

UNIVERSITY OF CALIFORNIA  
SANTA CRUZ

**SEARCH FOR  $WW$  AND  $WZ$  RESONANCE PRODUCTION IN  
 $\ell\nu qq$  FINAL STATES IN  $pp$  COLLISIONS AT  $\sqrt{s} = 13$  TEV WITH  
THE ATLAS DETECTOR**

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

**Natasha Woods**

December 2019

The Dissertation of Natasha Woods  
is approved:

---

Abraham Seiden, Chair

---

Mike Hance

---

Bruce Schumm

---

Quentin Williams  
Vice Provost and Dean of Graduate Studies

Copyright © by

Natasha Woods

2019

# Table of Contents

List of Figures	vi
List of Tables	xv
Abstract	xvii
Dedication	xviii
Acknowledgments	xix
<b>I Theoretical Motivation</b>	<b>2</b>
<b>1 The Standard Model of Particle Physics</b>	<b>3</b>
1.1 Introduction . . . . .	3
1.2 Quantum Field Theory . . . . .	3
1.3 $U(1)_{EM}$ Local Gauge Invariance . . . . .	4
1.4 Yang-Mills Gauge Theories . . . . .	7
1.5 Particles in the Standard Model . . . . .	8
1.6 Higgs Mechanism . . . . .	13
1.7 Electroweak Theory . . . . .	14
1.8 Quantum ChromoDynamics . . . . .	15
<b>2 Standard Model Successes and Limitations</b>	<b>20</b>
<b>3 New Physics Models with Diboson Resonances</b>	<b>23</b>
3.1 Randall Sundrum Bulk Model . . . . .	23
3.2 Extended Scalar Sector . . . . .	25
3.3 Simple Standard Model Extensions . . . . .	26

<b>II Experimental Setup</b>	<b>29</b>
<b>4 LHC</b>	<b>30</b>
4.1 LHC Layout and Design . . . . .	32
<b>5 The ATLAS Detector</b>	<b>37</b>
5.1 Coordinate System . . . . .	39
5.2 Inner Detector . . . . .	40
5.2.1 Pixel Detector . . . . .	43
5.2.2 Semiconductor Tracker . . . . .	43
5.2.3 Transition Radiation Tracker . . . . .	43
5.3 Calorimeters . . . . .	45
5.4 Muon Spectrometer . . . . .	49
5.5 Magnet System . . . . .	53
5.6 Trigger System . . . . .	54
<b>III Method</b>	<b>56</b>
<b>6 Dataset and Simulated Samples</b>	<b>57</b>
6.1 Dataset . . . . .	57
6.2 Simulated Samples . . . . .	60
6.3 Object Selection . . . . .	60
6.3.1 Electrons . . . . .	60
6.3.2 Muons . . . . .	61
6.3.3 small-R jets . . . . .	63
6.3.4 large-R jets . . . . .	66
6.3.5 Variable Radius jets . . . . .	69
6.3.6 MET/neutrinos . . . . .	69
6.3.7 Jet Flavor Tagging . . . . .	69
6.3.8 Overlap Removal . . . . .	70
<b>7 Event Selection and Categorization</b>	<b>71</b>
7.1 Pre-selection . . . . .	71
7.2 Trigger . . . . .	71
7.3 GGF/VBF RNN . . . . .	73
7.4 Topological Cuts . . . . .	78
7.5 Selection Acceptance times efficiency for Signal Events . . . . .	88
7.6 Background Estimate . . . . .	89
7.6.1 Control Regions . . . . .	89
7.6.2 Fake Lepton Backgrounds . . . . .	89

<b>8 Systematic Uncertainties</b>	<b>103</b>
8.1 Experimental Systematics . . . . .	103
8.2 Theory Systematics . . . . .	105
<b>9 Statistical Analysis</b>	<b>113</b>
9.1 Likelihood Function Definition . . . . .	113
9.2 Fit Configuration . . . . .	114
9.3 Best Fit $\mu$ . . . . .	116
9.4 Discovery Test . . . . .	117
9.5 Exclusion Limits . . . . .	118
<b>IV Results</b>	<b>120</b>
<b>10 Statistical Interpretation</b>	<b>121</b>
10.1 Discovery Tests . . . . .	121
10.2 Systematic Profiling and Correlations . . . . .	127
10.3 Expected and Measured Yields . . . . .	128
10.4 Limits . . . . .	142
<b>V Quark and Gluon Tagging</b>	<b>144</b>
<b>11 Prospects</b>	<b>145</b>
<b>12 <math>n_{trk}</math> Calibration</b>	<b>152</b>
<b>13 Application</b>	<b>158</b>
<b>VI Conclusion</b>	<b>163</b>
<b>14 Conclusions</b>	<b>164</b>
<b>Bibliography</b>	<b>165</b>

# List of Figures

1.1	The particles of the Standard Model. . . . .	10
1.2	Summary of how Standard Model particles interact with other Standard Model particles. . . . .	11
1.3	This figure shows the three dominant QCD interactions. From Ref. [14] . . . . .	17
1.4	Strength of the U(1), SU(2), and SU(3) gauge couplings as a function of the energy scale of the interaction ( $Q$ ). From Ref. [10] . . . . .	18
2.1	A comparison of cross section measurements at $\sqrt{s} = 7, 8, 13$ TeV from ATLAS compared to theoretical measurements. From Ref. [5]	22
3.1	Cartoon of RS Bulk Model . . . . .	24
4.1	Scaling of cross sections with $\sqrt{s}$ . Natasha: write more here . . . . .	31
4.2	LHC Layout. Natasha write more . . . . .	33
4.3	LHC Accelerator. Natasha write more . . . . .	35
5.1	Big picture layout of ATLAS detector. Natasha: write more . . . . .	38
5.2	Big picture layout of ATLAS detector. Natasha: write more . . . . .	38
5.3	A simplified schematic of how different particles interact and are detected within ATLAS. . . . .	39
5.4	Layout of ATLAS Inner Detector . . . . .	41
5.5	Layout of ATLAS ID Barrel System. . . . .	42
5.6	Overview of ATLAS electromagnetic and hadronic calorimeters. . . . .	46

5.7	Schematic of ECAL . . . . .	47
5.8	Schematic of HCAL . . . . .	48
5.9	Schematic of Muon Spectrometer [cite G35] . . . . .	51
5.10	Schematic of MDT chamber. [cite G35] . . . . .	52
5.11	Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35]. . . . .	52
5.12	Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35] . . . . .	53
5.13	Layout of ATLAS magnet systems. . . . .	54
6.1	Integrated luminosity for data collected from ATLAS from 2011 - 2018 . . . . .	58
6.2	Mean number of interactions per crossing for data collected from ATLAS from 2011 - 2018 . . . . .	59
6.3	This figure shows the breakdown of the muon reconstruction efficiency scale factor measured in $Z \rightarrow \mu\mu$ as a function of $p_T$ [4]. . . . .	63
6.4	[6] This diagram shows the calibration stages for EM jets. . . . .	66
6.5	The upper cut on $D_2$ (a) and jet mass window cut i.e. the upper and lower boundary of the mass (b) of the $W$ -tagger as a function of jet $p_T$ . Corresponding values for $Z$ -tagger are shown in (c) and (d). The optimal cut values for maximum significance are shown as solid markers and the fitted function as solid lines. Working points from $VV \rightarrow JJ$ [ATLAS-HDBS-2018-31-002] is also shown as dashed lines as a reference. Natasha reword? . . . . .	68
6.6	Natasha write caption . . . . .	68
7.1	RNN architecture. Natasha add caption . . . . .	75
7.2	RNN Score distribution for ggF and VBF signals and backgrounds. . . . .	76
7.3	ROC curve using k-fold validation for RNN. . . . .	77

7.4	Comparison of GGF Z' limits for different RNN score selections. The bottom panel shows the ratio of the upper limits set for different RNN cuts to the cut-based analysis. In this panel smaller numbers, indicate that the expected upper limit is smaller than the cut-based analysis, which is desired. . . . .	78
7.5	Event Categorization. Natasha write more. . . . .	81
7.6	Data MC comparison for the merged $WW$ HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	82
7.7	Data MC comparison for the merged $WW$ LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	83
7.8	Data MC comparison for the merged $WZ$ HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	84
7.9	Data MC comparison for the merged $WZ$ LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	85
7.10	Data MC comparison for the resolved $WW$ TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	86
7.11	Data MC comparison for the resolved $WZ$ TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background. . . . .	87

7.12 Selection acceptance times efficiency for the $W' \rightarrow WZ \rightarrow \ell\nu qq$ events from MC simulations as a function of the $W'$ mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the $WW \rightarrow \ell\nu J$ selection and the resolved regions of the $WW \rightarrow \ell\nu jj$ selection. . . . .	88
7.13 Selection acceptance times efficiency for the $G \rightarrow WW \rightarrow \ell\nu qq$ events from MC simulations as a function of the $G$ mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the $WW \rightarrow \ell\nu J$ selection and the resolved regions of the $WW \rightarrow \ell\nu jj$ selection. . . . .	89
7.14 The $E_T^{miss}$ distribution in MJCR for 2017 data in the electron channel(left), muon channel with W-boson pT < 150 GeV (center) and > 150 GeV (right). Multi-jet templates are calculated as remaining data components after excluding known MC . . . . .	91
7.15 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WW$ electron channel. The MJ template is obtained from the pre-MJ-fit. . . . .	92
7.16 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WW$ muon channel. The MJ template is obtained from the pre-MJ-fit. . . . .	93
7.17 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag electron channel. The MJ template is obtained from the pre-MJ-fit. . . . .	94
7.18 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag muon channel. The MJ template is obtained from the pre-MJ-fit. . . . .	95
7.19 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag electron channel. The MJ template is obtained from the pre-MJ-fit. . . . .	96

7.20 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the $WZ$ untag muon channel. The MJ template is obtained from the pre-MJ-fit.	97
7.21 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WW$ electron channel. The MJ template is obtained from the pre-MJ-fit.	98
7.22 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WW$ muon channel. The MJ template is obtained from the pre-MJ-fit.	99
7.23 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WZ$ electron channel. The MJ template is obtained from the pre-MJ-fit.	100
7.24 Postfit Data/MC comparison of distributions of $E_T^{miss}$ , $m_T^W$ , lepton and neutrino $p_T$ , $m_{\ell\nu jj}$ , lepton- $\nu$ angular distance in the VBF $WZ$ muon channel. The MJ template is obtained from the pre-MJ-fit.	101
8.1 The W/Z+jet systematics for the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The top subplot shows the nominal and variation distributions/bands, the middle shows the ratio of the two, and the final shows just the shape of the envelope (the final uncertainty).	107
8.2 The two-point generator comparison between Sherpa and MadGraph for the W/Z+jet samples in the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The normalization of the Madgraph sample is set to the Sherpa value to consider only shape effects. The bottom inlet shows the ratio of the two.	108
8.3 Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for $t\bar{t}$ sample.	109
8.4 Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for $t\bar{t}$ sample.	109

8.5	Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample. . . . .	110
8.6	Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample. . . . .	111
8.7	Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for $t\bar{t}$ sample. . . . .	112
8.8	Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for $t\bar{t}$ sample. . . . .	112
9.1	The HVT signal mass resolution as a function of mass fit with a straight line in the Resolved ggF region (left) and VBF (right) region. . . . .	116
9.2	The HVT signal mass resolution as a function of mass fit with a straight line in the Merged ggF region (left) and VBF (right) region. . . . .	116
10.1	These plots show the measured $p_0$ value as a function of resonance mass for HVT Z' DY production. . . . .	122
10.2	These plots show the measured $p_0$ value as a function of resonance mass for HVT Z' VBF production. . . . .	123
10.3	These plots show the measured $p_0$ value as a function of resonance mass for HVT W' DY production. . . . .	124
10.4	These plots show the measured $p_0$ value as a function of resonance mass for HVT W' VBF production. . . . .	125
10.5	These plots show the measured $p_0$ value as a function of resonance mass for the RS Graviton DY production. . . . .	126
10.6	Ranked systematics and their fitted values for WW DY (right) and VBF (left) selections. . . . .	127
10.7	Ranked systematics and their fitted values for WZ DY (right) and VBF (left) selections. . . . .	128
10.8	Correlations between systematics for WW DY (right) and VBF (left) selections. . . . .	128

10.9 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WW control regions. . . . .	136
10.10 This figure shows the distribution of $m_{\ell\nu qq}$ in the DY WZ control regions. . . . .	137
10.11 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WW control regions. . . . .	138
10.12 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ control regions. . . . .	139
10.13 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WW signal regions. . . . .	140
10.14 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Untag signal regions. . . . .	140
10.15 This figure shows the distribution of $m_{\ell\nu qq}$ in the GGF WZ Tag signal regions. . . . .	141
10.16 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions. . . . .	141
10.17 This figure shows the distribution of $m_{\ell\nu qq}$ in the VBF WZ Tag signal regions. . . . .	142
10.18 This figure shows theory, expected and observed limits for HVT $W'$ DY (left) and VBF (right) production. . . . .	142
10.19 This figure shows theory, expected and observed limits for HVT $Z'$ DY (left) and VBF (right) production. . . . .	143
10.20 This figure shows theory, expected and observed limits for RS Gravitons via DY production. . . . .	143
11.1 PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background. . . . .	147

11.2 The number of tracks in small-R jets in events passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets. These distributions are shown for 300, 500, and 700GeV Z' signals and the background. . . . .	147
11.3 The number of tracks in background small-R jets in events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one. . . . .	148
11.4 The number of tracks in small-R jets in 300GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one. . . . .	148
11.5 The number of tracks in small-R jets in 500GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one. . . . .	149
11.6 The number of tracks in small-R jets in 700GeV Z' events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one. . . . .	149

11.7 The number of tracks in leading small-R jets in background events passing the Resolved GGF WW Signal region selection vs. $\ln(p_T)$ for (a)Gluons (b) Quarks jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one. . . . .	150
11.8 ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the $\ln(p_T)$ . . . . .	150
11.9 The top panel shows the distribution of $m_{lvqq}$ with and without quark gluon tagging. The middle panel shows the ratio of the signals and backgrounds with and without quark gluon tagging. The bottom panel shows the change in $S/\sqrt{B}$ with quark gluon tagging. . . . .	151
13.1 The top panel shows the distribution of $m_{lvqq}$ with and without requiring jets to be true quarks. The middle panel shows the ratio of the signals and backgrounds with and without requiring jets to be true quarks. The bottom panel shows the change in $S/\sqrt{B}$ when requiring jets to be true quarks. . . . .	159
13.2 Unfolded and extracted $n_C$ qg dstbs. . . . .	160
13.3 PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background. . . . .	161
13.4 These figures show the impact of the uncertainties on the number of tracks in the leading jet in the sum of the background sample in the Resolved GGF WW SR (a) tracking efficiency (b) fake (c) PDF (d) ME (e) unfolding uncertainties. . . . .	162

# List of Tables

1.1	Representations of the SM fermions under $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetry group. $SU(2)_L$ gauge transformations allow one to go between rows and $SU(3)_C$ transformations allow one to go between columns in these fermion representations. [REWORD]	9
7.1	The list of triggers used in the analysis. . . . .	72
7.2	Summary of selection criteria used to define the signal region (SR), $W$ +jets control region ( $W$ CR) and $t\bar{t}$ control region ( $t\bar{t}$ CR) for merged 1-lepton channel. . . . .	80
7.3	The list of selection cuts in the resolved analysis for the $WW$ and $WZ$ signal regions (SR), $W$ +jets control region (WR) and $t\bar{t}$ control region (TR). . . . .	81
7.4	Definitions of “inverted” leptons used in multijet control region . .	91
7.5	Fit validation result in WCRs for 2015+16 data. The fit is done in various WCRs, in order to obtain the corresponding scale factors for MJ templates: ggF resolved WCR for the $WW \rightarrow lvqq$ selection, ggF resolved untagged WCR for the $WZ \rightarrow lvqq$ selection, ggF resolved tagged WCR for the $WZ \rightarrow lvqq$ selection, VBF resolved WCR for the $WW \rightarrow lvqq$ selection, and VBF resolved WCR for the $WZ \rightarrow lvqq$ selection. Post-fit event yields for electroweak processes and MJ contributions are shown. The SF column shows the corresponding normalization scale factors for electroweak processes from the fit. R.U. stands for relative uncertainty. . . . .	102

10.1	Expected and Measured for DY $WW$ $W+\text{jets}$ , $t\bar{t}$ control regions and signal regions. . . . .	129
10.2	Expected and Measured for DY $WZ$ $W+\text{jets}$ , $t\bar{t}$ tag and untag control regions. . . . .	130
10.3	Expected and Measured for DY $WZ$ $W+\text{jets}$ , $t\bar{t}$ tag and untag signal regions. . . . .	131
10.4	Expected and Measured for VBF $WW$ $W+\text{jets}$ , $t\bar{t}$ control regions and signal regions. . . . .	132
10.5	Expected and Measured for VBF $WZ$ $W+\text{jets}$ , $t\bar{t}$ control regions and signal regions. . . . .	133
10.6	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the DY $WW$ analysis region. . . . .	134
10.7	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the DY $WZ$ analysis region. . . . .	134
10.8	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the VBF $WW$ analysis region. . . . .	134
10.9	Fitted background normalizations for $t\bar{t}$ and $W+\text{jets}$ backgrounds for the VBF $WZ$ analysis region. . . . .	135

## Abstract

Search for  $WW$  and  $WZ$  resonance production in  $\ell\nu qq$  final states in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector

by

Natasha Woods

This thesis reviews a search for  $WW$  and  $WZ$  resonance production using data from  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector, corresponding to an integrated luminosity of  $139 \text{ fb}^{-1}$ . Diboson resonances are predicted in a number of Standard Model (SM) extensions, such as Extended Gauge Models, Extra dimensions, and technicolor models. This search looks for resonances where one  $W$  boson decays leptonically and the other  $W$  or  $Z$  boson decays hadronically. This search is sensitive to diboson resonance production via vector-boson fusion as well as quark-antiquark annihilation and gluon-gluon fusion mechanisms. No significant excess of events is observed with respect to the Standard Model backgrounds. As the dominant backgrounds in this search contain gluons, classifying jets as quark or gluon initiated would make this analysis more sensitive to new physics. Towards this end, this thesis considers the prospects for adding a quark gluon tagger based on the number of tracks in jets and reviews the calibration of the number of tracks in jets.

A loving dedication.

å

xviii

## **Acknowledgments**

Proper acknowledgments of everyone else who helped you graduate. Write later.

<sub>1</sub> another intro???

<sup>2</sup>

## Part I

<sup>3</sup>

# Theoretical Motivation

<sup>4</sup> **Chapter 1**

<sup>5</sup> **The Standard Model of Particle  
6 Physics**

<sup>7</sup> **1.1 Introduction**

<sup>8</sup> By determining the dynamics of the most elementary degrees of freedom, par-  
<sup>9</sup> ticle physics hopes to uncover the fundamental laws of the universe. The definition  
<sup>10</sup> of elementary has evolved through time and currently refers to matter and force  
<sup>11</sup> mediating particles: fermions and bosons, respectively. The Standard Model of  
<sup>12</sup> Particle Physics (SM) describes the quantum behavior of three of the four funda-  
<sup>13</sup> mental forces: weak, strong, and electromagnetic, via boson and fermion interac-  
<sup>14</sup> tions. Gravity is not included in the SM and still under investigation.

<sup>15</sup> **1.2 Quantum Field Theory**

<sup>16</sup> In the SM, forces (and particles) are represented as fields. In this context,  
<sup>17</sup> fields are mathematical objects that define a tensor (e.g. scalar, vector, etc) at  
<sup>18</sup> every point on a manifold, here the manifold is space-time. These fields obey laws

<sup>19</sup> dictated by Quantum Field Theory (QFT). Particles arise naturally in QFT as  
<sup>20</sup> quantized field excitations localized in spacetime.

<sup>21</sup> According to Noether's theorem, symmetries of a field give rise to conserved  
<sup>22</sup> quantities (e.g. time-translation invariance leads to energy conservation). Often  
<sup>23</sup> in the history of physics, a conserved quantity of a field is found and then the  
<sup>24</sup> underlying symmetry of the field is inferred. Gauge symmetries are symmetries  
<sup>25</sup> among the internal degrees of freedom of the field (components of the tensor),  
<sup>26</sup> which give rise to quantities associated with fields. By specifying the symmetries  
<sup>27</sup> of a system the dynamics and conserved quantities of the system may be succinctly  
<sup>28</sup> defined.

### <sup>29</sup> 1.3 $U(1)_{EM}$ Local Gauge Invariance

<sup>30</sup> The Lagrangian of Quantum Electrodynamics (QED) describes the electro-  
<sup>31</sup> magnetic force. QED may be derived by requiring local  $U(1)_{EM}$  gauge invariance  
<sup>32</sup> of the free dirac fermion Lagrangian,  $\psi$ :

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (1.1)$$

<sup>33</sup> This symmetry may be represented as a complex number with unit modulus,  
<sup>34</sup>  $e^{i\theta}$ .  $U(1)$  gauge invariance requires this gauge transformation of  $\psi$  will leave the  
<sup>35</sup> Lagrangian unchanged.

$$\psi(x) \rightarrow \psi'(x) = e^{i\theta(x)}\psi(x) \quad (1.2)$$

<sup>36</sup> NB: This transformation is a local gauge transformation as  $\theta$  depends on the  
<sup>37</sup> spacetime coordinate.

<sup>38</sup> By requiring this symmetry of the free Dirac fermion Lagrangian:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi \quad (1.3)$$

<sup>39</sup> The mass term is unaffected, but the kinetic term is modified due to  $\theta(x)$ .

$$\mathcal{L} \rightarrow \mathcal{L}' = i\bar{\psi}e^{-i\theta(x)}\gamma^\mu\partial_\mu\psi e^{i\theta(x)} - m\bar{\psi}e^{-i\theta(x)}\psi e^{i\theta(x)} \quad (1.4)$$

<sup>40</sup>

$$= i\psi\gamma^\mu(\partial_\mu\psi + i\psi\partial_\mu\theta) - m\bar{\psi}\psi \quad (1.5)$$

<sup>41</sup> The  $\partial_\mu\theta$  terms breaks the gauge invariance of the Lagrangian. By introducing a  
<sup>42</sup> new field,  $A_\mu$  we can recover the gauge invariance of the derivative. Now redefining  
<sup>43</sup> the derivative as the covariant derivative:

$$D_\mu\psi \equiv (\partial_\mu - iqA_\mu)\psi \quad (1.6)$$

<sup>44</sup> And letting  $A_\mu$  transform under  $U(1)$  as:

$$A_\mu \rightarrow A_\mu + \delta A_\mu \quad (1.7)$$

<sup>45</sup> The transformed covariant derivative becomes:

$$D_\mu\psi \rightarrow D_\mu\psi' = (\partial_\mu - iqA_\mu)\psi' \quad (1.8)$$

<sup>46</sup>

$$= (\partial_\mu - iq(A_\mu + \delta A_\mu))\psi e^{i\theta} \quad (1.9)$$

<sup>47</sup>

$$= e^{i\theta}D_\mu + ie^{i\theta}\psi(\partial_\mu\theta - q\delta A_\mu) \quad (1.10)$$

<sup>48</sup> The covariant derivative can be made gauage invariant by setting the last term  
<sup>49</sup> to zero.

$$\delta A_\mu = \frac{1}{q} \partial_\mu \theta \quad (1.11)$$

50 So now  $A_\mu$  transforms as:

$$A_\mu \rightarrow A_\mu + \frac{1}{q} \partial_\mu \theta \quad (1.12)$$

51 Finally, replacing the derivative with the covariant derivative the Dirac La-  
52 grangian we have:

$$\mathcal{L} = i\bar{\psi}\gamma^\mu D_\mu \psi - m\bar{\psi}\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.13)$$

53

$$= \mathcal{L}_{QED} \quad (1.14)$$

54 Here  $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ . This last term in the Lagrangian is the kinetic  
55 energy of the gauge boson field.

56 So we have derived the QED Lagrangian. By requiring the free Dirac La-  
57 grangian to be invariant under U(1) transformations we have generated a new  
58 gauge boson field,  $A_\mu$ , which describes the photon. As expected the photon inter-  
59 acts with fermions.

60 Stepping back, a global U(1) gauge symmetry of the free Dirac Lagrangian  
61 implies we cannot measure the absolute phase of a charged particle. A local U(1)  
62 gauge symmetry changes the phase of fields differently across space time. For this  
63 type of transformation to leave the Lagrangian invariant, we had to introduce an  
64 additional field,  $A_\mu$ , which "communicates" these phase changes across space-time.  
65 In less formal language this effectively means: if the field at one location changes,  
66 this change is conferred to other particles via  $A_\mu$ .

## 67 1.4 Yang-Mills Gauge Theories

68 Requiring  $U(1)_{EM}$  gauge invariance of the free Dirac Lagrangian gave us QED.

69 Requiring different gauge symmetries we can derive the structure of other inter-  
70 actions. Any gauge symmetry may be written as:

$$\psi_i \rightarrow \exp(i\theta^a T_{ij}^a) \psi_j \quad (1.15)$$

71 Here  $\theta$  is a dimensionless real parameter and  $T$  is the generator of the gauge  
72 symmetry group. With this the covariant derivative can be written as:

$$D_\mu \psi_i \equiv \partial_\mu \psi_i + ig A_\mu^a T_{ij}^a \psi_j \quad (1.16)$$

73 Then the gauge field must transform as:

$$A_\mu^a \rightarrow A_\mu^a - \frac{1}{g} \partial_\mu \theta^a - f^{abc} \theta^b A_\mu^c \quad (1.17)$$

74 Here  $f$  is the structure constant of the gauge group. The field strength tensor  
75 is given by:

$$F_{\mu\nu}^a \equiv \partial_\mu A_\nu^a - \partial_\nu A_\mu^a - g f^{abc} A_\mu^b A_\nu^c \quad (1.18)$$

76

$$F_{\mu\nu}^a \rightarrow F_{\mu\nu}^a - f^{abc} \theta^b F_{\mu\nu}^c \quad (1.19)$$

77 This gives the Yang-Mills Lagrangian:

$$\mathcal{L}_{YM} = -\frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a + i \bar{\psi}_i \gamma^\mu D_\mu \psi_i + m \bar{\psi}_i \psi_i \quad (1.20)$$

## 78 1.5 Particles in the Standard Model

79 The SM consists of fermions (half-integer spin matter constituents) and bosons  
80 (integer spin force mediators). Fermions are spinor representations of the Poincare  
81 group and can be further separated into leptons and quarks. Bosons are the result  
82 of requiring a particular symmetry among the spinor fields:

$$SU(3)_C \times SU(2)_L \times U(1)_Y \quad (1.21)$$

83  $SU(3)_C$  is the symmetry group of the strong force and generates eight gluon  
84 fields,  $G_\mu$ .  $SU(2)_L$  is the symmetry group of the Electroweak force and generates  
85 three electroweak boson fields. The mixing of this  $SU(2)_L$  and  $U(1)_Y$  gives rise  
86 to the photon field, where Y is the weak-hypercharge:

$$Y = 2(Q - T_3) \quad (1.22)$$

87 Q is the electromagnetic charge, and  $T_3$  is the z-component of the weak isospin.  
88 Weak isospin is the charge associated with the  $SU(2)_L$  symmetry. The correspond-  
89 ing covariant derivative is then:

$$D_\mu \phi \equiv (\partial_\mu + ig_1 B_\mu Y_{L/R} + [ig_2 W_\mu^\alpha T^\alpha]_L + [ig_3 G_\mu^\alpha \tau^\alpha]_C) \psi \quad (1.23)$$

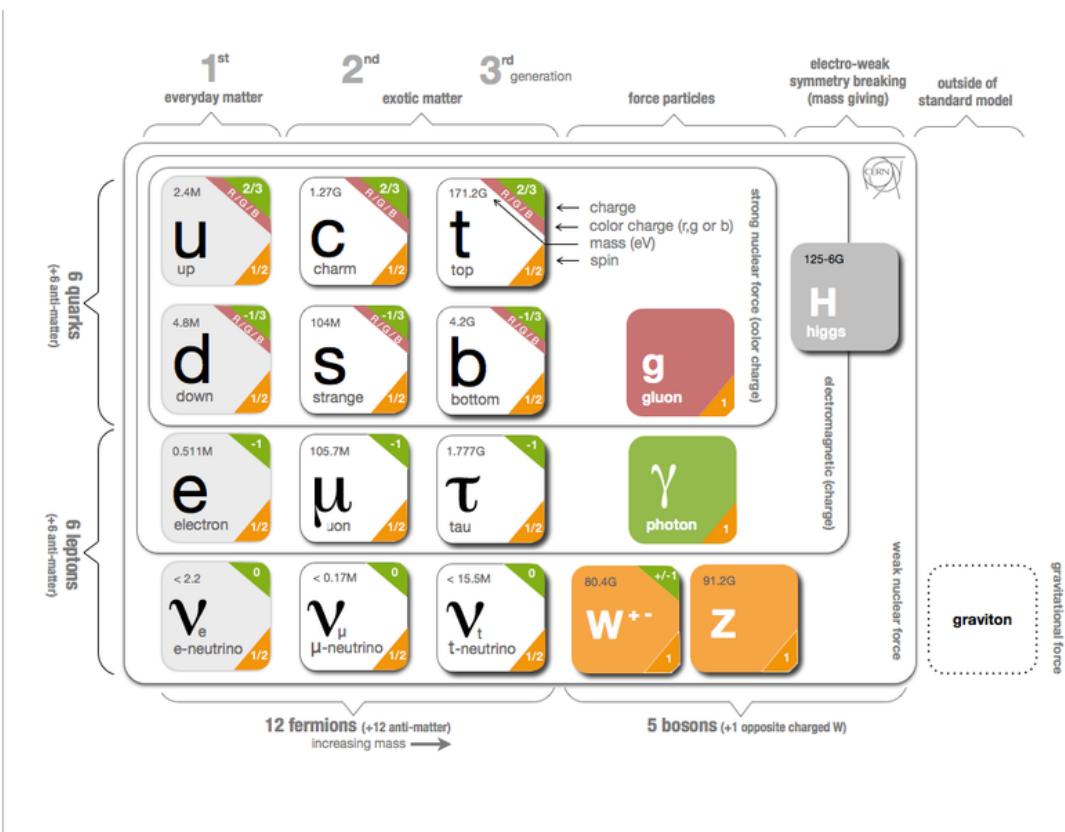
90 It is important to note that the gauge symmetry of the SM yields a particular  
91 structure of the fermion representations. So for a given fermion to interact with  
92 a given gauge field it must have a non-zero corresponding Noether charge for  
93 that gauge symmetry. If the corresponding Noether charge is zero, that fermion  
94 transforms as a singlet and does not participate in that gauge interaction.

95 Fermions are divided into quarks and leptons based on their transformations  
96 under  $SU(3)_C$ . Quarks transform as color triplets. Leptons transform as color

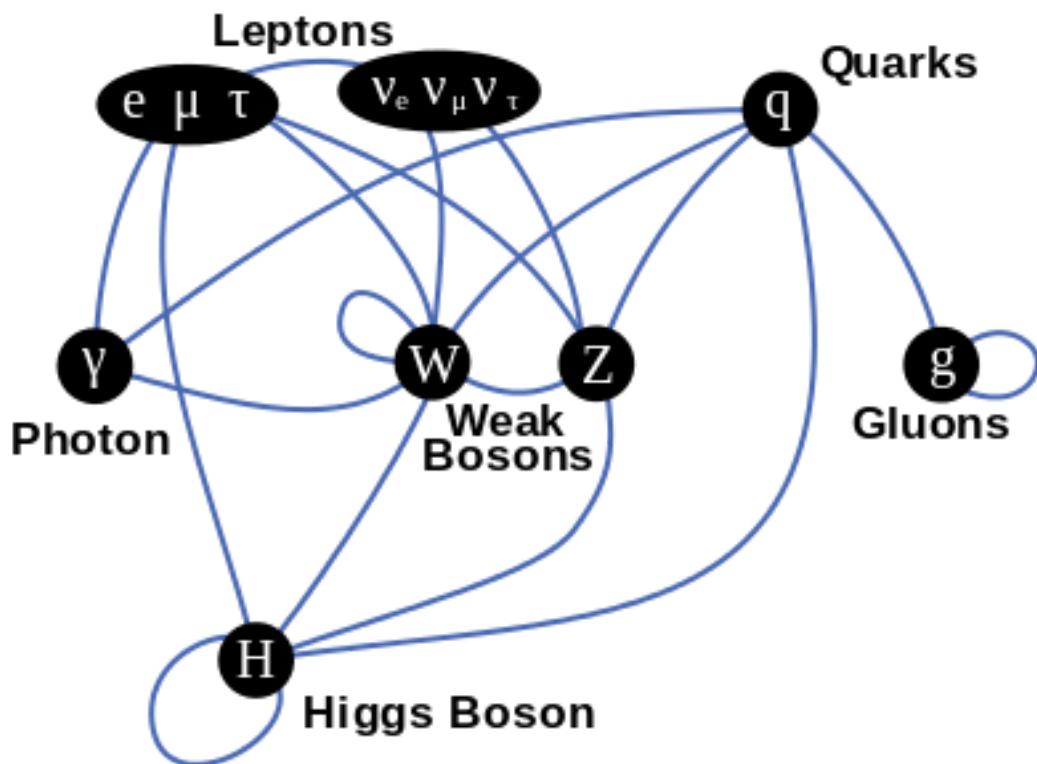
singlets and consequently do not interact with gluons. Fermions may be further  
 classified by their  $SU(2)_L$  interactions. Only the left-chiral part of fermions (denoted by L here) transform as  $SU(2)_L$  doublets, the right-chiral part forms singlets under this gauge. Lastly, all these groups of particles come in three generations, each a heavier copy of the previous, but with differing flavor quantum numbers.  
 This is summarized in Table 1.1 and shown in Figures 1.1 and 1.2.

SM Fermion Gauge Group	First Generation	Second Generation	Third Generation	$(SU(3)_C, SU(2)_L, U(1)_Y)$ Representations
Left-handed quarks	$\begin{pmatrix} u_L^r & u_L^g & u_L^b \\ d_L^r & d_L^g & d_L^b \end{pmatrix}$	$\begin{pmatrix} c_L^r & c_L^g & c_L^b \\ s_L^r & s_L^g & s_L^b \end{pmatrix}$	$\begin{pmatrix} t_L^r & t_L^g & t_L^b \\ b_L^r & b_L^g & b_L^b \end{pmatrix}$	$(3, 2, \frac{1}{6})$
Right-handed quarks	$(u_R^r, u_R^g, u_R^b)$ $(d_R^r, d_R^g, d_R^b)$	$(c_R^r, c_R^g, c_R^b)$ $(s_R^r, s_R^g, s_R^b)$	$(t_R^r, t_R^g, t_R^b)$ $(b_R^r, b_R^g, b_R^b)$	$(3, 1, \frac{2}{3})$ $(3, 1, -\frac{1}{3})$
Left-handed leptons	$\begin{pmatrix} \nu_e^L \\ e_L \end{pmatrix}$	$\begin{pmatrix} \mu_e^L \\ \mu_L \end{pmatrix}$	$\begin{pmatrix} \tau_e^L \\ \tau_L \end{pmatrix}$	$(1, 2, -\frac{1}{2})$
Right-handed leptons	$e_R$	$\mu_R$	$\tau_R$	$(1, 1, -1)$

**Table 1.1:** Representations of the SM fermions under  $SU(3)_C \times SU(2)_L \times U(1)_Y$  gauge symmetry group.  $SU(2)_L$  gauge transformations allow one to go between rows and  $SU(3)_C$  transformations allow one to go between columns in these fermion representations. [REWORD]



**Figure 1.1:** The particles of the Standard Model.



**Figure 1.2:** Summary of how Standard Model particles interact with other Standard Model particles.

<sup>103</sup> Now we can understand the SM Lagrangian density as a Yang-Mills theory  
<sup>104</sup> with the gauge group:  $SU(3)_C \times SU(2)_L \times U(1)_Y$  with an additional  $SU(2)$  complex  
<sup>105</sup> scalar Higgs field doublet that will be discussed later.

$$\begin{aligned} \mathcal{L}_{SM} = & \underbrace{-\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^aW^{a\mu\nu} - \frac{1}{4}G_{\mu\nu}^\alpha G^{\alpha\mu\nu}}_{\text{Kinetic Energies and Self-Interactions of Gauge Bosons}} \\ & + \underbrace{\bar{L}_i \gamma^\mu (i\partial_\mu - \frac{1}{2}g_1 Y_{iL} B_\mu - \frac{1}{2}g_2 \sigma^a W_\mu^a) L_i}_{\text{Kinetic Energies and EW Interactions of Left-handed Fermions}} \\ & + \underbrace{\bar{R}_i \gamma^\mu (i\partial_\mu - \frac{1}{2}g_1 Y_{iR} B_\mu) R_i}_{\text{Kinetic Energies and EW Interactions of Right-Handed Fermions}} \\ & + \underbrace{\frac{ig_3}{2} \bar{Q}_j \gamma^\mu \lambda^\alpha G_\mu^\alpha Q_j}_{\text{Strong Interactions between Quarks and Gluons}} \\ & + \underbrace{\frac{1}{2} |(i\partial_\mu - \frac{1}{2}g_1 B_\mu - \frac{1}{2}g_2 \sigma^a W_\mu^a)\Phi|^2 - V(\Phi)}_{\text{Electroweak Boson Masses and Higgs Couplings}} \\ & - \underbrace{(y_{kl}^d \bar{L}_k \Phi R_l + y_{kl}^u \bar{R}_k \tilde{\Phi} L_l + h.c.)}_{\text{Fermion Mass terms and Higgs Couplings}} \end{aligned}$$

<sup>106</sup> Here several abstract spaces are being spanned:

- <sup>107</sup> – a spans the three  $SU(2)_L$  gauge fields with generators expanded in Pauli  
<sup>108</sup> matrices,  $T^\alpha = \frac{1}{2}\sigma^\alpha$
- <sup>109</sup> –  $\alpha$  spans the eight  $SU(3)_C$  gauge fields, with generators expanded in Gell-  
<sup>110</sup> Mann matrices,  $\tau^\alpha = \frac{1}{2}\lambda^\alpha$
- <sup>111</sup> – L/R represent left and right projections of Dirac fermion fields. The Strong  
<sup>112</sup> interaction is not chiral, so  $Q = L+R$

<sub>113</sub> –  $\mu$  and  $\nu$  are four-vector indices

<sub>114</sub> –  $i, j, k$  are summed over the three generations of SM particles.

## <sub>115</sub> 1.6 Higgs Mechanism

<sub>116</sub> The SM Lagrangian without the addition of a Higgs field does not allow for  
<sub>117</sub> gauge boson and fermion mass terms:  $\frac{1}{2}m_A^2 A_\mu A_\mu$  and  $m(\bar{\psi}\psi)$ , as these terms are  
<sub>118</sub> not gauge invariant. By introducing the Higgs field, mass terms for these particles  
<sub>119</sub> may be included in a gauge invariant way. This field is a complex doublet with a  
<sub>120</sub> potential  $V(\Phi)$ :

$$\Psi = \begin{pmatrix} \Phi^\dagger \\ \Phi^0 \end{pmatrix} \quad (1.24)$$

<sub>121</sub>  $V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda |\Phi^\dagger \Phi|^2 \quad (1.25)$

<sub>122</sub> The minima of this field occurs for  $|\Phi| = \sqrt{\frac{\mu^2}{2\lambda}} \equiv \frac{v}{2}$ . This yields degenerate  
<sub>123</sub> minima, this symmetry is broken by choosing a specific minima (a.k.a. sponta-  
<sub>124</sub> neous symmetry breaking). By convention  $\Phi_{min} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$  is chosen. This means  
<sub>125</sub> the ground state of the Higgs field (Higgs vacuum) is non-zero,  $\sqrt{\frac{-\mu^2}{\lambda}}$ . The Higgs  
<sub>126</sub> Field may now be expanded around this new ground state:

$$\Phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.26)$$

<sub>127</sub> This non-zero Higgs vacuum now generates mass terms for the gauge bosons  
<sub>128</sub> from the following term in the Lagrangian:

$$|(-\frac{1}{2}g_1B_\mu - \frac{1}{2}g_2\sigma^aW_\mu^a)\Phi|^2 = \frac{1}{2}m_W^2W_\mu^+W^{-\mu} + \frac{1}{2}m_Z^2Z_\mu Z^\mu \quad (1.27)$$

129 where:

$$W_\mu^\pm \equiv \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2) \quad (1.28)$$

$$\begin{aligned} \text{130} \quad Z_\mu &\equiv \frac{1}{\sqrt{g_1^2 + g_2^2}}(g_2W_\mu^2 - g_1B_\mu) \end{aligned} \quad (1.29)$$

$$\begin{aligned} \text{131} \quad m_W &= \frac{vg_2}{\sqrt{2}} \end{aligned} \quad (1.30)$$

$$\begin{aligned} \text{132} \quad m_Z &= \frac{v}{\sqrt{2}}\sqrt{g_1^2 + g_2^2} \end{aligned} \quad (1.31)$$

133 The Higgs field also generates a mass term for the Higgs boson and self-  
134 interactions for the Higgs boson.

## 135 1.7 Electroweak Theory

136  $SU(2)_L$  generates  $W^\pm, W^0$  gauge bosons, which would be massless if  $SU(2)_L$   
137 was a perfect symmetry. These bosons are massive as this symmetry is broken.

