table B.2: The kinematic properties of the selected particles for the two highest S_T observed events in the $\mu\tau_h$ channel of the top squark search.

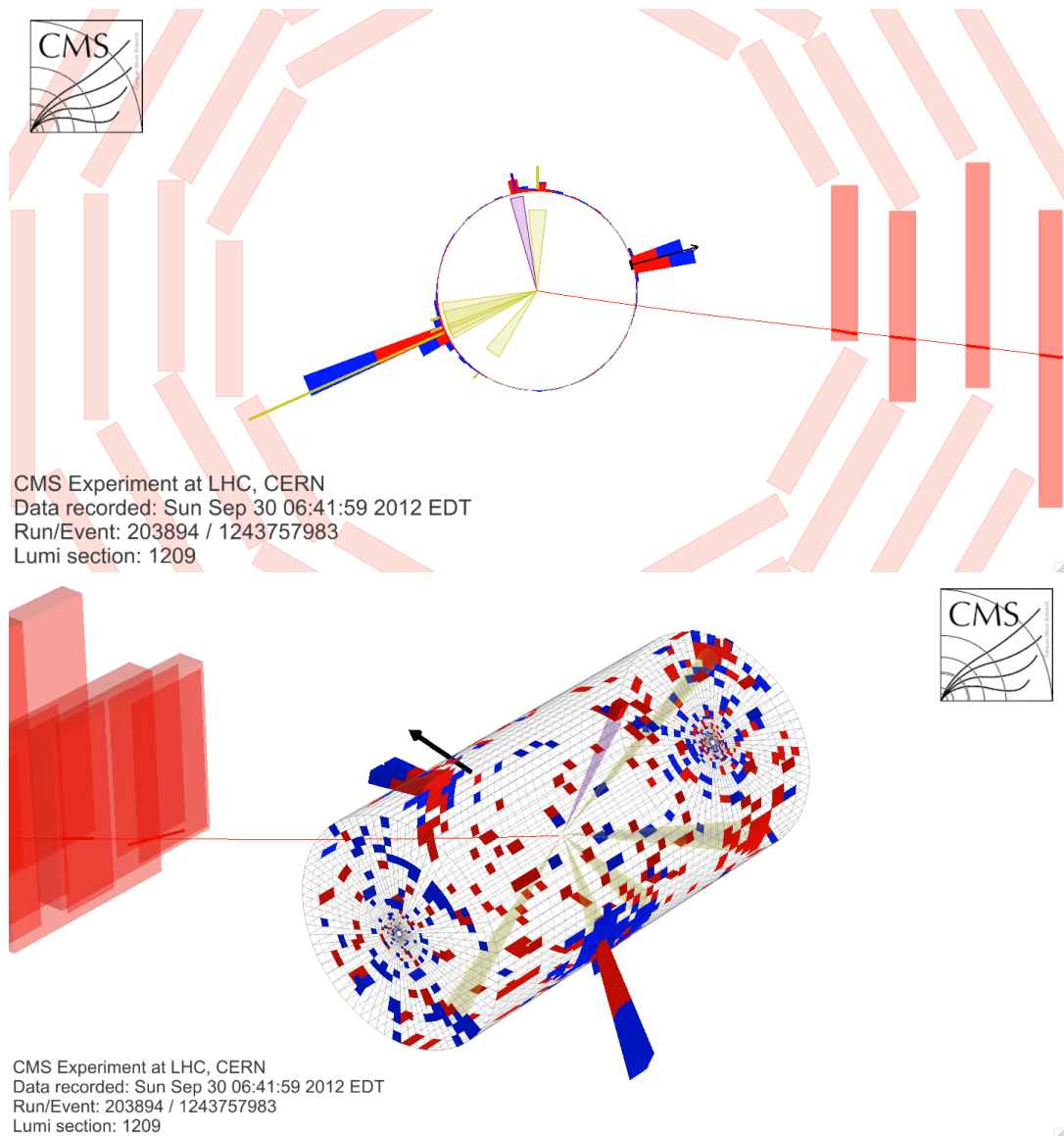


Figure B.3: A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the highest S_T observed event in the $\mu\tau_h$ channel of the top squark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.