138 The mass eigenstates,  $Z$  and  $\gamma$  given by:

$$\begin{pmatrix} Z_\mu \\ A_\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & -\sin\theta_W \\ \sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} W_\mu^3 \\ B_\mu \end{pmatrix} \quad (1.32)$$

139 Here  $\theta_W$  is the Weinberg angle given by:

$$\cos\theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \frac{m_W}{m_Z} \quad (1.33)$$

## <sup>140</sup> 1.8 Quantum ChromoDynamics

<sup>141</sup> As mentioned earlier the Strong Force, which binds the proton together, is  
<sup>142</sup> mediated by gluons. Quantum Chromodynamics is the QFT which describes the  
<sup>143</sup> interactions of quarks and gluons via  $SU(3)_C$  symmetry. QCD contains features  
<sup>144</sup> not present in Electroweak Interactions due to  $SU(3)_C$  generators not commuting  
<sup>145</sup> (a.k.a.  $SU(3)_C$  is a non-abelian group). For example, in QCD there is color  
<sup>146</sup> confinement and asymptotic freedom due to the structure constants being non-  
<sup>147</sup> zero. Requiring  $SU(3)_C$  local gauge invariance implies:

$$\psi(x) \rightarrow \psi(x)' = \exp[i g_S \alpha(x) \cdot \hat{T}] \psi(x) \quad (1.34)$$

<sup>148</sup> where  $\alpha(x)$  is the local phase function,  $g_S$  is the strong coupling constant, and  
<sup>149</sup>  $\hat{T}$  are the eight generators of  $SU(3)$  (note  $\hat{T}^a = \frac{1}{2}\lambda^a a$ , where  $\lambda^a$  are the Gell-Mann  
<sup>150</sup> matrices). As the Gell-Mann matrices are 3x3, this means  $\psi$  has three degrees of  
<sup>151</sup> freedom under these  $SU(3)$  rotations. So we represent  $\psi$  under  $SU(3)$  rotations  
<sup>152</sup> as:

$$\psi = \begin{pmatrix} \psi_{red} \\ \psi_{green} \\ \psi_{blue} \end{pmatrix} \quad (1.35)$$

<sup>153</sup> Consequently, particle fields transforming under  $SU(3)$  rotations have three  
<sup>154</sup> components which physicists describe as color components (red, green, and blue).  
<sup>155</sup> A particle's corresponding antiparticle has the corresponding anticolor. This color  
<sup>156</sup> is the "charge" of QCD and is conserved under  $SU(3)$  rotations. Combining colors,  
<sup>157</sup> color neutral states (e.g. red and antired, or red, green and blue) may be created.  
<sup>158</sup> For the Free Dirac Lagrangian to remain invariant under  $SU(3)$  transformations,  
<sup>159</sup> we must again postulate a boson field that modifies the derivative. The gluon

<sub>160</sub> field tensor is given by ( $\alpha = 1, \dots, 8$ ):

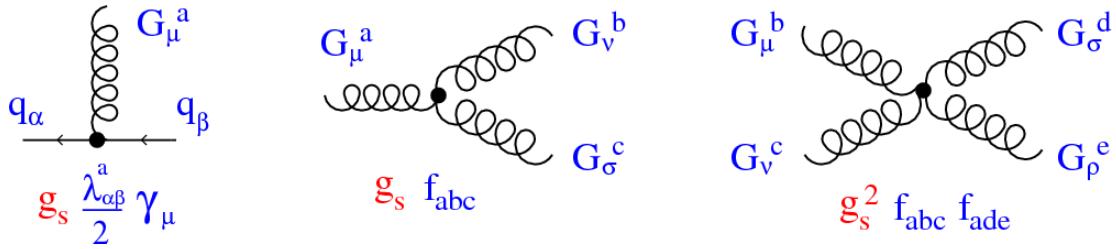
$$G_{\mu\nu}^k = \partial^\mu G_\alpha^\nu - \partial^\nu G_\alpha^\mu - g_S f^{\alpha\beta\gamma} G_\beta^\mu G_\gamma^\nu \quad (1.36)$$

<sub>161</sub> Here  $f^{\alpha\beta\gamma}$  are the structure constants of  $SU(3)$ . Combining all this gives the  
<sub>162</sub> QCD Lagrangian:

$$\mathcal{L}_{QCD} = \bar{\psi}_q i\gamma^\mu (D_\mu)_{ij} \psi^{qj} - m \bar{\psi}^{qi} \psi_{qi} - \frac{1}{4} G_{\mu\nu}^\alpha G^{\alpha\mu\nu} \quad (1.37)$$

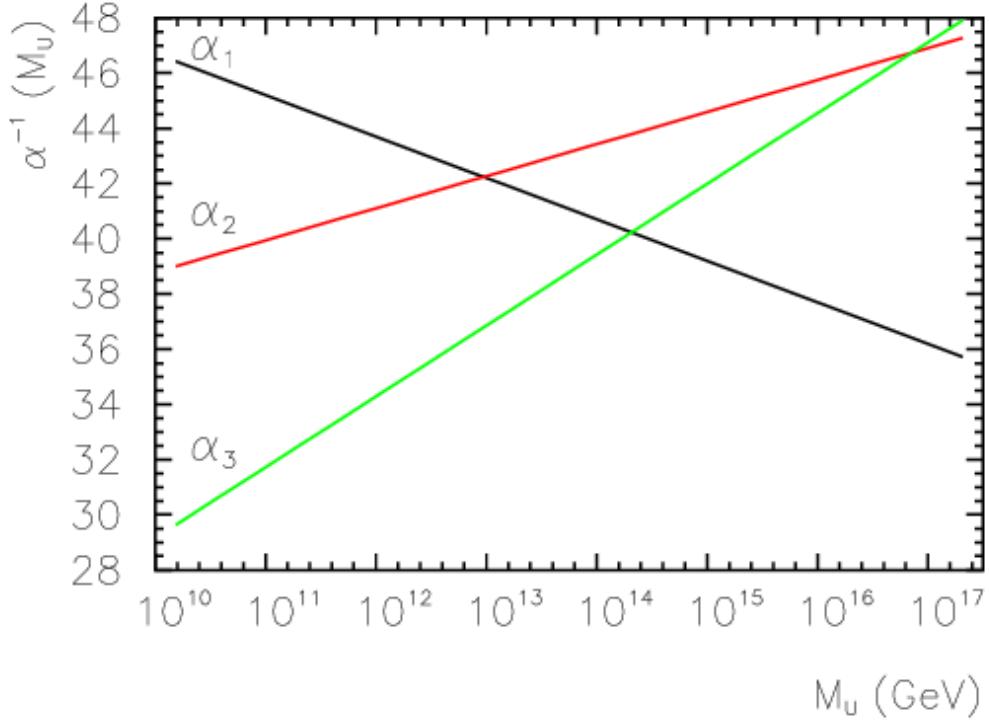
<sub>163</sub> Here  $i$  are the color indices, and  $q$  are the quark flavors. It is important to  
<sub>164</sub> note that quarks transform under the fundamental representation of  $SU(3)$ , while  
<sub>165</sub> gluons transform under the adjoint representation. This means quarks carry a  
<sub>166</sub> single color charge (red, green, blue, antired, antigreen, antiblue) and gluons carry  
<sub>167</sub> a color and anticolor charge.

<sub>168</sub> Figure 1.3 shows the three dominant QCD interactions. Since gluons carry  
<sub>169</sub> color charge, they interact with one another. This does not occur in QED, as  
<sub>170</sub> photons do not have electric charge and therefore do not interact with each other.  
<sub>171</sub> In QED, a bare electron's effective charge is largest closest to the electron and  
<sub>172</sub> decreases as a function of distance. This is because the QED vacuum fills with  
<sub>173</sub> particle antiparticle pairs spontaneously, which screen the charge of the bare elec-  
<sub>174</sub> tron. The larger the distance from the electron, the smaller the effective charge  
<sub>175</sub> and therefore the weaker the force. So for a pair of electrons, as the distance  
<sub>176</sub> between them increases the repulsive force decreases and they may be observed  
<sub>177</sub> separately.



**Figure 1.3:** This figure shows the three dominant QCD interactions. From Ref. [14]

178        As the distance from a quark increases it's effective color charge increases due  
 179        to the vacuum polarization in QCD. Color charge grows as the distance from  
 180        the source increases (a.k.a. color is anti-screened in QCD). In this way, strong  
 181        interactions become stronger at large distances (low momenta interactions). At  
 182        small distances (large momenta interactions) strong interactions are significantly  
 183        weaker and considered nearly free. This effect of referred to as asymptotic freedom.  
 184        At large distances, a quark's effective charge is large and the strong force is more  
 185        significant. This force becomes so strong that quarks form colorless bound states  
 186        instead of remaining free particles. This effect is known as color confinement.  
 187        This running of all SM fields is shown in Figure 1.4.



**Figure 1.4:** Strength of the U(1), SU(2), and SU(3) gauge couplings as a function of the energy scale of the interaction ( $Q$ ). From Ref. [10]

Commonly the change in a particle's effective charge under a given force is quantified with  $\beta(r) \equiv -\frac{de(r)}{d\ln r}$ , where  $e(r)$  is the effective charge of a given particle under a force. In QED this function is positive but in QCD this function is negative leading to confinement and asymptotic freedom. Moreover, one can calculate how the coupling ( $\alpha$ ) of a force varies with energies. (More deeply this amounts to incorporating renormalization and vacuum polarization in the boson propagators).

For QCD this is:

$$\alpha_S(x) = \frac{\alpha_S(\mu^2)}{1 + \beta_0 \alpha_S(\mu^2) \ln(Q^2/\mu^2)} \quad (1.38)$$

195

$$\beta_0 = \frac{11N_c - 2n_f}{12\pi} \quad (1.39)$$

196 where  $Q$  is the momentum of the the force is probed at,  $\mu^2$  is the renormalization scale.

198 As stated previously, quarks and gluons have not been observed in isolation.  
 199 Instead they form bound colorless states. Hadronization is the process by which  
 200 quarks and gluons form hadrons. The process of hadronization is still an active  
 201 area of research. One qualitative description is show in Figure [Natasha add  
 202 figure]. In this figure, as two quarks separate the color field between them is  
 203 restricted to a tube with energy density of 1GeV/fm. As they separate further,  
 204 the energy in the color field increases, until there is enough energy to produce  
 205  $q\bar{q}$  pairs, which breaks the color field. This process repeats until quarks and  
 206 antiquarks have low enough energy to form colorless hadrons. The resulting spray  
 207 of hadrons is called a jet.

208 Since quarks and gluons carry different color charges, their respective jets have  
 209 different properties. As quarks carry only a single color charge (vs. gluons which  
 210 have color and anticolor charge), so their jets have less constituent particles. More  
 211 precisely, the Altarelli-Parisi splitting functions [3] contain a factor  $C_A$  for gluon  
 212 radiation off a gluon and  $C_F$  for gluon radiation off a quark ( $C_A/C_F = 9/4$ ). These  
 213 color factors are the prefactor in the Feynman diagrams for these processes [1],  
 214 which leads to gluon jets having more constituents and therefore more tracks than  
 215 quark jets. Gluon jets also tend to have a larger radius with lower momentum  
 216 constituents than quarks. There are many novel techniques to distinguish quarks  
 217 from gluons. For this study the number of charged particles will be focused on.

<sup>218</sup> **Chapter 2**

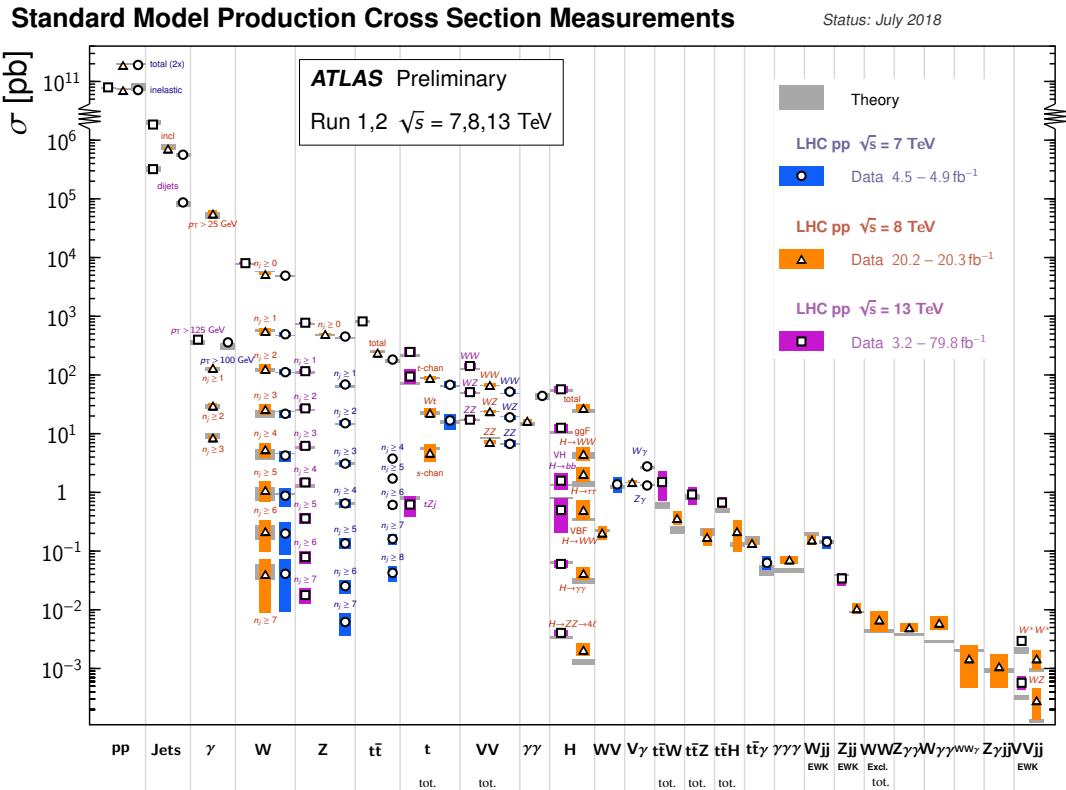
<sup>219</sup> **Standard Model Successes and  
Limitations**

<sup>221</sup> The Standard Model has consistently described much of reality to an extreme  
<sup>222</sup> degree of accuracy. It has predicted cross sections for strong and electroweak pro-  
<sup>223</sup> cesses that span over ten order of magnitude correctly [see Fig 2.1] and contains  
<sup>224</sup> no known logical inconsistencies. Despite the strength and reality of the Stan-  
<sup>225</sup> dard Model, it still fails to describe aspects of reality and suffers from aesthetic  
<sup>226</sup> issues. To date, dark matter and energy comprise 95% of the universe, but are  
<sup>227</sup> not accounted for in the SM. Additionally, neutrinos are known to have mass but  
<sup>228</sup> are massless in the SM. There are mechanisms for introducing massive neutrinos  
<sup>229</sup> in the SM, but these mechanisms create hierarchy problems.

<sup>230</sup> Possibly the most significant aesthetic issue is the hierarchy between the elec-  
<sup>231</sup> troweak and Planck scales. The electroweak scale is the scale of electroweak  
<sup>232</sup> symmetry breaking. The Planck scale is the scale where the gravitational force  
<sup>233</sup> is comparable in strength to the other forces. The Planck scale is where the SM  
<sup>234</sup> breaks down, as there is not an experimentally verified theory of quantum gravity,  
<sup>235</sup> and at this scale gravity cannot be ignored (like it can at the electro-weak scale).

<sup>236</sup> These scales differ by  $\sim 30$  orders of magnitude. Understanding the difference  
<sup>237</sup> in these energy scales, may help explain the weakness of gravity at electroweak  
<sup>238</sup> scales, and possibly a QFT for gravity. (NB: This hierarchy can also be framed in  
<sup>239</sup> terms of the corrections to the Higgs mass, which depend on the UV cutoff scale -  
<sup>240</sup> where the SM is suppose to break, which is taken at the Planck scale. This leads  
<sup>241</sup> the quantum corrections to the Higgs mass to force the Higgs mass to  $10^{18}$  TeV.)

<sup>242</sup> These stark contrasts in scales may indicate that a more fundamental theory  
<sup>243</sup> exists. It is hoped that such a theory would explain and motivate some of the ad-  
<sup>244</sup> hoc features of the SM. In particular, there currently are no experimentally verified  
<sup>245</sup> explanations of why there are three generations of fermions, the values of the 19  
<sup>246</sup> SM parameters (6 quark masses, 3 charged lepton masses, 3 gauge couplings,  
<sup>247</sup> Higgs parameters  $(\mu^2, \lambda)$ ), the structure of the fermion representations, etc.



**Figure 2.1:** A comparison of cross section measurements at  $\sqrt{s} = 7,8,13$  TeV from ATLAS compared to theoretical measurements. From Ref. [5]

<sup>248</sup> **Chapter 3**

<sup>249</sup> **New Physics Models with**  
<sup>250</sup> **Diboson Resonances**

<sup>251</sup> **3.1 Randall Sundrum Bulk Model**

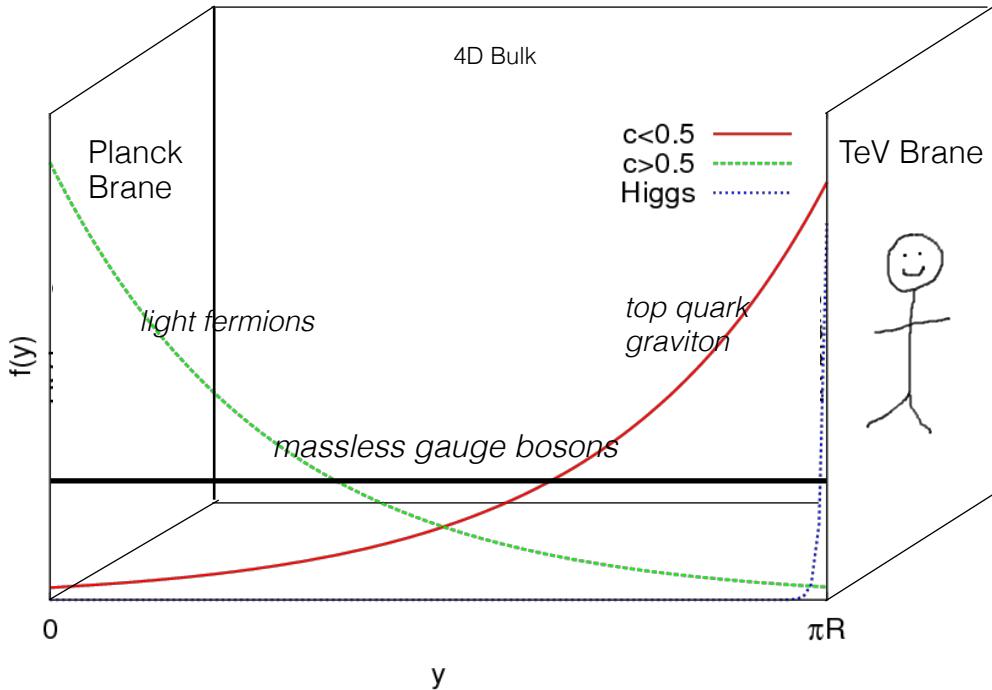
<sup>252</sup> The electroweak-planck hierarchy may be explained by the existence of extra  
<sup>253</sup> dimensions, like the 5D Randall Sundrum Bulk Model ([15], [2]). In this model,  
<sup>254</sup> there is one extra warped spatial dimension,  $y$ , with a metric:

$$ds^2 = e^{-2k|y|} \eta_{\mu\nu} dx^\mu dx^\nu + dy^2 \quad (3.1)$$

<sup>255</sup> where  $e^{-k|y|}$  is the warp factor of the extra dimension, which is compactified on  
<sup>256</sup> a  $S^1/Z_2$  orbifold (a.k.a. a circle where  $y \rightarrow -y$ ). This can be visualized as every  
<sup>257</sup> point in space time having a line extending from it a distance  $L$ , representing  
<sup>258</sup> this fifth dimension. At the end of this line is the Planck brane. This fourth  
<sup>259</sup> spatial dimension separates two 4-D branes: Planck brane and TeV brane. We  
<sup>260</sup> live on the TeV brane, as shown in Figure 3.1. The Higgs field (and to a lesser  
<sup>261</sup> degree the top quark and graviton fields) is localized near the TeV Brane, while

262 the light fermion fields are localized more near the Planck brane. Fundamental  
 263 parameters are set on the Planck brane. The warp factor may be scaled away from  
 264 all dimensionless SM terms by field redefinitions. However, the only dimensionful  
 265 parameter,  $m_H^2 = v^2$  is rescaled by  $\tilde{v} \sim e^{-kL} M_{Pl} \sim 1\text{TeV}$  for  $kL \sim 35$ , explaining  
 266 why gravity is so weak on the TeV brane. Also, by localizing the light fermion  
 267 fields near the Planck brane and top and graviton fields near the TeV brane, the  
 268 light quarks will have smaller masses.

269 The two free parameters of this theory are  $M_{Pl}$  and  $k$ . Based on this RS Bulk  
 270 model, all SM particles should have Kaluza-Klein (KK) excitations. In particular,  
 271 the graviton would have KK excitations that prefer to decay to WW or ZZ, which  
 272 is why this analysis searches for RS Gravitons.



**Figure 3.1:** Cartoon of RS Bulk Model

## <sup>273</sup> 3.2 Extended Scalar Sector

<sup>274</sup> A further striking asymmetry of the SM is the simplicity of the scalar sector in  
<sup>275</sup> comparison to the boson and fermion sectors. To date, the scalar sector has only  
<sup>276</sup> one member, the Higgs boson. Therefore, it is natural to posit an extension to the  
<sup>277</sup> scalar sector. From a theoretical standpoint this could also help generate baryon  
<sup>278</sup> asymmetry through additional sources of CP violation. This analysis searches for  
<sup>279</sup> a simple extension to the scalar sector as proposed in Ref. [16]. The extended  
<sup>280</sup> scalar sector includes a real Higgs singlet ( $S$ ) and complex  $SU(2)_L$  doublet ( $\Phi$ )  
<sup>281</sup> (the SM Higgs), where mass eigenstates are mixtures of the fields.  $S$  has a vev of  
<sup>282</sup>  $v$  and  $\Phi$  has a vev of  $x$ . This then gives a Lagrangian of:

$$\mathcal{L} \supset (D^\mu \Phi)^\dagger D_\mu \Phi + \partial^\mu S \partial_\mu S - m^2 \Phi^\dagger \Phi - \mu^2 S^2 + \lambda_1 (\Phi^\dagger \Phi)^2 + \lambda_2 S^4 + \lambda_3 \Phi^\dagger \Phi S^2 \quad (3.2)$$

<sup>283</sup> The mass eigenstates of the scalar sector are then mixtures of  $S$  and  $\Phi$  and  
<sup>284</sup> the free parameters of the theory are  $m_H$ ,  $\sin \alpha$ , and  $\tan \beta = v/x$ . The fields are  
<sup>285</sup> then given by:

$$\Phi \equiv \begin{pmatrix} 0 \\ \frac{\tilde{h}+v}{\sqrt{2}} \end{pmatrix} \quad (3.3)$$

<sup>286</sup>

$$S \equiv \frac{h' + x}{\sqrt{2}} \quad (3.4)$$

<sup>287</sup> Diagonalizing the mass matrix leads to the mass eigenstates  $h$  (discovered  
<sup>288</sup> Higgs boson) and  $H$  (the physical particles):

$$\begin{pmatrix} h \\ H \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \quad (3.5)$$

289 This suppressed h and H production and SM H couplings:

$$BR_{H \rightarrow SM} = \sin^2 \alpha \times \frac{\Gamma_{SM, H \rightarrow SM}}{\Gamma_{tot}} \quad (3.6)$$

290 Moreover, in the case that  $m_H > m_h$ ,  $H \rightarrow hh$  is possible. This further suppresses  
291  $H \rightarrow VV/ff$ . This search is most sensitive to  $H \rightarrow WW$ .

### 292 3.3 Simple Standard Model Extensions

293 The RS Bulk model is motivated by resolving the hierarchy problem. Extending the Scalar sector is a natural space to look for new physics due to the complexity of fermion and boson groups. There are many other interesting and well motivated frameworks, but there is a lack of completely predictive models, due to model flexibility (free parameters). Therefore it is hard for experimentalists to know which theories to search for in data. However, as seen in [13], a "Simplified Model" approach may be taken. In the search for reasonably narrow width particles, as in this search, the search is not sensitive to all the details and free parameters of the theory. Generally such searches are only sensitive to the resonance mass and its interactions. Therefore, a theory's Lagrangian may be reduced to only retain this information (mass parameters and couplings). Experimental results using this framework may then be reinterpreted in a given theory.

305 In the simplified approach, the new resonance searched for is represented as a real vector field in the adjoint representation of  $SU(2)_L$  with vanishing hypercharge. This results in one neutral and two charged bosons. Defined as:

$$V^\pm = \frac{V_\mu^1 \mp iV_\mu^2}{\sqrt{2}} \quad (3.7)$$

308  $V_\mu^0 = V_\mu^3 \quad (3.8)$

309        The SM Lagrangian is then augmented with the additional terms:

$$\mathcal{L} \supset -\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]}_a + \frac{m_V^2}{2}V_\mu^a V^{a\mu} + ig_V c_H V_\mu^a H^\dagger \tau^a \overset{\leftrightarrow}{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a J_F^{\mu a} \quad (3.9)$$

310        In order the terms represent: the kinetic,  $V$  mass, Higgs- $V$  interaction, and  
311         $V$ -left-handed fermion interaction terms. Phenomenologically the three physical  
312        particles this predicts are degenerate, where  $V$  couples most strongly to  $VV$ , via  
313        the  $g_V$  coupling factor. The dominant production modes are DY and VBF.

314        Two versions of HVT are considered, Model A and B. Model A is a weakly  
315        coupled model where  $g_V \sim 1$ , like the extended gauge symmetry discussed in Ref  
316        . [16]. Model B is a strongly coupled model, where  $1 < g_V < 4\pi$ . The width  
317        of the resonance grows with  $g_V$  so for this narrow resonance search only  $g_V$  is  
318        chosen to be less than 6 (so  $\Gamma/M < 10\%$ ). More precisely, the coupling of these  
319        resonances to fermions scales as  $g_f = g^2 c_F/g_V$ , where  $g$  is the SM  $SU(2)_L$  gauge  
320        coupling and  $c_F$  is the free parameter (expected to be of order 1 for Model A and  
321        B). This then means that for Model B the coupling to fermions is more suppressed  
322        than for Model A, leading to a smaller DY production rate and BR to fermionic  
323        final states. The coupling of  $V$  to SM bosons scales as  $g_H = g_V c_H$ , where  $c_H$  is  
324        a free parameter on the order of 1 for Model A and B. So for small values of  $g_V$   
325        (i.e. Model A - weakly coupled theories) the BR to gauge bosons is smaller than  
326        for Model B. So weakly coupled vectors have large production cross sections and  
327        decay prominently to leptons or jets, while strongly coupled vectors are produced  
328        less and decay predominantly to gauge bosons.

329        Vectors in Model A and B are generally produced via quark-anti-quark annihi-  
330        lation. The more rare production via vector-boson-fusion is considered by setting  
331         $g_H = 1$  and  $g_F = 0$ . In Model B diboson final states are enhanced as stated  
332        previously due to  $g_H$  and moreover the BR to  $WZ$ ,  $WH$ ,  $WW$ , and  $ZH$  are the

<sup>333</sup> same.

<sup>334</sup> In summary,  $V$  couples most strongly to left-handed fermions and  $VV$  dependent on  $g_V$ .

336

## Part II

337

## Experimental Setup

<sup>338</sup> **Chapter 4**

<sup>339</sup> **LHC**

<sup>340</sup> The Large Hadron Collider (LHC) is the highest-energy particle collider in the  
<sup>341</sup> world. It was designed to expand the frontier of high energy particle collisions in  
<sup>342</sup> energy and luminosity. This enables LHC experiments to test the Standard Model  
<sup>343</sup> and search for new physics at higher energies than tested with previous colliders.  
<sup>344</sup> Collisions at higher energies not only produce more massive particles but also  
<sup>345</sup> more weakly interacting particles. Fig. 4.1 shows production cross sections for  
<sup>346</sup> various processes at hadron colliders. The rate for electroweak physics processes  
<sup>347</sup> including  $W$  and  $Z$  scale with the center-of-momentum energy,  $\sqrt{s}$ .

<sup>348</sup> The LHC consists of a 26.7 km (17 miles) ring, approximately 100 m un-  
<sup>349</sup> derground, outside Geneva, Switzerland. Counter-circulating proton (and occa-  
<sup>350</sup> sionally heavy ions) beams collide inside four experiments along the beam line:  
<sup>351</sup> ATLAS, CMS, LHCb, ALICE. ATLAS and CMS are general purpose detectors de-  
<sup>352</sup> signed to explore the high energy frontier. LHCb is designed to study the physics  
<sup>353</sup> of  $b$ -quarks. ALICE specializes in studying heavy ion collisions.

## proton - (anti)proton cross sections

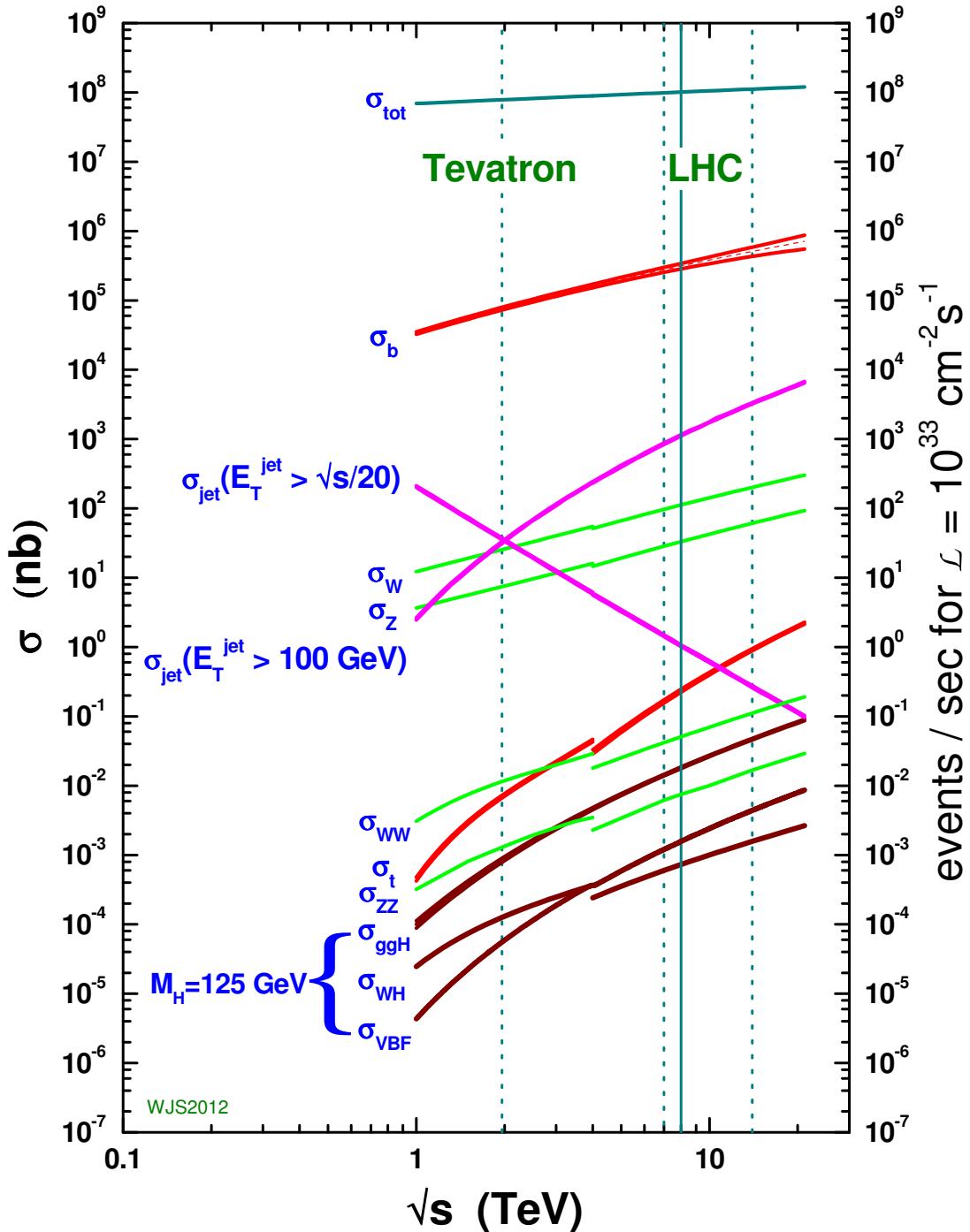
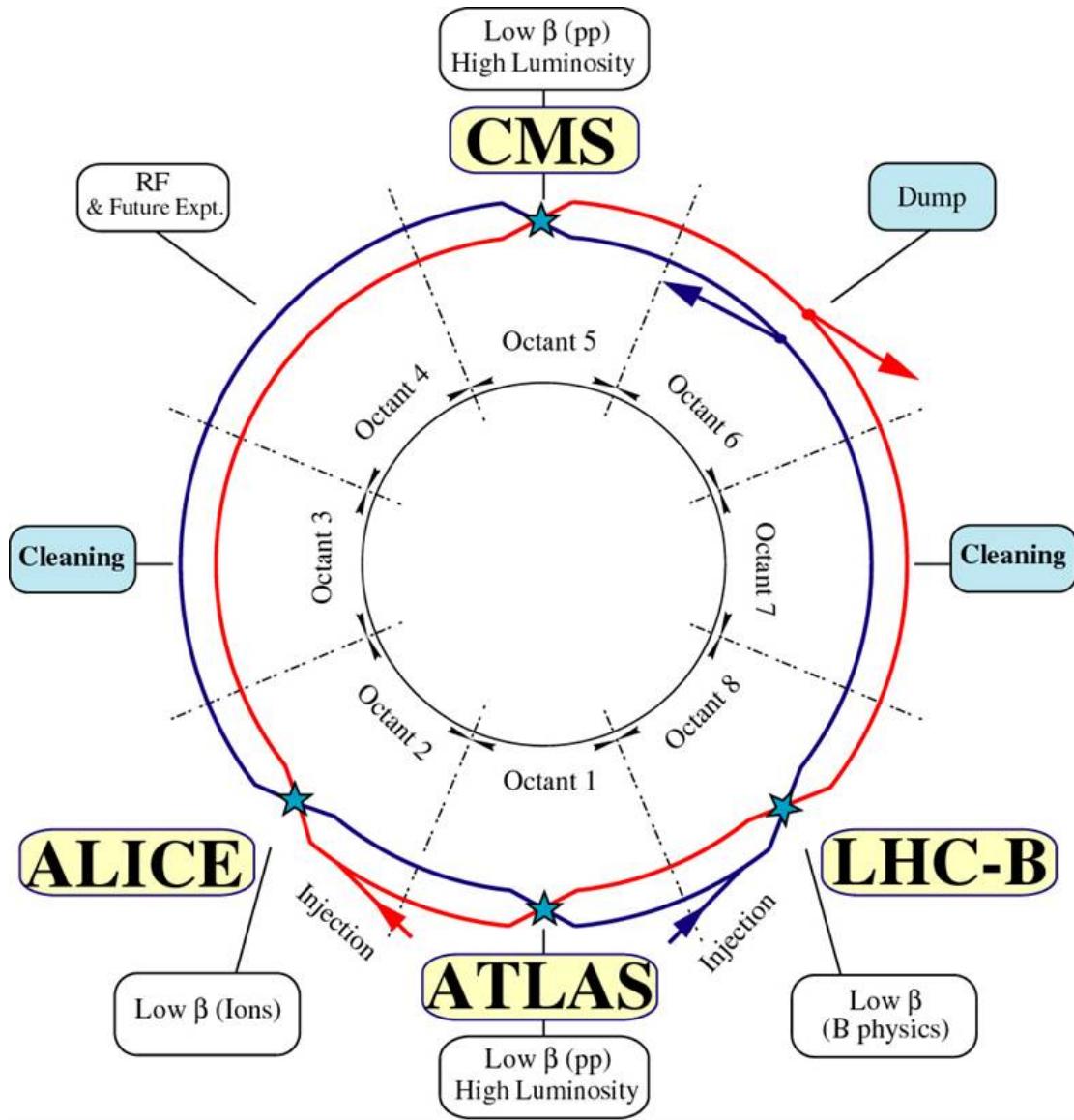


Figure 4.1: Scaling of cross sections with  $\sqrt{s}$ . Natasha: write more here

354        The first proton beams circulated in September, 2008. Nine days later an elec-  
355        trical fault lead to mechanical damage and liquid helium leaks in the collider. This  
356        incident delayed further operation until November 2009, when the LHC became  
357        the world's highest energy particle collider, at 1.18TeV per beam. This first oper-  
358        ational run continued until 2013, reaching 7 and 8 TeV collision energies. During  
359        this run a particle with properties consistent with the Standard Model Higgs bo-  
360        son was discovered. The next run began after a two year shutdown after upgrades  
361        to the LHC and experiments. This run lasted from 2013 to 2018 reaching 13 TeV  
362        collision energies. This analysis uses data from the second operational run.

## 363        **4.1 LHC Layout and Design**

364        The layout of the LHC is shown in Figure 4.2. The red and blue lines in the  
365        figure represent the counter-circulating proton beams. The LHC is divided into  
366        eight octants. Octant 4 contains the RF cavities that accelerate the protons and  
367        octant 6 contains the beam dump system. Octants 3 and 7 house the collimation  
368        systems for beam cleaning. The beams collide inside the four aforementioned  
369        experiments. Each octant contains a curved and straight section. The LHC  
370        magnets are built with NbTi superconductors cooled with super-fluid Helium to  
371        2K, creating a 8.3T magnetic field to bend the proton beams.

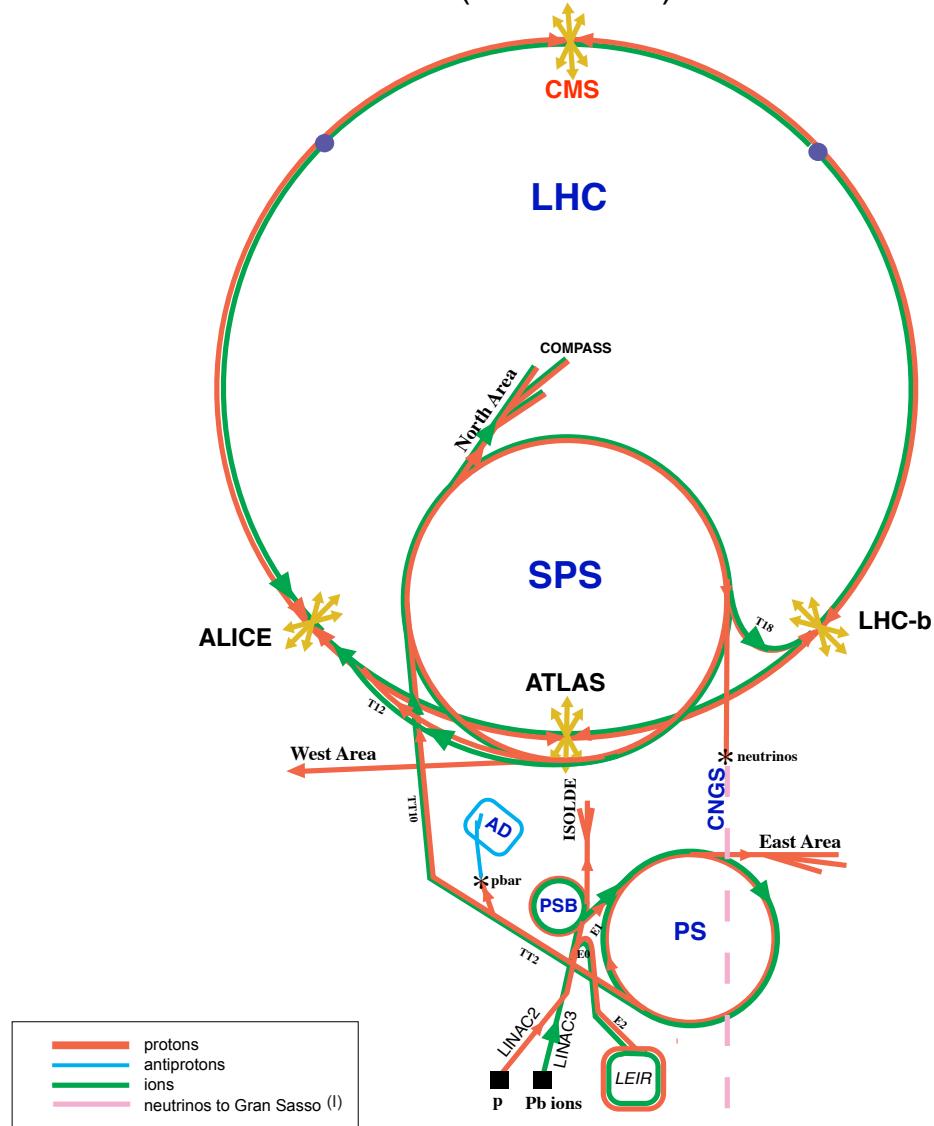


**Figure 4.2:** LHC Layout. Natasha write more

372        Four sequential particle accelerators are used to accelerate protons from rest  
 373        as shown in Figure 4.3. First, Hydrogen gas is ionized to produce protons which  
 374        are then accelerated to 50 MeV using Linac 2, a linear accelerator. The result-  
 375        ing proton beam is then passed to three circular particle accelerators: Proton  
 376        Synchrotron Booster, Proton Synchrotron, and Super Proton Synchrotron (SPS),

<sup>377</sup> accelerating protons to 1.4, 25, and 450 GeV, respectively. Once the protons exit  
<sup>378</sup> the SPS, they are injected into the LHC at octant 2 and 8. Each proton bunch  
<sup>379</sup> contains  $\sim 10^{11}$  protons. The spacing between bunches is 25 ns, which means  
<sup>380</sup> each beam contains 3564 bunches. However, some bunches are left empty due  
<sup>381</sup> to injection and safety requirements, yielding 2808 bunches per beam. Once the  
<sup>382</sup> proton beams are injected they are accelerated to 13 TeV.

## CERN Accelerators (not to scale)



LHC: Large Hadron Collider

SPS: Super Proton Synchrotron

AD: Antiproton Decelerator

ISOLDE: Isotope Separator OnLine DEvice

PSB: Proton Synchrotron Booster

PS: Proton Synchrotron

LINAC: LINear ACcelerator

LEIR: Low Energy Ion Ring

CNGS: Cern Neutrinos to Gran Sasso

Rudolf LEY, PS Division, CERN, 02.09.96  
Revised and adapted by Antonella Del Rosso, ETT Div.,  
in collaboration with B. Desforges, SL Div., and  
D. Manglunki, PS Div. CERN, 23.05.01

Figure 4.3: LHC Accelerator. Natasha write more

383        As many new physics models predict cross-sections below the weak scale it was  
384        important to design the LHC to be capable of collecting enough data, by running  
385        in high luminosity conditions. The machine luminosity depends only on beam  
386        parameters:

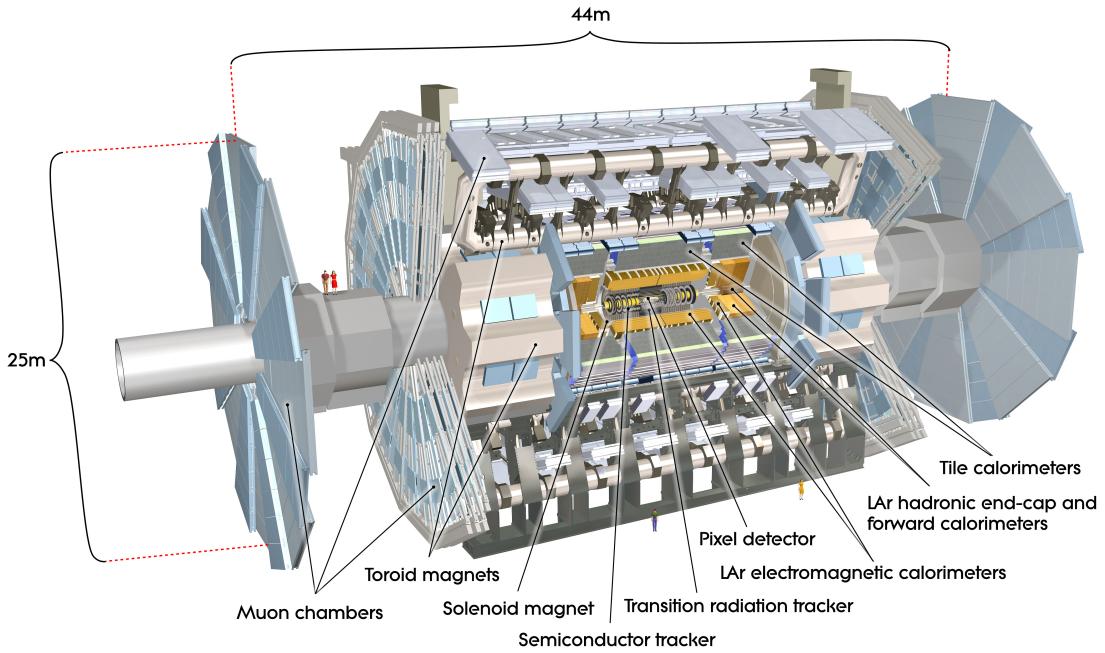
$$L = \frac{N_p^2 f}{4\epsilon\beta^*} F \quad (4.1)$$

387        where  $N_p$  is the number of protons per bunch,  $f$  is the bunch crossing frequency,  
388         $\epsilon$  is the transverse beam emittance,  $\beta^*$  is the amplitude function at the collision  
389        point, and  $F$  is the geometric luminosity reduction factor due to the beams crossing  
390        at an angle (rather than head-on).