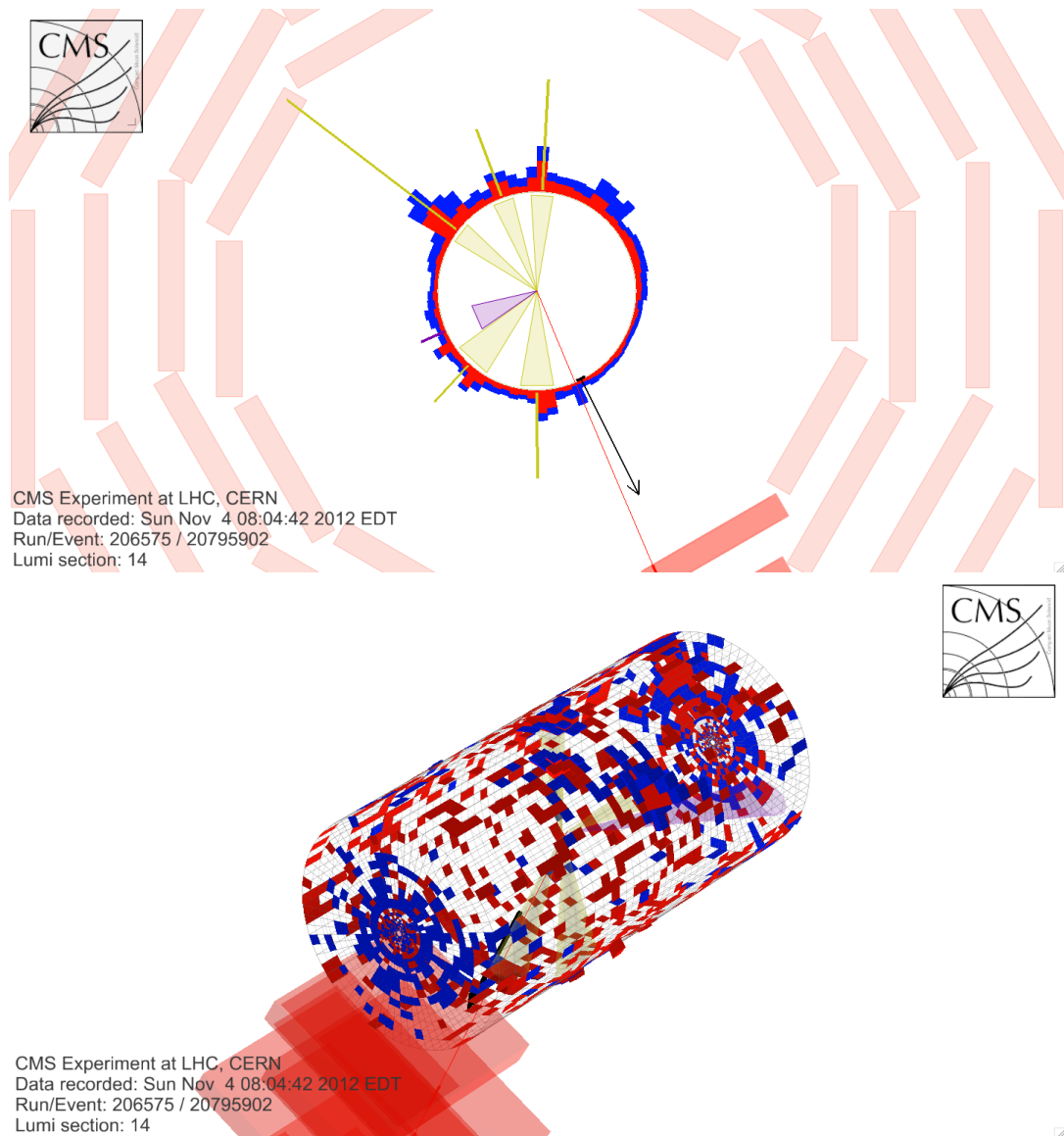


Figure B.4: A two-dimensional display in the transverse (r - ϕ) plane (top) and a three-dimensional display (bottom) for the second-highest S_T observed event in the $\mu\tau_h$ channel of the top squark search. The red line represents the muon, and the associated red rectangles represent the muon chamber hits. The purple cone represents the hadronic tau, and the yellow cones represent the jets. The black arrow indicates the \cancel{E}_T in the event, while the ECAL and HCAL energy deposits are represented as red and blue towers, respectively.

Appendix C: Table of Monte Carlo Datasets

Tables C.1 and C.2 list all MC samples for the LQ signal and SM background processes, respectively. These samples are centrally produced by the CMS collaboration and stored in a database called the Data Aggregation System (DAS). The location of each sample in DAS is given by the “Dataset Name” field in the tables.

Signal Processes			
M_{LQ} [GeV]	Dataset Name	σ [pb]	α_s order
200	/LQToTauB_M-200_beta-1_TuneZ2star_8TeV-pythia6	17.4	NLO
250	/LQToTauB_M-250_beta-1_TuneZ2star_8TeV-pythia6	5.26	NLO
300	/LQToTauB_M-300_beta-1_TuneZ2star_8TeV-pythia6	1.89	NLO
350	/LQToTauB_M-350_beta-1_TuneZ2star_8TeV-pythia6	0.77	NLO
400	/LQToTauB_M-400_beta-1_TuneZ2star_8TeV-pythia6	0.342	NLO
450	/LQToTauB_M-450_beta-1_TuneZ2star_8TeV-pythia6	0.163	NLO
500	/LQToTauB_M-500_beta-1_TuneZ2star_8TeV-pythia6	0.082	NLO
550	/LQToTauB_M-550_beta-1_TuneZ2star_8TeV-pythia6	0.0431	NLO
600	/LQToTauB_M-600_beta-1_TuneZ2star_8TeV-pythia6	0.0235	NLO
650	/LQToTauB_M-650_beta-1_TuneZ2star_8TeV-pythia6	0.0132	NLO
700	/LQToTauB_M-700_beta-1_TuneZ2star_8TeV-pythia6	0.00761	NLO
750	/LQToTauB_M-750_beta-1_TuneZ2star_8TeV-pythia6	0.00448	NLO
800	/LQToTauB_M-800_beta-1_TuneZ2star_8TeV-pythia6	0.00269	NLO
850	/LQToTauB_M-850_beta-1_TuneZ2star_8TeV-pythia6	0.00164	NLO
900	/LQToTauB_M-900_beta-1_TuneZ2star_8TeV-pythia6	0.00101	NLO
950	/LQToTauB_M-950_beta-1_TuneZ2star_8TeV-pythia6	0.000634	NLO
1000	/LQToTauB_M-1000_beta-1_TuneZ2star_8TeV-pythia6	0.000401	NLO

Table C.1: The full list of MC samples for the LQ signal used in the analysis. Each dataset name should be followed by */Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM*. Cross sections σ are given at the specified perturbative order in α_s for each process.

Background Processes			
Process	Dataset Name	σ [pb]	α_s order
$t\bar{t}$	/TTJets_MassiveBinDECAY_TuneZ2star_8TeV-madgraph-tauola	234	NNLL
$t\bar{t} \rightarrow b\ell\nu b\ell\nu$	/TTJets_FullLeptMGDecays_8TeV-madgraph-tauola	24.56	NNLL
$t\bar{t} \rightarrow b\ell\nu bqq'$	/TTJets_SemiLeptMGDecays_8TeV-madgraph-tauola	102.51	NNLL
$t\bar{t} \rightarrow bqq' bqq'$	/TTJets_HadronicMGDecays_8TeV-madgraph	106.96	NNLL
$t \rightarrow b\ell\nu$ (s -channel)	/T_s-channel_TuneZ2star_8TeV-powheg-tauola	1.76	NNLL
$t \rightarrow b\ell\nu$ (t -channel)	/T_t-channel_TuneZ2star_8TeV-powheg-tauola	30.7	NNLL
$t \rightarrow X$ (tW)	/T_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1	NNLL
$\bar{t} \rightarrow b\ell\nu$ (s -channel)	/Tbar_s-channel_TuneZ2star_8TeV-powheg-tauola	3.79	NNLL
$\bar{t} \rightarrow b\ell\nu$ (t -channel)	/Tbar_t-channel_TuneZ2star_8TeV-powheg-tauola	56.4	NNLL
$\bar{t} \rightarrow X$ (tW)	/Tbar_tW-channel-DR_TuneZ2star_8TeV-powheg-tauola	11.1	NNLL
W + jets	/WJetsToLNu_TuneZ2Star_8TeV-madgraph-tarball	37509.0	NNLO
W + 1 jet	/W1JetsToLNu_TuneZ2Star_8TeV-madgraph	6440.58	LO
W + 2 jets	/W2JetsToLNu_TuneZ2Star_8TeV-madgraph	2087.225	LO
W + 3 jets	/W3JetsToLNu_TuneZ2Star_8TeV-madgraph	619.0113	LO
W + 4 jets	/W4JetsToLNu_TuneZ2Star_8TeV-madgraph	255.2378	LO
Z + jets	/DYJetsToLL_M-50_TuneZ2Star_8TeV-madgraph-tarball	3503.7	NNLO
Z + 1 jet	/DY1JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	666.3	LO
Z + 2 jets	/DY2JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	214.97	LO
Z + 3 jets	/DY3JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	60.69	LO
Z + 4 jets	/DY4JetsToLL_M-50_TuneZ2Star_8TeV-madgraph	27.36	LO
WW	/WW_TuneZ2star_8TeV_pythia6_tauola	55.47	NLO
WZ	/WZ_TuneZ2star_8TeV_pythia6_tauola	33.59	NLO
ZZ	/ZZ_TuneZ2star_8TeV_pythia6_tauola	8.27	NLO

Table C.2: The full list of MC samples for the SM background used in the analysis. Each dataset name should be followed by */Summer12_DR53X-PU_S10_START53_V7A-v1/AODSIM*. Cross sections σ are given at the specified perturbative order in α_s for each process.

Appendix D: CMS Collaboration

Bibliography

- [1] MissMJ et al., “Standard Model of Elementary Particles”. Creative Commons Attribution 3.0 Unported License, Oct, 2014.
http://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg. xi, 6
- [2] CMS Collaboration, “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”, *Phys. Lett. B* **716** (2012) 30, doi:10.1016/j.physletb.2012.08.021, arXiv:1207.7235. 7
- [3] ATLAS Collaboration, “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”, *Phys. Lett. B* **716** (2012) 1, doi:10.1016/j.physletb.2012.08.020, arXiv:1207.7214. 7
- [4] E. Drexler, “Elementary particle interactions in the Standard Model”. Creative Commons CC0 1.0 Universal Public Domain Dedication, May, 2014. http://en.wikipedia.org/wiki/File:Elementary_particle_interactions_in_the_Standard_Model.png. xi, 11
- [5] L. Susskind, “The gauge hierarchy problem, technicolor, supersymmetry, and all that”, *Phys. Rept.* **104** (1984) 181, doi:10.1016/0370-1573(84)90208-4. 11
- [6] D. E. Morrissey, T. Plehn, and T. M. Tait, “Physics searches at the LHC”, *Phys. Rept.* **515** (2012) 1, doi:10.1016/j.physrep.2012.02.007, arXiv:0912.3259. 11, 12
- [7] D. Clowe et al., “A Direct Empirical Proof of the Existence of Dark Matter”, *ApJL* **648** (2006) L109, doi:10.1086/508162, arXiv:astro-ph/0608407. 12
- [8] K. A. Olive et al., “The Review of Particle Physics”, *Chin. Phys. C* **38** (2014) 090001. vii, 12, 26, 72, 73