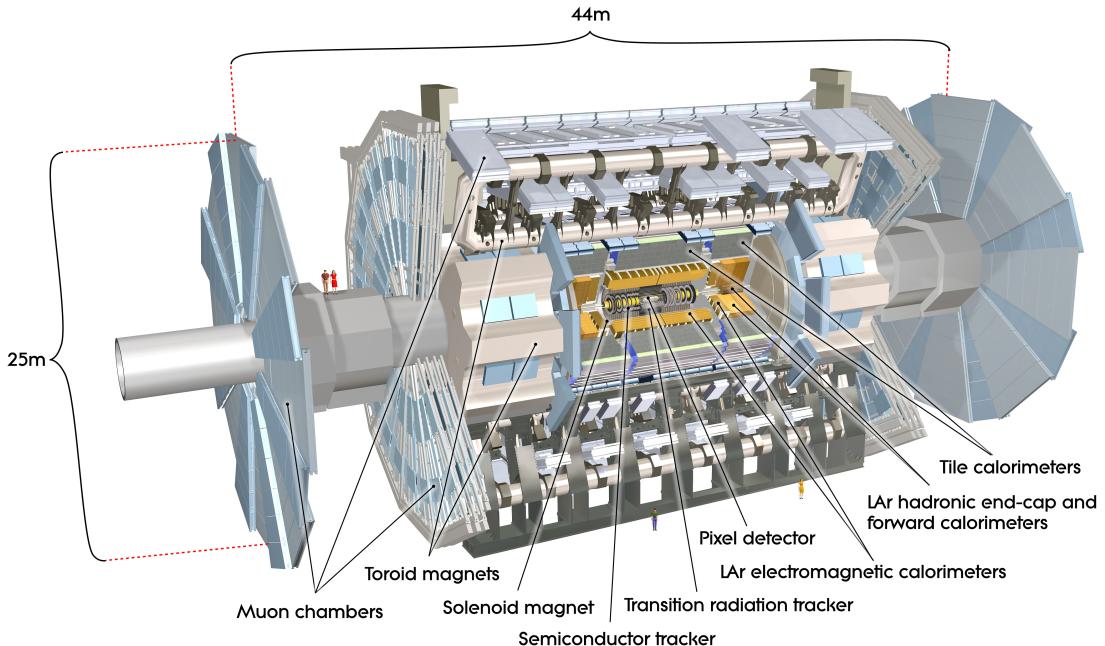
<sup>391</sup> **Chapter 5**

<sup>392</sup> **The ATLAS Detector**

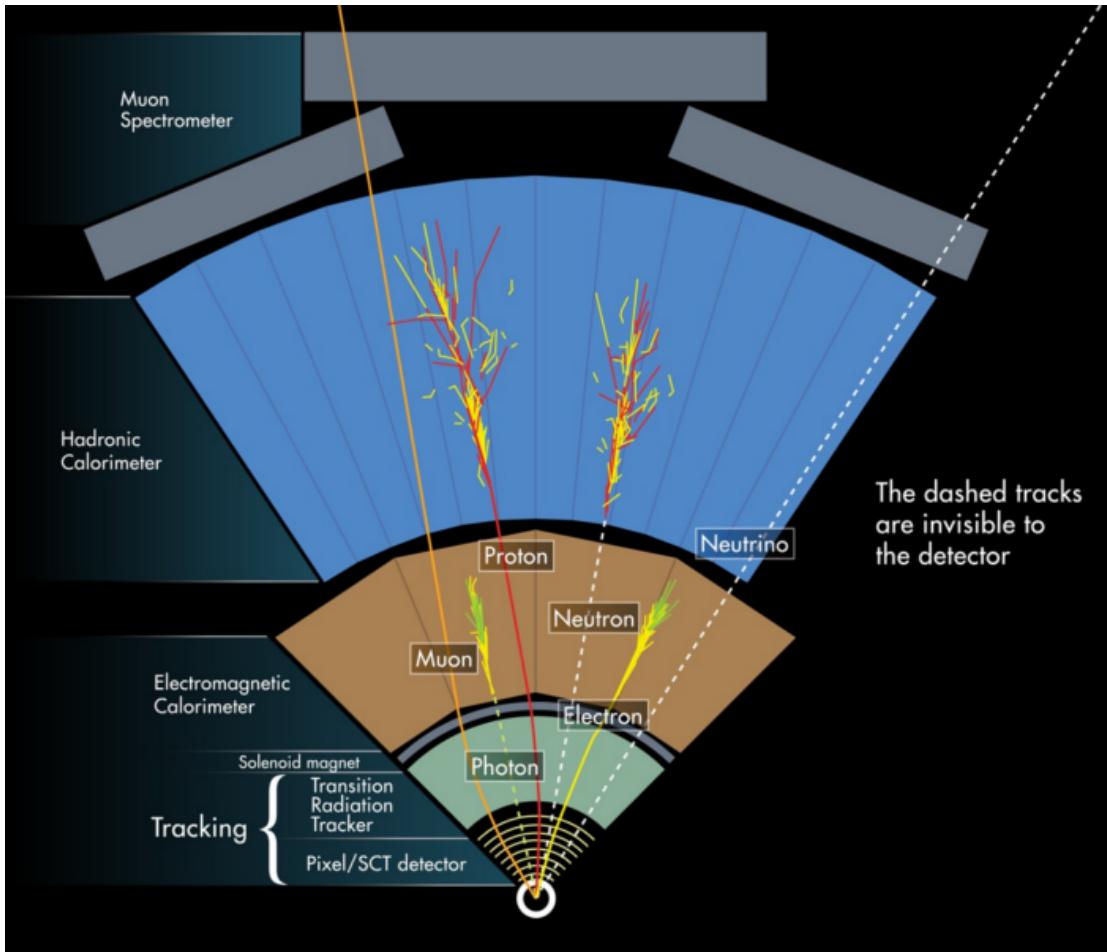
<sup>393</sup>      The ATLAS detector measures the position, momentum and energy of parti-  
<sup>394</sup> cles produced in the proton collisions by using magnetic fields, silicon detectors,  
<sup>395</sup> sampling calorimeters, and gaseous wire detectors. It is located approximately  
<sup>396</sup> 100 m underground at Point-1 around the LHC beam line and weighs 7000 metric  
<sup>397</sup> tons. The detector is 46 m long, 25 m high, 25 m wide as shown in Figure 5.2.  
<sup>398</sup> The detector can be divided into three subsystems: the Inner Detector (ID), the  
<sup>399</sup> Calorimeters, and the Muon Spectrometer (MS). Figure 5.3 shows an overview of  
<sup>400</sup> how different particles interact in the detector.



**Figure 5.1:** Big picture layout of ATLAS detector. Natasha: write more



**Figure 5.2:** Big picture layout of ATLAS detector. Natasha: write more



**Figure 5.3:** A simplified schematic of how different particles interact and are detected within ATLAS.

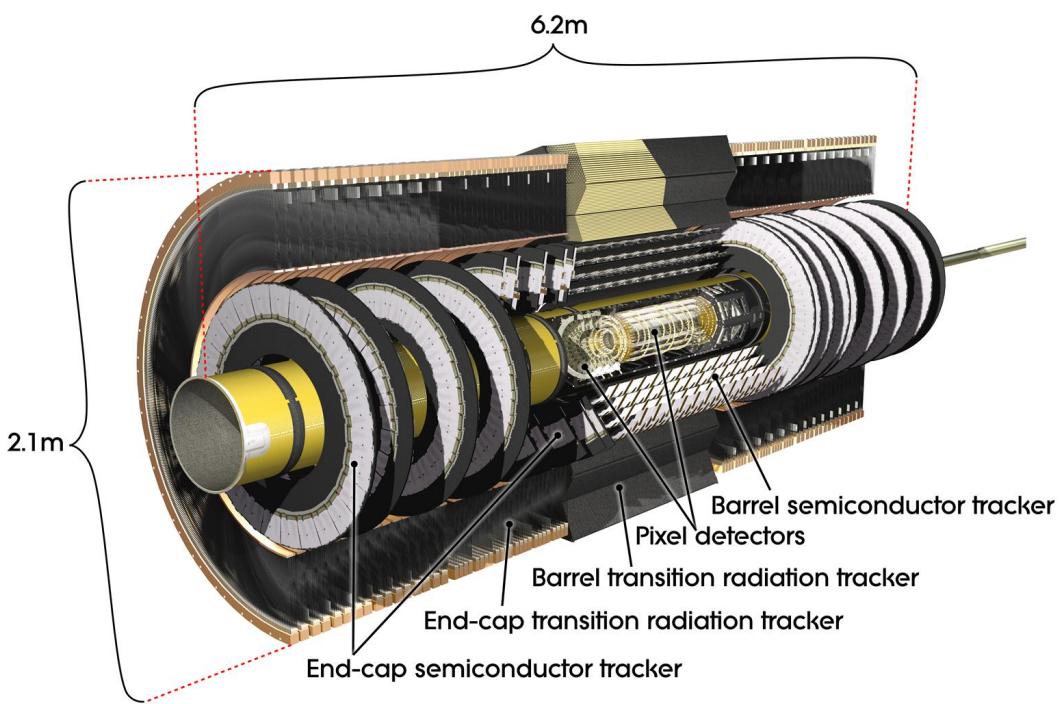
## 401 5.1 Coordinate System

402     The trajectory of particles within ATLAS is measured relative to the nominal  
 403 interaction point. The  $z$ -axis points along the beam line, such that when the  
 404 LHC is viewed from above, the counter-clockwise circulating beam points along  
 405 the positive- $z$  direction. The  $x - y$  plane is transverse to the beam line, with the  
 406 positive  $x$ -axis pointing towards the center of the LHC ring. The positive  $y$ -axis  
 407 points vertically upward. The azimuthal angle,  $\phi$ , is the angular distance about

408 the  $z$ -axis, with  $\phi = 0$  along the  $x$ -axis. The polar angle from the  $z$ -axis is denoted  
409 as  $\theta$ . However, this quantity is not Lorentz invariant, like rapidity,  $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$ ,  
410 where  $E$  is the energy of the particle considered, and  $p_z$ , is it's momentum along  
411 the  $z$ -axis. Pseudo-rapidity is preferred as  $\Delta\eta$  is invariant under boosts along  $z$   
412 and particle production is approximately invariant under  $\eta$ . For massless particles,  
413 rapidity and a related quantity, pseudorapidity, are the identical. The pseudora-  
414 pidity is defined as:  $\eta = -\ln \tan(\frac{\theta}{2})$ . This quantity is preferred as it is purely a  
415 geometric quantity, independent of particle energy. Angular separation between  
416 particles in ATLAS are given by  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ . The distance from the  
417 beamline is given by  $r = \sqrt{x^2 + y^2}$

## 418 5.2 Inner Detector

419 The Inner Detector (ID) was designed to identify and reconstruct vertices,  
420 distinguish pions from electrons, and measure the momentum of charged particles.  
421 The ID uses three different technologies for particle reconstruction: the Pixel  
422 Detector, Semiconductor Tracker (SCT), and the Transition Radiation Tracker  
423 (TRT), shown in Figure 5.4 and 5.5. The entire ID is immersed in a 2T solenoidal  
424 magnetic field parallel to the  $+z$ -axis, causing charged particles to bend in the  
425 transverse-plane, allowing particle momentum measurements.



**Figure 5.4:** Layout of ATLAS Inner Detector

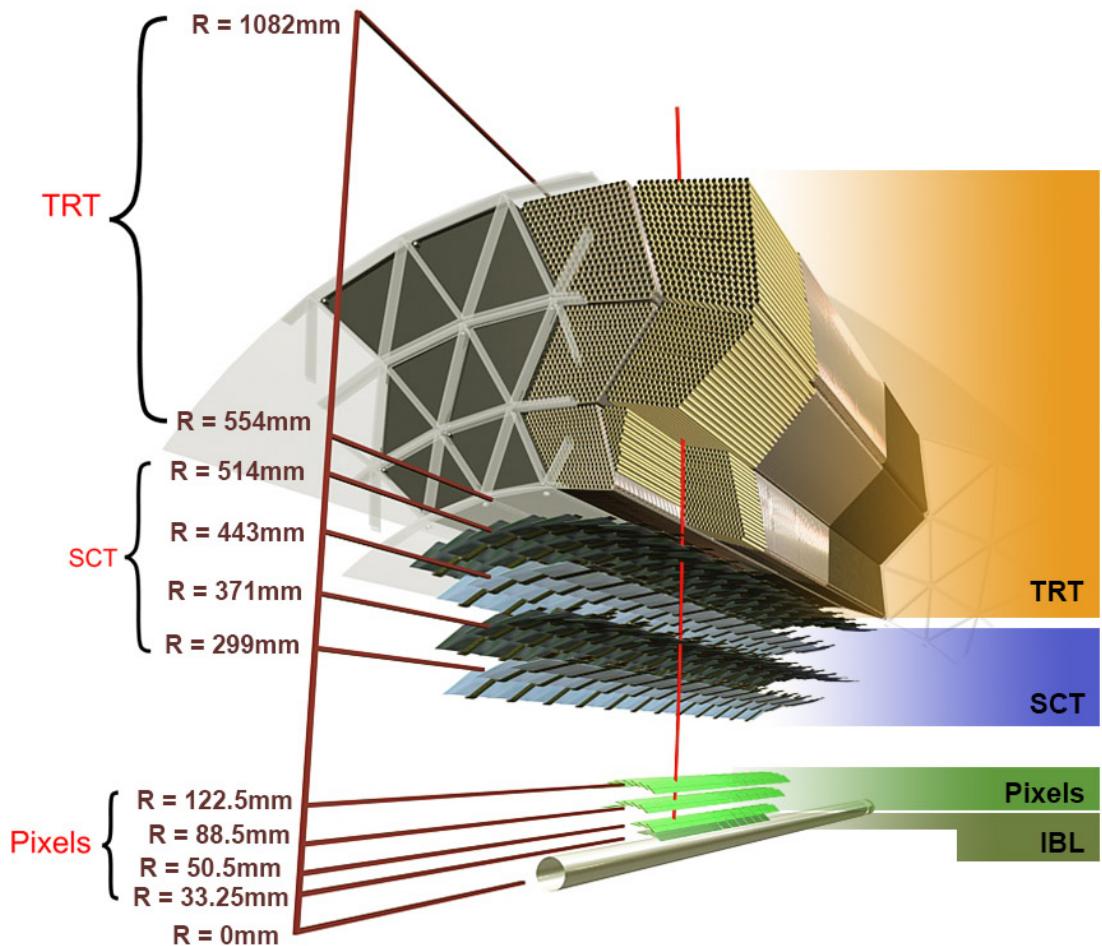


Figure 5.5: Layout of ATLAS ID Barrel System.

426 **5.2.1 Pixel Detector**

427 The pixel detector consists of four barrel layers between  $r = 32.7$  and  $122.5$   
428 mm, extending to  $|z| = 400.5$  mm. The remaining detectors are arranged in bar-  
429 rels and forward and backward rings. The innermost pixel barrel, the Insertable  
430 b-Layer (IBL), only extends to  $|z| = 332$  mm. The pixel detectors closer to the  
431 beam line (larger  $\eta$  values) consists of six parallel cylindrical rings of pixel de-  
432 tectors transverse to the beam line. The entire pixel detector consists of 1744  
433 identical pixel sensors each with 46080 readout channels, totaling about 80 mil-  
434 lion individual pixels. Most of the pixel sensors are  $50 \times 400 \mu\text{m}^2$ . Each pixel has  
435 a position resolution of  $14 \mu\text{m}$  in  $\phi$  and  $115 \mu\text{m}$  in the  $z$  direction.

436 **5.2.2 Semiconductor Tracker**

437 The SCT is located outside the pixel detector and has the same barrel and  
438 endcap geometry as the pixel detector. SCT sensors are  $80 \mu\text{m} \times 12$  cm with  
439 a  $80 \mu\text{m}$  strip pitch. In the barrel the strips are parallel to the  $z$ -axis and are  
440 segmented in  $\phi$ . In the endcaps, the strips extend radially. Sensors are grouped in  
441 modules containing two layers of strips rotated 40 mrad with respect to each other.  
442 This offset allows for the two-dimensional position of a track to be determined by  
443 identifying the crossing point of the strips that registered a hit. SCT modules  
444 measure tracks with an accuracy of  $17 \mu\text{m}$  in  $r - \phi$  and  $580 \mu\text{m}$  in  $z(r)$  in the  
445 barrel (end-cap) region.

446 **5.2.3 Transition Radiation Tracker**

447 The transition radiation tracker (TRT), enveloping the SCT, is a gaseous  
448 straw-tube tracker mainly used for electron/pion track separation. Each straw  
449 is 4 mm in diameter and filled with a Xe- $\text{CO}_2$ - $\text{O}_2$  gas mixture. An anode wire at

450 the center of the straw is held at ground potential, while the walls of the straw  
451 are kept at -1.4kV. When a charged particle passing through the TRT ionizes the  
452 gaseous mixture, the resulting ions form an avalanche on the anode wire with a  
453 gain of  $\sim 10^4$ . The signal from the anode wire is then digitized and discriminated.  
454 Signals passing a low threshold cutoff are used to distinguish noise from tracks.  
455 Signals passing a high threshold cutoff are sensitive to transition radiation (TR).  
456 TR photons are emitted when charged particles pass between materials with dif-  
457 ferent dielectric constants. The probability that a charged particle with energy  $E$   
458 and mass  $m$  passing between two materials emits a TR photon in the keV range  
459 is proportional to  $\gamma = E/m$ . In the TRT staws these often then convert via the  
460 photoelectric effect, causing a large avalanche triggering the high-threshold. Since  
461 electrons have a smaller mass than pions, electron tracks are more likely to trig-  
462 ger the high threshold. This then provides discrimination between electrons and  
463 charged hadrons.

464 The barrel region of the TRT extends from  $r = 563\text{-}1066$  mm and  $|z| < 712$   
465 mm. Barrel Straws are 144 cm long (divided  $\sim \eta \approx 0$ ) and orientated parallel to  
466 the beam direction. End-cap straws extend radially and are 37 cm long. There  
467 are 53,544 straws in the barrel and 160,000 straws in the end-caps. Radiator mats  
468 of polypropylene/polyethylene fibers in the barrel are aligned perpendicular to the  
469 barrel straws (with holes for the straws to pass through). In the end-cap region,  
470 radiator foils are layered between the radial TRT straws.

471 The arrival time of the signal pulse is sensitive to the distance between the  
472 charged particle track and the anode wire and allows for a hit resolution of  $130\mu\text{m}$ .  
473 The TRT extends to  $|\eta| = 2.0$  and provides about 36 hits per track.

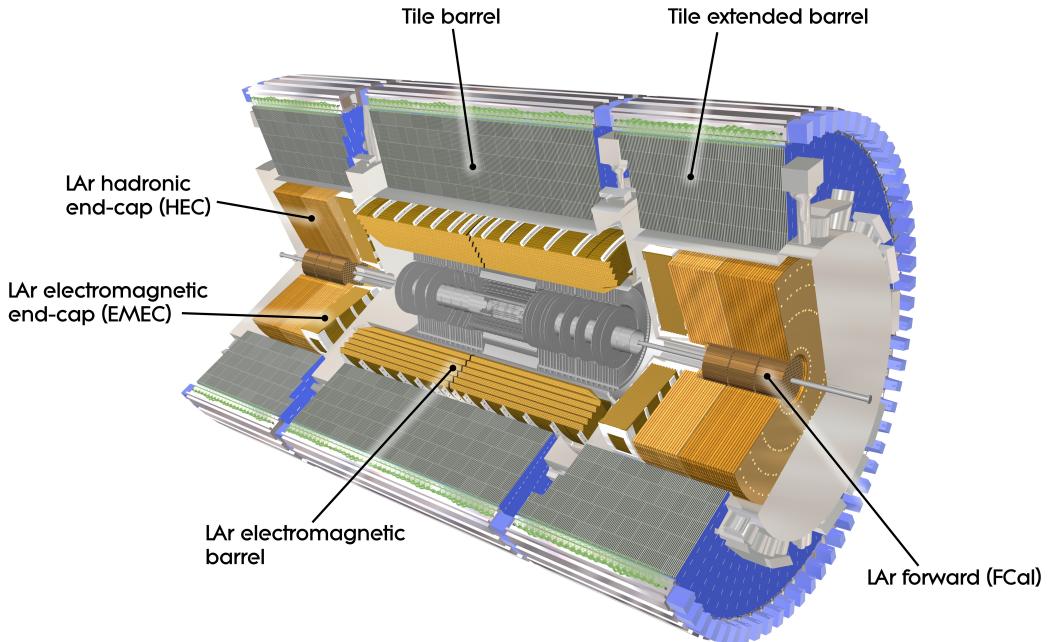
474 **5.3 Calorimeters**

475 The ATLAS electromagnetic and hadronic calorimeters (EMC and HCAL,  
476 respectively) absorb and measure the energy of high energy hadrons, photons,  
477 and electrons with  $|\eta| < 4.9$ . Both systems use sampling calorimeters which  
478 consist of alternating layers of dense absorbing and active layers. In the absorbing  
479 layer particles interact and lose energy, creating showers. These showers are then  
480 detected and measured in the active layer. The amount of charge measured in the  
481 active material scales with the energy of the incident particle, and thus provides a  
482 measurement of the particle's energy. An overview of the layout of the calorimeter  
483 system is shown in Figure 5.6.

484 The EMC measures and contains the energy of electromagnetically interacting  
485 particles. It consists of layered accordion-shaped Lead absorber plates and elec-  
486 trodes immersed in liquid Argon with 170k channels.. Using accordion-shaped  
487 electrode and absorbers ensures  $\phi$  symmetry and coverage. The EMC is com-  
488 posed of a barrel part ( $|\eta| < 1.475$ ), two end-caps ( $1.375 < |\eta| < 3.2$ ), and a  
489 presampler ( $|\eta| < 1.8$ ). The presampler, containing only liquid Argon, corrects  
490 for upstream energy losses of electrons and photons. The EMC barrel is segmented  
491 into three layers. The first layer has finest segmentation with readout cells ex-  
492 tending  $\Delta\eta \times \Delta\phi = 0.025/8 \times 0.1$ . This provides a precise shower measurements  
493 used to separate prompt photons from  $\pi^0 \rightarrow \gamma\gamma$  decays. The second layer has  
494 coarser segmentation and is approximately 16 radiation lengths long. A radiation  
495 length is the average distance an electron travels before losing all but  $1/e$  of its  
496 energy to bremsstrahlung. The last layer is the most coarse and measures the tail  
497 of the electromagnetic shower. A schematic of the ECAL is shown in Figure 5.7.

498 The hadronic calorimeter located outside the EMC and is used to contain  
499 and measure the energy of hadronically interacting particles. It consists of a tile

500 calorimeter (TileCal), hadronic end-cap calorimeter (HEC), and liquid Argon for-  
 501 ward calorimeter (FCAL). TileCal is located behind the LAr EMC and uses steel  
 502 absorbers and liquid Argon as the active material. TileCal consists of three barrel  
 503 layers in the central and forward regions, extending up to  $|\eta| < 1.7$ . Photons  
 504 generated from hadronic interactions are collected via wavelength-shifting fibers  
 505 connected to photomultiplier tubes, as shown in Figure 5.8. The HEC lies behind  
 506 the EMC endcap wheels. It uses copper absorbers and liquid Argon as the active  
 507 material and covers  $1.5 < |\eta| < 3.2$ . Finally, the FCAL covers  $3.1 < |\eta| < 4.9$   
 508 and consists of three modules all using liquid Argon as the active material. The  
 509 first module uses copper absorber and was designed for electromagnetic measure-  
 510 ments. The second and third modules consist of tungsten absorber and are used  
 511 to measure the kinematics of hadronically interacting particles. A schematic of  
 512 the HCAL is shown in Figure 5.8.



**Figure 5.6:** Overview of ATLAS electromagnetic and hadronic calorimeters.

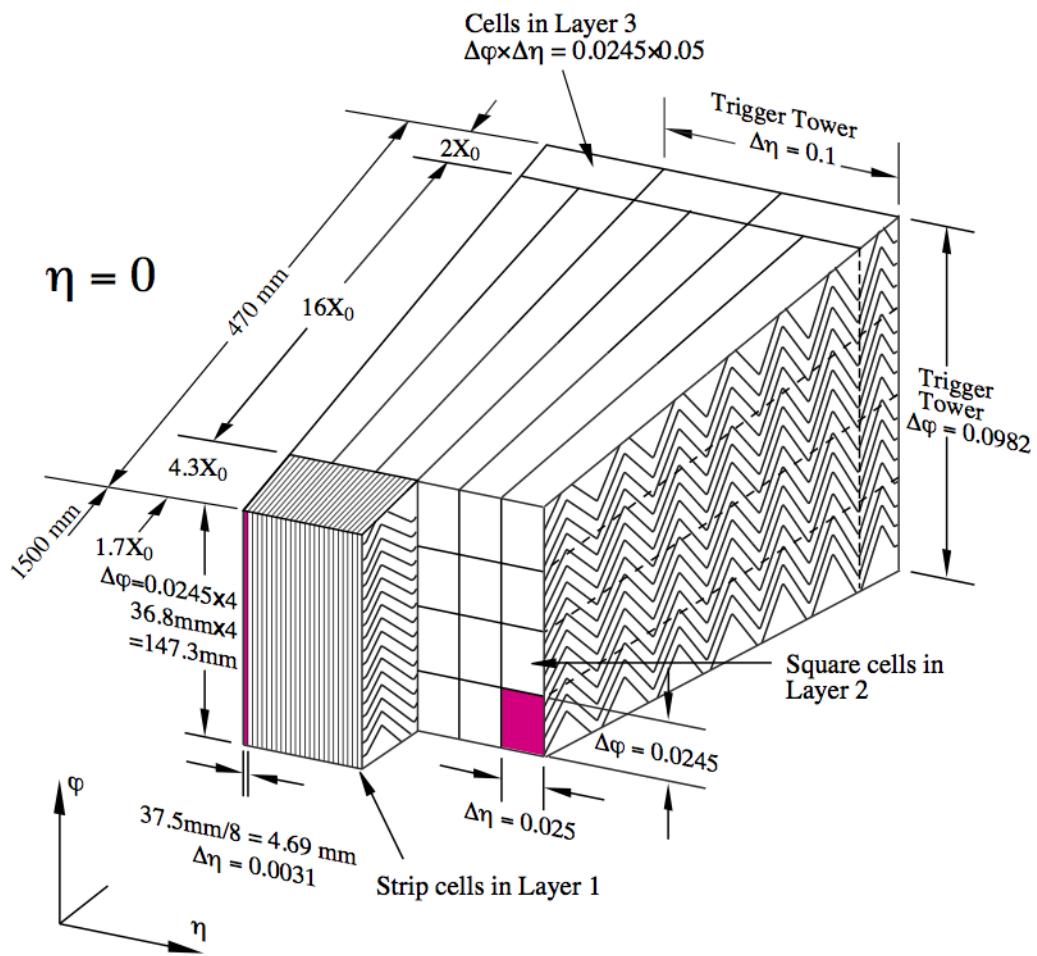


Figure 5.7: Schematic of ECAL.

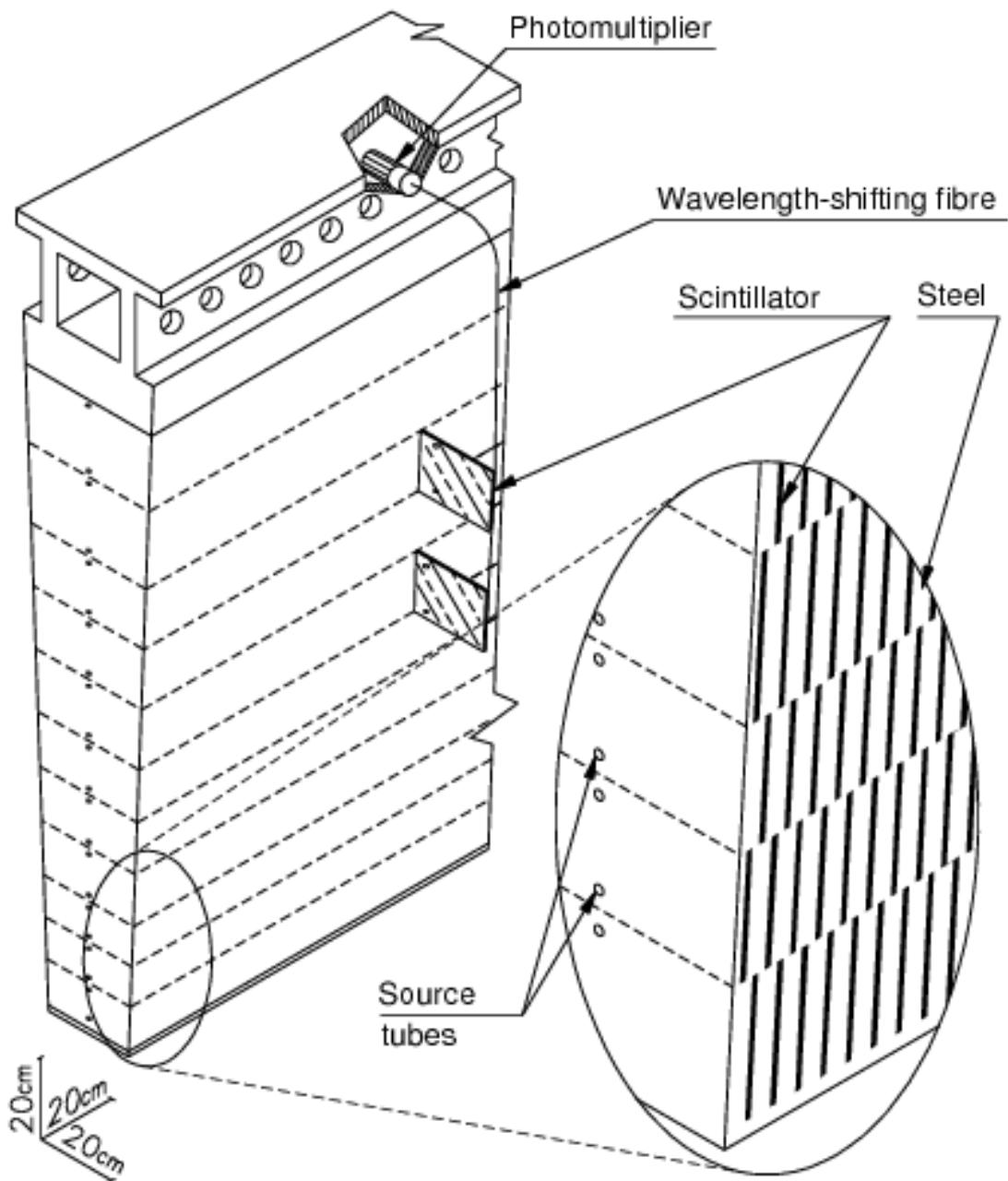


Figure 5.8: Schematic of HCAL.

513        The energy resolution of the calorimeter subsystems are:

514         $\frac{\sigma_E}{E} = \frac{10\%}{\sqrt{E}} \oplus 0.7\% \text{ EMC}$

515         $\frac{\sigma_E}{E} = \frac{50\%}{\sqrt{E}} \oplus 3\% \text{ hadronic barrel}$

516         $\frac{\sigma_E}{E} = \frac{100\%}{\sqrt{E}} \oplus 10\% \text{ hadronic end-cap}$

## 517        5.4 Muon Spectrometer

518        The muon spectrometer (MS) is the outermost detector system in ATLAS.

519        Muons with a  $p_T > 4$  GeV are energetic enough to reach the MS. To measure the  
520        momentum of these muons barrel and end-cap toroid magnets are used covering  
521         $|\eta| < 1.4$  and  $1.6 < |\eta| < 2.7$ . For  $1.4 < |\eta| < 1.6$ , a combination of the barrel  
522        and end-cap toroidal magnetic fields bend muon trajectories. The detector in the  
523        barrel region form three concentric rings at  $R = 5, 7.5, 10$ m and are segmented  
524        in  $\phi$  to accommodate the magnets. The end-cap region consists of three circular  
525        planes perpendicular to  $z$  and located at  $|z| = 7.4, 14, 21.5$ m from the interaction  
526        region. An additional detector at  $|z| = 10.8$ m covers the transition region between  
527        the barrel and end-cap.

528        The MS readout consists of four subsystems: Monitored Drift Tubes (MDT),  
529        Cathode Strip Chambers (CSC), Resistive Plate Chambers (RPC), and Thin Gap  
530        Chambers (TGC). The first two subsystems are used primarily for measuring  
531        muon track parameters, while the RPC and TGC subsystems are used for muon  
532        triggering. A schematic of this system is shown in Figure 5.9.

533        The MDT subsystem consists of precision tracking chambers for  $|\eta| < 2.7$ ,  
534        except for the inner most end-cap layer ( $2.0 < |\eta| < 2.7$ ), where CSCs are used.  
535        The basic unit of MDT chambers are thin walled Aluminum tubes with a diameter  
536        of 3 cm and length of 0.9-6.2 m. These tubes are filled with a mixture of Ar-CO<sub>2</sub>  
537        gas with a 50μm W-Rn wire running down the center of the tube, which is kept at

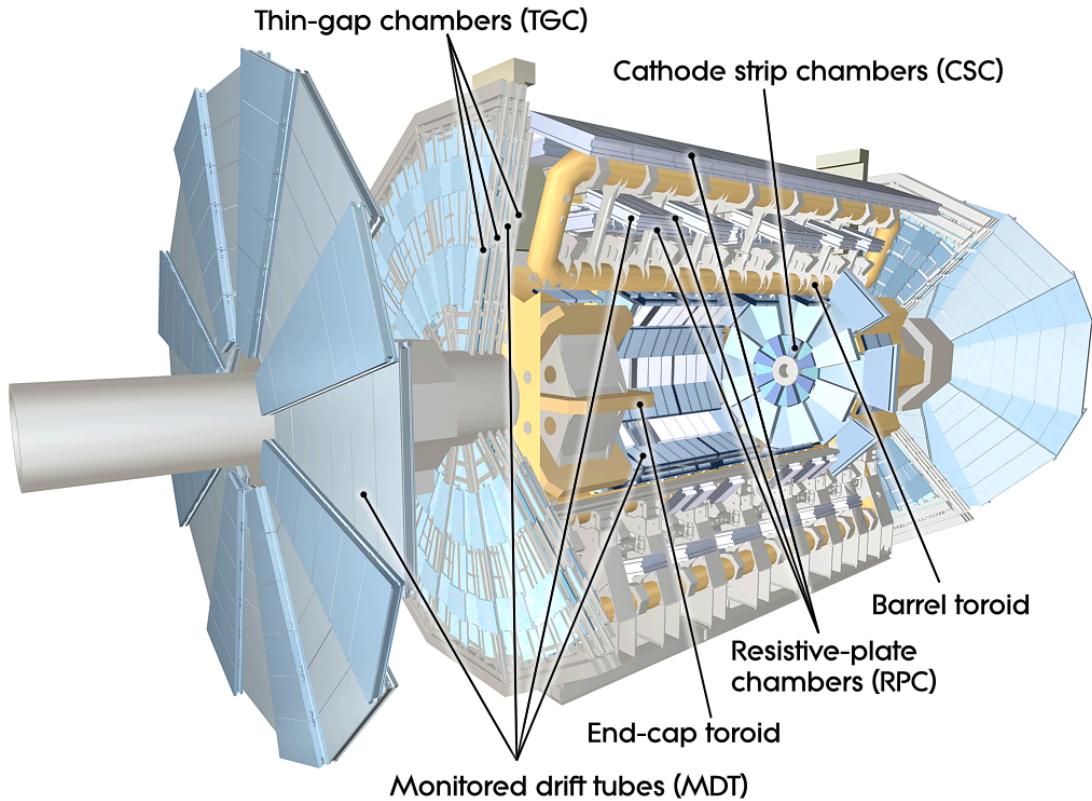
538 3080 V. Since the maximum drift time of these chambers is  $\sim 700$  ns, they are not  
539 used for triggering. MDT chambers consist of 3-4 layers of tubes mounted on a  
540 rectangular support system, as seen in Figure 5.10, orientated along  $\phi$  to measure  
541 the coordinate in the bending plane of the magnetic field with a resolution of 35  
542  $\mu\text{m}$ .

543 The MDT subsystem can only handle hit rates below  $150\text{Hz}/\text{cm}^2$ . For this  
544 reason, CSCs are used in the innermost end-cap layer where hit rates are larger.  
545 CSCs can handle hit rates up to  $1000\text{Hz}/\text{cm}^2$ . CSC are multiwire proportional  
546 chambers. These chambers are filled with a Ar- $\text{CO}_2$  gas mixture and evenly spaced  
547 wires kept at 1900 V. These wires are orientated in the radial direction but not  
548 read out. Instead on one side of the cathode are copper strips parallel to the wires,  
549 measuring  $\eta$ , while on the other side of the cathode are strips parallel to the wires  
550 measuring  $\phi$ . The width between strips is approximately 1.5 mm providing a  
551 resolution of 60  $\mu\text{m}$  in the bending-plane and 5 mm in the non-bending plane.

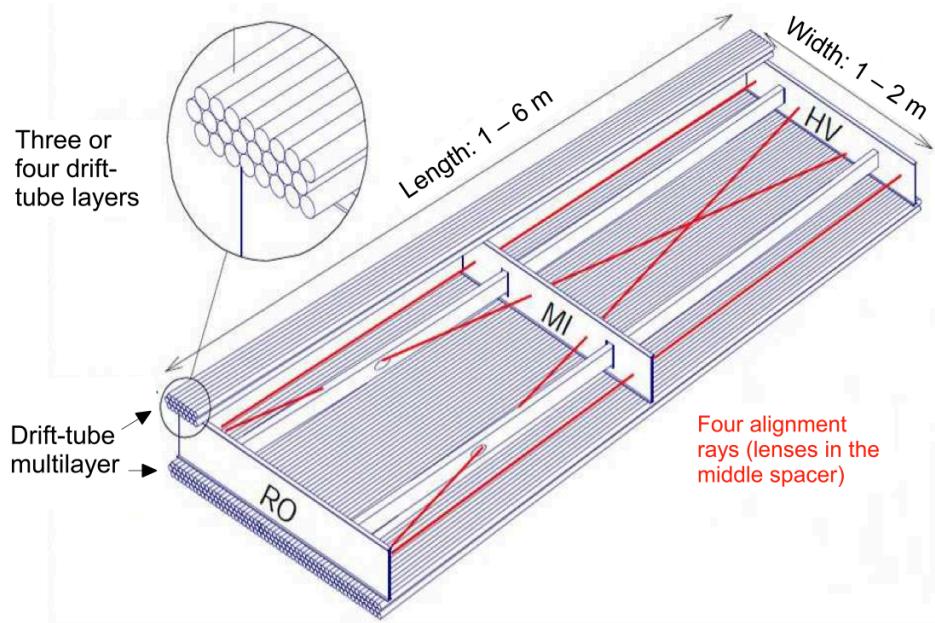
552 Since the CSC and MDT systems do not have prompt timing signals, the RPC  
553 and TGC systems are used for triggering. The RPC system is used in the barrel  
554 region ( $|\eta| < 1.05$ ). RPC consist of two parallel resistive plates separated by a 2  
555 mm insulated spacer with 100 mm spacing kept at 9.8 kV, as shown in Figure 5.11.  
556 A gaseous mixture of  $\text{C}_2\text{H}_2\text{F}_4$ ,  $\text{C}_4\text{H}_{10}$ , and  $\text{SF}_6$  fills the space between the two  
557 plates. Metallic strips on the outer faces of the plates are used to read out signals  
558 produced by the gas ionizing. The middle barrel layer consists of two layers of  
559 RPCs on either side of the MDT layer and one layer on the outermost MDT  
560 layer. Each layer contains two orthogonal sets of metallic strips providing  $\eta$  and  
561  $\phi$  measurements. The timing resolution of RPCs is 1.5 ns, and therefore may be  
562 used to identify bunch crossings.

563 Finally, the TGCs are used in the end-cap regions and are primarily used to

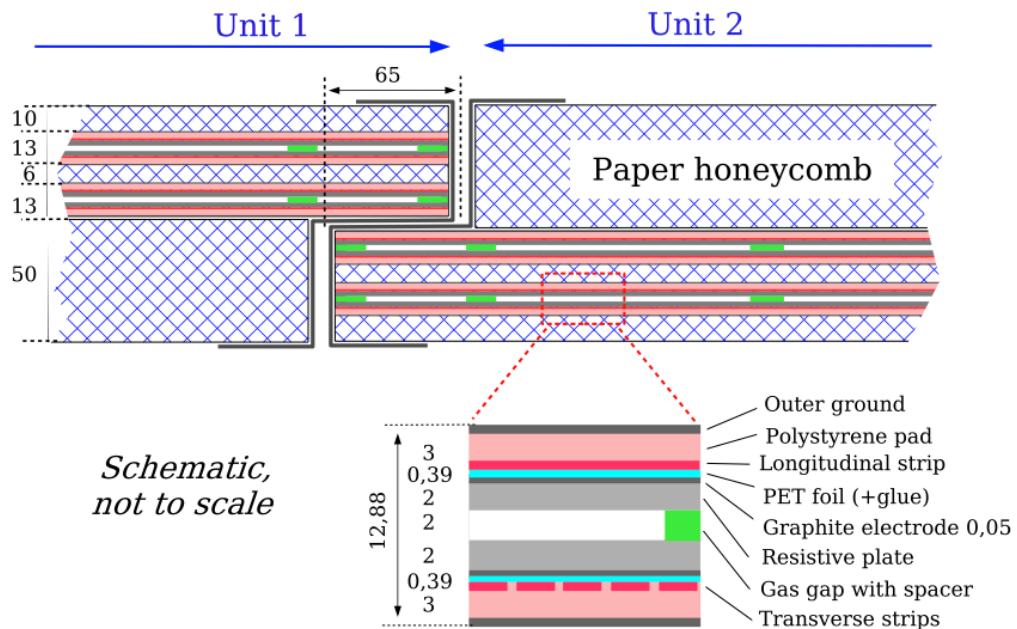
564 provide L1 trigger decisions and  $\phi$  measurements. TGCs are multi-wire propor-  
 565 tional chambers consisting of arrays of gold-coated tungsten wires placed between  
 566 two cathode planes. These wires are separated by 1.8 mm and cathodes are 1.4 mm  
 567 from the wires. Orthogonal to the wires, on the opposite side of the cathode plane  
 568 are copper strips held at 2900 V. The chambers are filled with a mixture of  $CO_2$   
 569 and n-pentane gas, the latter acts as a quenching gas to prevent avalanches initi-  
 570 ated by secondary  $\gamma$ -rays from the primary avalanche. Figure 5.12 is a schematic  
 571 of a TGC. The timing resolution of TGCs is less than 25 ns and therefore they  
 572 are used for bunch crossing measurements.



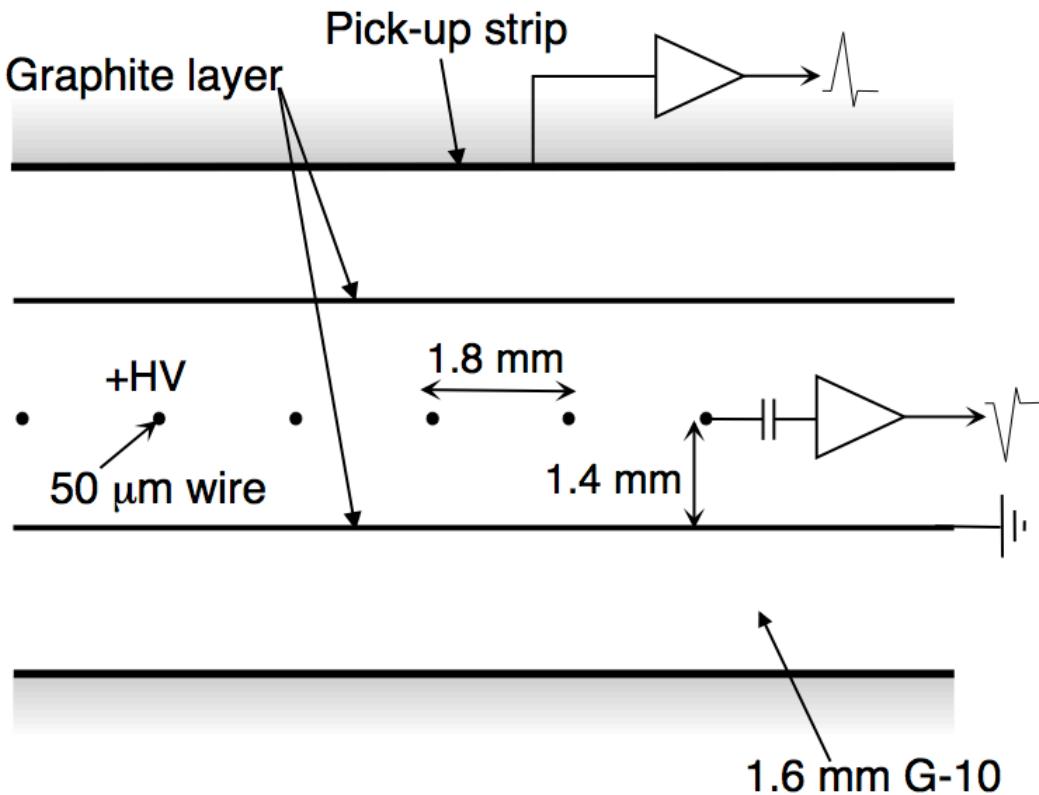
**Figure 5.9:** Schematic of Muon Spectrometer [cite G35]



**Figure 5.10:** Schematic of MDT chamber. [cite G35]



**Figure 5.11:** Schematic of RPC chamber, which is used for triggering in the central region of the detector [cite G35].



**Figure 5.12:** Schematic of TGC chamber, which is used for triggering in the muon end-cap region. [cite G35]

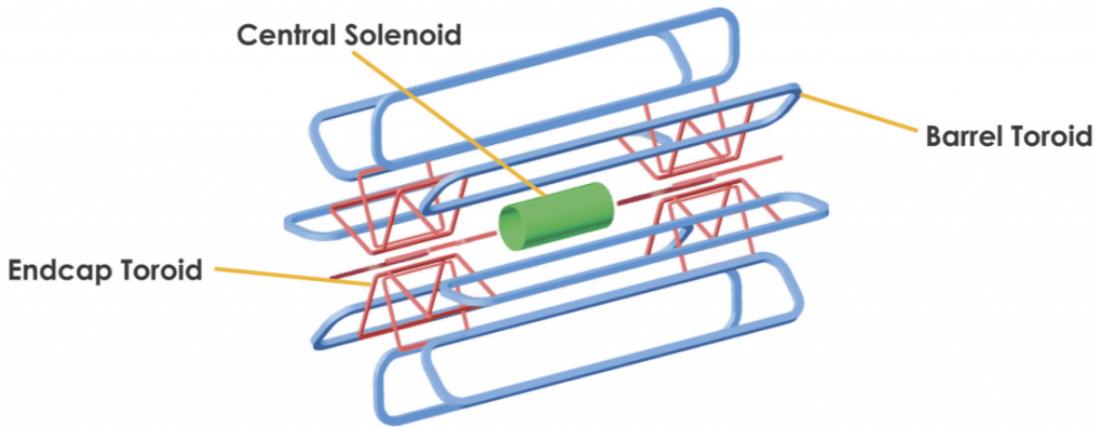
## 5.5 Magnet System

A particles with charge,  $q$ , and velocity  $v$ , moving in magnetic field,  $B$ , experiences a force,  $F = qv \times B$ . This force can cause charged particles to have a curved trajectory in magnetic fields, which the ID and MS use to determine the particles  $p_T$ . The central solenoid provides the magnetic field for the ID and the toroidal magnets provide the magnetic field for the MS.

The layout of the magnet system is shown in Figure 5.13. The central solenoid consists of a single-layer Al-stabilized NbTi conductor coil wound inside an Al

581 support cylinder. The solenoid is 5.8 m long, 50 cm thick and has an inner radius  
582 of 1.23 m. It is cooled to 4.5 K to reach superconducting temperatures and shares  
583 the liquid argon calorimeter vacuum vessel to minimize material in the detector.  
584 A current of 7.730kA produces a 1.998 T solenoidal magnetic field, pointing in  
585 the  $+z$  direction.

586 The toroidal magnet system consists of a barrel and two end-cap toroidal  
587 magnets used to create a magnetic field outside the calorimeters that is orientated  
588 along  $\phi$ . Each barrel toroid is 25.3 m long with an inner and outer diameter of 9.4  
589 and 20.1 m and weighs 830 tonnes. Endcap toroids are 5 m long with an inner and  
590 outer radius of 1.65 and 10.7 m. Both toroid systems use Al-stabilized Nb/Ti/Cu  
591 conductors. The magnetic field strength in the barrel and endcap regions are 0.5  
592 and 1 T, respectively.



**Figure 5.13:** Layout of ATLAS magnet systems.

## 593 5.6 Trigger System

594 Since collisions occur every 25 ns and reading out all detector channels and  
595 storing that information is not currently feasible (would require saving 60 million  
596 megabytes per second), the majority of events are not kept for analysis. ATLAS

597 uses a multi-stage trigger system to select approximately 1,000 of the 1.7 billion  
598 collisions that occur each second (corresponding to a rate of 1 kHz from the 40  
599 MHz proton collision rate). The first stage of the trigger system is the hardware  
600 level (L1) trigger. This trigger reduces the event rate to  $\sim$ 100 kHz by identifying  
601 Regions-of-Interest (ROIs) containing high  $p_T$  leptons, photons, jets, or  $E_T^{miss}$  by  
602 using information from RPCs, TGCs, and calorimeters to make a  $2.5 \mu\text{s}$  decision.  
603 This information is then passed to a high-level trigger (HLT) which further de-  
604 creases event rates to  $\sim$  1 kHz. The HLT uses finer granularity measurements  
605 from the MS and ID to perform simplified offline reconstruction to decide which  
606 events to keep.

## **Part III**

607

## **Method**

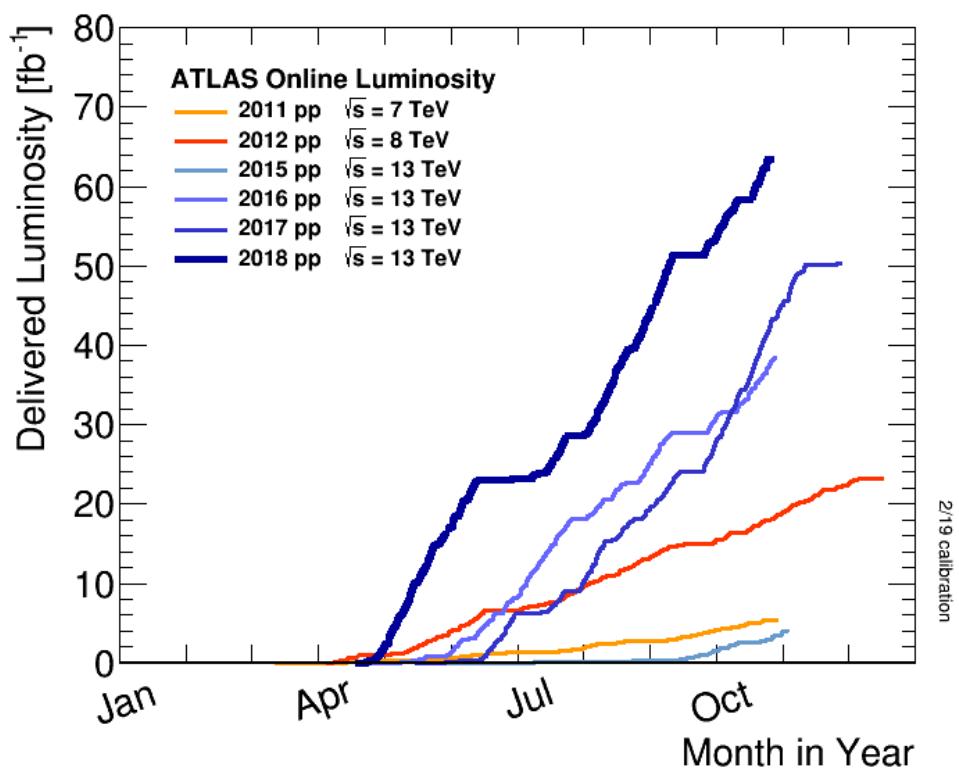
608

609 **Chapter 6**

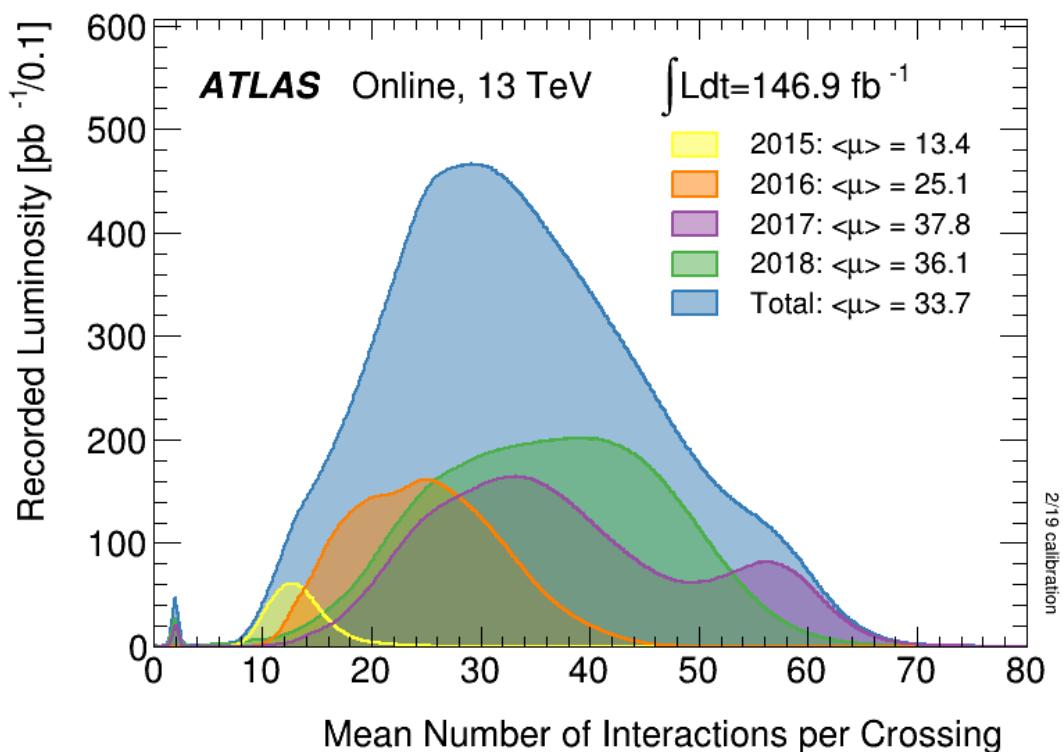
610 **Dataset and Simulated Samples**

611 **6.1 Dataset**

612 This analysis uses  $pp$  collision data collected from 2015 to 2018 at  $\sqrt{s} = 13$   
613 TeV, corresponding to 139/fb of data as shown in Figure 6.1 and 6.2. From this  
614 dataset, only those events in which the tracker, calorimeters, and muon spectrom-  
615 eter have good data quality are used. For a given event, the solenoid and toroidal  
616 magnets must also be operating at their nominal field strengths. In addition to  
617 this, events must pass further quality checks to reject events where detector sub-  
618 systems may have failed. These selections reject events that containing LAr noise  
619 bursts, saturation in the electromagnetic calorimeter, TileCal errors, and failures  
620 in event recovery due to tracker failures. Events with information missing from  
621 subsystems (usually due to busy detector conditions) are rejected. Events must  
622 also contain a primary vertex with at least two associated tracks, where the pri-  
623 mary vertex is selected as the vertex with the largest  $\sum p_T^2$  over tracks associated  
624 with the vertex and  $p_T > 0.5$  GeV.



**Figure 6.1:** Integrated luminosity for data collected from ATLAS from 2011 - 2018



**Figure 6.2:** Mean number of interactions per crossing for data collected from ATLAS from 2011 - 2018

625 **6.2 Simulated Samples**

626 Samples are simulated in order to model backgrounds, evaluate signal ac-  
627 ceptance, optimize event selection and estimate systematic and statistical uncer-  
628 tainties. The dominant backgrounds for this analysis are  $W/Z + \text{jets}$ , diboson  
629 ( $WZ/WW$ ),  $t\bar{t}$ , single top and multijet production.

630  $W/Z+\text{jet}$  events are simulated using Sherpa 2.2.1 at NLO [cite [29]] and merged  
631 with the Sherpa parton shower using the ME+PS@NLO prescription [11]. These  
632 events are then normalized to NNLO cross sections. The  $t\bar{t}$  and single-top back-  
633 grounds are generated with Powheg-Box with NNPDF3.0NLO PDF sets in the  
634 matrix element calculation [cite[35]]. For all processes, the parton shower, frag-  
635 mentation, and underlying event are simulated using Pythia 8.320 with the A14  
636 tune set[cite[ATL-PHYS-PUB-2014-02]]. Diboson processes are generated using  
637 Sherpa 2.2.1.

638 Signal samples are simulated using MadGraph 5-2.2.2 [cite 42] and Pythia  
639 8.186 with NNPDF230LO. RS Graviton samples are generated with  $k/M_{PL}=1$ .  
640 HVT Model A (B) samples are simulated with  $g_V = 1(3)$ , as the difference in the  
641 width of the samples is smaller than detector resolution. Model C is generated by  
642 setting  $g_H = 1$  and  $g_f = 0$  to model VBF production of HVT bosons. Signals are  
643 generated for masses between 300 GeV and 5 TeV.

644 **6.3 Object Selection**

645 **6.3.1 Electrons**

646 Electrons are reconstructed from electromagnetic showers in the LAr EM  
647 calorimeter. During reconstruction cells of  $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$  are grouped  
648 into  $3 \times 5$  clusters. These clusters are then scanned for local maxima that seed

649 electron clusters. These clusters must then be matched to ID track from the PV.  
650 This requirement minimizes non-prompt electron and fake electron backgrounds.  
651 Electrons must pass identification and isolation requirements. Electron identifica-  
652 tion (loose, medium, tight) classification is based on a multivariate discriminant  
653 that identifies electrons using a likelihood based method. For this analysis tight  
654 electrons are used. Electrons are also required to be isolated. The electrons are  
655 considered isolated if the quotient of the sum of the transverse momentum (of  
656 calorimeter energy deposits) in a cone around the electron of size  $\Delta R = 0.2$  and  
657 the transverse momentum of the electron to be less than  $0.015 * p_T$  or 3.5 GeV,  
658 whichever is smaller. This requirement rejects non-prompt photons and other  
659 fake leptons. Electrons in this analysis are also required to have  $p_T > 30$  GeV and  
660  $|\eta| < 2.47$ . Electrons are also required to have  $p_T > 30$  GeV.

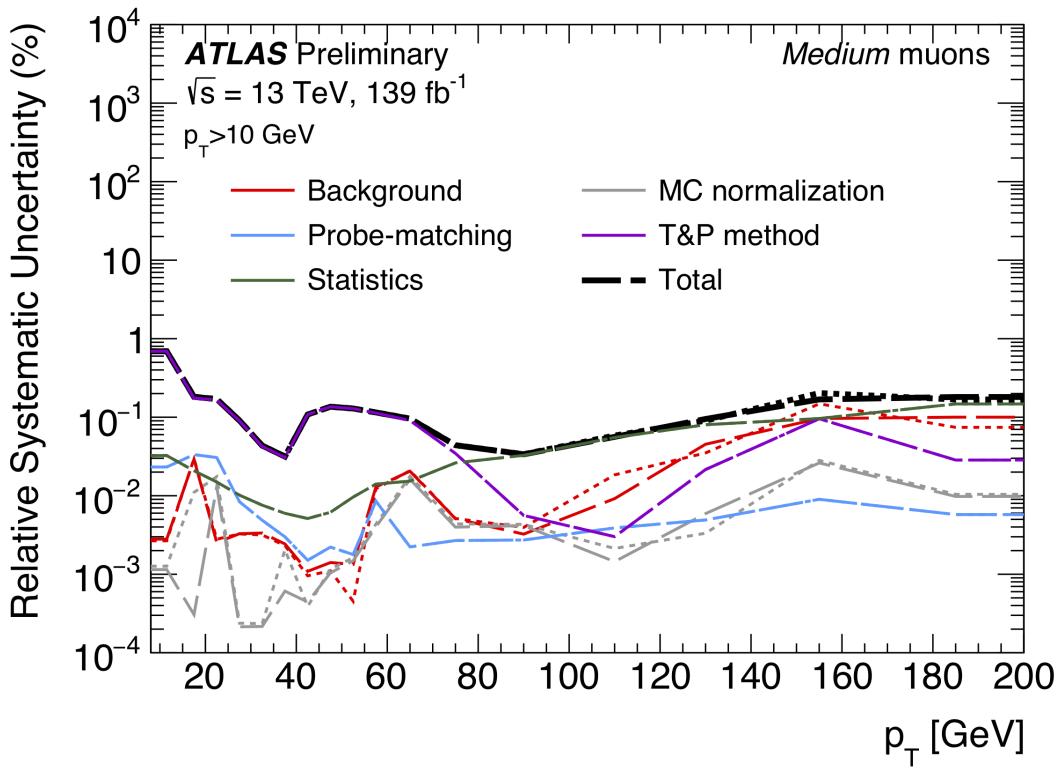
661 Electrons are calibrated to determine data-driven scale factors using  $J/\Psi \rightarrow$   
662  $ee$ ,  $Z \rightarrow ee$ ,  $Z \rightarrow \ell\ell\gamma$  processes. These corrections account for the non-uniform  
663 response of the detector which introduces modeling and reconstruction uncertain-  
664 ties.

### 665 6.3.2 Muons

666 As muons traverse the entire detector, they are reconstructed from ID and MS  
667 tracks. For this analysis the muon identification and isolation working points are  
668 chosen to minimize the contributions from non-prompt muons. Towards this end,  
669 the medium muon identification working point is used. For this working point,  
670 two types of reconstructed muons are used: combined and extrapolated muons  
671 (CB and ME, respectively). For CB muons, ID and MS tracks are reconstructed  
672 independently and a combined track fit is performed by adding or removing MS  
673 tracks to improve the fit quality. ME muons are reconstructed from only MS

674 tracks with hits in at least two layers, which ensures the track originates from the  
675 PV. ME muons extend the acceptance for muon reconstruction outside the ID  
676 from  $2.5 < |\eta| < 2.7$ . The medium identification working point uses CB and ME  
677 tracks. CB tracks must have at least 3 hits in two MDT layers. ME tracks are  
678 required to have at least three MDT/CSC hits. To further minimize contributions  
679 from fake muons, the selected muons are required to be isolated from other tracks,  
680 as muons from  $W, Z$  decays are often isolated from other particles. To insure the  
681 selected muons are isolated, the scalar sum of the transverse momentum of tracks  
682 in a cone of  $\Delta R = 0.3$  compared to the transverse momentum of the muon must  
683 be less then 0.06. Muons are also required to have  $p_T > 30$  GeV.

684 Muons are calibrated using well-studied resonances  $J/\Psi \rightarrow \mu\mu$  (low- $p_T$ ),  $Z \rightarrow$   
685  $\mu\mu$  (high- $p_T$ ). Figure 6.3 shows the combined muon  $p_T$  uncertainty from this  
686 calibration. The total systematic uncertainty is less then 1% for all  $p_T$  ranges  
687 considered in this analysis.



**Figure 6.3:** This figure shows the breakdown of the muon reconstruction efficiency scale factor measured in  $Z \rightarrow \mu\mu$  as a function of  $p_T$  [4].

### 688 6.3.3 small-R jets

689 Small-R jets are used to reconstruct the hadronically decaying  $W/Z$  candi-  
 690 dates in the resolved analysis and VBF jets. These jets are less boosted and  
 691 therefore spatially separated and reconstructed separately. These jets are con-  
 692 structed from topologically connected clusters of calorimeter cells (topoclusters),  
 693 seeded from calorimeter cells with energy deposits significantly above the noise  
 694 threshold. These cells are then used as inputs to the  $anti - k_t$  algorithm [12] with  
 695 a distance = 0.4, here called small-R jets. These jets are calibrated to compensate  
 696 and account for biases from jet reconstruction.

697 The jet energy is calibrated sequentially as shown in Figure 6.4. After the jet

698 direction is corrected to point to the PV, the energy of the jet is corrected. First,  
699 the jet energy is corrected to account for pileup contributions based on the  $p_T$  and  
700 area of the jet (these corrections are extracted from a  $pp \rightarrow jj$  sample). Following  
701 this, another pileup correction is applied that scales with  $\mu$  and  $N_{PV}$ .

702 MC-based corrections are then applied that are meant to transform the jet  
703 energy and  $\eta$  back to truth level. Therefore, these corrections account for the  
704 non-compensating nature of the ATLAS calorimeters and inhomogeneity of the  
705 detector. Following this, the Global Sequential Calibration is applied that re-  
706 duces flavor dependence of jet calibrations and accounts for energy leakage of jets  
707 outside the calorimeters. Finally, in-situ corrections are applied that account for  
708 differences in jet responses between data and simulation ( $\gamma/Z+jet$  and multijet  
709 samples are used). These differences can be due to mismodelling of the hard  
710 scatter event, pile-up, jet formation, etc.

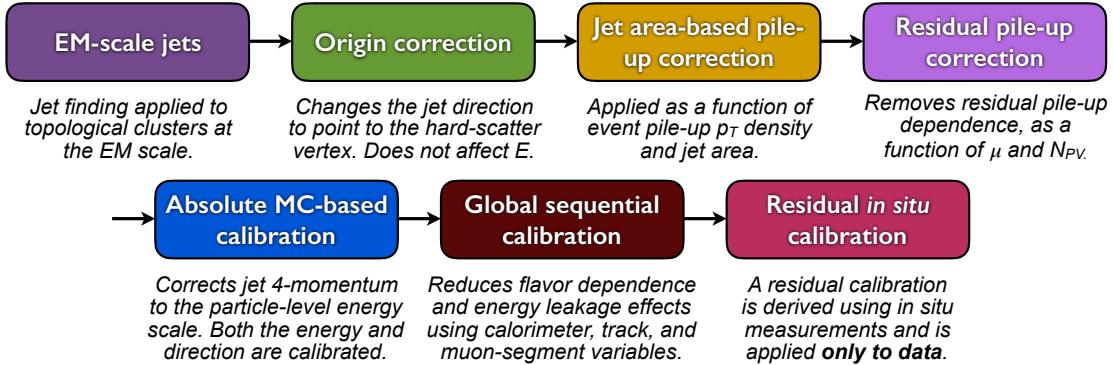
711 Jets used in this analysis must have  $p_T > 30$  GeV and  $|\eta| < 2.5$ . To further  
712 reduce fake jets the jet-vertex-tagger (JVT) is used to reject pile-up jets [cite 43  
713 P]. The JVT uses two track-based variables, corrJVF and  $R_{p_T}$  to calculate the  
714 likelihood that the jet originated from the PV. The corrJVF compares the scalar  
715 sum of the  $p_T$  of tracks associated with the jet and PV to the scalar sum of the  
716  $p_T$  of tracks associated with the jet. This variable also includes a correction that  
717 reduces the dependency of corrJVF with the number of reconstructed vertices in  
718 the event. The other discriminant,  $R_{p_T}$ , is given by the ratio of the scalar sum  
719 of the  $p_T$  of tracks associated with the jet and PV to the  $p_T$  of the jet. Both of  
720 these variables peak around zero for pileup jets, as these jets are unlikely to have  
721 tracks associated with the PV. JVT cuts are applied to all jets with  $p_T > 120$   
722 GeV. Central jets ( $|\eta| < 2.4$ ) are required to have a  $JVT > 0.59$  and forward jets  
723 ( $2.4 < |\eta| < 2.5$ ) are required to have  $JVT > 0.11$ .

724 To further reject fake jets, jets must pass quality requirements based on the  
725 following variables ([cite P42]):

- 726 -  $f_Q^{LAr}$ : fraction of energy of jet's LAr cells with poor signal shape
- 727 -  $f_Q^{HEC}$ : fraction of energy of jet's HEC cells with poor signal shape
- 728 -  $E_{neg}$ : sum of cells with negative energy
- 729 -  $f_{EM}$ : fraction of jet's energy deposited in EM calorimeter
- 730 -  $f_{HEC}$ : fraction of jet's energy deposited in HEC calorimeter
- 731 -  $f_{max}$ : maximum energy fraction in any single calorimeter layer
- 732 -  $f_{ch}$ : ratio of the scalar sum of the  $p_T$  of a jet's charged tracks to the jet's  $p_T$

733 Jets selected for the resolved analysis must pass one of the following criteria:

- 734 -  $f_{HEC} > 0.5$  and  $|f_Q^{HEC}| > 0.5$  and  $\langle Q \rangle > 0.8$
- 735 -  $|E_{neg}| > 60$  GeV
- 736 -  $f_{EM} > 0.95$  and  $f_Q^{LAr} > 0.8$  and  $\langle Q \rangle > 0.8$  and  $|\eta| < 2.8$
- 737 -  $f_{max} > 0.99$  and  $|\eta| < 2$
- 738 -  $f_{EM} < 0.05$  and  $f_{ch} < 0.05$  and  $|\eta| < 2$
- 739 -  $f_{EM} < 0.05$  and  $|\eta| > 2$



**Figure 6.4:** [6] This diagram shows the calibration stages for EM jets.

### 740 6.3.4 large-R jets

741 Large-R ( $\Delta R = 1.0$ ) jets are used to reconstruct the high- $p_T W/Z \rightarrow qq$  candi-  
 742 dates in the merged analysis. Track-Calorimeter Clusters (TCCs) are used to reconstruct  
 743 these jets [cite ANA 50]. These jets are constructed via a pseudo particle flow  
 744 method using ID tracks matched to calorimeter clusters. The angular resolution  
 745 of the calorimeter degrades sharply with jet  $p_T$ , but the jet energy resolution  
 746 improves. The tracker has excellent angular resolution which improves with  $p_T$ .  
 747 Therefore, by matching tracks to jets, TCCs have more precise energy and angular  
 748 resolution than jets constructed from calorimeter information only. These jets are  
 749 required to have  $p_T > 200$  GeV,  $|\eta| < 2.0$  and  $m_J > 50$  GeV.

750 TCC jets are trimmed as detailed in [cite ANA 45], which suppresses pileup  
 751 and soft radiation in the jet, the jet mass is calculated as the four-vector sum  
 752 of the jet's constituents (assuming massless constituents). The jet mass peaks  
 753 around the  $W/Z$  boson mass for  $W/Z \rightarrow qq$  jets, and more broadly for quark and  
 754 gluon induced jets.

755 These jets are then tagged as  $W/Z$  jets if they pass the jet mass and  $D_2$   
 756 cuts. The jet substructure variable  $D_2$  is given by the ratio of energy correlation  
 757 functions based on energies and pair-wise angles of a jet's constituents [cite ANA

758 46, 47]:

$$D_2^{\beta=1} = E_{CF3} \left( \frac{E_{CF1}}{E_{CF2}} \right)^3 \quad (6.1)$$

759 Where the energy correlation functions are defined as:

$$E_{CF1} = \sum_i p_{T,i} \quad (6.2)$$

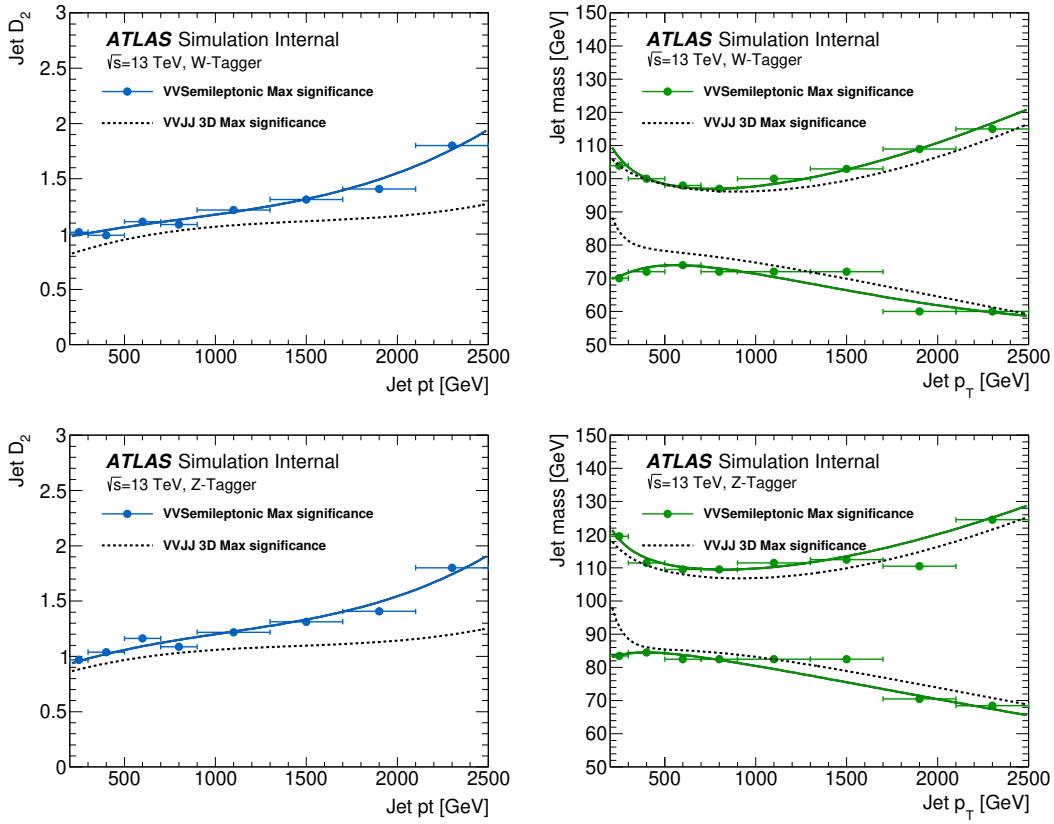
760

$$E_{CF2} = \sum_{ij} p_{T,i} p_{T,j} \Delta R_{ij} \quad (6.3)$$

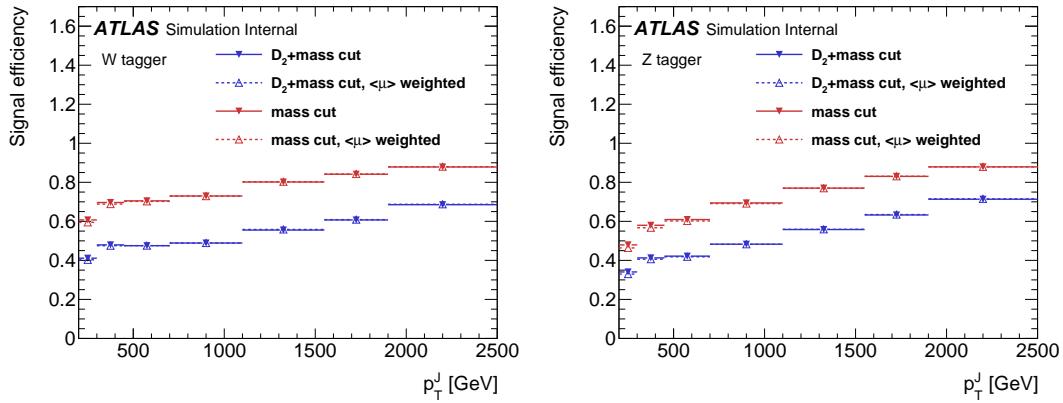
761

$$E_{CF3} = \sum_{ijk} p_{T,i} p_{T,j} p_{T,k} \Delta R_{ij} \Delta R_{jk} \Delta R_{ki} \quad (6.4)$$

762 A two-dimensional optimization of the jet mass and  $D_2$  thresholds was per-  
763 formed to provide maximum sensitivity for this analysis. Figure 6.5 shows the  
764 optimized thresholds on  $D_2$  and jet mass as a function of  $p_T$ . Figure 6.6 shows  
765 the efficiency of the optimized  $W/Z$  taggers as a function of jet  $p_T$ .



**Figure 6.5:** The upper cut on  $D_2$  (a) and jet mass window cut i.e. the upper and lower boundary of the mass (b) of the  $W$ -tagger as a function of jet  $p_T$ . Corresponding values for  $Z$ -tagger are shown in (c) and (d). The optimal cut values for maximum significance are shown as solid markers and the fitted function as solid lines. Working points from  $VV \rightarrow JJ$ [ATLAS-HDBS-2018-31-002] is also shown as dashed lines as a reference. Natasha reword?



**Figure 6.6:** Natasha write caption

766 **6.3.5 Variable Radius jets**

767 Variable-radius (VR) track jets are used to identify b-quark induced jets within  
768 the catchment area of large-R jets [cite ANA 52]. These jets use a  $p_T$  dependent  
769 cone size defined as:

$$R_{eff}(p_{T,i}) = \frac{\rho}{p_{T,i}} \quad (6.5)$$

770 for building jets from ID tracks with an anti- $k_t$  algorithm. For this analysis  $\rho = 30$   
771 GeV and an upper and lower limit on cone size are set to 0.02 and 0.4, respectively.  
772 Collinear VR jets are possible, so track jets that are not separated by the the  
773 smaller jet's cone size are not used. These jets are also required to have  $p_T > 10$   
774 GeV and  $|\eta| < 2.5$ .

775 **6.3.6 MET/neutrinos**

776 As neutrinos are uncharged and colorless they do not leave tracks or jets in the  
777 detector. For this reason, neutrinos are reconstructed calculated the  $E_T^{miss}$ , the  
778 negative vector sum of  $p_T$  all the physics objects and an extra "soft" term. The  
779 "soft" term accounts for energy deposits not associated with any of the objects in  
780 the event. For this analysis the soft term is given by the sum  $p_T$  of all ID tracks  
781 not associated with objects in the event. The selected tracks must be matched to  
782 the primary vertex, which decreases pile-up contamination [cite G 217 218]. The  
783 tight working point is used [Natasha look up what this means].

784 **6.3.7 Jet Flavor Tagging**

785 To further classify events, the small radius jets originating from a b-quark are  
786 classified using a multivariate  $b$ -tagging algorithm (BDT), MV2c10 [cite G 210  
787 199]. This algorithm uses the impact parameters of the jet's ID tracks, secondary

788 vertices (if they exist), and reconstructed flight paths of  $b$  and  $c$  hadrons in the  
789 jet to determine if the jet was induced by a  $b$ -quark. For this analysis the 85%  
790 efficient working point of this algorithm is used giving  $c$ ,  $\tau$ , and light-flavor jet  
791 rejection of 3, 8, and 34 respectively in a simulated  $t\bar{t}$ .

### 792 6.3.8 Overlap Removal

793 Reconstructed jets and leptons in this analysis can arise from the same energy  
794 deposits. For instance, a cluster of energy from an electron can also be a valid  
795 calorimeter seed for a jet. To mitigate this confusion of multiple objects originating  
796 from a single jet or lepton overlapping objects are removed via a procedure referred  
797 to as overlap removal. In this procedure the separation of the two objects,  $\Delta(R) =$   
798  $\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  determines which object is removed from the event.

799 The overlap selections used in this analysis are:

- 800 - when an electron shares a track with another electron with the lower  $p_T$   
801 electron is rejected, as it is more likely to be a fake electron
- 802 - when a muon and electron share a track the muon is rejected if it is a  
803 calo-muon, otherwise the electron is rejected
- 804 - when  $\Delta R < 0.2$  for an electron and jet, the jet is rejected to maximize signal  
805 acceptance
- 806 - when  $\Delta R > 0.2$  for an electron and jet, the electron is rejected as likely  
807 originated from decays within the jet
- 808 - when  $\Delta R < \min(0.4, 0.04 + 10\text{GeV}/p_T^\mu)$  the muon is rejected, again maximiz-  
809 ing signal acceptance, otherwise the jet is rejected
- 810 - when  $\Delta R < 1.0$  for the a large-R jet and electron, the jet is rejected

811 **Chapter 7**

812 **Event Selection and**

813 **Categorization**

814 **7.1 Pre-selection**

815 Before applying topological cuts to suppress backgrounds and reduce data  
816 size in this search, preselection cuts are applied which include trigger and event  
817 requirements. Events must contain exactly one tight lepton (no additional loose  
818 leptons), the  $p_T^{\ell\nu} > 75$  GeV, and there must be at least two small-R jets or one  
819 large-R jet.

820 **7.2 Trigger**

821 The data were collected using the lowest unprescaled single-lepton or  $E_T^{miss}$   
822 triggers, as summarized in Table ???. Since the muon term is not considered in the  
823 trigger  $E_T^{miss}$  calculation, the  $E_T^{miss}$  trigger is fully efficient to events with high- $p_T$   
824 muons. For this reason, the  $E_T^{miss}$  trigger is used for events where  $p_T^\mu > 150$  GeV, to  
825 compensate for the poor efficiency of the single muon trigger above  $p_T^\mu > 150$  GeV.

**Table 7.1:** The list of triggers used in the analysis.

Data-taking period	$e\nu qq$ channel	$\mu\nu qq$ ( $p_T(\mu\nu) < 150$ GeV) channel	$\mu\nu qq$ ( $p_T(\mu\nu) > 150$ GeV) channel
2015	HLT_e24_lhmedium_L1EM20 OR HLT_e60_lhmedium OR HLT_e120_lhloose	HLT_mu20_iloose_L1MU15 OR HLT_mu50	HLT_xe70
2016a (run $< 302919$ ) $(L < 1.0 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	HLT_e26_lhtight_nod0_ivarloose OR HLT_e60_lhmedium_nod0 OR HLT_e140_lhloose_nod0 HLT_e300_etcut	HLT_mu26_ivarmedium OR HLT_mu50	HLT_xe90_mht_L1XE50
2016b (run $\geq 302919$ ) $(L < 1.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1})$	same as above	same as above	HLT_xe110_mht_L1XE50
2017	same as above	same as above	HLT_xe110_pufit_L1XE55
2018	same as above	same as above	HLT_xe110_pufit_xe70_L1XE50

826    **7.3 GGF/VBF RNN**

827    To classify events as originating from GGF/DY or VBF production a recursive  
828    neural network (RNN [17]) is used. This approach is more powerful than a cut-  
829    based classification as it improves signal efficiency and analysis sensitivity by  
830    exploiting correlations between variables that the RNN learns. In particular, a  
831    RNN architecture is ideal as it can handle variable numbers of jets in the events.

832    The RNN uses the four-momentum of candidate VBF jets to classify events  
833    as VBF or GGF topologies. As sometimes jets will be incorrectly reconstructed  
834    the number of jets in the event are expected to vary across the inputs samples.  
835    VBF candidate jets are identified by removing jets from the event that are likely  
836    from  $W/Z \rightarrow qq$ . For the resolved regime this means removing the two leading  
837    small-R jets from the VBF candidate jet list. For the merged regime this means  
838    removing small-R jets separated by less than 1.0 in  $dR$  from the large-R jet. VBF  
839    candidate jets are also required to be within  $|\eta| < 4.5$ . From the list of remaining  
840    VBF candidate jets, the two highest- $p_T$  jets are chosen.

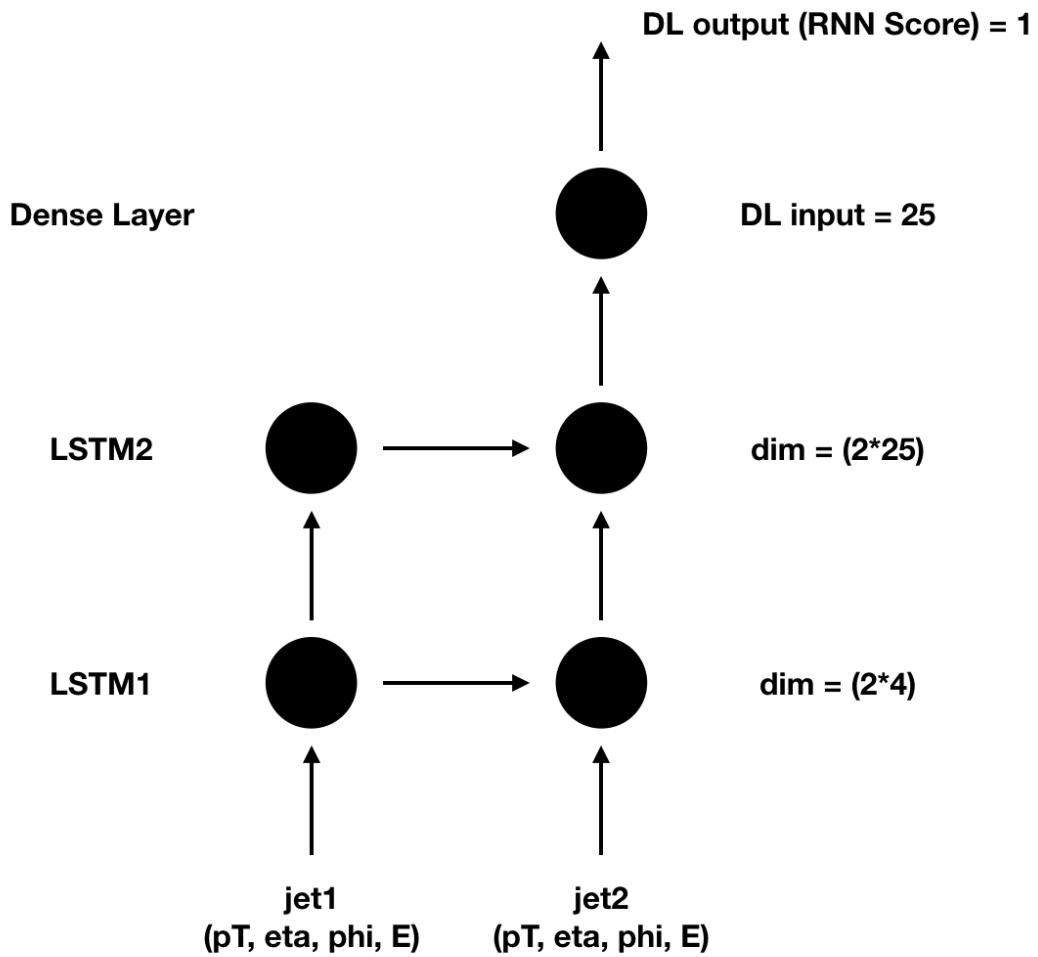
841    The architecture of the RNN is show in Figure 7.1. LSTMs are a type of  
842    RNN that extract meaningful information and can retain it (unlike other neural  
843    networks architectures). This is useful for VBF event classification for events with  
844    two jets, where using the kinematic properties of both jets (and their correlations)  
845    will lead to more efficient event classification.