- [9] H. Georgi, H. R. Quinn, and S. Weinberg, “Hierarchy of Interactions in Unified Gauge Theories”, *Phys. Rev. Lett.* **33** (Aug, 1974) 451, doi:10.1103/PhysRevLett.33.451. 12
- [10] H. P. Nilles, “Supersymmetry, supergravity and particle physics”, *Phys. Rept.* **110** (1984) 1, doi:10.1016/0370-1573(84)90008-5. 12
- [11] H. E. Haber and G. L. Kane, “The search for supersymmetry: Probing physics beyond the standard model”, *Phys. Rept.* **117** (1985) 75, doi:10.1016/0370-1573(85)90051-1. 12
- [12] J. A. Evans and Y. Kats, “LHC coverage of RPV MSSM with light stops”, *JHEP* **4** (2013) doi:10.1007/JHEP04(2013)028, arXiv:1209.0764. 13
- [13] M. E. Peskin and D. V. Schroeder, “An Introduction To Quantum Field Theory”. Westview Press, 1995. 13, 20
- [14] J. C. Pati and A. Salam, “Lepton number as the fourth “color””, *Phys. Rev. D* **10** (1974) 275, doi:10.1103/PhysRevD.10.275. 13
- [15] H. Georgi and S. L. Glashow, “Unity of All Elementary-Particle Forces”, *Phys. Rev. Lett.* **32** (1974) 438, doi:10.1103/PhysRevLett.32.438. 13
- [16] J. L. Hewett and T. G. Rizzo, “Low-Energy Phenomenology of Superstring Inspired E_6 Models”, *Phys. Rept.* **183** (1989) 193, doi:10.1016/0370-1573(89)90071-9. 14
- [17] B. Gripaios, “Composite leptoquarks at the LHC”, *JHEP* **1002** (2010) 045, doi:10.1007/JHEP02(2010)045, arXiv:0910.1789. 14
- [18] E. Eichten and K. Lane, “Dynamical breaking of weak interaction symmetries”, *Phys. Lett. B* **90** (1980) 125, doi:10.1016/0370-2693(80)90065-9. 14
- [19] W. Buchmüller, R. Rückl, and D. Wyler, “Leptoquarks in lepton-quark collisions”, *Phys. Lett. B* **191** (1987) 442, doi:10.1016/0370-2693(87)90637-X. [Erratum: doi:10.1016/S0370-2693(99)00014-3]. 14, 18
- [20] S. Davidson, D. Bailey, and B. A. Campbell, “Model independent constraints on leptoquarks from rare processes”, *Z. Phys. C* **61** (1994) 613, doi:10.1007/BF01552629, arXiv:hep-ph/9309310. 15, 20
- [21] M. Leurer, “A comprehensive study of leptoquark bounds”, *Phys. Rev. D* **49** (1994) 333, doi:10.1103/PhysRevD.49.333, arXiv:hep-ph/9309266. 17, 20, 140
- [22] J. L. Hewett and T. G. Rizzo, “Much ado about leptoquarks: A comprehensive analysis”, *Phys. Rev. D* **56** (Nov, 1997) 5709, doi:10.1103/PhysRevD.56.5709, arXiv:hep-ph/9703337. 17, 20, 140