846    In this RNN architecture, the VBF candidates are first passed to a masking  
847    layer which checks the number of jets in the event. If there is only one jet, only one  
848    LSTM layer is used. The output of masking is then passed to a Long Short-Term  
849    Memory (LSTM) cell, with a tanh activation function. This output is passed to  
850    a dropout layer, that has a probability of 0.3 to completely forget the output of  
851    the LSTM. Dropout is a regularization method, that prevents over-fitting. The

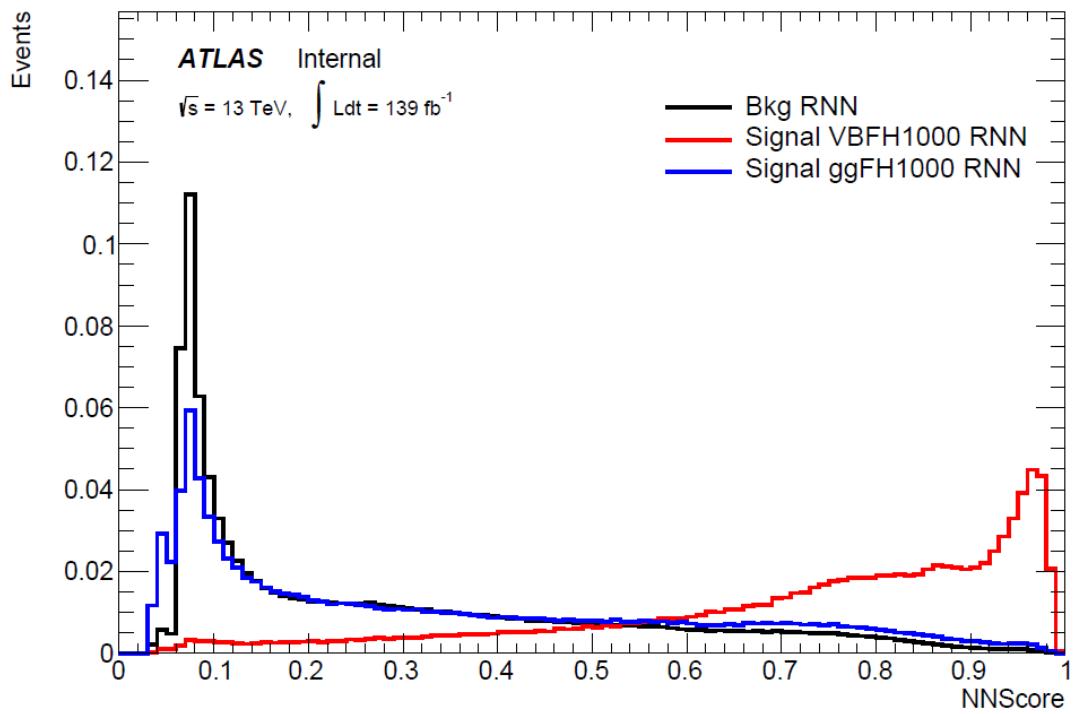
852 output of the dropout layer is then passed to the second LSTM and then through  
853 another dropout layer with a probability of 0.3.

854 The weights and other parameters of the network are learned by training the  
855 network with VBF and GGF signals over 200 epochs with an Adam Optimizer  
856 [natasha add reference]. The training is truncated if the network parameters are  
857 unchanged after ten iterations. The training, testing and validation sets are 56,  
858 30, and 14 percent of the input samples, respectively. Figure [add INT figure  
859 32] shows the loss function of the network as a function of training epochs. The  
860 validation test set has a smaller loss function as dropout was not applied. Figure  
861 7.3 shows the ROC curve for the RNN using k-fold cross validation.

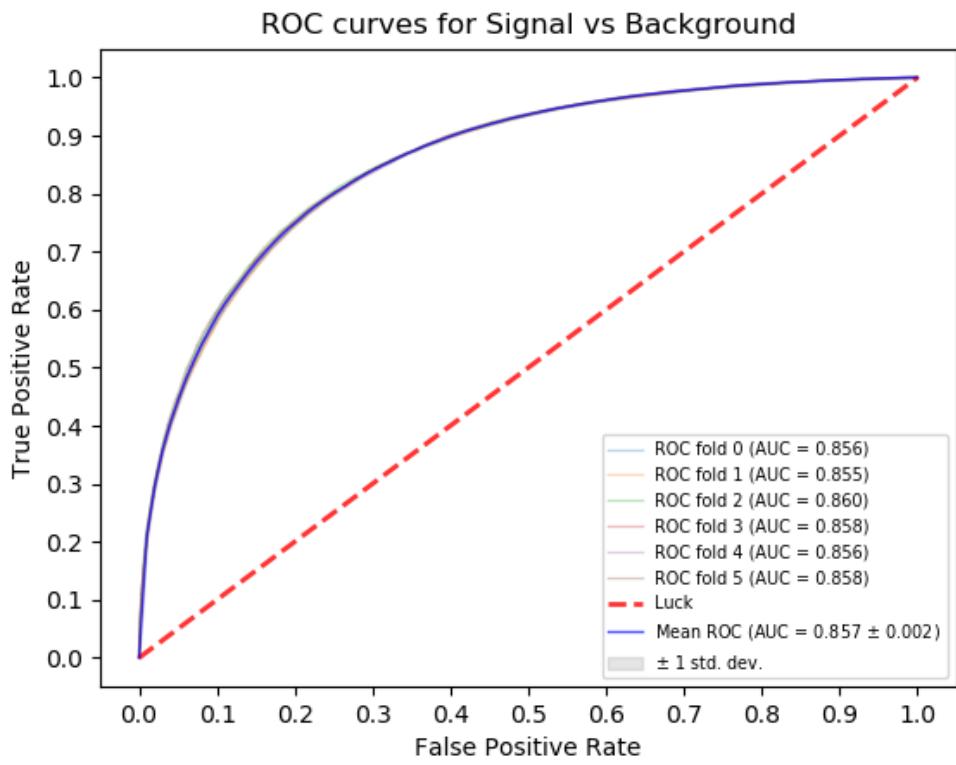
862 Finally this output is passed to a dense layer [natasha ask antonio about this]  
863 and then to a sigmoid activation layer, leading to an overall RNN score. Figure 7.2  
864 shows the RNN discriminant for backgrounds, GGF signals, and VBF signals. The  
865 RNN score is  $\sim 0$  for GGF and background processes and  $\sim 1$  for VBF processes.  
866 Figure 7.4 shows the limits for various signal processes based on the RNN cut  
867 applied. Requiring the RNN score to be  $> 0.8$  was chosen as it provided the best  
868 significance (and signal efficiency) across for this final state and the  $\nu\nu qq$  and  $\ell\ell qq$   
869 channels this channel will be combined with for future publications.



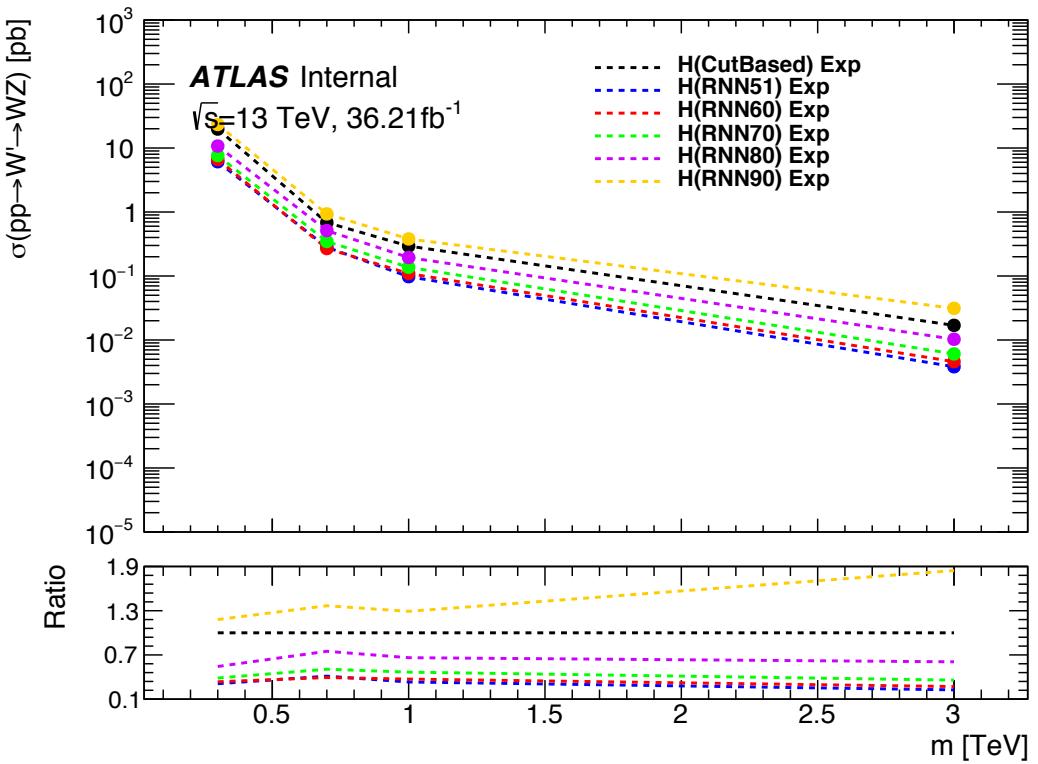
**Figure 7.1:** RNN architecture. Natasha add caption



**Figure 7.2:** RNN Score distribution for ggF and VBF signals and backgrounds.



**Figure 7.3:** ROC curve using k-fold validation for RNN.



**Figure 7.4:** Comparison of GGF Z' limits for different RNN score selections. The bottom panel shows the ratio of the upper limits set for different RNN cuts to the cut-based analysis. In this panel smaller numbers, indicate that the expected upper limit is smaller than the cut-based analysis, which is desired.

## 870 7.4 Topological Cuts

871 Once an event is classified as VBF or GGF via the RNN, it must pass other  
 872 topological cuts that maximize  $S/\sqrt{B}$ . First, to efficiently select events with a  
 873  $W \rightarrow \ell\nu$  candidate exactly one tight lepton is required and  $E_T^{miss} > 100(60)$   
 874 GeV and  $p_{T,\ell\nu} > 200(75)$  GeV in the merged (resolved) analysis to suppress the  
 875 multi-jet background.

876 For the merged analysis, in addition to the  $W \rightarrow \ell\nu$  and  $W/Z \rightarrow J$  selections  
 877 above, the  $\min(p_{T,\ell\nu}, p_{T,J})/m_{WV} > 0.35(0.25)$  for the GGF (VBF) category. To  
 878 reduce  $t\bar{t}$  contamination events with the signal region and  $W+jets$  control region

879 events with at least one b jet with  $\Delta R > 1.0$  from the large-R jet are excluded.  
 880 High purity signal regions require the  $D_2$  and  $W/Z$  mass window cut to be passed,  
 881 whereas the low purity region only requires the  $W/Z$  mass window cut to be  
 882 passed. Finally for events to be classified as tagged the large-R jet must contain  
 883 exactly two b-tagged jets. Untagged events must have no more than one b-tagged  
 884 jet matched to the large-R jet. These selections are shown in Table 7.2. The  
 885 distributions for the variables used in merged analysis for top control regions are  
 886 shown in Figure 7.6-7.9.

887 Events failing the merged selection are then re-analyzed in the resolved cat-  
 888 egory. To enhance resolved signals, the event should contain two high- $p_T$  boson  
 889 candidates that are back-to-back in the  $\phi$  as shown by the cuts in Table 18. Again  
 890 to suppress the  $t\bar{t}$  background in the WCR and SR events are required to have no  
 891 additional b-jets.

892 The  $WV$  system mass,  $m_{WV}$  is reconstructed from the lepton, neutrino, and  
 893 hadronically-decaying boson candidate. The momentum of the neutrino along the  
 894  $z$ -direction is obtained by constraining the  $W$  boson mass of the lepton neutrino  
 895 system to be  $80.3 \text{ GeV}/c^2$ . For complex solutions to this constraint,  $p_Z$  is taken  
 896 as either the real component of the solution or the one with the smaller absolute  
 897 value of the two real solutions. For the resolved analysis,  $m_{WV}$  is reconstructed  
 898 by constraining the  $W(Z)$  dijet system:

$$p_{T,jj}^{corr} = p_{T,jj} \times \frac{m_{W/Z}}{m_{jj}} \quad (7.1)$$

899

$$m_{jj}^{corr} = m_{W/Z} \quad (7.2)$$

900 where  $m_{jj}$  and  $m_{W/Z}$  are the reconstructed invariant mass of the hadronically-  
 901 decaying  $W/Z$  boson and the PDG values of the  $W/Z$  boson masses, respectively.

**Table 7.2:** Summary of selection criteria used to define the signal region (SR),  $W$ +jets control region ( $W$  CR) and  $t\bar{t}$  control region ( $t\bar{t}$  CR) for merged 1-lepton channel.

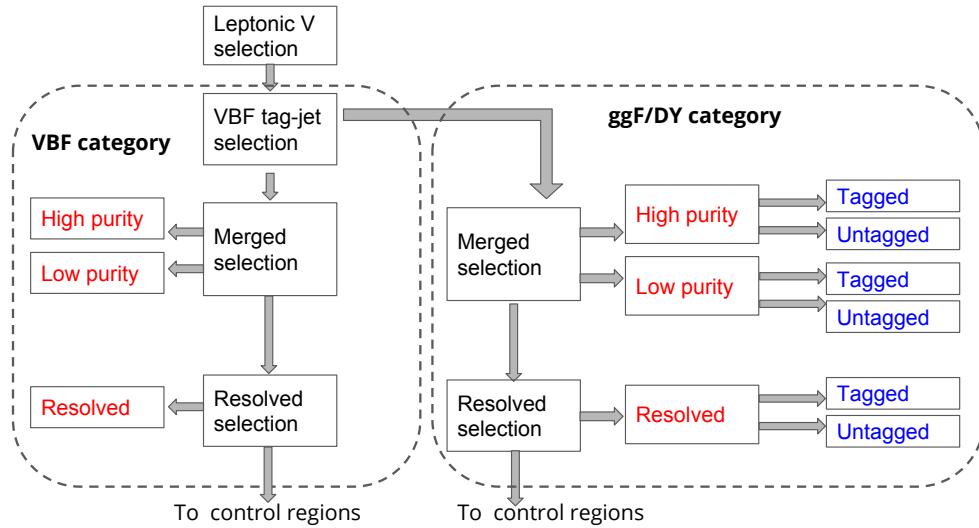
Selection	SR		$W$ CR (WR)		$t\bar{t}$ CR (TR1)		
	HP	LP	HP	LP	HP	LP	
$W \rightarrow \ell\nu$	Num of Tight leptons			1			
	Num of Loose leptons			0			
	$E_T^{\text{miss}}$			> 100 GeV			
	$p_T(\ell\nu)$			> 200 GeV			
$W/Z \rightarrow J$	Num of large- $R$ jets			$\geq 1$			
	$D_2$ cut	pass	fail	pass	fail	pass	fail
	$W/Z$ mass window cut	pass	pass	fail	fail	pass	pass
	Numb. of associated VR track jets $b$ -tagged	For $Z \rightarrow J$ : $\leq 1$ ( $= 2$ ) for untagged (tagged) category					
	$\min(p_{T,\ell\nu}, p_{T,J}) / m_{WV}$	> 0.35(0.25) for DY/ggF (VBF) category					
	Top-quark veto	Num of $b$ -tagged jets outside of large- $R$ jet	0			$\geq 1$	
Pass VBF selection		no (yes) for DY/ggF (VBF) category					

902 A summary of the resolved selections is shown in Table 7.3. The distributions for  
903 the variables used in the resolved analysis in the TCR are shown in Figure 7.10,  
904 7.11.

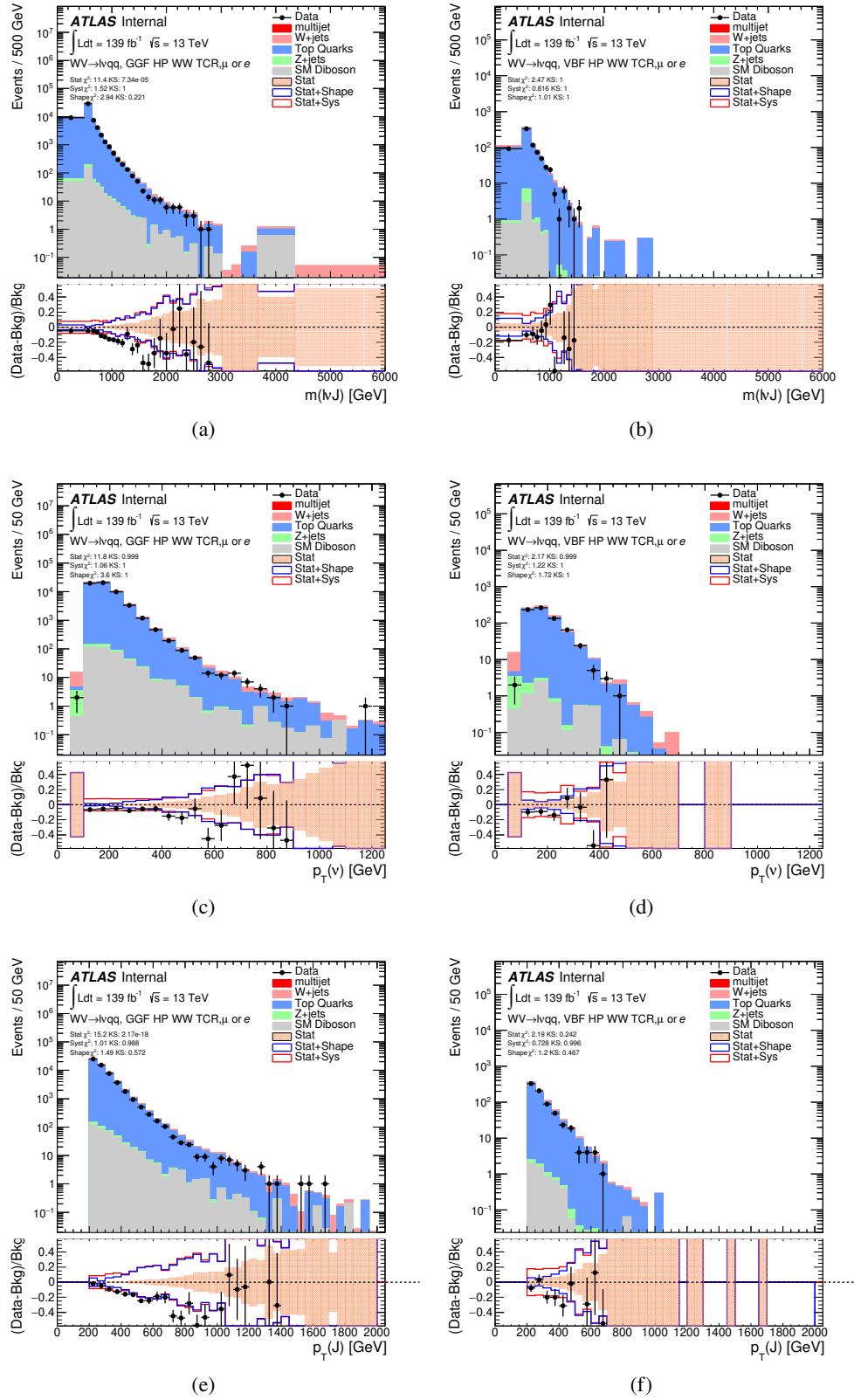
905 The analysis cutflow is shows in Figure 7.5. Events classified as VBF events  
906 are classified as Merged High purity, low purity or resolved signal region selections  
907 sequentially. If the event does not pass any of these selections but passes a VBF  
908 control region selection it is classified as a VBF CR event. If the event fails the  
909 VBF selection it is then checked if it passes the Merged High purity, Low purity  
910 or resolved signal region selections (NB: for the  $WZ$  decay modes all the regions  
911 have tagged and untagged categories). If the event fails all the GGF signal region  
912 selections, it is then kept for GGF control region selections, if it passes those  
913 selections.

**Table 7.3:** The list of selection cuts in the resolved analysis for the  $WW$  and  $WZ$  signal regions (SR),  $W+\text{jets}$  control region (WR) and  $t\bar{t}$  control region (TR).

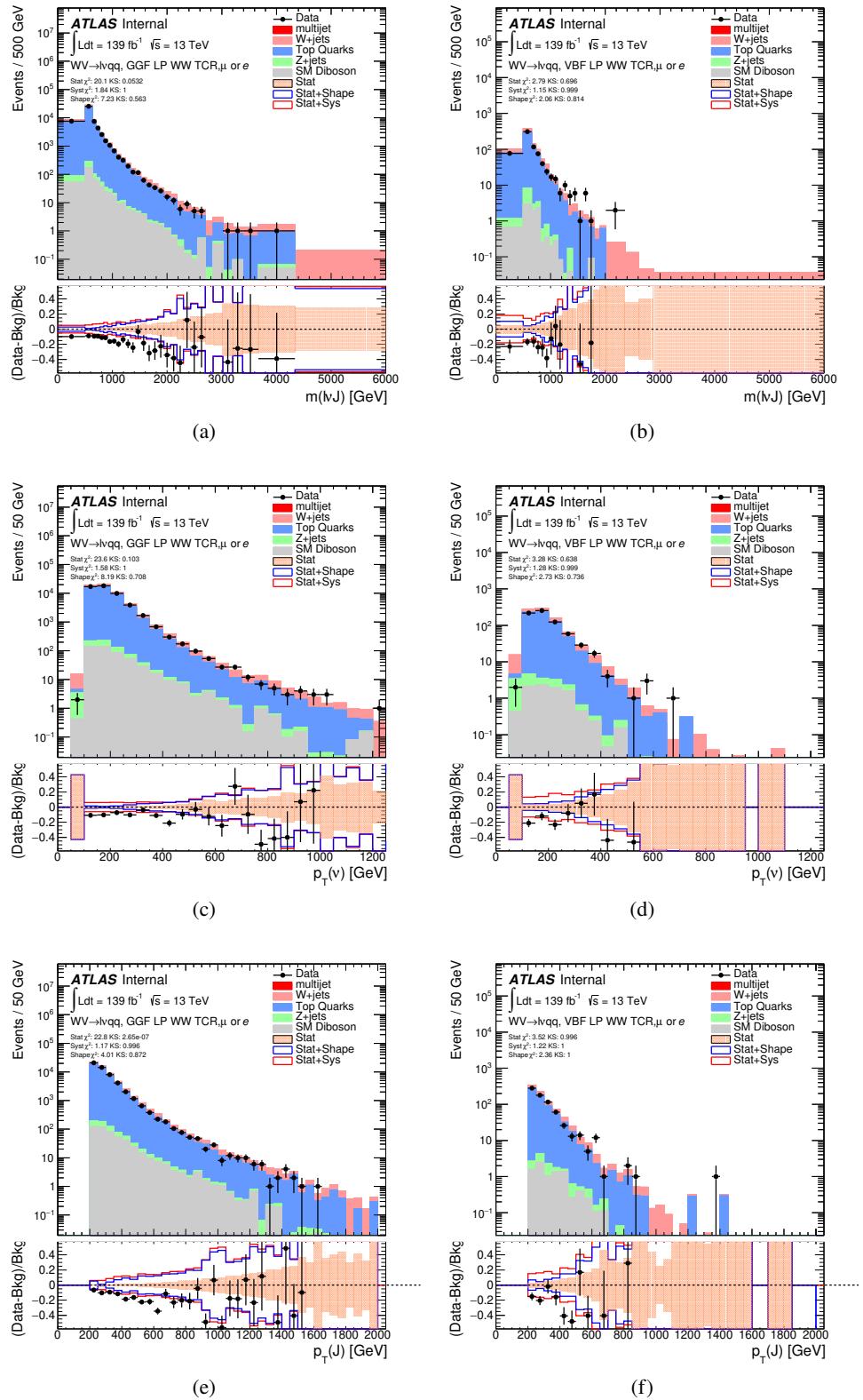
cuts	SR	$W$ CR (WR)	$t\bar{t}$ CR (TR1)
$W \rightarrow \ell\nu$	Number of Tight leptons	1	
	Number of Loose leptons	0	
	$E_T^{\text{miss}}$	$> 60 \text{ GeV}$	
	$\cancel{p}_T(\ell\nu)$	$> 75 \text{ GeV}$	
$W/Z \rightarrow jj$	Number of small-R jets	$\geq 2$	
	Leading jet $p_T$	$> 60 \text{ GeV}$	
	Subleading jet $p_T$	$> 45 \text{ GeV}$	
	$Z \rightarrow q\bar{q}$ $W \rightarrow q\bar{q}$	$78 < m_{jj} < 105 \text{ GeV}$ $68 < m_{jj} < 98 \text{ GeV}$	$50 < m_{jj} < 68 \text{ GeV}$ or $105 < m_{jj} < 150 \text{ GeV}$
Topology cuts	Num. of $b$ -tagged jets	For $Z \rightarrow jj$ : $\leq 1$ ( $= 2$ ) for untagged (tagged) category	
	$\Delta\phi(j, \ell)$	$> 1.0$	
	$\Delta\phi(j, E_T^{\text{miss}})$	$> 1.0$	
	$\Delta\phi(j, j)$	$< 1.5$	
	$\Delta\phi(\ell, E_T^{\text{miss}})$	$< 1.5$	
Topo veto	$\min(p_{T,\ell\nu}, p_{T,jj}) / m_{WW}$	$> 0.35(0.25)$ for DY/ggF (VBF) category	
	Number of additional $b$ -tagged jets	0	$\geq 1$
Pass VBF selection		no (yes) for DY/ggF (VBF) category	



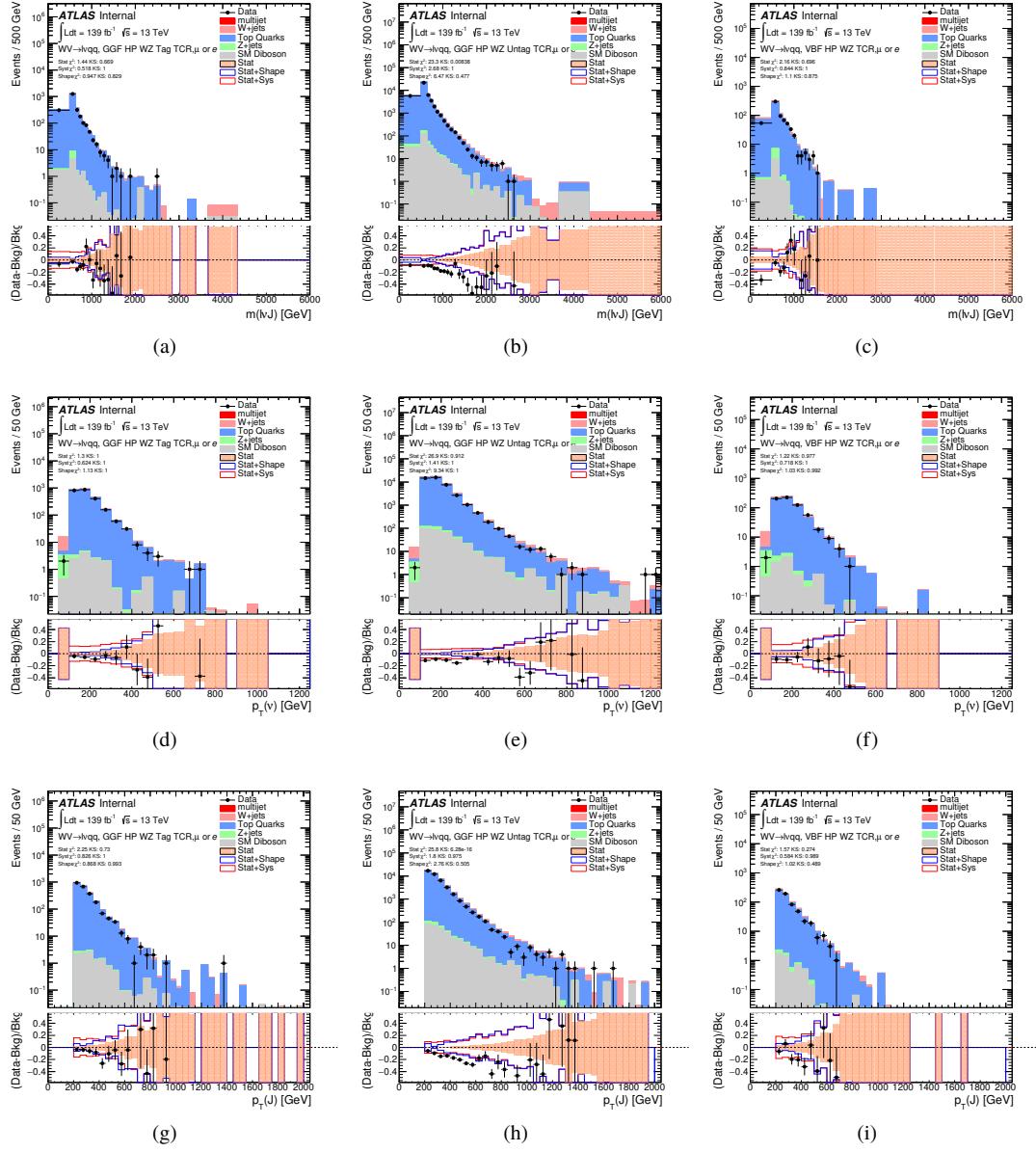
**Figure 7.5:** Event Categorization. Natasha write more.



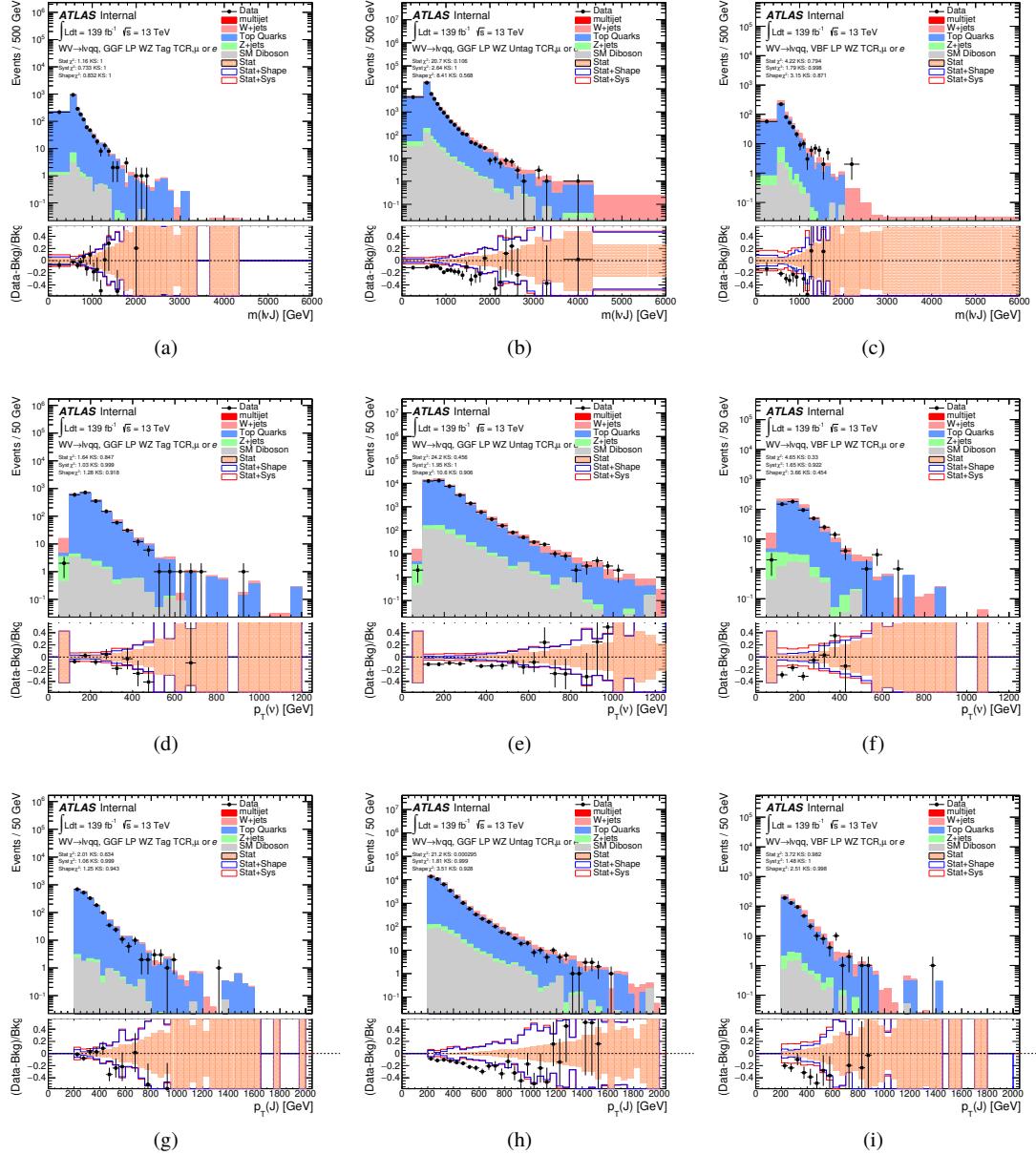
**Figure 7.6:** Data MC comparison for the merged  $WW$  HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



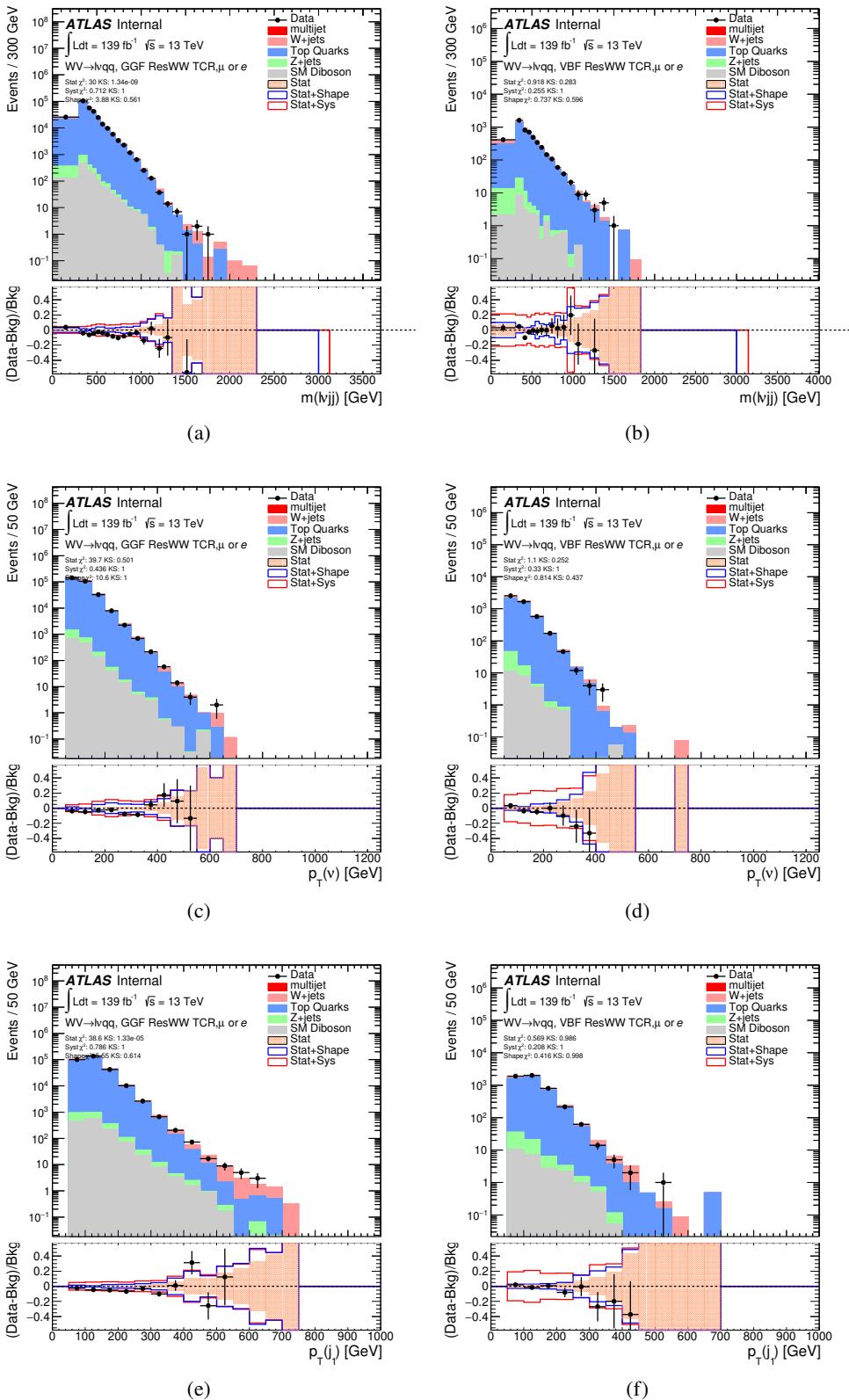
**Figure 7.7:** Data MC comparison for the merged  $WW$  LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.<sup>83</sup>



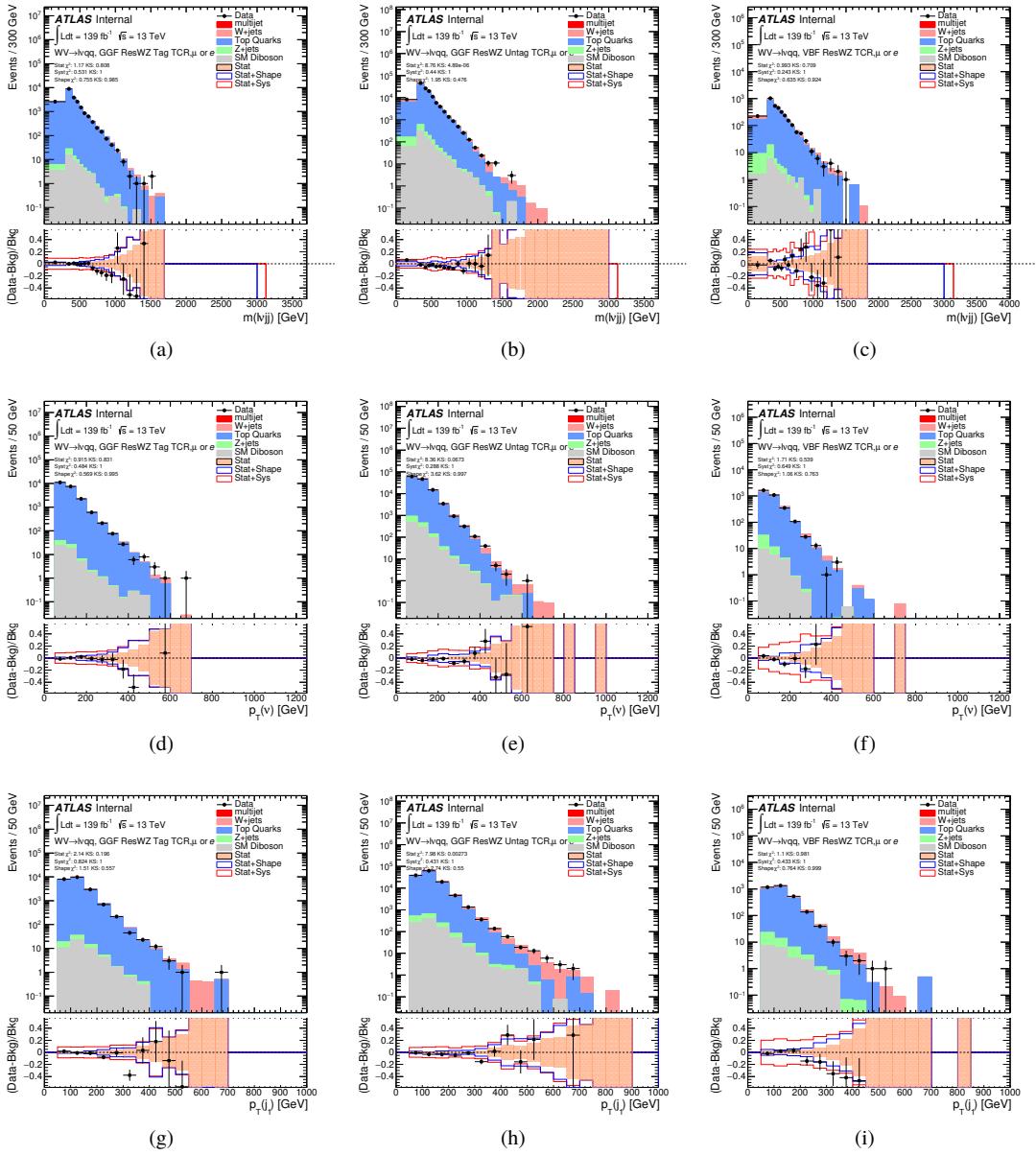
**Figure 7.8:** Data MC comparison for the merged  $WZ$  HP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



**Figure 7.9:** Data MC comparison for the merged  $WZ$  LP TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.



**Figure 7.10:** Data MC comparison for the resolved  $WW$  TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.<sup>86</sup>



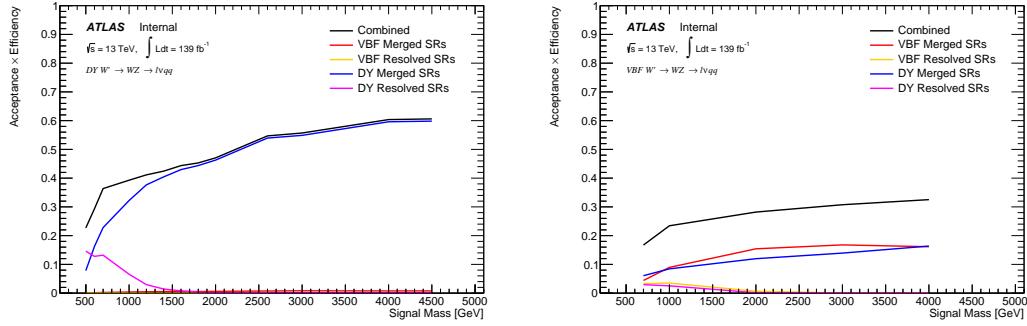
**Figure 7.11:** Data MC comparison for the resolved  $WZ$  TCR. The bottom panel shows the ratio of the difference between data and simulation to simulation. The red bands include the all systematic and statistical uncertainties on the background.

914    **7.5 Selection Acceptance times efficiency for Sig-**  
 915    **nal Events**

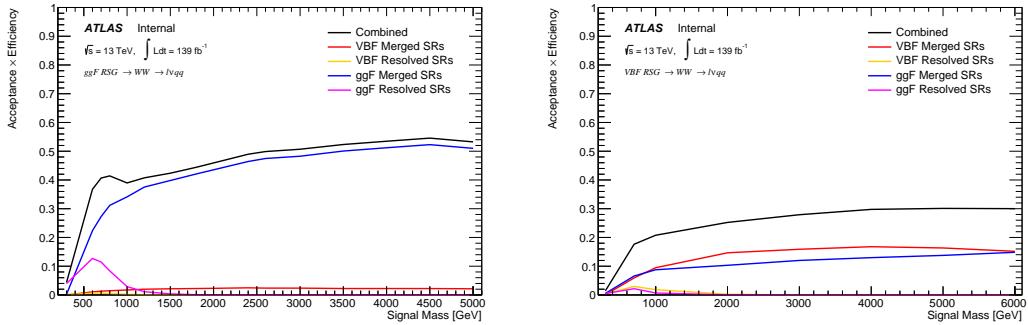
916    The acceptance times efficiency for the signal region selection is defined as:

$$A \cdot \epsilon = \frac{N_{\text{events selected}}^{\text{truth}}}{N_{\text{events generated}}^{\text{truth}}} \cdot \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events selected}}^{\text{truth}}} = \frac{N_{\text{events selected}}^{\text{reco}}}{N_{\text{events generated}}^{\text{truth}}} \quad (7.3)$$

917    The distributions of  $A \cdot \epsilon$  as a function of the resonance mass for the different spin  
 918    models are shown in Figures 7.13 - ??.