- [23] D. E. Acosta and S. K. Blessing, “Leptoquark searches at HERA and the Tevatron”, *Annu. Rev. Nucl. Part. Sci.* **49** (1999) 389, doi:10.1146/annurev.nucl.49.1.389. 17, 20, 140
- [24] J. Blümlein, E. Boos, and A. Kryukov, “Leptoquark pair production in hadronic interactions”, *Z. Phys. C* **76** (1997) 137, doi:10.1007/s002880050538, arXiv:hep-ph/9610408. 17, 140
- [25] J. Pumplin et al., “New Generation of Parton Distributions with Uncertainties from Global QCD Analysis”, *JHEP* **10** (2002) 012, doi:10.1088/1126-6708/2002/07/012, arXiv:hep-ph/0201195. 17
- [26] D. Stump et al., “Inclusive jet production, parton distributions, and the search for new physics”, *JHEP* **7** (2003) 046, doi:10.1088/1126-6708/2003/10/046, arXiv:hep-ph/0303013. 17
- [27] M. Krämer, T. Plehn, M. Spira, and P. M. Zerwas, “Pair production of scalar leptoquarks at the CERN LHC”, *Phys. Rev. D* **71** (Mar, 2005) 057503, doi:10.1103/PhysRevD.71.057503, arXiv:hep-ph/0411038. 17, 84, 140
- [28] CMS Collaboration, “Search for Pair Production of Third-Generation Leptoquarks and Top Squarks in pp Collisions at $\sqrt{s} = 7$ TeV”, *Phys. Rev. Lett.* **110** (2013) 081801, doi:10.1103/PhysRevLett.110.081801, arXiv:1210.5629. 19, 20, 139
- [29] ATLAS Collaboration, “Search for third generation scalar leptoquarks in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector”, *JHEP* **06** (2013) 033, doi:10.1007/JHEP06(2013)033, arXiv:1303.0526. 19, 139
- [30] S. P. Martin, “A Supersymmetry Primer”, *Adv. Ser. Direct. High Energy Phys.* **21** (2010) 1, doi:10.1142/9789814307505_0001, arXiv:hep-ph/9709356. xi, 20, 22, 26
- [31] R. Barbier et al., “R-Parity-violating supersymmetry”, *Phys. Rept.* **420** (2005) 1, doi:10.1016/j.physrep.2005.08.006, arXiv:hep-ph/0406039. xi, 23, 27, 29
- [32] M. Papucci, J. T. Ruderman, and A. Weiler, “Natural SUSY Endures”, *JHEP* **09** (2012) 035, doi:10.1007/JHEP09(2012)035, arXiv:1110.6926. 25, 27, 140
- [33] U. Ellwanger, C. Hugonie, and A. M. Teixeira, “The Next-to-Minimal Supersymmetric Standard Model”, *Phys. Rept.* **496** (2010) 1, doi:10.1016/j.physrep.2010.07.001, arXiv:0910.1785. 27
- [34] CMS Collaboration, “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ with the CMS experiment”, *Phys. Rev. Lett.* **111** (Sep, 2013) 101804, doi:10.1103/PhysRevLett.111.101804, arXiv:1307.5025. 27

- [35] LHCb Collaboration, “Measurement of the $B_s^0 \rightarrow \mu^+ \mu^-$ branching fraction and search for $B^0 \rightarrow \mu^+ \mu^-$ decays at the LHCb experiment”, *Phys. Rev. Lett.* **111** (Sep, 2013) 101805, doi:10.1103/PhysRevLett.111.101805, arXiv:1307.5024. 27
- [36] CMS Collaboration, “Summary of comparison plots in simplified models spectra for the 8 TeV dataset”, (2014). xi, 28
- [37] W. Beenakker et al., “Squark and gluino hadroproduction”, *Int. J. Mod. Phys. A* **26** (2011) 2637, doi:10.1142/S0217751X11053560, arXiv:1105.1110. 29
- [38] J. Evans and Y. Kats, “LHC searches examined via the RPV MSSM”, in *The 2013 European Physical Society Conference on High Energy Physics*. SISSA, 2013. arXiv:1311.0890. 30
- [39] L. Evans and P. Bryant, “LHC Machine”, *JINST* **3** (2008) S08001, doi:10.1088/1748-0221/3/08/S08001. 31
- [40] CMS Collaboration, “The CMS experiment at the CERN LHC”, *JINST* **3** (2008) S08004, doi:10.1088/1748-0221/3/08/S08004. 31
- [41] J.-L. Caron, “LHC Layout”, (September, 1997). AC Collection. Legacy of AC. Pictures from 1992 to 2002. xi, 33
- [42] J.-L. Caron, “The LHC injection complex”, (May, 1993). AC Collection. Legacy of AC. Pictures from 1992 to 2002. xi, 34
- [43] S. Dailier, “LHC Dipole”, (July, 1998). AC Collection. Legacy of AC. Pictures from 1992 to 2002. xi, 35
- [44] CMS Collaboration, “Public CMS Luminosity Information”, (Jan, 2014). 35, 81
- [45] CMS Collaboration, “Operation and performance of the CMS tracker”, *JINST* **9** (2014) C03005, doi:10.1088/1748-0221/9/03/C03005, arXiv:1402.0675. 37
- [46] CMS Collaboration, “2012 ECAL detector performance plots”, CMS Detector Performance Summary CMS-DP-2013-007, CERN, 2013. 39
- [47] CMS Collaboration, “Performance of the CMS hadron calorimeter with cosmic ray muons and LHC beam data”, *JINST* **5** (2010) T03012, doi:10.1088/1748-0221/5/03/T03012, arXiv:0911.4991. 42
- [48] J. Freeman, “Silicon photomultipliers for the CMS hadron calorimeter”, *Nucl. Instrum. Meth. A* **617** (2010) 393, doi:10.1016/j.nima.2009.10.132. Proceedings of the 11th Pisa Meeting on Advanced Detectors. 42