**Figure 7.12:** Selection acceptance times efficiency for the  $W' \rightarrow WZ \rightarrow \ell\nu qq$  events from MC simulations as a function of the  $W'$  mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the  $WW \rightarrow \ell\nu J$  selection and the resolved regions of the  $WW \rightarrow \ell\nu jj$  selection.



**Figure 7.13:** Selection acceptance times efficiency for the  $G \rightarrow WW \rightarrow \ell\nu qq$  events from MC simulations as a function of the  $G$  mass for (a) Drell-Yan and (b) VBF production, combining the merged HP and LP signal regions of the  $WW \rightarrow \ell\nu J$  selection and the resolved regions of the  $WW \rightarrow \ell\nu jj$  selection.

## 919 7.6 Background Estimate

### 920 7.6.1 Control Regions

921 To more accurately model the two dominant backgrounds in this analysis,  
 922  $W+jets$  and  $t\bar{t}$ , control regions are constructed for each. These control regions  
 923 are dominated by these processes and used to extract normalization factors in  
 924 the final likelihood fit that are then used in the signal region estimates. For the  
 925  $t\bar{t}$  control region the event must contain at least one such b jet. The WCR is  
 926 constructed using the  $m_{jj/J}$  mass window sidebands. All other backgrounds are  
 927 estimated using simulation, except fake lepton backgrounds, which are derived  
 928 using a data-driven method.

### 929 7.6.2 Fake Lepton Backgrounds

930 Backgrounds in this analysis containing real leptons (e.g.  $W/Z+jets$ , diboson,  
 931  $t\bar{t}$ , single- $t$ ) are well-modeled with simulated samples and constrained with data  
 932 from CRs. However, the fake lepton background (also referred to as the multijet  
 933 background) is not well-modeled with simulation. For this reason, the multijet

background is extracted from data. Heavy flavor decay products, jets, and converted photons can be mistakenly reconstructed as electrons. Fake electrons often arise from jet fakes while non-prompt muons usually arise from heavy flavor decay. For this analysis, these fake electrons generally fail the electron ID criteria and fake muons fail the muon isolation requirement. Therefore, to derive the multijet template shape the SR and CR selections and inverted lepton requirements are used as seen in Table 7.4. NB: by inverting the lepton isolation/identification criteria the SRs and CRs are orthogonal.

The template shape of the MJ background is determined by using a multijet validation region (MJVR) that requires the inverted lepton isolation/identification requirement and the two signal jets to satisfy the  $m_{jj}$  requirement used in the  $W+jets$  CRs. The  $E_T^{miss}$  distribution in MJCR is shown in Figure 7.14 for 2017 data. The template is then extracted by subtracting the data in the MJVR from the electroweak background processes. The resulting template and electroweak backgrounds are then fit to data. In this fit, the  $E_T^{miss}$  distribution compared to data to extract electroweak background, multijet electron and muon background normalizations. The fitted scale factors from this MJVR template are then applied in the MJCR template. The electron and muon background normalizations in the MJCR template are parameters in the final simultaneous fit. Technically, there should be a separate template for every CR and SR, but some MJ regions have insufficient statistics to do this. Additionally, the shapes for the MJ templates for VBF and ggF regions are found to be compatible within statistical uncertainty. Therefore, the sample MJ template used for VBF and ggF CR/SRs, but with different pre-MJ-fit scale factors.

This template method was validated using WCR and full Run 2 data. The results of the fit are shown in Table 7.5. The multijet contribution in the muon

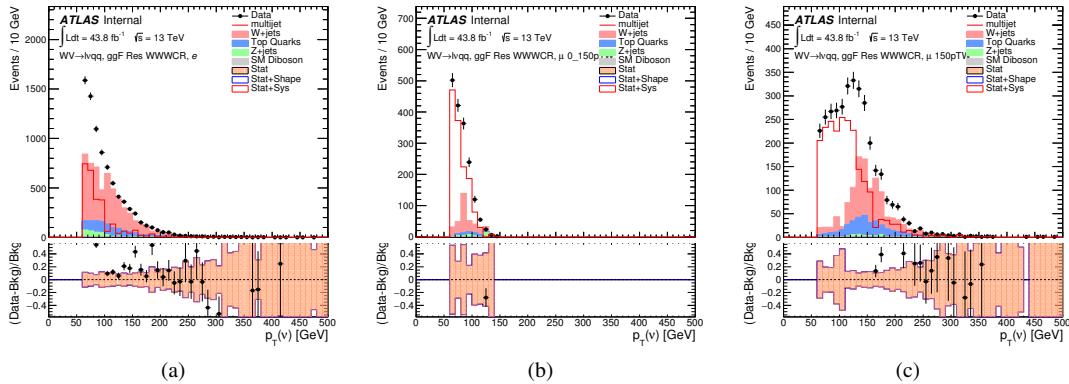
960 channel for  $p_T^W > 150$  GeV is consistent with zero, and therefore neglected in  
 961 the final fit. Applying the extracted normalization factor to MJVR in WCRs for  
 962 various kinematic variables such as  $E_T^{miss}$ ,  $W$  transverse mass, lepton  $p_T$ , and the  
 963 invariant mass as show in Figures 7.15 -7.24. These figures show good agreement  
 964 between the data and background estimate.

**Table 7.4:** Definitions of “inverted” leptons used in multijet control region

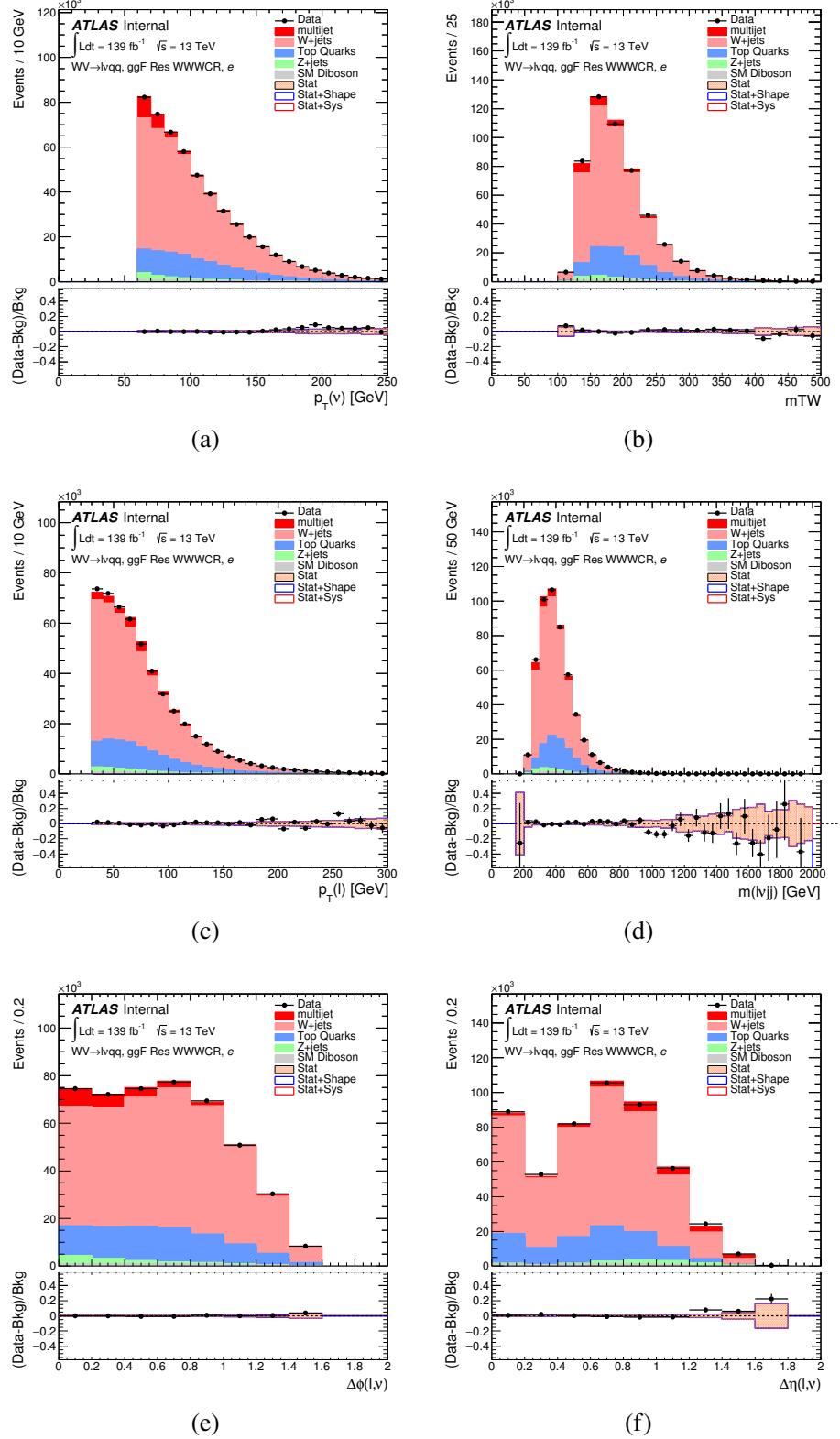
	Criterion	signal lepton	inverted lepton
Electron	ID	TightLH	MediumLH !TightLH
	Calo Isolation	FixedCutHighPtCaloOnlyIso	FixedCutHighPtCaloOnlyIso
Muon	ID	WHSignalMuon	WHSignalMuon
	Track Isolation	FixedCutTightTrackOnlyIso	!FixedCutTightTrackOnlyIso $ptvarcone30/pt < 0.07^*$

\*Only applied to events with  $pTW < 150\text{GeV}$

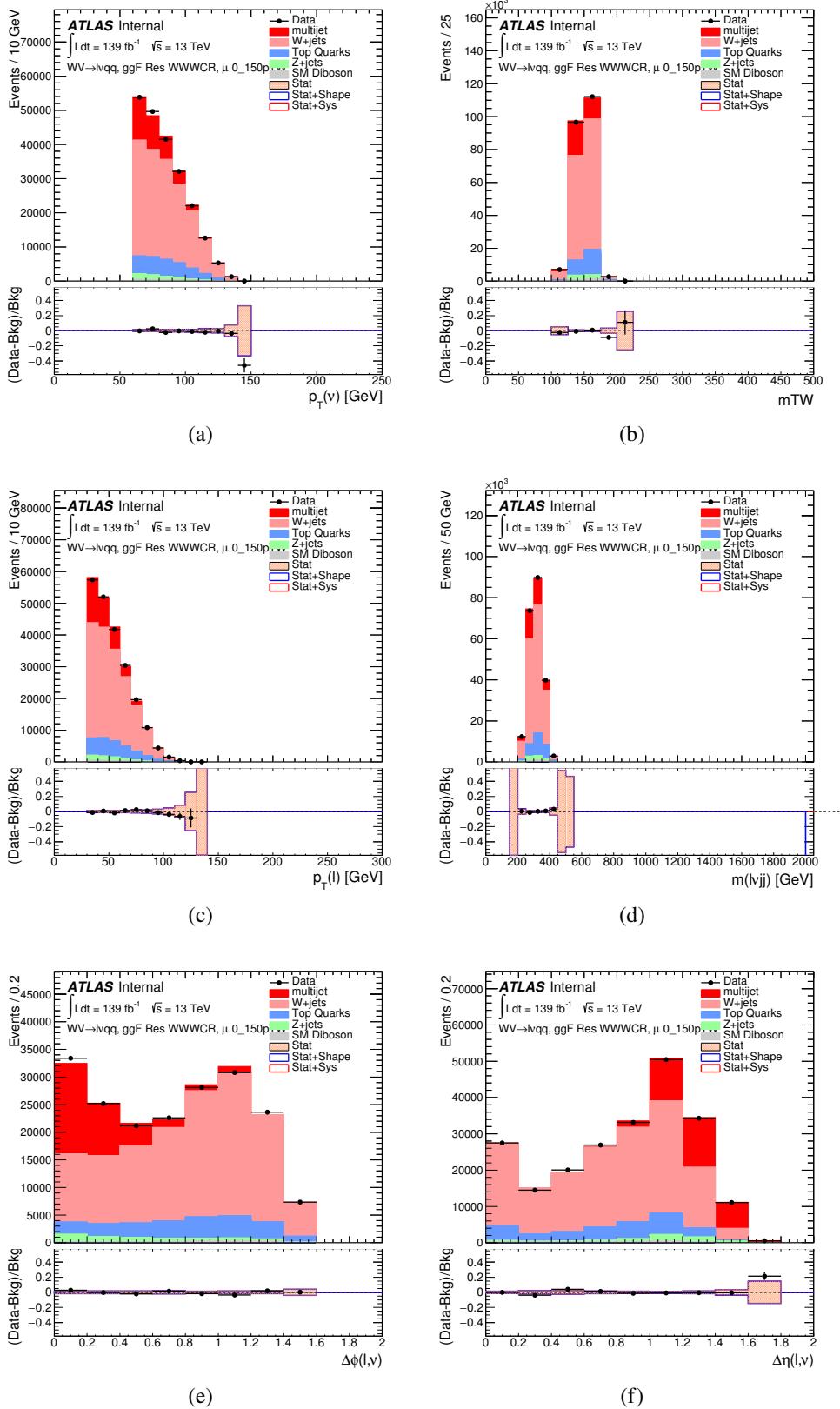
965



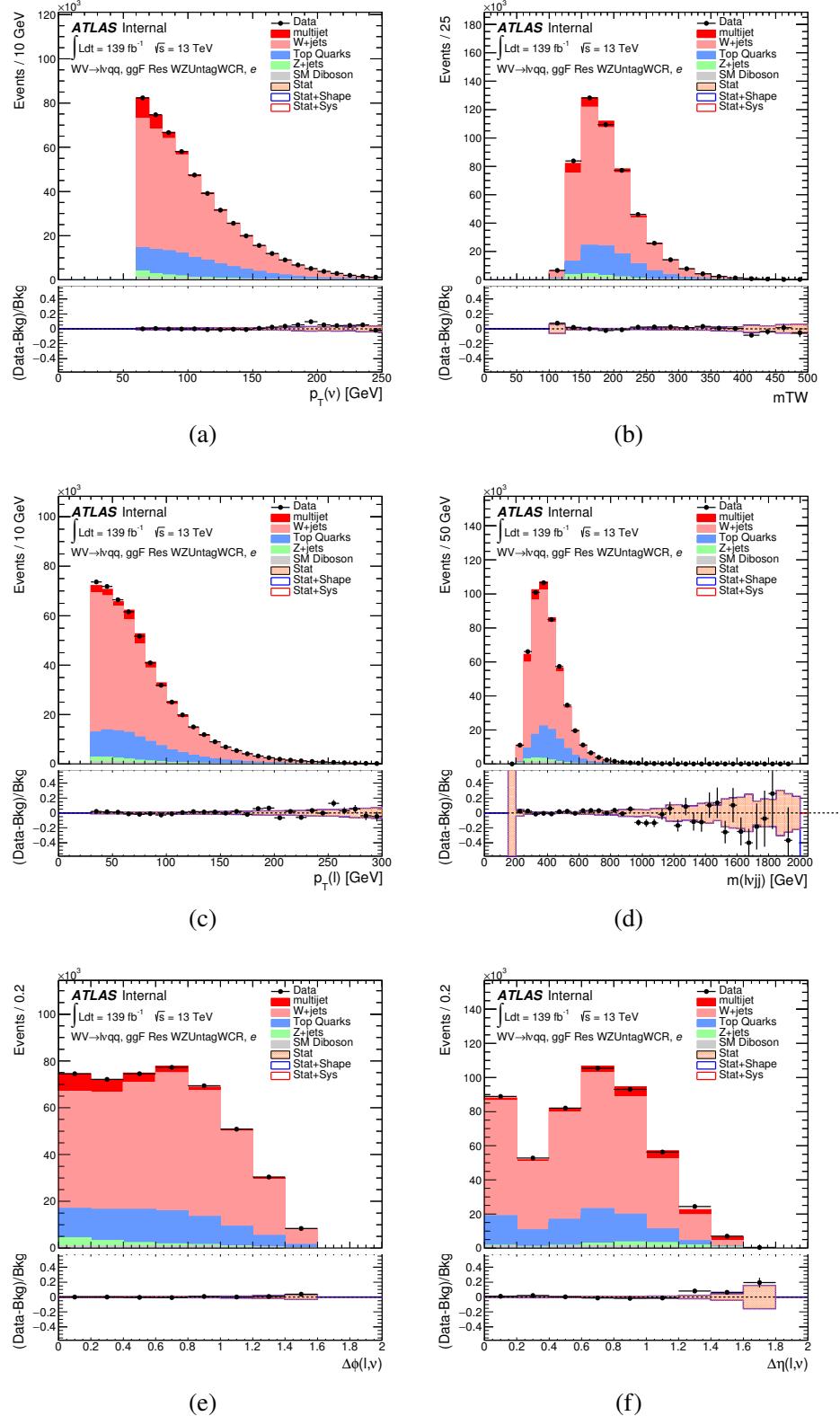
**Figure 7.14:** The  $E_T^{miss}$  distribution in MJCR for 2017 data in the electron channel(left), muon channel with  $W$ -boson  $pT < 150$  GeV (center) and  $> 150$  GeV (right). Multi-jet templates are calculated as remaining data components after excluding known MC



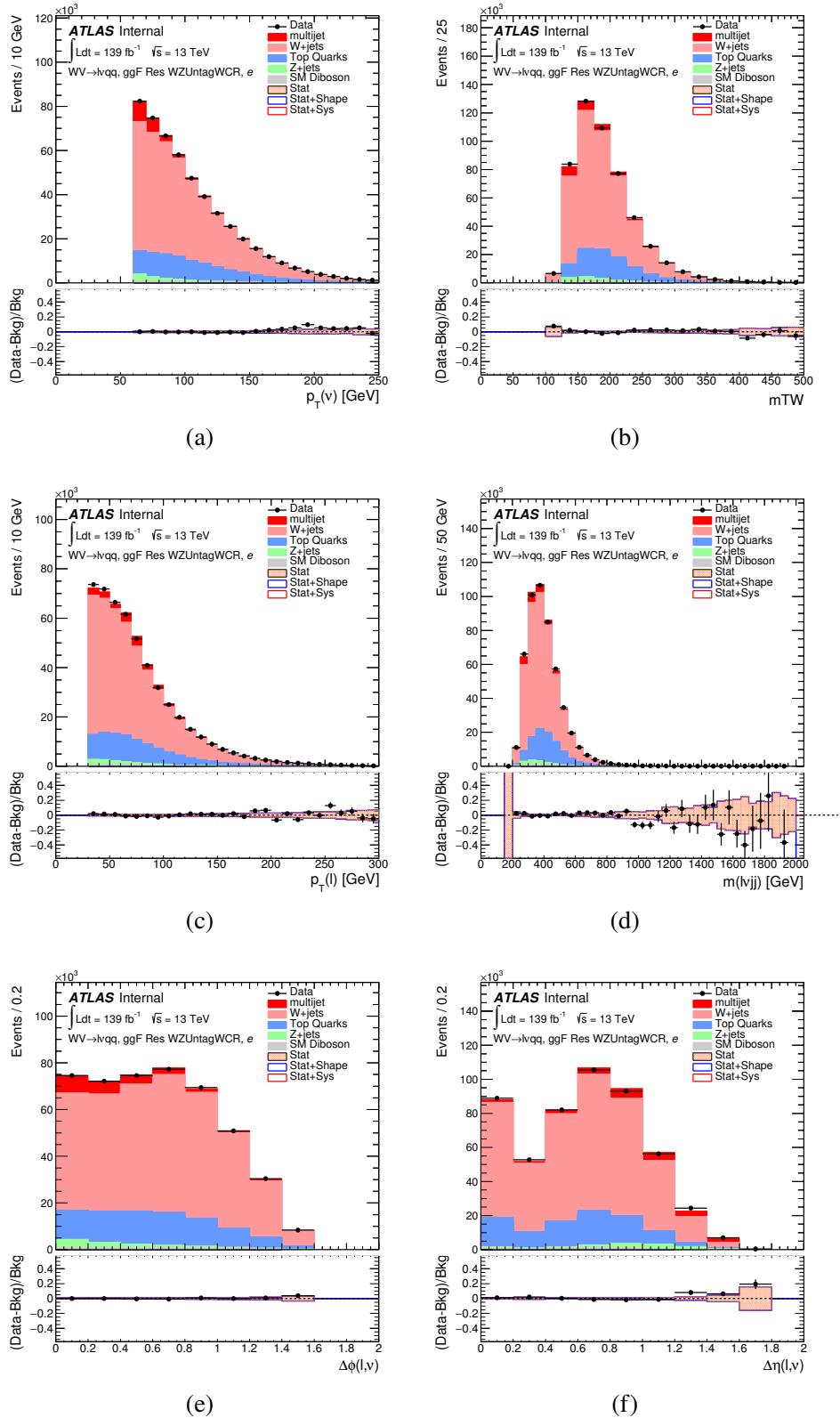
**Figure 7.15:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WW$  electron channel. The MJ template is obtained from the pre-MJ-fit.



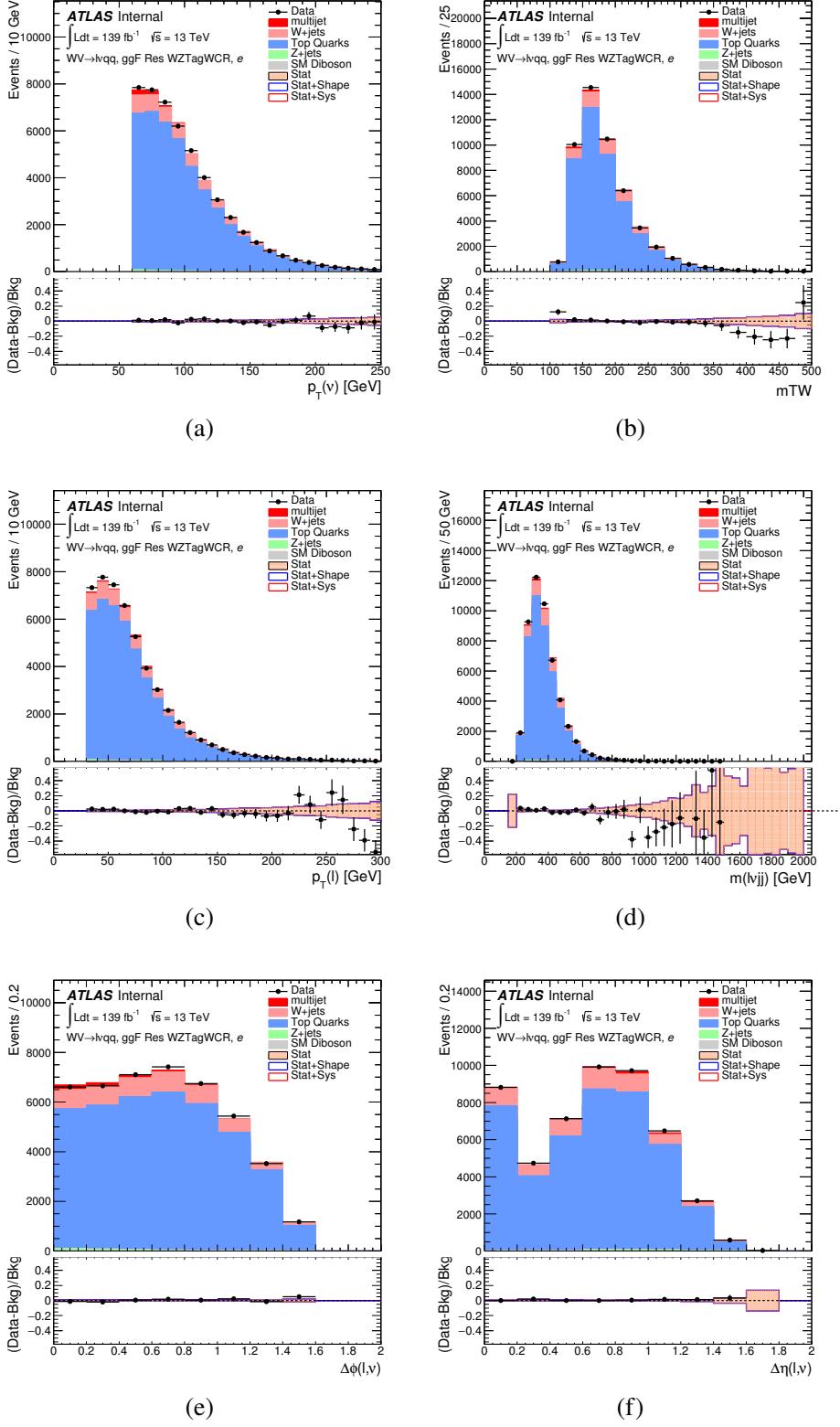
**Figure 7.16:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WW$  muon channel. The MJ template is obtained from the pre-MJ-fit.



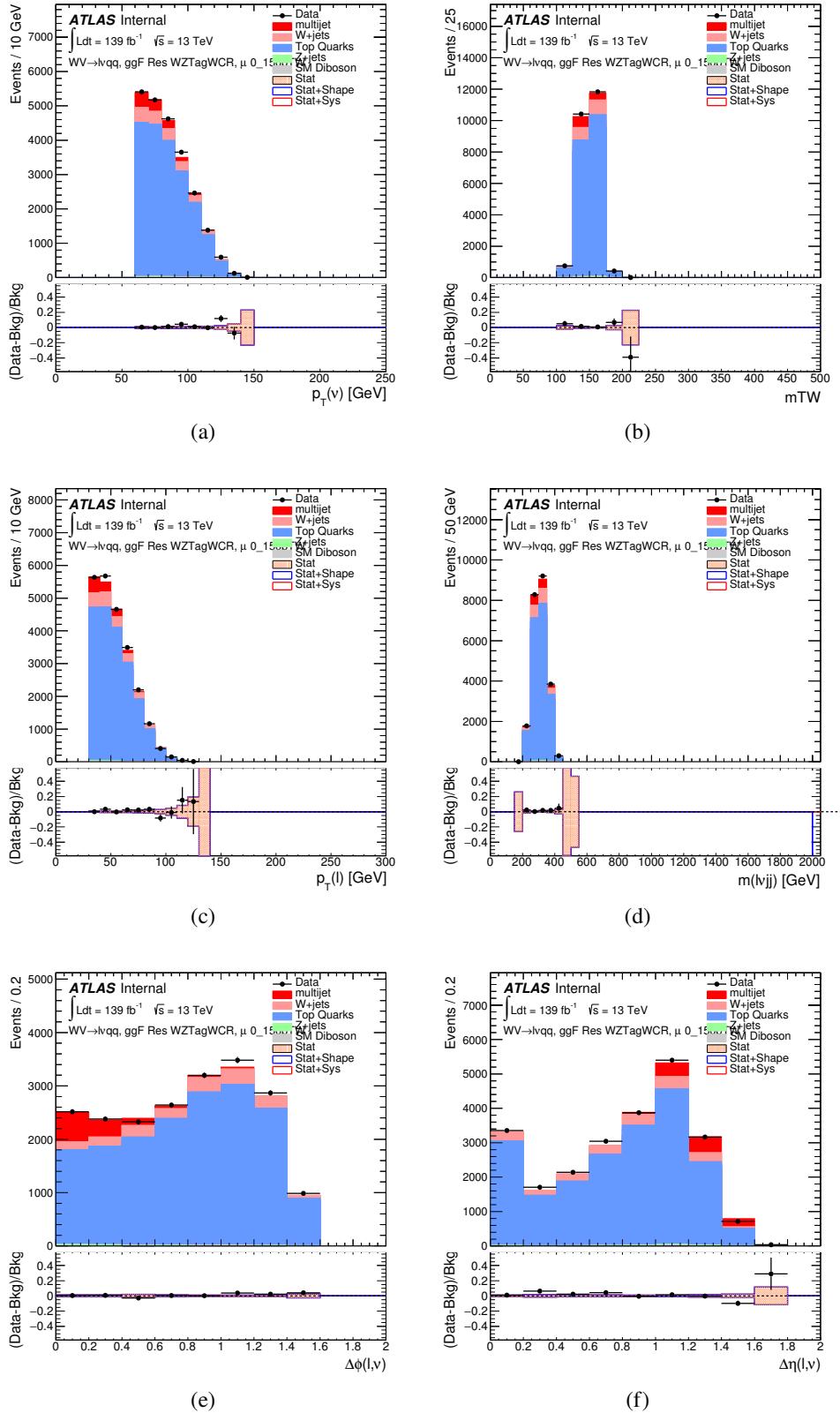
**Figure 7.17:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag electron channel. The MJ template is obtained from the pre-MJ-fit.



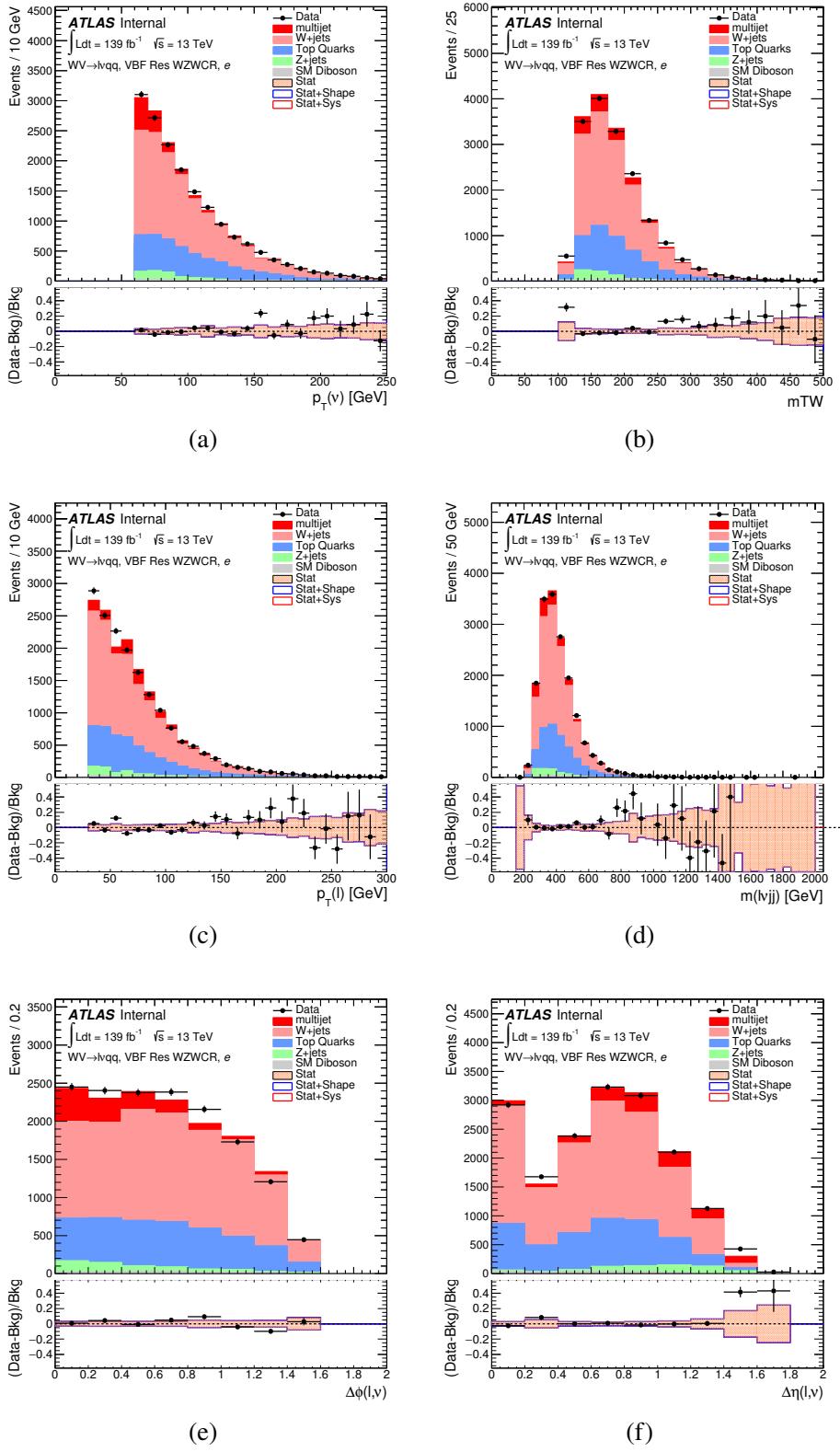
**Figure 7.18:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag muon channel. The MJ template is obtained from the pre-MJ-fit.



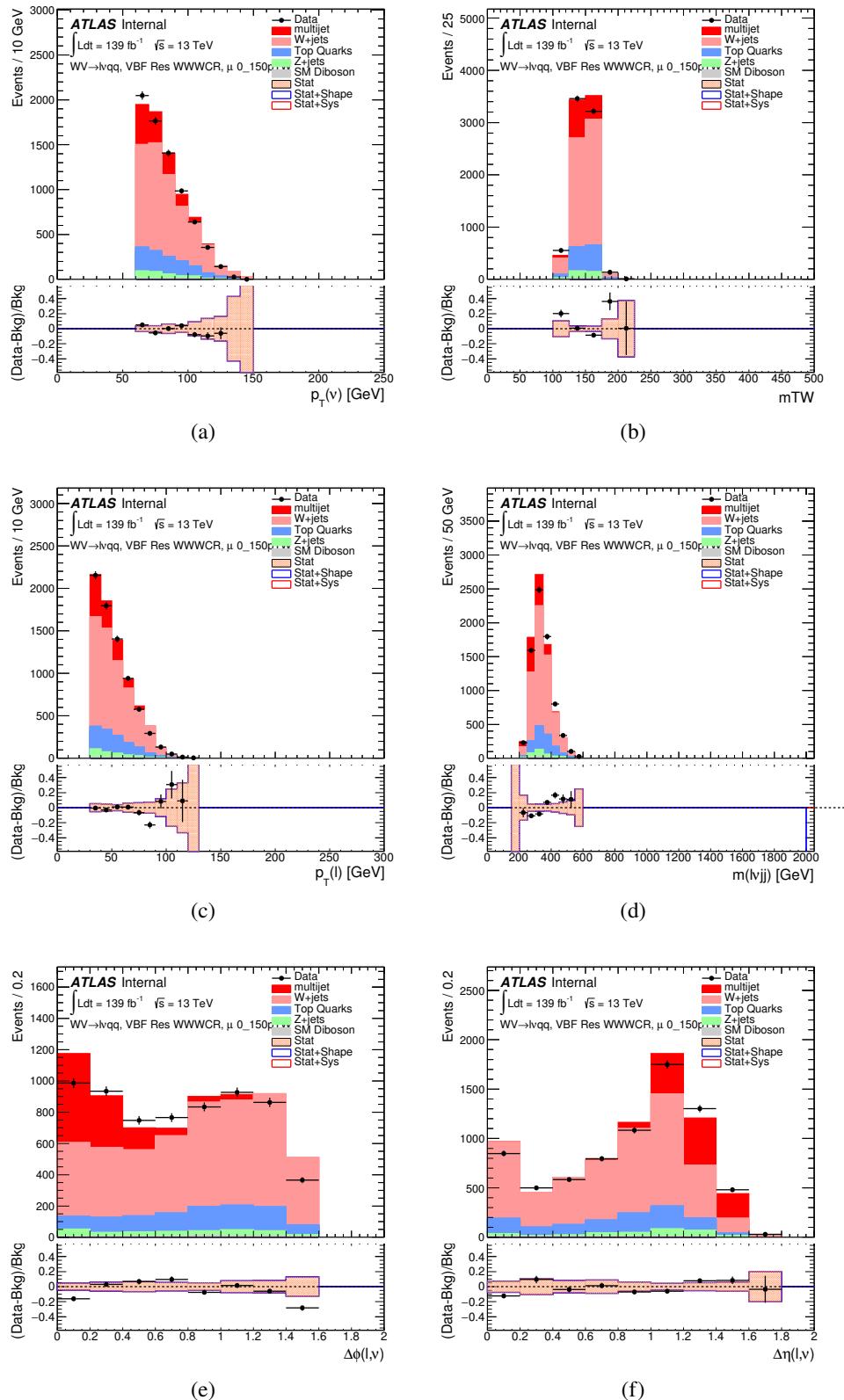
**Figure 7.19:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag electron channel. The MJ template is obtained from the pre-MJ-fit.



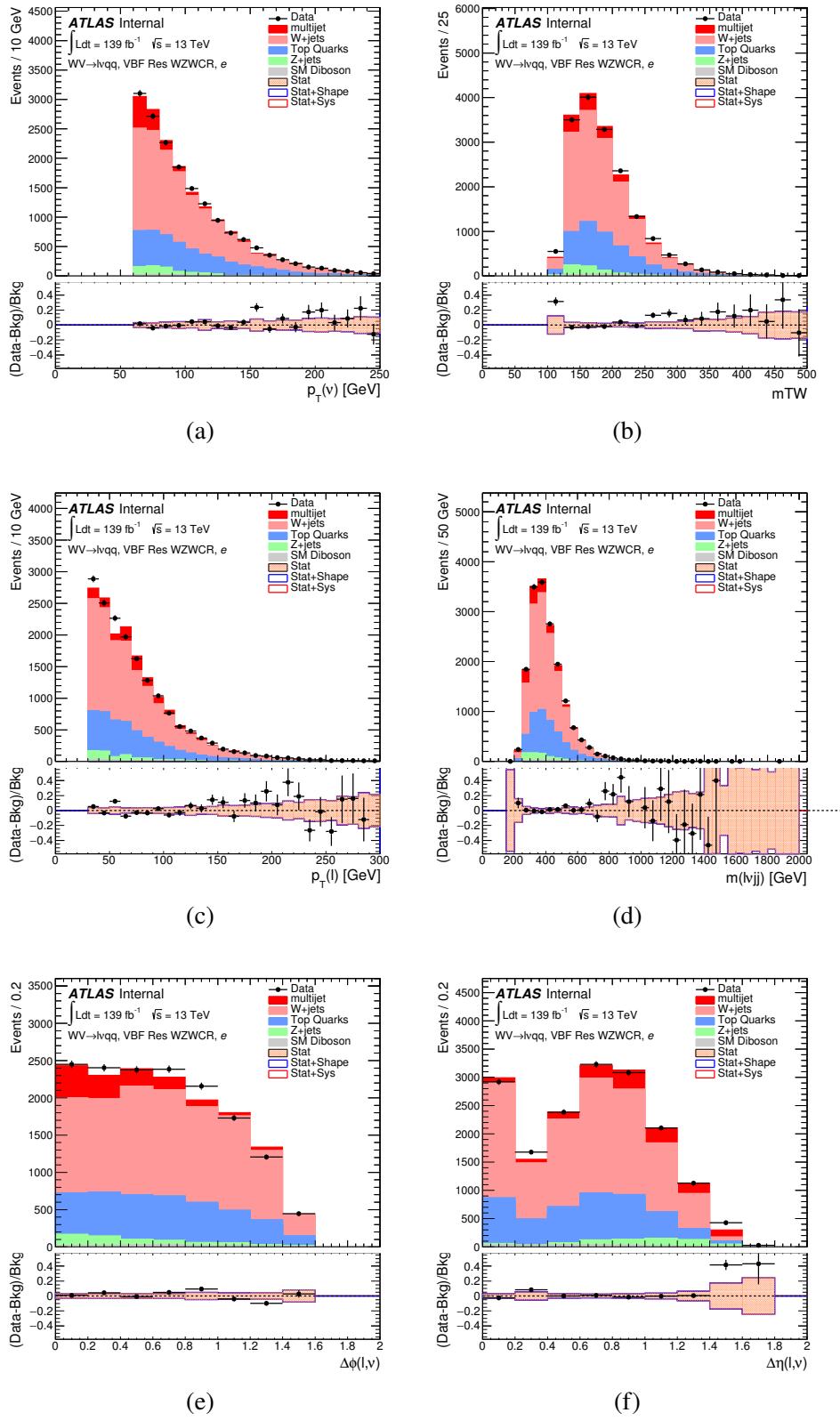
**Figure 7.20:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the  $WZ$  untag muon channel. The MJ template is obtained from the pre-MJ-fit.



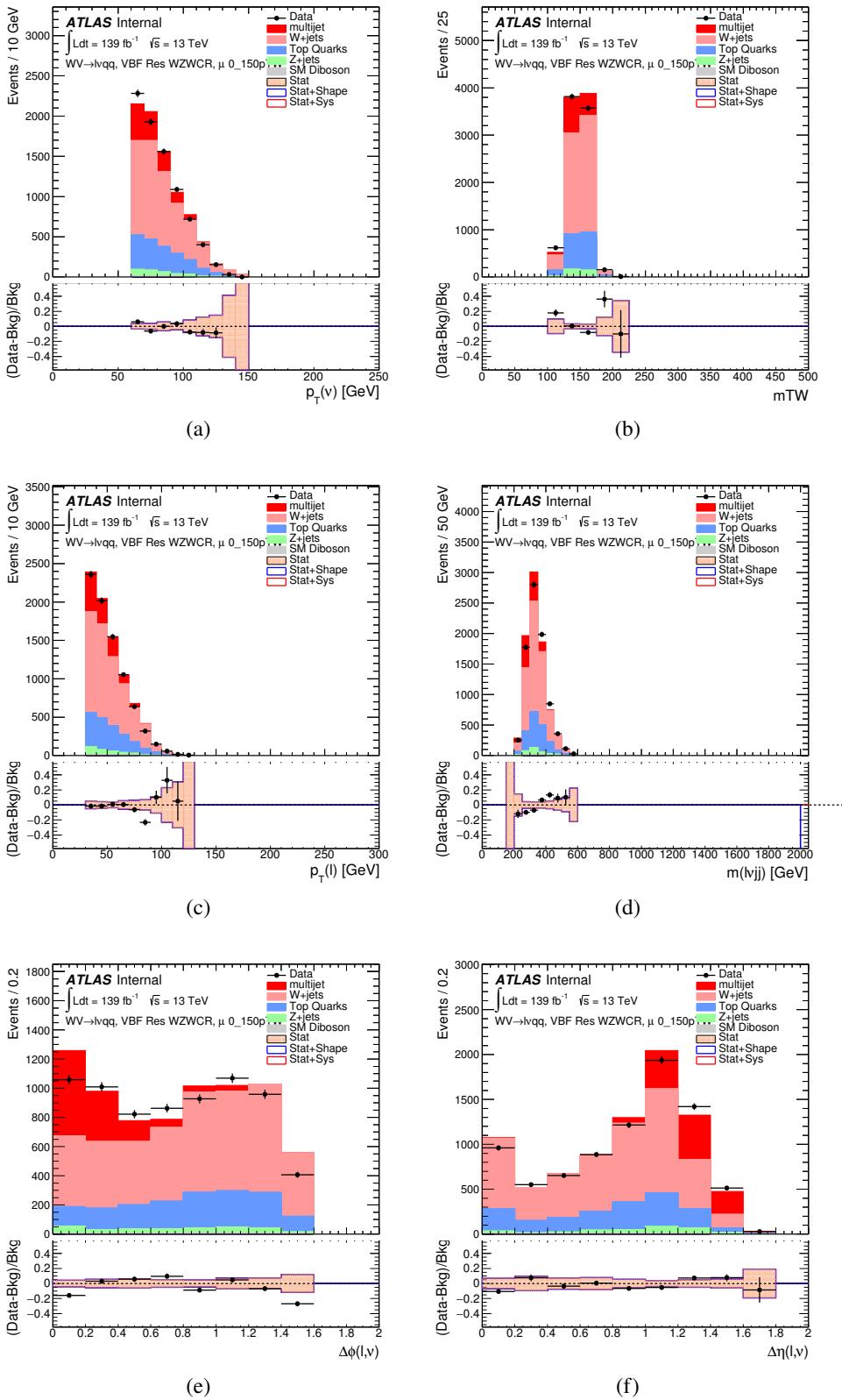
**Figure 7.21:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WW$  electron channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 7.22:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{\ell\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WW$  muon channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 7.23:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WZ$  electron channel. The MJ template is obtained from the pre-MJ-fit.



**Figure 7.24:** Postfit Data/MC comparison of distributions of  $E_T^{miss}$ ,  $m_T^W$ , lepton and neutrino  $p_T$ ,  $m_{l\nu jj}$ , lepton- $\nu$  angular distance in the VBF  $WZ$  muon channel. The MJ template is obtained from the pre-MJ-fit.

Full Run 2  
ggF Res WWWCR

Sample	Yield	R.U.	SF
Top&W	$645040 \pm 1971.68$	0.31%	0.998
Z&VV	24075.9		fixed
MJ_el	$24156.3 \pm 1224.62$	5.06%	3.973
MJ_mu	$35528.5 \pm 923.94$	2.60%	9.019

ggF Res WZ01bWCR

Sample	Yield	R.U.	SF
Top&W	$644690 \pm 1981.4$	0.31%	0.997
Z&VV	24075.9		fixed
MJ_el	$24366.5 \pm 1232.69$	5.05%	3.874
MJ_mu	$35528.5 \pm 921.27$	2.58%	8.746

ggF Res WZ2bWCR

Sample	Yield	R.U.	SF
Top&W	$71236.5 \pm 688.74$	0.97%	1.031
Z&VV	518.5		fixed
MJ_el	$595.63 \pm 449.34$	75.44%	0.094
MJ_mu	$1196.9 \pm 222.13$	18.56%	0.294

VBF Res WWWCR

Sample	Yield	R.U.	SF
Top&W	$19032.3 \pm 364.43$	1.91%	0.928
Z&VV	1091.63		fixed
MJ_el	$1425.73 \pm 214.42$	15.03%	0.235
MJ_mu	$1281.36 \pm 157.21$	11.83%	0.314

VBF Res WZWCR

Sample	Yield	R.U.	SF
Top&W	$21341.8 \pm 392.21$	1.84%	0.942
Z&VV	1111.75		fixed
MJ_el	$1413.76 \pm 230.36$	16.29%	0.225
MJ_mu	$1281.36 \pm 157.21$	12.27%	0.314

**Table 7.5:** Fit validation result in WCRs for 2015+16 data. The fit is done in various WCRs, in order to obtain the corresponding scale factors for MJ templates: ggF resolved WCR for the  $WW \rightarrow lvqq$  selection, ggF resolved untagged WCR for the  $WZ \rightarrow lvqq$  selection, ggF resolved tagged WCR for the  $WZ \rightarrow lvqq$  selection, VBF resolved WCR for the  $WW \rightarrow lvqq$  selection, and VBF resolved WCR for the  $WZ \rightarrow lvqq$  selection. Post-fit event yields for electroweak processes and MJ contributions are shown. The SF column shows the corresponding normalization scale factors for electroweak processes from the fit. R.U. stands for relative uncertainty.

966 **Chapter 8**

967 **Systematic Uncertainties**

968 This section describes the sources of systematic uncertainties considered in  
969 this analysis. These uncertainties are divided into three categories: experimental  
970 uncertainties, background modeling uncertainties, and theoretical uncertainties on  
971 signal processes. In the statistical analysis each systematic uncertainty is treated  
972 as a nuisance parameter estimated on the  $m_{VV}$  distribution.

973 **8.1 Experimental Systematics**

974 The uncertainty on the integrated luminosity of the dataset used is 1.7% and  
975 a systematic in the final fit. This uncertainty was calculated using  $x - y$  beam  
976 separation scans [ref P55].

977 An additional source of systematic uncertainty is assigned to the pileup mod-  
978eling in MC samples. This ensures simulated detector response and particle re-  
979construction conditions are as similar as possible. The distribution of the average  
980number of interactions per bunch crossing applied to simulation is called the  $\mu$  pro-  
981file. The pileup modeling uncertainty is accounted for by re-weighting simulated  
982events so the average number of interactions per bunch crossing varies within its

983 uncertainty due to systematics from vertex reconstruction [ref ATL-COM-SOFT-  
984 2015-119]. The associated re-weighting factors are propagated through the entire  
985 analysis chain to construct a systematic uncertainty on  $m_{VV}$ .

986 The single-lepton and  $E_T^{miss}$  triggers used are not fully efficient, so scale factors  
987 are applied to simulation to more accurately model the data. These scale factors  
988 are given by the ratio of the distribution of offline objects before trigger selection  
989 and after trigger selection. The associated uncertainty on these scale factors are  
990 used in the final fit.

991 Uncertainties on small-R jet energy scale and resolution are measured in-situ  
992 by calculating the response between data and simulation. This analysis uses a  
993 reduced set of JES and JER uncertainties (totaling 30 and 8 systematics, re-  
994 spectively). These reduced sets of systematics are calculated using a principal  
995 component analysis, yield largely uncorrelated independent systematics. These  
996 uncertainties on JES and JER account for the dependence on  $p_T$ ,  $\eta$ ,  $\mu$ , flavor re-  
997 sponse and global sequential corrections. Systematic uncertainties associated with  
998  $b$ -tagging are also considered. These systematics are evaluated as uncertainties on  
999 the scale factor which account for the difference in  $b$ -tagging efficiencies in data  
1000 and MC, and the flavor dependence (between b, c, and light jets).

1001 The uncertainty on the  $p_T$  scale of the large-R jets is determined by comparing  
1002 the jet's  $p_T^{calo}$  to  $p_T^{track}$  in di-jet simulation and data. In addition to this uncertain-  
1003 ties from tracking, modeling (Pythia vs Herwig), and statistical constraints are  
1004 also calculated. The large-R jet  $p_T$  resolution is given by smearing the jet  $p_T$  with  
1005 a Gaussian with a 2% width.

1006 The  $W/Z$ -tagging efficiency cannot be evaluated using the Rtrk method as the  
1007 TCC algorithm uses track measurements to reconstruct jet substructure variables.  
1008 In order to avoid this potential bias, the  $W/Z$ -tagging is estimated in data using

1009 a control sample and corrected by comparing it with simulation. The efficiency  
1010 to  $W/Z$ -induced signal is estimated by a  $t\bar{t}$  control sample, while the efficiency  
1011 to single- $q/g$  background is estimated using a dijet sample. The effects of experi-  
1012 mental and theoretical uncertainties on the efficiency scale factor are determined  
1013 by taking the ratio of efficiencies in data and simulation. By taking this ratio the  
1014 uncertainties not arising for jet mass and  $D_2$  cancel.

1015 Lepton identification, reconstruction, isolation systematic uncertainties are de-  
1016 termined by reconstructing the  $Z$  mass peak with a tag and probe method. The  
1017 lepton energy and momentum scales are also measured with the  $Z$  mass peak.  
1018 Additionally, the track-to-vertex association efficiency is used for muons.

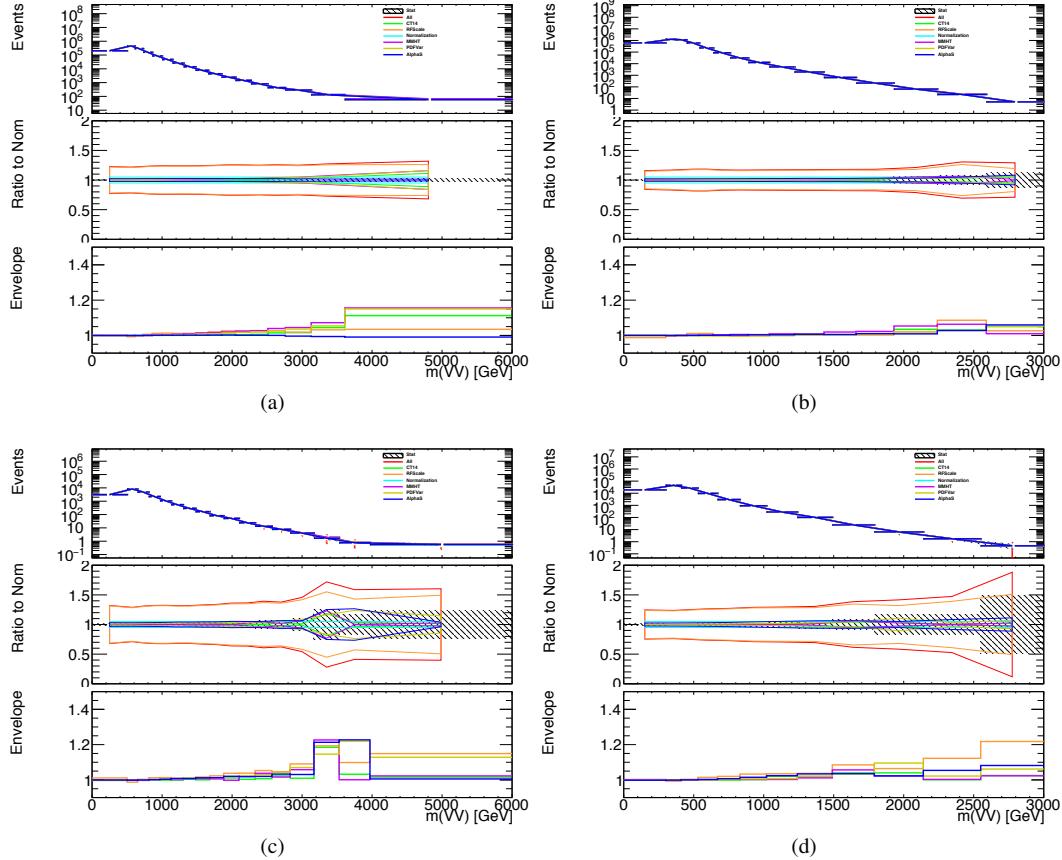
1019 As  $E_T^{miss}$  is calculated using all the physics objects in the event, all those objects  
1020 associated errors result in an uncertainty on  $E_T^{miss}$ . Additionally, the unassociated  
1021 tracks used to construct  $E_T^{miss}$  contribute to the uncertainty on  $E_T^{miss}$ .

## 1022 8.2 Theory Systematics

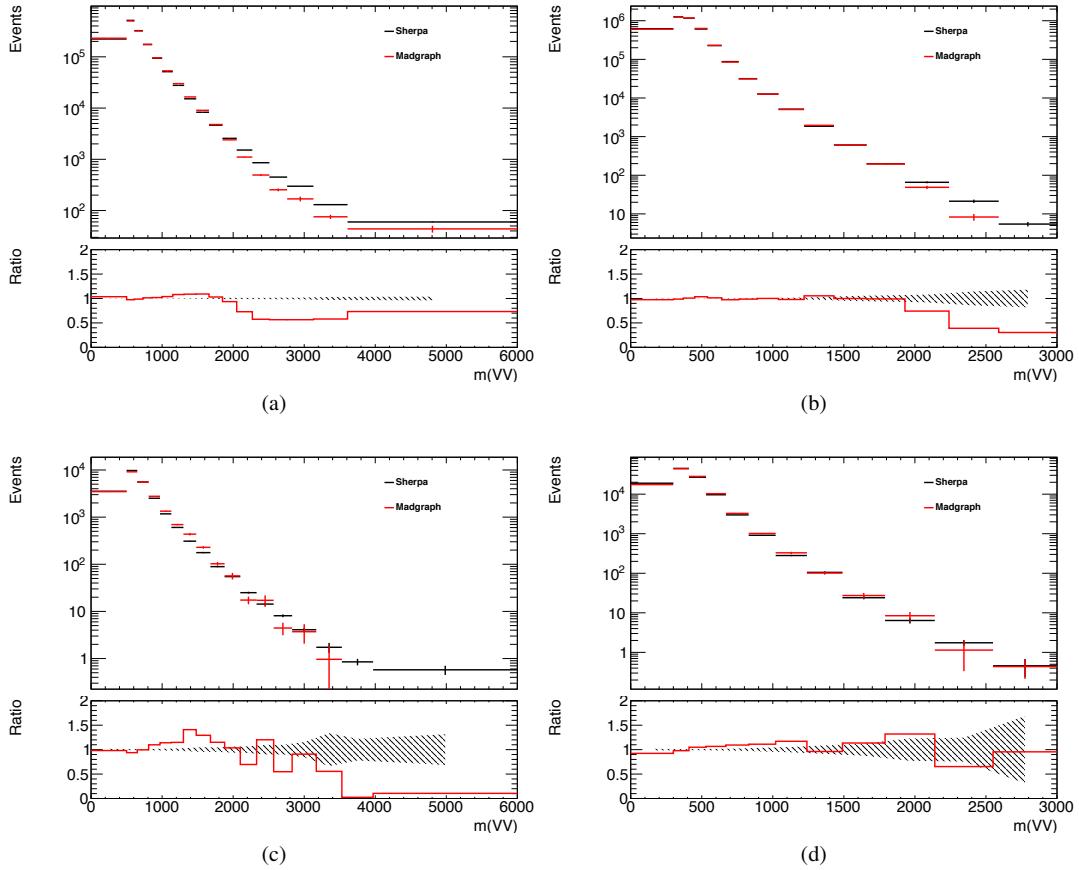
1023 Theoretical uncertainties for signal and background processes arise from un-  
1024 certainties in the parameters used in Monte Carlo simulation. In particular for  
1025 the  $t\bar{t}$ ,  $W/Z+jets$ , diboson backgrounds and signal samples, the QCD scale, PDF,  
1026 generator and hadronization uncertainties were evaluated. To assess the QCD  
1027 scale uncertainty the renormalization and factorization scales were scaled up and  
1028 down by a factor of two at the event generation stage of sample production. Un-  
1029 certainties due to the choice of the parton distribution functions were evaluated by  
1030 re-weighting samples from the nominal PDF to a set of error PDFs which account  
1031 for the uncertainty of the fits used to produce the PDF set. In addition to this,  
1032 samples are re-weighted to different PDF sets to account for the arbitrariness of  
1033 the PDF choice. The difference between the  $m_{WV}$  distributions using different

1034 event generators is assessed by comparing samples generated with different gen-  
1035 erators. Similarly, the uncertainty in hadronization models is accounted for by  
1036 comparing samples created using different hadronization models (e.g.  $t\bar{t}$  Powheg  
1037 is compared to AMC@NLO,  $W + jets$  compares Sherpa and MadGraph+Pythia  
1038 samples). Figures 8.2 - 8.8 show the impact of these uncertainties on the  $t\bar{t}$  and  
1039  $W/Z + jets$  backgrounds. Additionally, contributions to the diboson background  
1040 for the VBF analysis were found to be small and were accounted for by including  
1041 a 5(10)% systematic in the diboson normalization in the final fit.

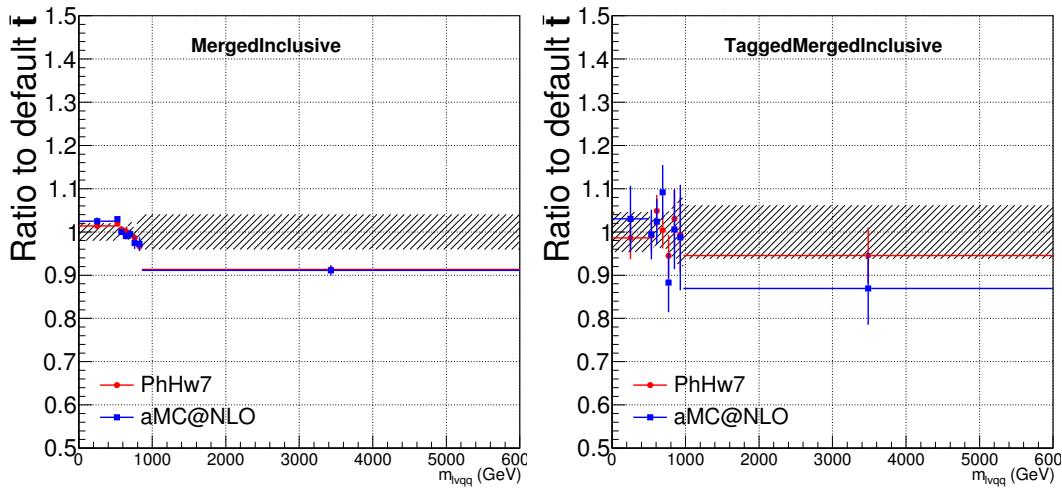
1042 The normalization of the  $t\bar{t}$  and  $W+jets$  processes impact the multijet tem-  
1043 plate shape. The impact of these normalizations was assessed by including a  
1044 shape systematic on the multijet background from varying the  $t\bar{t}$  and  $W+jets$   
1045 normalization factors.



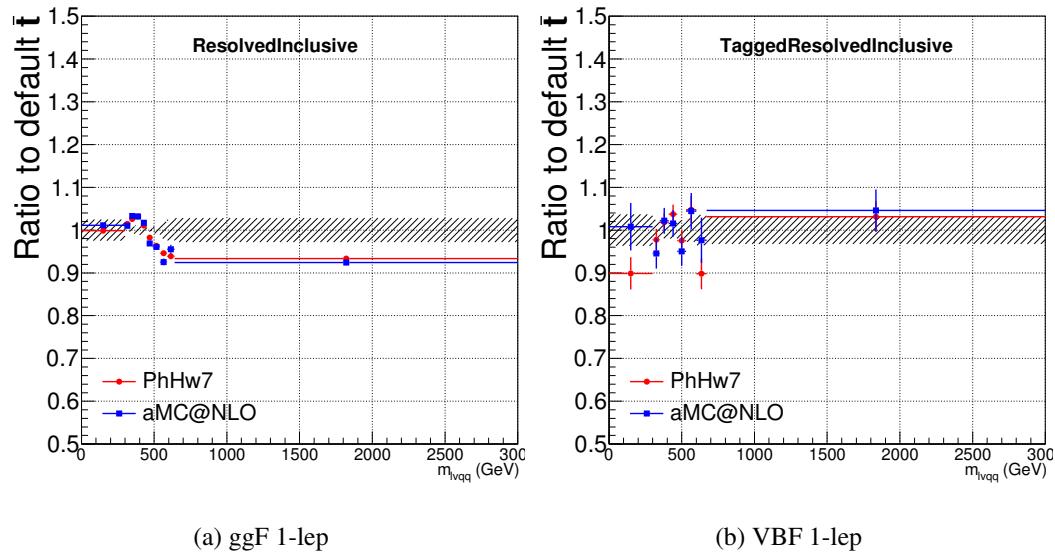
**Figure 8.1:** The  $W/Z + \text{jet}$  systematics for the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The top subplot shows the nominal and variation distributions/bands, the middle shows the ratio of the two, and the final shows just the shape of the envelope (the final uncertainty).



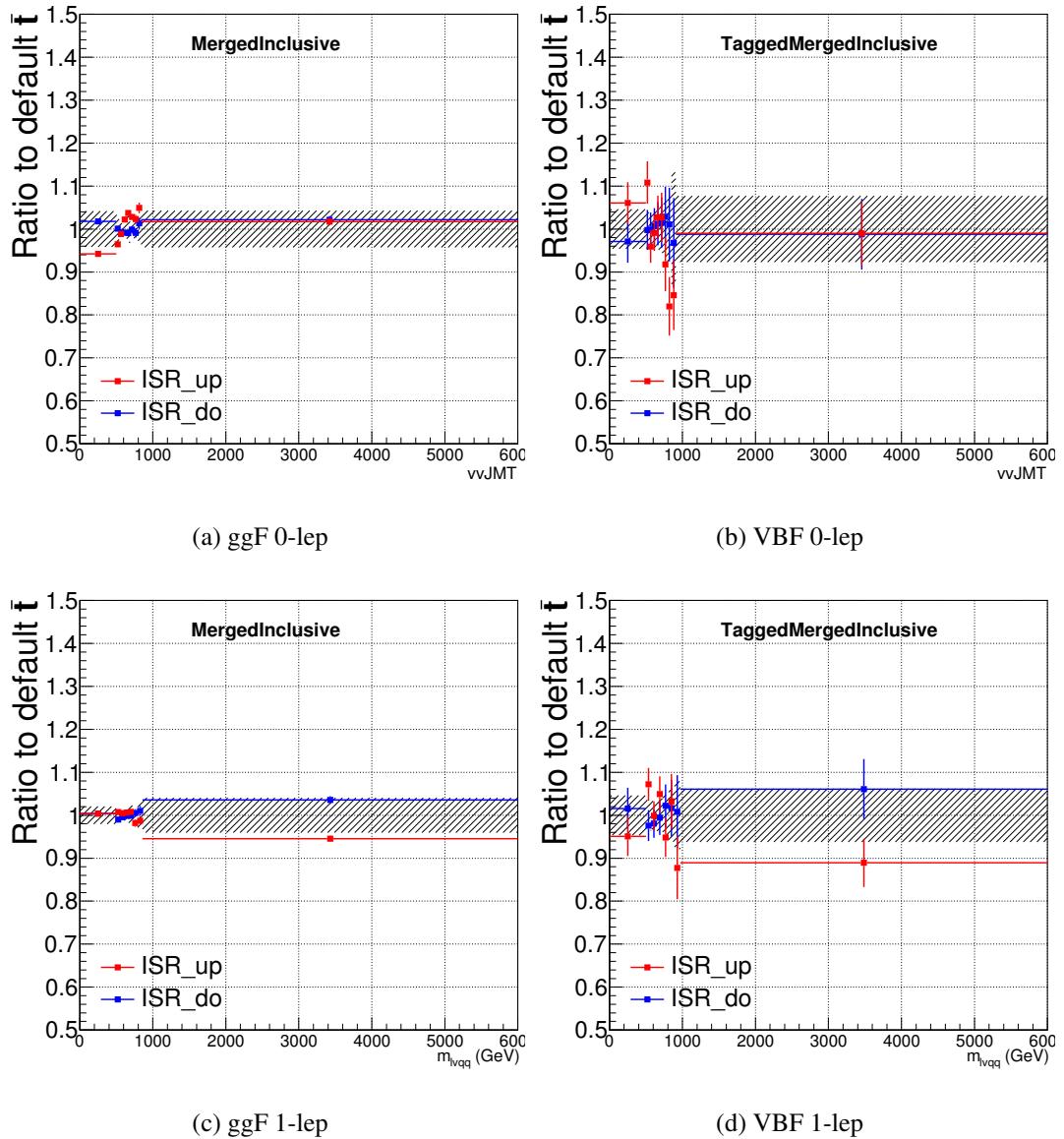
**Figure 8.2:** The two-point generator comparison between Sherpa and MadGraph for the  $W/Z + \text{jet}$  samples in the a) Merged ggF, b) Resolved ggF, c) Merged VBF, and d) Resolved VBF regions. The normalization of the Madgraph sample is set to the Sherpa value to consider only shape effects. The bottom inset shows the ratio of the two.



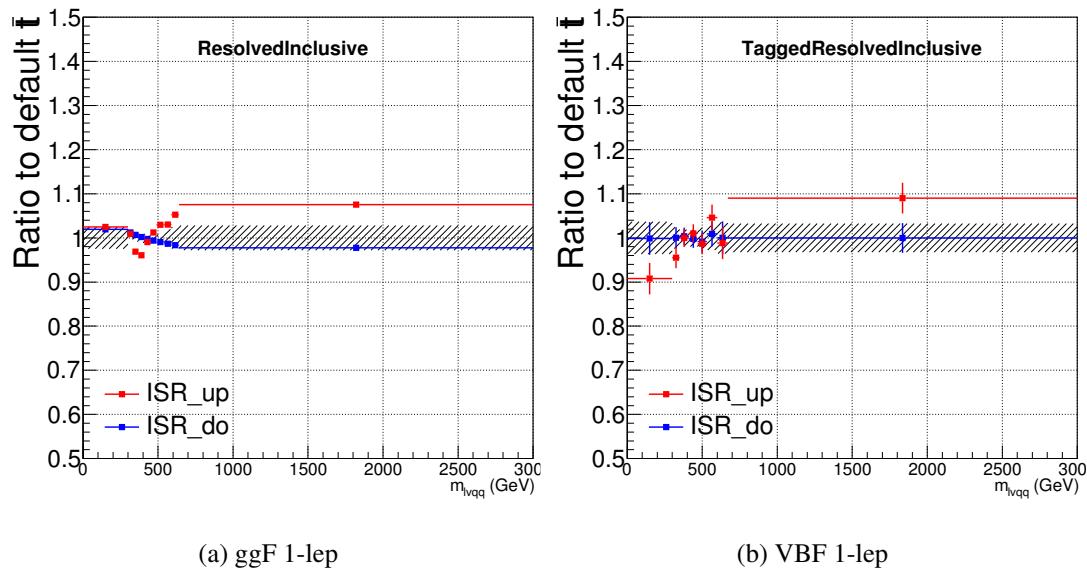
**Figure 8.3:** Ratio between the variations of generator (red) and hadronization (blue) variations for the Merged regime for  $t\bar{t}$  sample.



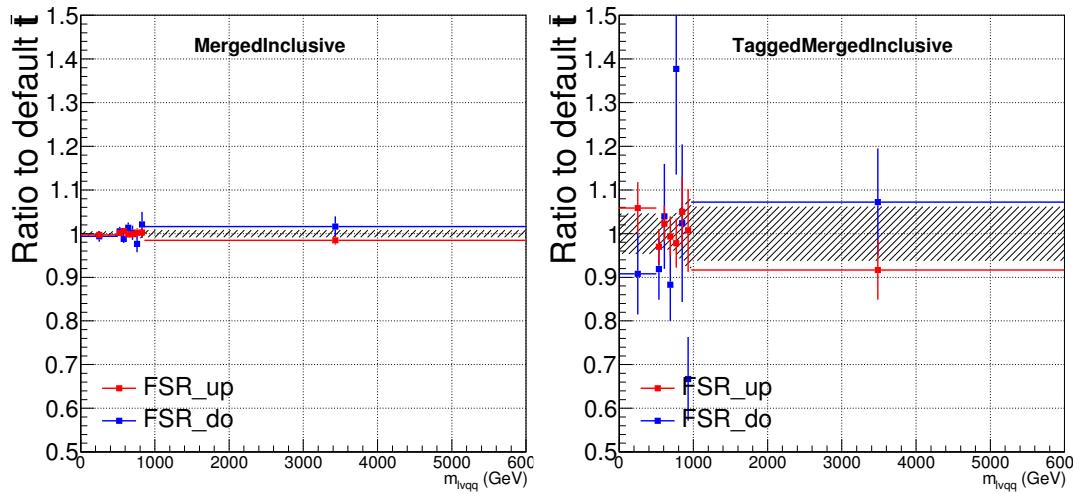
**Figure 8.4:** Ratio between the variations of generator (red) and hadronization (blue) variations for the Resolved regime for  $t\bar{t}$  sample.



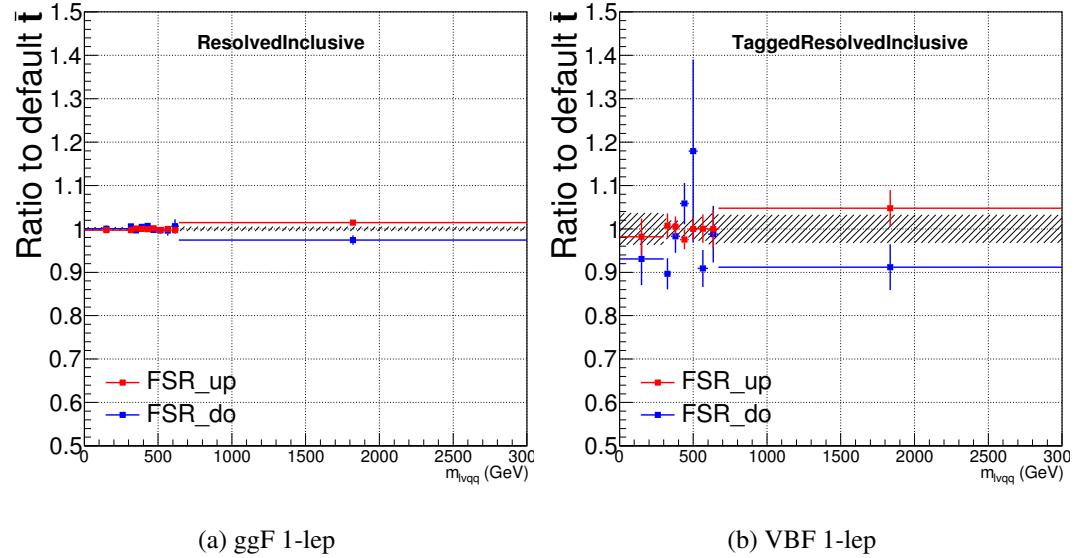
**Figure 8.5:** Ratio between the variations of ISR up (red) and down (blue) variations for the Merged regime for  $t\bar{t}$  sample.



**Figure 8.6:** Ratio between the variations of ISR up (red) and down (blue) variations for the Resolved regime for  $t\bar{t}$  sample.



**Figure 8.7:** Ratio between the variations of FSR up (red) and down (blue) variations for the Merged regime for  $t\bar{t}$  sample.



**Figure 8.8:** Ratio between the variations of FSR up (red) and down (blue) variations for the Resolved regime for  $t\bar{t}$  sample.

1046 **Chapter 9**

1047 **Statistical Analysis**

1048 To determine the compatibility of the data collected with the proposed reso-  
1049 nances a statistical procedure based on a likelihood function is used. A discovery  
1050 test is used to measure the compatibility of the observed data with the back-  
1051 ground only hypothesis. If the observed data is sufficiently incompatible with the  
1052 background only hypothesis, this could indicate a discovery. In the absence of  
1053 discovery, upper limits on the signal strength parameter,  $\mu$ , are assessed using the  
1054 CLs method.

1055 **9.1 Likelihood Function Definition**

1056 The likelihood function is product of Poisson probabilities for all analysis bins  
1057 and systematic constraints:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_c \prod_i \frac{(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))^{n_{ci}}}{n_{ci}!} e^{-(\mu s_{ci}(\boldsymbol{\theta}) + b_{ci}(\boldsymbol{\theta}))} \prod_k (\theta'_k | \theta_k) \quad (9.1)$$

1058 Here  $c$  are the analysis channels considered and  $i$  runs over all the  $m_{\ell\nu qq}$  bins  
1059 used in the fit. The signal strength parameter,  $\mu$ , multiplies the expected signal

1060 yield in each analysis bin,  $s_{ci}$ . The background content for channel  $c$  and bin  $i$  is  
1061 given by  $b_{ci}$ . The dependence of signal and background predictions on system-  
1062 atic uncertainties is described by the aforementioned set of nuisance parameters  
1063  $\boldsymbol{\theta}$ , which are parameterized by Gaussian or log-normal priors denoted here as  
1064  $\theta_k$ . Statistical uncertainties of the simulated bin contents are also included as  
1065 systematic uncertainties. Most systematics are correlated among all the analysis  
1066 regions and considered to be independent from each other. The validity of this  
1067 assumption is checked by evaluating the covariance of nuisance parameters.

## 1068 9.2 Fit Configuration

1069 The binning of  $m_{\ell\nu qq}$  in signal regions for likelihood fit is determined by the  
1070 statistical uncertainty of signal mass width. For each signal mass point, the signal  
1071 mass resolution is given by the fitted Gaussian width of the  $m_{\ell\nu qq}$ . The fitted  
1072 signal widths are then fit to a line to give a parameterized signal mass width,  
1073 as shown in Figures 9.1 and 9.2. Bin widths are set first to this parameterized  
1074 signal mass resolution. Then if the statistical uncertainty of the data or simulated  
1075 background is more than 50%, bins are merged until the statistical uncertainty is  
1076 less than 50%. All control regions contain only a single bin.

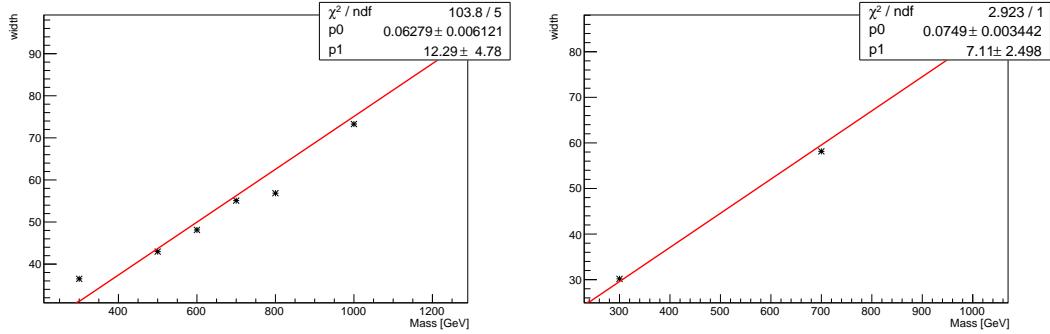
1077 For this analysis, each signal model is fit in the Merged and Resolved chan-  
1078 nels for the relevant signal production mode simultaneously. The  $W + \text{jets}$  and  $t\bar{t}$   
1079 normalizations are given by the best fit values in the overall fit and these fitted  
1080 normalizations are then applied to those backgrounds in the SRs.

1081 Systematics may be affected by low statistics, leading to unsmooth  $m_{VV}$  dis-  
1082 tributions with unphysically large fluctuations. This can lead to artificial pulls  
1083 and constraints in the fit. To remove such issues a multi-step smoothing pro-  
1084 cedure is applied to all systematic variation distributions in all regions. First,

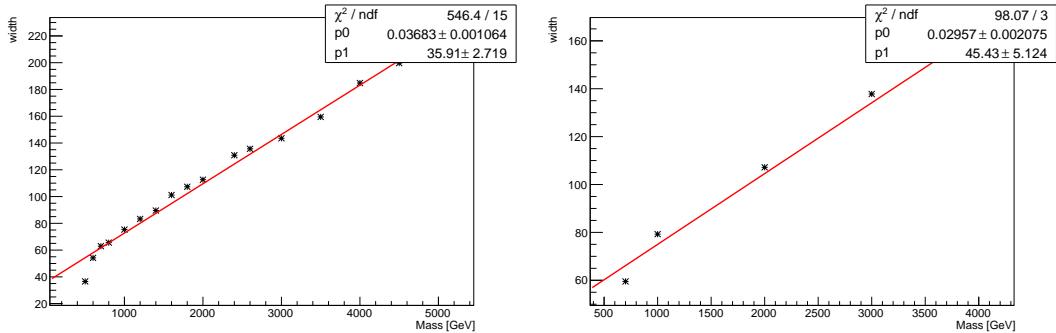
1085 distributions are rebinned until the statistical error per bin is at least 5%. Next  
1086 all local extrema are identified. The bins around smallest extrema are iteratively  
1087 merged until only four local extrema remain. Then distributions are rebinned so  
1088 that statistical uncertainties in each bin are < 5%.

1089 For some systematics, up and down variations may be in the same direction  
1090 with respect to the nominal distributions. This causes the variations to not cover  
1091 the nominal choice, and the interpretation of the confidence interval is skewed as  
1092 the nominal distribution should be bracketed by the up and down variations. This  
1093 asymmetry may also lead to unconstrained systematics in the fit. To handle such  
1094 asymmetric systematics, if the up and down variation for a given systematic are in  
1095 the same direction for at least three  $m_{VV}$  bins the variation is averaged for those  
1096 bins. The averaging procedure replaces bin-by-bin the up and down variation bins  
1097 by  $b_{\pm}^{new} = b_{nom} \pm \frac{|b_+ - b_-|}{2}$ , where  $b_{nom}$  is the nominal bin content and  $b_{\pm}$  are the  
1098 original up and down variation bin content. The same procedure is also applied to  
1099 any variations where the integral of the difference between the up/down variation  
1100 and the nominal distribution is twice that of the other down/up variation, further  
1101 ensuring variations are symmetric around the nominal distribution.

1102 Finally, systematics that have a negligible effect on the  $m_{VV}$  distribution are  
1103 not considered in the fit. Shape systematics where no bin in the variational dis-  
1104 tribution deviates more than 1% from the nominal distribution (after normalizing  
1105 all histograms to the nominal) are not included in the fit. Also, statistical bin  
1106 uncertainties < 1% are ignored.



**Figure 9.1:** The HVT signal mass resolution as a function of mass fit with a straight line in the Resolved ggF region (left) and VBF (right) region.



**Figure 9.2:** The HVT signal mass resolution as a function of mass fit with a straight line in the Merged ggF region (left) and VBF (right) region.

### 1107 9.3 Best Fit $\mu$

1108     The best fit signal strength parameter is denoted by  $\hat{\mu}$  and calculated by  
 1109     maximizing the likelihood function with respect to all systematics and  $\mu$ . The  
 1110     corresponding set of systematics that maximize the likelihood are given by  $\hat{\mu}$ .  
 1111     The first term in the likelihood is maximized when the expected number of signal  
 1112     and background events is equal to the number of events in data ( $n_{ci} = \mu s_{ci} +$   
 1113      $b_{ci}$ ). Thus, by maximizing the likelihood, the fit determines values of  $\mu$  and  $\theta$   
 1114     that give the best agreement between expected and measured event yields. The

1115 second term in the likelihood is a penalty term which decreases the likelihood  
1116 when systematics are shifted from their nominal values. This prevents the fit  
1117 from profiling systematics in unphysical ways to maximize the likelihood. The  
1118 uncertainty on  $\mu$  is calculated by varying  $\mu$  up and down until the natural log of  
1119 the likelihood function shifts by one-half.

## 1120 9.4 Discovery Test

1121 To determine if the observed dataset is consistent with tested signal model a  
1122 likelihood ratio is constructed:

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \quad (9.2)$$

1123 The denominator in this equation is the maximized value of  $\mathcal{L}$  over all system-  
1124 atics and  $\mu$ . The numerator is the maximized likelihood over all systematics for  
1125 a given  $\mu$  value, where the maximized systematics are given by  $\hat{\mu}_\mu$ . To test for  
1126 the existence of signal the observed dataset the null hypothesis ( $H_0$ ) is defined as  
1127 the background only hypothesis and the alternate hypothesis includes signal and  
1128 background ( $H_1$ ). This test quantifies the compatibility of observed data with  
1129  $H_0$  by calculating a p-value representing the probability of observing data as dis-  
1130 crepant or more than the observed data under the  $H_0$ . The test statistic used to  
1131 calculate this p-value is given by ( $r_0$ ):

$$r_0 = \begin{cases} -2 \ln \lambda(0), \hat{\mu} > 0 \\ +2 \ln \lambda(0), \hat{\mu} < 0 \end{cases} \quad (9.3)$$

1132 The expected distribution of the the test statistic under  $H_0$  ( $f(r_0|0)$ ) is used to  
1133 calculate the p-value:

$$p_0 = \int_{r_0, obs}^{\infty} f(r_0 | 0) dr_0 \quad (9.4)$$

1134        Small p-values indicate the observed data is poorly described by  $H_0$ . This  
 1135        equivalent Z-score of a given p-value is usually used to further quantify the agree-  
 1136        ment between the observed data and  $H_0$ . The Z-score is given by the number of  
 1137        standard deviations away from the mean of a Gaussian distribution, the integral  
 1138        of the upper tail of the distribution would equal the p-value. Mathematically:

$$Z = \Phi^{-1}(1 - p_0) \quad (9.5)$$

1139        where  $\Phi$  is the Gaussian cumulative distribution function. The statistical  
 1140        significance of these tests are expressed as the  $Z$ -score. In particle physics,  $3\sigma$  is  
 1141        considered evidence for new phenomena and  $5\sigma$  is the threshold for discovery.

## 1142        9.5 Exclusion Limits

1143        In the absence of discovery, upper limits on the signal strength,  $\mu$  are set using  
 1144        the CLs method [cite P60]. The test statistic for this test,  $q_\mu$ , is constructed as:

$$\tilde{\lambda}_\mu = \begin{cases} \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\theta})}, \hat{\mu} > 0 \\ \frac{\mathcal{L}(\mu, \hat{\theta}_\mu)}{\mathcal{L}(0, \hat{\theta}_0)}, \hat{\mu} < 0 \end{cases} \quad (9.6)$$

$$\tilde{q}_\mu = \begin{cases} -2 \ln \tilde{\lambda}(\mu), \hat{\mu} < \mu \\ +2 \ln \tilde{\lambda}(\mu), \hat{\mu} > \mu \end{cases} \quad (9.7)$$

1145        As defined, larger values of  $q_\mu$  correspond to increasing incompatibility between  
 1146        the observed data and the background + signal hypothesis. The observed value  
 1147        of the test statistic,  $q_{\mu, obs}$ , is then compared to its expected distribution,  $f$ , to

1148 calculate p-values to assess the likelihood of the background+signal hypothesis.

1149 Using these distributions,  $CL_s$  values are computed as:

$$1150 \quad CL_{s+b} = \int_{q_{\mu,obs}}^{\infty} f(q_{\mu}|\mu) dq_{\mu} \quad (9.8)$$

$$1151 \quad CL_b = \int_{q_0^{obs}}^{\infty} f(q_{\mu}|\mu = 0) dq_{\mu} \quad (9.9)$$

$$1152 \quad CL_s = \frac{CL_{s+b}}{CL_b} \quad (9.10)$$

1153  $CL_{s+b}$  is the p-value for the signal + background hypothesis and  $CL_b$  is the  
1154 p-value for the background only hypothesis. The  $CL_s$  value is interpreted as  
1155 the probability to observe the background + signal hypothesis normalized to the  
1156 probability of background-only hypothesis. Normalizing by  $CL_b$  prevents setting  
1157 artificially strong exclusion limits due to downward fluctuations in data.

1158 For a given signal hypothesis,  $\mu$  values are scanned simultaneously over all  
1159  $m_{WV}$  bins to find the  $\mu$  value that yields  $CL_s=0.05$ , meaning the likelihood of  
1160 finding data more incompatible with the signal+background hypothesis (relative  
1161 to the background only hypothesis) is 5%. The 95% upper limit on the cross  
1162 section is then calculated as the product of the  $\mu$  value found, branching ratio,  
1163 and theory cross section.