- [49] CMS Collaboration, “Performance of the CMS Level-1 Trigger”, *PoS ICHEP2012* (2013) 508, [arXiv:1302.2469](#). 51
- [50] D. Trocino, “The CMS High Level Trigger”, *J. Phys. Conf.* **513** (2014) 012036, [doi:10.1088/1742-6596/513/1/012036](#). 51
- [51] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2012 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-12-001, CERN, Geneva, 2012. 52
- [52] CMS Collaboration, “CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update”, CMS Physics Analysis Summary CMS-PAS-LUM-13-001, CERN, Geneva, 2013. 52
- [53] V. Balagura, “Notes on van der Meer Scan for Absolute Luminosity Measurement”, *Nucl. Instrum. Meth. A* **654** (2011) 634, [doi:10.1016/j.nima.2011.06.007](#), [arXiv:1103.1129](#). 52
- [54] D. Barney and E. Quigg, “Interactive Slice of the CMS detector”, (July, 2010). xii, 54
- [55] CMS Collaboration, “CMS Physics: Technical Design Report Volume 1: Detector Performance and Software”, Technical Report CMS-TDR-8-1, CERN, Geneva, 2006. 53
- [56] A. Martin, W. Stirling, R. Thorne, and G. Watt, “Parton distributions for the LHC”, *Eur. Phys. J. C* **63** (2009) 189, [doi:10.1140/epjc/s10052-009-1072-5](#), [arXiv:0901.0002](#). xii, 56
- [57] J. M. Campbell, J. W. Huston, and W. J. Stirling, “Hard interactions of quarks and gluons: a primer for LHC physics”, *Rep. Prog. Phys.* **70** (2007), no. 1, 89, [doi:10.1088/0034-4885/70/1/R02](#). xii, 56, 58
- [58] T. Sjöstrand, S. Mrenna, and P. Skands, “PYTHIA 6.4 physics and manual”, *JHEP* **05** (2006) 026, [doi:10.1088/1126-6708/2006/05/026](#), [arXiv:hep-ph/0603175](#). 57, 83, 84
- [59] 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”, (2014). [arXiv:1405.0301](#). 57, 83
- [60] P. Nason, “A new method for combining NLO QCD with shower Monte Carlo algorithms”, *Journal of High Energy Physics* **2004** (2004), no. 11, 040, [doi:10.1088/1126-6708/2004/11/040](#), [arXiv:hep-ph/0409146](#). 58
- [61] 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* **1006** (2010) 043, [doi:10.1007/JHEP06\(2010\)043](#), [arXiv:1002.2581](#). 58

- [62] Z. Was, “TAUOLA the library for τ lepton decay, and KKMC/KORALB/KORALZ...status report”, *Nucl. Phys. B, Proc. Suppl.* **98** (2001) 96, doi:10.1016/S0920-5632(01)01200-2. 58
- [63] S. Agostinelli et al., “Geant4 a simulation toolkit”, *Nucl. Instrum. Meth. A* **506** (2003), no. 3, 250 – 303, doi:10.1016/S0168-9002(03)01368-8. 58
- [64] J. Allison et al., “Geant4 developments and applications”, *IEEE Trans. Nucl. Sci.* **53** (February, 2006) 270 –278, doi:10.1109/TNS.2006.869826. 58
- [65] R. Brun and F. Rademakers, “ROOT – An object oriented data analysis framework”, *Nucl. Instrum. Meth. A* **389** (1997) 81, doi:10.1016/S0168-9002(97)00048-X. See also <http://root.cern.ch>. 59
- [66] 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. 59, 61
- [67] D. Giordano and G. Sguazzoni, “CMS reconstruction improvements for the tracking in large pile-up events”, *J. Phys. Conf.* **396** (2012), no. 2, 022044, doi:10.1088/1742-6596/396/2/022044. vii, 60, 61
- [68] CMS Collaboration, “Tracking and Primary Vertex Results in First 7 TeV Collisions”, CMS Physics Analysis Summary CMS-PAS-TRK-10-005, CERN, Geneva, 2010. 62
- [69] CMS Collaboration, “Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET”, CMS Physics Analysis Summary CMS-PAS-PFT-09-001, CERN, 2009. 62, 70
- [70] CMS Collaboration, “Commissioning of the Particle-Flow Reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-002, CERN, 2010. 62, 68, 70
- [71] CMS Collaboration, “Commissioning of the particle-flow event reconstruction with leptons from J/Ψ and W decays at 7 TeV”, CMS Physics Analysis Summary CMS-PAS-PFT-10-003, CERN, 2010. 62
- [72] F. Beaudette, “The CMS Particle Flow Algorithm”, in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013)*, p. 295. École Polytechnique, Paris, 2013. arXiv:1401.8155. 62
- [73] W. Adam, R. Frühwirth, A. Strandlie, and T. Todorov, “Reconstruction of electrons with the Gaussian-sum filter in the CMS tracker at the LHC”, *J. Phys. G* **31** (2005) N9, doi:10.1088/0954-3899/31/9/N01. 64

- [74] CMS Collaboration, “Electron reconstruction and identification at $\sqrt{s} = 7$ TeV”, CMS Physics Analysis Summary CMS-PAS-EGM-10-004, CERN, Geneva, 2010. 65
- [75] S. Baffioni et al., “Electron reconstruction in CMS”, *Eur. Phys. J. C* **49** (2007), no. 4, 1099, doi:10.1140/epjc/s10052-006-0175-5. 65
- [76] CMS Collaboration, “Cut Based Electron ID”, (2013). 65
- [77] CMS Collaboration, “SW Guide Egamma Shower Shape”, (2011). 65
- [78] CMS Collaboration, “Performance of muon identification in pp collisions at $\sqrt{s} = 7$ TeV”, CMS Physics Analysis Summary CMS-PAS-MUO-10-002, CERN, Geneva, 2010. 67
- [79] G. P. Salam, “Towards jetography”, *Eur. Phys. J. C* **67** (2010) 637, doi:10.1140/epjc/s10052-010-1314-6, arXiv:0906.1833. 69
- [80] 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. xii, 69, 71
- [81] CMS Collaboration, “Determination of jet energy calibration and transverse momentum resolution in CMS”, *JINST* **6** (2011) 11002, doi:10.1088/1748-0221/6/11/P11002, arXiv:1107.4277. 71, 85, 130
- [82] N. Saoulidou, “Particle Flow Jet Identification Criteria”, CMS Analysis Note CMS-AN-2010-003, CERN, Geneva, Jun, 2010. 71
- [83] CMS Collaboration, “Jet Identification”, (2012). 72
- [84] CMS Collaboration, “Performance of t-lepton reconstruction and identification in CMS”, *JINST* **7** (2012) P01001, doi:10.1088/1748-0221/7/01/P01001. 72
- [85] C. Calabria, “Tau trigger, reconstruction and identification at CMS”, CMS Conference Report CMS-CR-2012-344, CERN, Geneva, Nov, 2012. 72
- [86] The Tau Physics Object Group, “Performance of tau reconstruction and identification in pp collisions at $\sqrt{s} = 8$ TeV”, CMS Analysis Note CMS-AN-2014-008, CERN, Geneva, Jun, 2014. 74, 128
- [87] CMS Collaboration, “Tau ID Performance Plots”, CMS Detector Performance Summary CMS-DP-2014-015, CERN, Geneva, Apr, 2014. xii, 75
- [88] CMS Collaboration, “Identification of b-quark jets with the CMS experiment”, *JINST* **8** (2013) P04013, doi:10.1088/1748-0221/8/04/P04013, arXiv:1211.4462. 77

- [89] CMS Collaboration, “Performance of b tagging at $\sqrt{s} = 8$ TeV in multijet, ttbar and boosted topology events”, CMS Physics Analysis Summary CMS-PAS-BTV-13-001, CERN, Geneva, 2013. 77
- [90] CMS Collaboration, “Performance of the missing transverse energy reconstruction by the CMS experiment in $\sqrt{s} = 8$ TeV pp data”, (2014). [arXiv:1411.0511](#). Submitted to *JINST*. 79
- [91] CMS Collaboration, “Missing transverse energy performance of the CMS detector”, *JINST* **6** (2011) P09001, [doi:10.1088/1748-0221/6/09/P09001](#), [arXiv:1106.5048](#). 80
- [92] CMS Collaboration, “Public CMS Data Quality Information”, (Jun, 2013). xii, 82
- [93] S. Frixione, P. Nason, and C. Oleari, “Matching NLO QCD computations with parton shower simulations: the POWHEG method”, *JHEP* **11** (2007) 070, [doi:10.1088/1126-6708/2007/11/070](#), [arXiv:0709.2092](#). 83
- [94] S. Alioli, P. Nason, C. Oleari, and E. Re, “NLO single-top production matched with shower in POWHEG: *s*- and *t*-channel contributions”, *JHEP* **09** (2009) 111, [doi:10.1088/1126-6708/2009/09/111](#), [arXiv:0907.4076](#). [Erratum: [doi:10.1007/JHEP02\(2010\)011](#)]. 83
- [95] E. Re, “Single-top *Wt*-channel production matched with parton showers using the POWHEG method”, *Eur. Phys. J. C* **71** (2011) 1547, [doi:10.1140/epjc/s10052-011-1547-z](#), [arXiv:1009.2450](#). 83
- [96] N. Kidonakis, “Differential and total cross sections for top pair and single top production”, in *Proc. of XX Int. Workshop on Deep-Inelastic Scattering and Related Subjects*, p. 831. 2012. [arXiv:1205.3453](#). [doi:10.3204/DESY-PROC-2012-02/251](#). 83
- [97] K. Melnikov and F. Petriello, “Electroweak gauge boson production at hadron colliders through $\mathcal{O}(\alpha_s^2)$ ”, *Phys. Rev. D* **74** (2006) 114017, [doi:10.1103/PhysRevD.74.114017](#), [arXiv:hep-ph/0609070](#). 84
- [98] J. M. Campbell, R. K. Ellis, and C. Williams, “Vector boson pair production at the LHC”, *JHEP* **07** (2011) 018, [doi:10.1007/JHEP07\(2011\)018](#), [arXiv:1105.0020](#). 84
- [99] CMS Collaboration, “Reference muon id and isolation efficiencies”, (2013). 85
- [100] The Egamma ID Group, “Tag and probe methodology for analyses using electrons and photons”, CMS Analysis Note CMS-AN-2012-116, CERN, Geneva, May, 2012. 85

- [101] CMS Collaboration, “Single Electron efficiencies and Scale-Factors for cut based Id - 2012”, (2013). 85
- [102] CMS Collaboration, “TauID: recommendation from the Tau POG”, (2013). 85, 130
- [103] CMS Collaboration, “Methods to apply b-tagging efficiency scale factors”, (2013). 85
- [104] M. Marionneau and C. Veelken, “Study of \cancel{E}_T Performance in events containing one Z boson decaying into two muons in 2012 Data”, CMS Analysis Note CMS-AN-2012-333, CERN, Geneva, Feb, 2013. 85
- [105] A. Apyan et al., “Search for Higgs to Tau Tau in the Muon-Tau and Electron-Tau Channels”, CMS Analysis Note CMS-AN-2013-178, CERN, Geneva, Oct, 2013. 118
- [106] M. Marionneau, “Missing transverse energy studies for events with W and Z bosons decaying into electrons channels using 2010 data”, CMS Analysis Note CMS-AN-2010-359, CERN, Geneva, Feb, 2011. 121
- [107] CMS Collaboration, “Measurement of the inclusive W and Z production cross sections in pp collisions at $\sqrt{s} = 7$ TeV with the CMS experiment”, *JHEP* **1110** (2011) 132, doi:10.1007/JHEP10(2011)132, arXiv:1107.4789. 121
- [108] CMS Collaboration, “Electron performance with 19.6 fb⁻¹ of data collected at $\sqrt{s} = 8$ TeV with the CMS detector”, CMS Detector Performance Summary CMS-DP-2013-003, CERN, Geneva, Mar, 2013. 128
- [109] CMS Collaboration, “Single Muon efficiencies in 2012 Data”, CMS Detector Performance Summary CMS-DP-2013-009, CERN, Geneva, Mar, 2013. 128
- [110] E. Barberis, D. Baumgartel, and D. Morse, “Search for Pair-production of Second generation Leptoquarks in 8 TeV proton-proton collisions”, CMS Analysis Note CMS-AN-2012-481, CERN, Geneva, Aug, 2014. 129
- [111] CMS Collaboration, “Measurement of Inclusive W and Z Boson Production Cross Sections in pp Collisions at $\sqrt{s} = 8$ TeV”, *Phys. Rev. Lett.* **112** (May, 2014) 191802, doi:10.1103/PhysRevLett.112.191802, arXiv:1402.0923. 129
- [112] CMS Collaboration, “Measurement of the W⁺W⁻ and ZZ production cross sections in pp collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett. B* **721** (2013) 190, doi:10.1016/j.physletb.2013.03.027, arXiv:1301.4698. 129
- [113] CMS and ATLAS Collaboration, “Combination of single top-quark cross-sections measurements in the t-channel at $\sqrt{s} = 8$ TeV with the ATLAS and CMS experiments”, CMS Physics Analysis Summary CMS-PAS-TOP-2012-002, CERN, Geneva, Sep, 2013. 129

- [114] CMS Collaboration, “8 TeV Jet Energy Corrections and Uncertainties based on 19.8 fb⁻¹ of data in CMS”, CMS Detector Performance Summary CMS-DP-2013-033, CERN, Geneva, Oct, 2013. 130
- [115] CMS Collaboration, “Search for pair production of third-generation scalar leptoquarks and top squarks in proton-proton collisions at $\sqrt{s} = 8$ TeV”, *Phys. Lett. B* (2014) doi:10.1016/j.physletb.2014.10.063, arXiv:1408.0806. 131
- [116] A. L. Read, “Presentation of search results: the CL_s technique”, *J. Phys. G* **28** (2002) 2693, doi:10.1088/0954-3899/28/10/313. xiv, 131, 143
- [117] T. Junk, “Confidence level computation for combining searches with small statistics”, *Nucl. Instrum. Meth. A* **434** (1999) 435, doi:10.1016/S0168-9002(99)00498-2, arXiv:hep-ex/9902006. 131
- [118] ATLAS and CMS Collaborations, LHC Higgs Combination Group, “Procedure for the LHC Higgs boson search combination in Summer 2011”, Technical Report ATL-PHYS-PUB-2011-11, CMS-NOTE-2011-005, CERN, 2011. 131
- [119] CMS Collaboration, “Search for top-squark pair production in the single-lepton final state in pp collisions at $\sqrt{s} = 8$ TeV”, *Eur. Phys. J. C* **73** (2013) 2677, doi:10.1140/epjc/s10052-013-2677-2, arXiv:1308.1586. xiv, 132, 138, 139
- [120] CMS Collaboration, “Projected Performance of an Upgraded CMS Detector at the LHC and HL-LHC: Contribution to the Snowmass Process”, CMS Note CMS-NOTE-2013-002, CERN, Geneva, 2013. arXiv:1307.7135. 140
- [121] A. L. Read, “Modified frequentist analysis of search results (the CLs method)”, CERN Report CERN-OPEN-2000-005, CERN, 2000. 141