# Part IV

1163

# Results

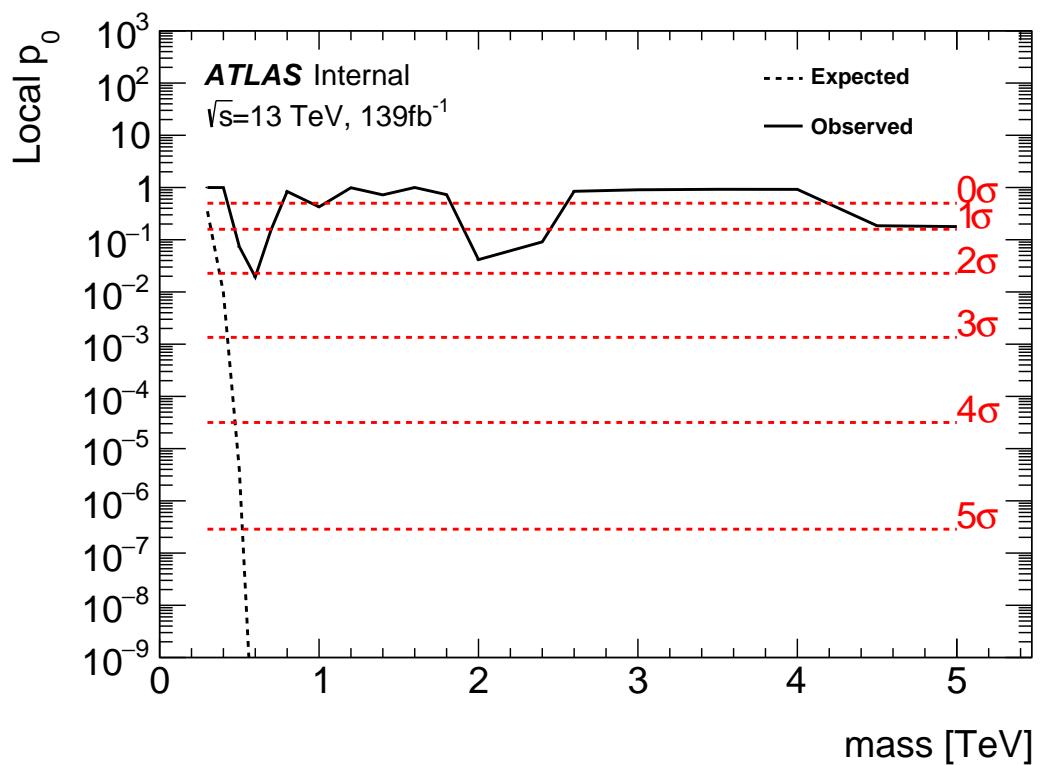
1164

1165 **Chapter 10**

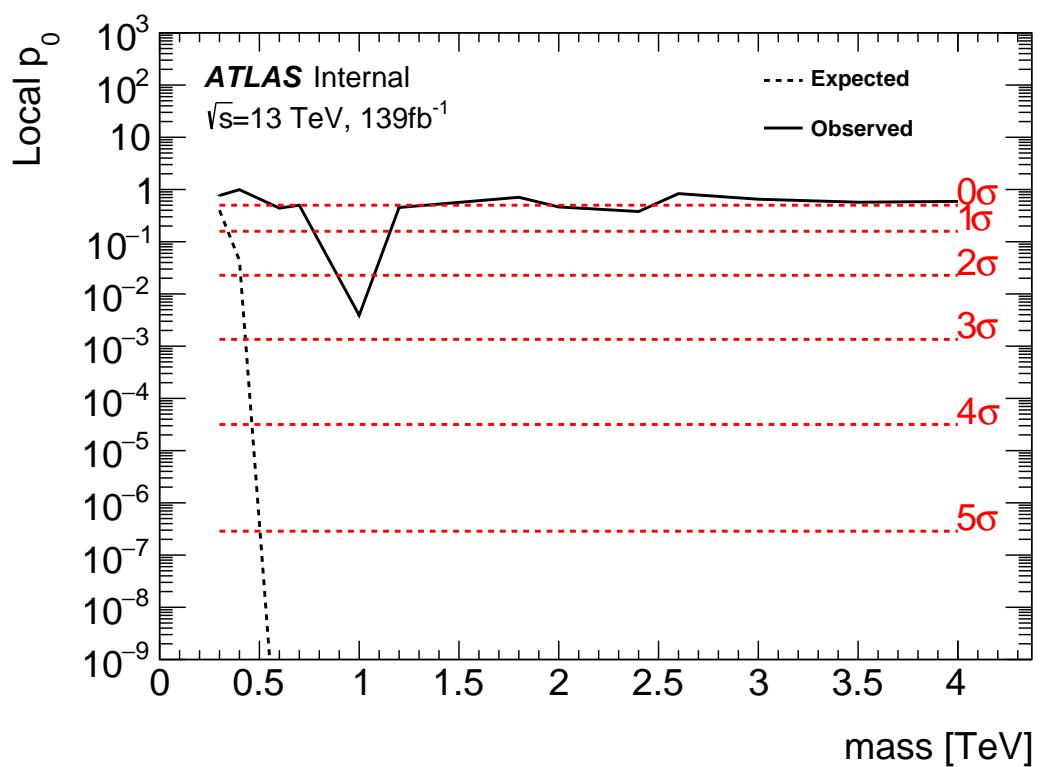
1166 **Statistical Interpretation**

1167 **10.1 Discovery Tests**

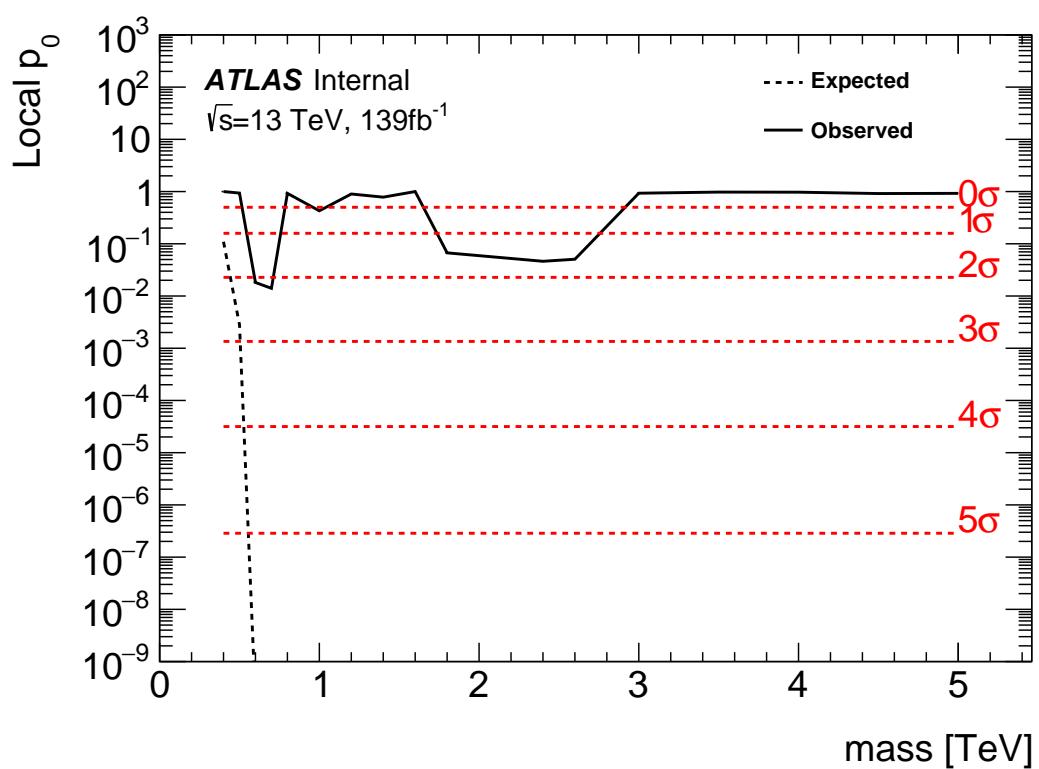
1168 To test for the existence of signal in the observed dataset, the discovery tests  
1169 discussed earlier are used to calculate p-values as a function of resonance mass.  
1170 The results of these tests are shown in Figures 10.1 - 10.5. Across the different  
1171 DY signals the largest excesses are  $\sim 2.2\sigma$  at 600 GeV and  $1.8\sigma$  at 2 TeV. The  
1172 largest excesses for VBF signals are  $< 2.5\sigma$  at for 1 TeV resonances. As these  
1173 deviations do not constitute discoveries, upper limits on  $\mu$  are calculated.



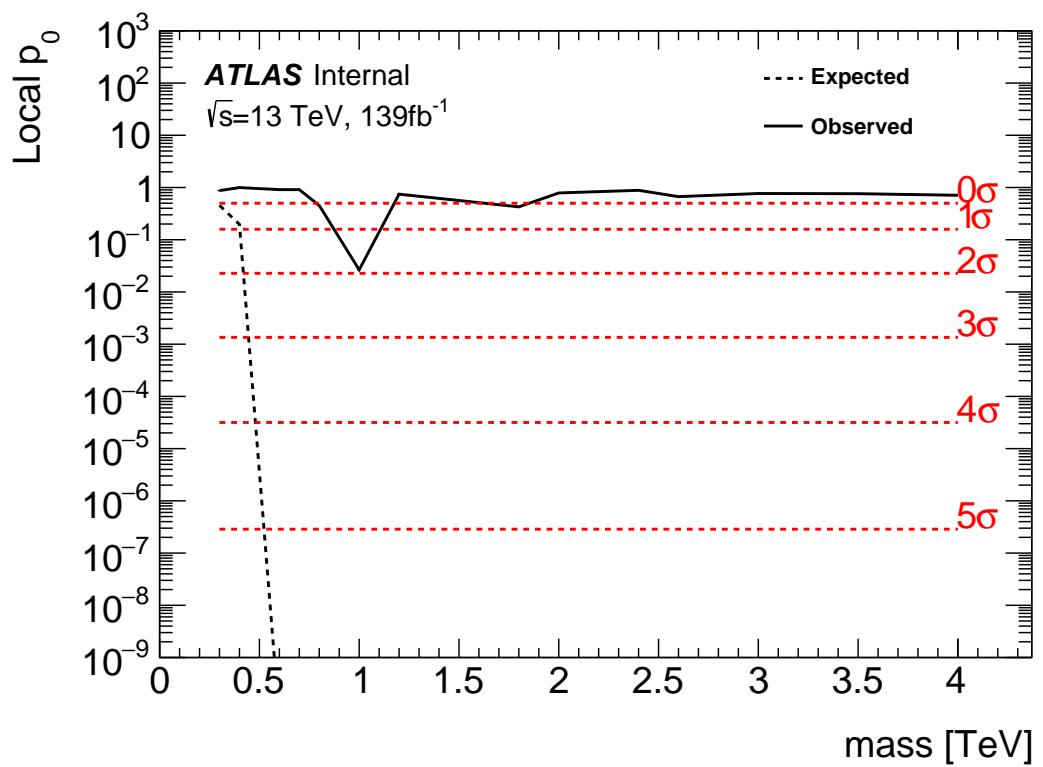
**Figure 10.1:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT Z' DY production.



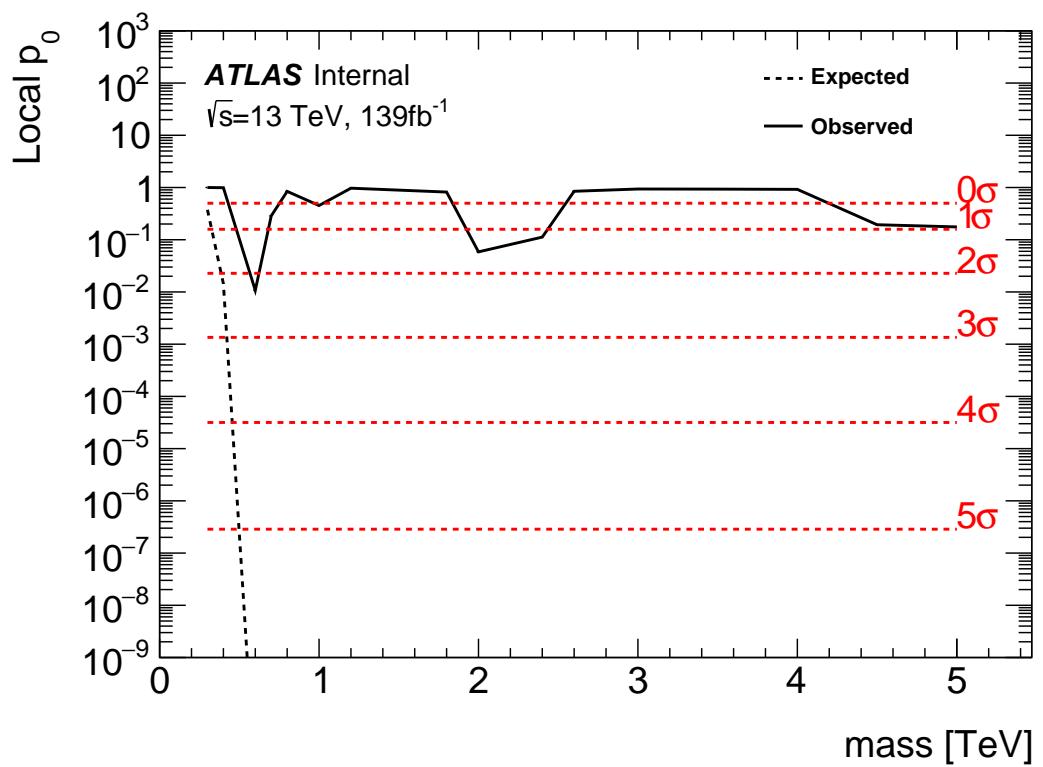
**Figure 10.2:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT Z' VBF production.



**Figure 10.3:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT W' DY production.



**Figure 10.4:** These plots show the measured  $p_0$  value as a function of resonance mass for HVT W' VBF production.

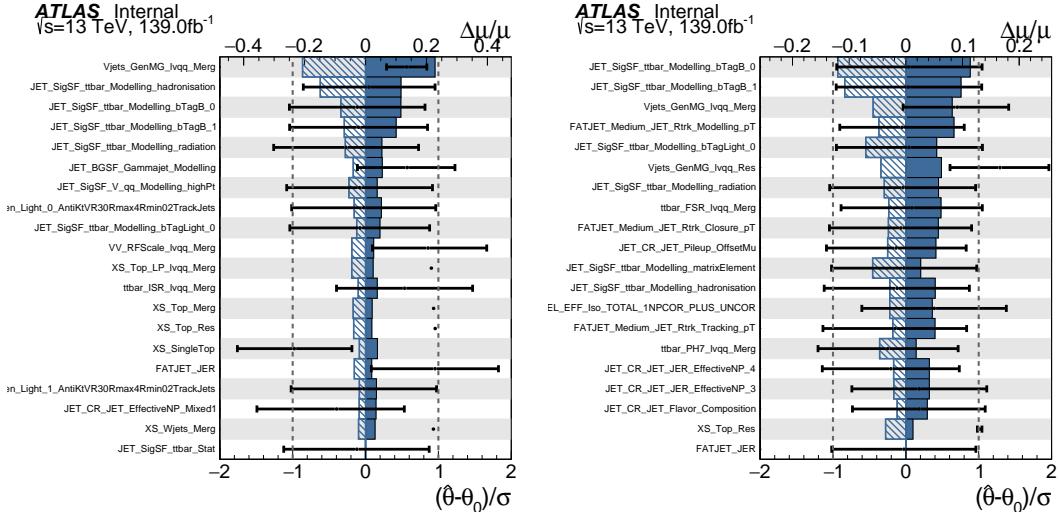


**Figure 10.5:** These plots show the measured  $p_0$  value as a function of resonance mass for the RS Graviton DY production.

## 1174 10.2 Systematic Profiling and Correlations

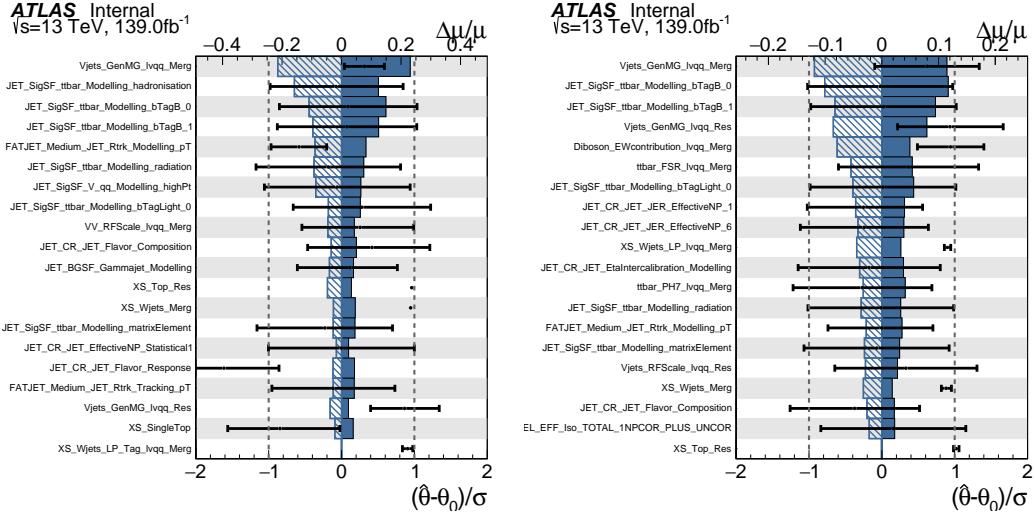
1175 The ranked systematics and their fitted values are shown for the different anal-  
 1176 ysis regions in Figure 10.6 - 10.8. Note that background normalizations for  $W+jets$   
 1177 and  $t\bar{t}$  are left free to float in the fit. This means the nominal normalization val-  
 1178 ues are at one and the uncertainties are not plotted in the ranked plots. Overall,  
 1179 systematics are not pulled outside their uncertainties, especially for highly ranked  
 1180 nuisance parameters.

1181 The correlation between systematics are shown in Figures ???. Correlations  
 1182 between background normalization are expected. The remaining systematic cor-  
 1183 relations are not very strong or unexpected.

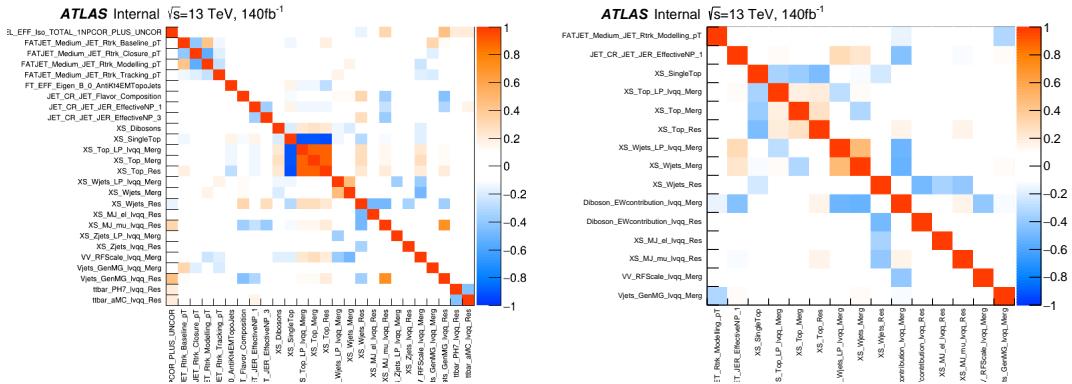


1184 **Figure 10.6:** Ranked systematics and their fitted values for  $WW$  DY (right) and  
 1185 VBF (left) selections.

1186



**Figure 10.7:** Ranked systematics and their fitted values for  $WZ$  DY (right) and VBF (left) selections.



**Figure 10.8:** Correlations between systematics for  $WW$  DY (right) and VBF (left) selections.

### 1184 10.3 Expected and Measured Yields

1185 The yield tables for the four analysis regions are shown in Tables ?? - ?? . The  
 1186 fitted background normalizations are shown in Tables ??-?? . The control region  
 1187  $m_{\ell\nu qq}$  distributions are shown in Figures 10.9 - 10.12. The signal region  $m_{\ell\nu qq}$   
 1188 distributions are shown in Figures 10.13 - ?? .

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$16507.83 \pm 2314.87$
Muon Multi-jet	-	-	$19977.12 \pm 2816.06$
Diboson	$1833.41 \pm 177.78$	$3323.93 \pm 320.92$	$9147.67 \pm 961.63$
Single-top	$2160.62 \pm 402.34$	$3551.09 \pm 660.00$	$20058.36 \pm 3817.26$
$t\bar{t}$	$15518.86 \pm 338.22$	$24069.54 \pm 453.15$	$138866.23 \pm 1989.71$
$W+jets$	$40141.57 \pm 357.79$	$88113.06 \pm 487.87$	$673200.38 \pm 4120.53$
$Z+jets$	$778.83 \pm 78.93$	$1765.54 \pm 179.10$	$16570.50 \pm 1672.71$
Total	$60433.29 \pm 664.92$	$120823.16 \pm 1006.99$	$894328.12 \pm 7247.12$
Data	60264.00	120852.00	895362.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$421.11 \pm 37.98$	$550.44 \pm 53.10$	$996.87 \pm 119.63$
Single-top	$4691.44 \pm 846.11$	$3466.26 \pm 631.03$	$16848.71 \pm 3258.26$
$t\bar{t}$	$38945.18 \pm 848.77$	$33836.95 \pm 637.04$	$224226.14 \pm 3212.76$
$W+jets$	$2258.34 \pm 20.13$	$6564.78 \pm 36.35$	$23466.41 \pm 143.63$
$Z+jets$	$66.35 \pm 6.72$	$213.26 \pm 21.63$	$846.66 \pm 85.47$
Total	$46382.43 \pm 1199.25$	$44631.70 \pm 899.23$	$266384.78 \pm 4580.43$
Data	46354.00	44629.00	266443.00
	WW SR	LP SR	Resolved 1-lepton SR
Electron Multi-jet	-	-	$10788.40 \pm 1512.85$
Muon Multi-jet	-	-	$15759.50 \pm 2221.53$
Diboson	$4990.30 \pm 376.50$	$3901.07 \pm 313.22$	$16971.29 \pm 1523.77$
Single-top	$3117.71 \pm 565.07$	$2176.46 \pm 400.52$	$20422.85 \pm 3731.94$
$t\bar{t}$	$13785.77 \pm 302.14$	$11005.12 \pm 207.41$	$126965.25 \pm 1819.66$
$W+jets$	$24718.56 \pm 223.72$	$60080.66 \pm 333.12$	$444133.56 \pm 2719.02$
$Z+jets$	$478.18 \pm 48.46$	$1226.69 \pm 124.44$	$11686.32 \pm 1179.69$
Total	$47090.52 \pm 777.65$	$78389.98 \pm 654.22$	$646727.19 \pm 5963.98$
Data	47330.00	78380.00	645610.00

**Table 10.1:** Expected and Measured for DY  $WW$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

	HP Untagged WCR	LP Untagged WCR	Resolved Untagged WCR
Electron Multi-jet	-	-	$15080.03 \pm 2277.99$
Muon Multi-jet	-	-	$27347.10 \pm 2950.07$
Diboson	$1508.48 \pm 154.20$	$2758.24 \pm 284.50$	$9038.55 \pm 728.69$
Single-top	$1756.59 \pm 306.69$	$2913.18 \pm 515.93$	$20511.74 \pm 3523.47$
$t\bar{t}$	$13134.00 \pm 238.30$	$21815.37 \pm 334.98$	$140157.77 \pm 2636.96$
$W+jets$	$40654.84 \pm 333.65$	$87657.76 \pm 501.96$	$665909.12 \pm 4420.62$
$Z+jets$	$768.72 \pm 77.97$	$1759.87 \pm 178.96$	$16512.46 \pm 1673.23$
Total	$57822.63 \pm 540.40$	$116904.42 \pm 862.16$	$894556.75 \pm 7492.20$
Data	57699.00	117306.00	895362.00
	HP Tagged WCR	LP Tagged WCR	Resolved Tagged WCR
Electron Multi-jet	-	-	$384.58 \pm 57.11$
Muon Multi-jet	-	-	$602.93 \pm 190.12$
Diboson	$30.22 \pm 4.69$	$48.95 \pm 7.16$	$264.64 \pm 28.24$
Single-top	$308.44 \pm 56.19$	$371.59 \pm 69.43$	$5752.39 \pm 1029.97$
$t\bar{t}$	$1683.82 \pm 48.73$	$2041.48 \pm 70.00$	$58431.49 \pm 614.30$
$W+jets$	$583.55 \pm 75.37$	$1109.45 \pm 85.78$	$11891.68 \pm 903.01$
$Z+jets$	$13.19 \pm 1.34$	$23.06 \pm 2.34$	$324.74 \pm 32.85$
Total	$2619.22 \pm 106.00$	$3594.53 \pm 130.90$	$77652.45 \pm 1514.89$
Data	2565.00	3546.00	77973.00
	HP Untagged TCR	LP Untagged TCR	Resolved Untagged TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$289.45 \pm 28.45$	$346.78 \pm 35.85$	$650.85 \pm 65.56$
Single-top	$3107.99 \pm 538.03$	$2250.64 \pm 385.41$	$9606.87 \pm 1698.22$
$t\bar{t}$	$30992.40 \pm 562.33$	$26954.21 \pm 413.89$	$91893.59 \pm 1728.91$
$W+jets$	$2236.29 \pm 18.35$	$4874.03 \pm 27.91$	$16122.97 \pm 107.03$
$Z+jets$	$71.54 \pm 7.26$	$155.50 \pm 15.81$	$577.71 \pm 58.54$
Total	$36697.66 \pm 779.03$	$34581.16 \pm 567.59$	$118851.98 \pm 2427.40$
Data	36677.00	34573.00	118928.00
	HP Tagged TCR	LP Tagged TCR	Resolved Tagged TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$9.72 \pm 1.13$	$8.75 \pm 1.16$	$34.06 \pm 4.98$
Single-top	$105.87 \pm 20.65$	$119.66 \pm 22.68$	$656.89 \pm 132.96$
$t\bar{t}$	$1904.75 \pm 50.61$	$1483.86 \pm 47.05$	$17965.33 \pm 188.87$
$W+jets$	$32.36 \pm 4.28$	$85.74 \pm 6.96$	$489.01 \pm 37.13$
$Z+jets$	$1.27 \pm 0.13$	$1.93 \pm 0.20$	$19.14 \pm 1.94$
Total	$2053.98 \pm 54.84$	$1699.93 \pm 52.70$	$19164.43 \pm 234.01$
Data	2047.00	1708.00	19143.00

**Table 10.2:** Expected and Measured for DY  $WZ$   $W+jets$ ,  $t\bar{t}$  tag and untag control regions.

	HP Untagged SR	LP Untagged SR	Resolved Untagged SR
Electron Multi-jet	-	-	$7782.17 \pm 1175.56$
Muon Multi-jet	-	-	$17004.81 \pm 1834.40$
Diboson	$3041.17 \pm 273.77$	$2266.35 \pm 212.79$	$14724.12 \pm 1224.31$
Single-top	$2123.28 \pm 373.83$	$1379.35 \pm 240.92$	$18336.88 \pm 3082.47$
$t\bar{t}$	$11678.86 \pm 213.63$	$8906.34 \pm 136.88$	$112669.24 \pm 2122.46$
$W+jets$	$22741.32 \pm 191.47$	$41726.76 \pm 240.56$	$342934.00 \pm 2280.21$
$Z+jets$	$442.03 \pm 44.84$	$849.79 \pm 86.42$	$9271.83 \pm 939.52$
Total	$40026.65 \pm 546.81$	$55128.59 \pm 432.90$	$522723.03 \pm 5131.71$
Data	40193.00	54735.00	521813.00
	HP Tagged SR	LP Tagged SR	Resolved Tagged SR
Electron Multi-jet	-	-	$199.22 \pm 29.58$
Muon Multi-jet	-	-	$393.43 \pm 124.06$
Diboson	$102.58 \pm 11.59$	$65.44 \pm 8.05$	$624.07 \pm 58.10$
Single-top	$178.21 \pm 33.62$	$155.53 \pm 28.95$	$3470.39 \pm 617.48$
$t\bar{t}$	$1017.93 \pm 31.95$	$706.76 \pm 26.20$	$38189.30 \pm 401.91$
$W+jets$	$325.58 \pm 41.62$	$575.36 \pm 43.29$	$6161.96 \pm 467.71$
$Z+jets$	$7.81 \pm 0.80$	$11.62 \pm 1.19$	$183.36 \pm 18.55$
Total	$1632.11 \pm 63.39$	$1514.70 \pm 58.86$	$49221.74 \pm 884.06$
Data	1699.00	1559.00	48919.00

**Table 10.3:** Expected and Measured for DY  $WZ$   $W+jets$ ,  $t\bar{t}$  tag and untag signal regions.

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$898.48 \pm 137.82$
Muon Multi-jet	-	-	$601.46 \pm 182.74$
Diboson	$107.45 \pm 45.20$	$166.87 \pm 68.11$	$292.10 \pm 235.29$
Single-top	$78.19 \pm 18.22$	$132.71 \pm 31.93$	$879.82 \pm 216.89$
$t\bar{t}$	$400.71 \pm 28.35$	$569.70 \pm 48.88$	$5067.51 \pm 155.69$
$W+jets$	$864.49 \pm 63.44$	$1940.80 \pm 89.41$	$18563.70 \pm 408.99$
$Z+jets$	$19.51 \pm 2.00$	$46.63 \pm 4.77$	$795.20 \pm 80.89$
Total	$1470.35 \pm 84.89$	$2856.71 \pm 126.74$	$27098.28 \pm 594.01$
Data	1495.00	2898.00	27120.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$14.95 \pm 6.61$	$27.57 \pm 14.12$	$24.33 \pm 20.32$
Single-top	$68.31 \pm 16.17$	$58.93 \pm 13.56$	$278.60 \pm 73.04$
$t\bar{t}$	$496.60 \pm 31.72$	$401.23 \pm 32.13$	$3834.49 \pm 104.60$
$W+jets$	$50.68 \pm 4.19$	$144.02 \pm 7.86$	$450.01 \pm 11.87$
$Z+jets$	$1.32 \pm 0.14$	$5.35 \pm 0.55$	$29.96 \pm 3.07$
Total	$631.87 \pm 36.45$	$637.10 \pm 38.44$	$4617.39 \pm 129.77$
Data	636.00	634.00	4615.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	$596.34 \pm 91.52$
Muon Multi-jet	-	-	$481.01 \pm 144.48$
Diboson	$148.84 \pm 48.64$	$181.42 \pm 67.30$	$395.52 \pm 318.06$
Single-top	$79.49 \pm 19.80$	$56.82 \pm 14.89$	$782.07 \pm 190.79$
$t\bar{t}$	$338.42 \pm 24.14$	$236.80 \pm 20.88$	$4261.70 \pm 138.98$
$W+jets$	$501.13 \pm 39.36$	$1347.76 \pm 64.50$	$11445.73 \pm 291.49$
$Z+jets$	$9.25 \pm 0.95$	$28.77 \pm 2.95$	$567.66 \pm 57.94$
Total	$1077.13 \pm 69.93$	$1851.57 \pm 96.73$	$18530.03 \pm 523.88$
Data	1096.00	1846.00	18530.00

**Table 10.4:** Expected and Measured for VBF  $WW$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

	HP WCR	LP WCR	Resolved WCR
Electron Multi-jet	-	-	$870.00 \pm 132.75$
Muon Multi-jet	-	-	$618.45 \pm 196.90$
Diboson	$92.92 \pm 41.77$	$145.90 \pm 64.26$	$228.62 \pm 114.62$
Single-top	$71.13 \pm 16.29$	$118.82 \pm 27.98$	$1209.87 \pm 281.64$
$t\bar{t}$	$427.80 \pm 29.72$	$509.19 \pm 46.57$	$6860.87 \pm 254.83$
$W+jets$	$871.68 \pm 64.22$	$2020.67 \pm 93.54$	$19088.50 \pm 442.10$
$Z+jets$	$19.58 \pm 2.01$	$47.39 \pm 4.85$	$800.19 \pm 82.02$
Total	$1483.11 \pm 83.79$	$2841.97 \pm 125.92$	$29676.50 \pm 644.96$
Data	1495.00	2898.00	29755.00
	HP TCR	LP TCR	Resolved TCR
Electron Multi-jet	-	-	-
Muon Multi-jet	-	-	-
Diboson	$10.12 \pm 4.51$	$12.73 \pm 6.55$	$14.23 \pm 7.49$
Single-top	$51.57 \pm 12.31$	$35.07 \pm 8.17$	$169.21 \pm 44.54$
$t\bar{t}$	$470.06 \pm 28.97$	$298.99 \pm 25.28$	$2414.75 \pm 75.42$
$W+jets$	$49.64 \pm 4.17$	$109.69 \pm 6.16$	$378.22 \pm 12.05$
$Z+jets$	$1.28 \pm 0.13$	$4.81 \pm 0.50$	$17.62 \pm 1.83$
Total	$582.67 \pm 32.07$	$461.30 \pm 28.05$	$2994.03 \pm 88.75$
Data	584.00	459.00	3001.00
	HP SR	LP SR	Resolved SR
Electron Multi-jet	-	-	$444.65 \pm 67.99$
Muon Multi-jet	-	-	$397.29 \pm 125.59$
Diboson	$109.66 \pm 44.13$	$112.28 \pm 46.45$	$265.75 \pm 139.43$
Single-top	$63.16 \pm 15.20$	$48.02 \pm 11.56$	$872.16 \pm 205.00$
$t\bar{t}$	$348.95 \pm 24.34$	$190.68 \pm 17.75$	$5134.25 \pm 193.57$
$W+jets$	$467.21 \pm 37.12$	$973.73 \pm 47.91$	$10226.83 \pm 254.67$
$Z+jets$	$8.15 \pm 0.84$	$23.62 \pm 2.43$	$558.48 \pm 57.25$
Total	$997.13 \pm 64.42$	$1348.33 \pm 70.06$	$17899.41 \pm 432.98$
Data	1018.00	1313.00	17826.00

**Table 10.5:** Expected and Measured for VBF  $WZ$   $W+jets$ ,  $t\bar{t}$  control regions and signal regions.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.905^{+0.0166}_{-0.0166}$
XS_Top_Merg	$0.936^{+0.0199}_{-0.0199}$
XS_Top_Res	$0.957^{+0.0134}_{-0.0134}$
XS_Wjets_LP_lvqq_Merg_binned	$0.884^{+0.00489}_{-0.00489}$
XS_Wjets_Merg	$0.931^{+0.00831}_{-0.00831}$
XS_Wjets_Res	$1.03^{+0.00628}_{-0.00628}$

**Table 10.6:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the DY  $WW$  analysis region.

Background	Fitted Normalization
XS_Top_LP_Tag_lvqq_Merg_binned	$0.973^{+0.0333}_{-0.0333}$
XS_Top_LP_lvqq_Merg_binned	$0.894^{+0.0135}_{-0.0135}$
XS_Top_Merg	$0.893^{+0.016}_{-0.016}$
XS_Top_Res	$0.965^{+0.0179}_{-0.0179}$
XS_Top_Tag_lvqq_Merg_binned	$0.954^{+0.0276}_{-0.0276}$
XS_Top_Tag_lvqq_Res_binned	$0.999^{+0.0105}_{-0.0105}$
XS_Wjets_LP_Tag_lvqq_Merg_binned	$0.912^{+0.0703}_{-0.0703}$
XS_Wjets_LP_lvqq_Merg_binned	$0.876^{+0.00502}_{-0.00502}$
XS_Wjets_Merg	$0.948^{+0.00779}_{-0.00779}$
XS_Wjets_Res	$1.01^{+0.00673}_{-0.00673}$
XS_Wjets_Tag_lvqq_Merg_binned	$0.906^{+0.117}_{-0.117}$
XS_Wjets_Tag_lvqq_Res_binned	$1.2^{+0.0904}_{-0.0904}$

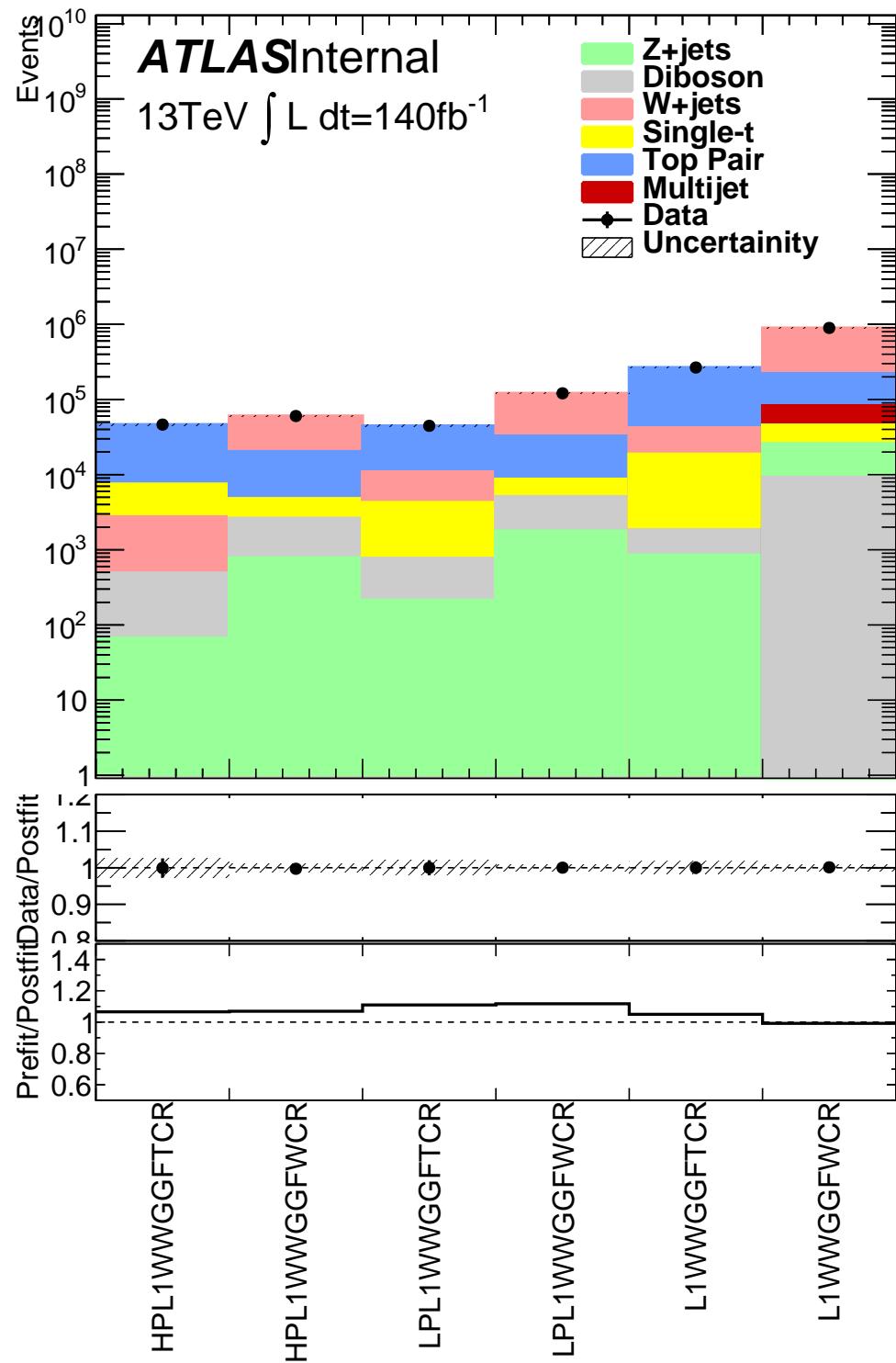
**Table 10.7:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the DY  $WZ$  analysis region.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.79^{+0.0673}_{-0.0673}$
XS_Top_Merg	$0.888^{+0.061}_{-0.061}$
XS_Top_Res	$1.01^{+0.0311}_{-0.0311}$
XS_Wjets_LP_lvqq_Merg_binned	$0.88^{+0.0423}_{-0.0423}$
XS_Wjets_Merg	$0.881^{+0.0677}_{-0.0677}$
XS_Wjets_Res	$0.932^{+0.0202}_{-0.0202}$

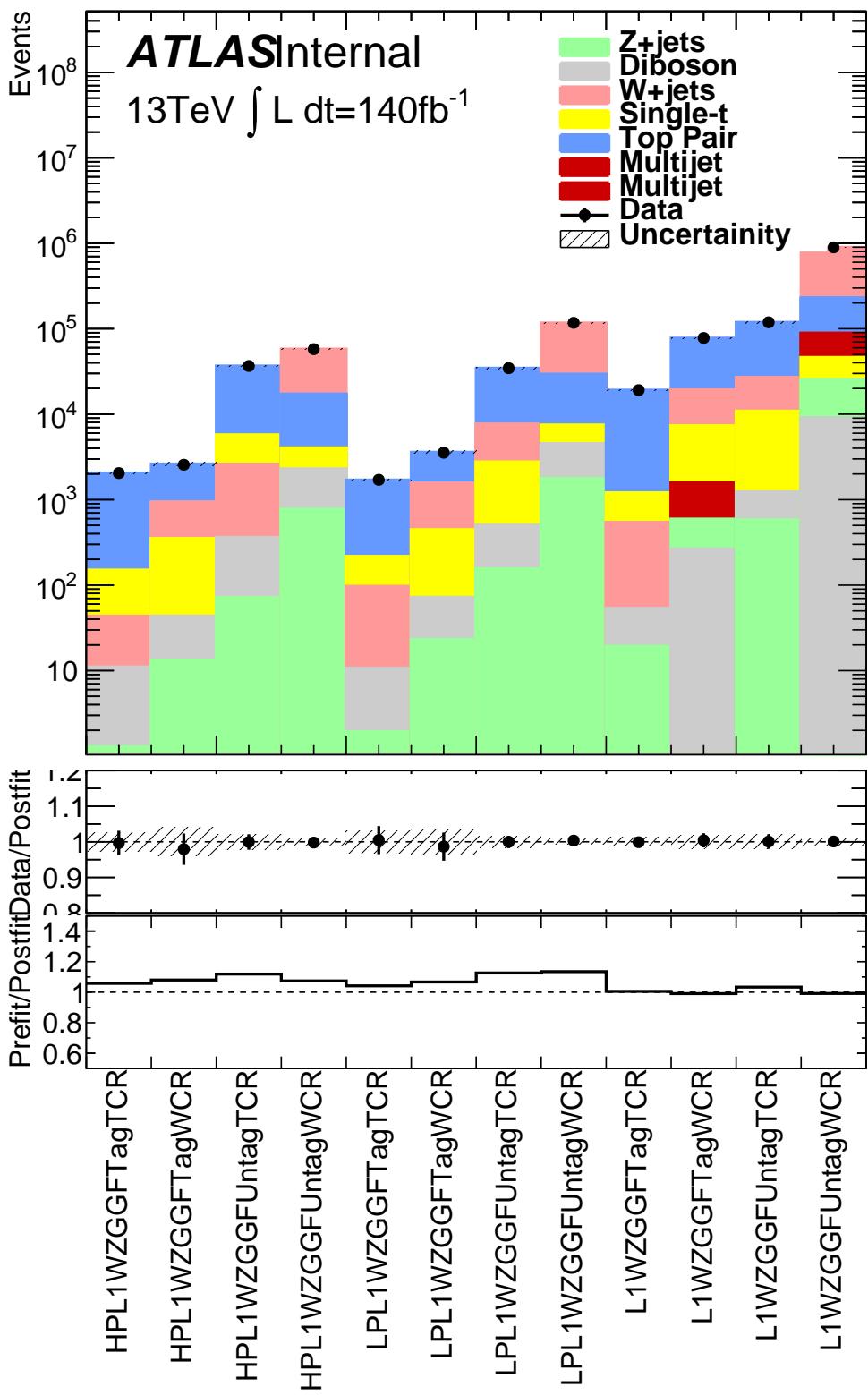
**Table 10.8:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the VBF  $WW$  analysis region.

Background	Fitted Normalization
XS_Top_LP_lvqq_Merg_binned	$0.708^{+0.064}_{-0.064}$
XS_Top_Merg	$0.958^{+0.0644}_{-0.0644}$
XS_Top_Res	$1.02^{+0.038}_{-0.038}$
XS_Wjets_LP_lvqq_Merg_binned	$0.9^{+0.0438}_{-0.0438}$
XS_Wjets_Merg	$0.883^{+0.0685}_{-0.0685}$
XS_Wjets_Res	$0.945^{+0.0219}_{-0.0219}$

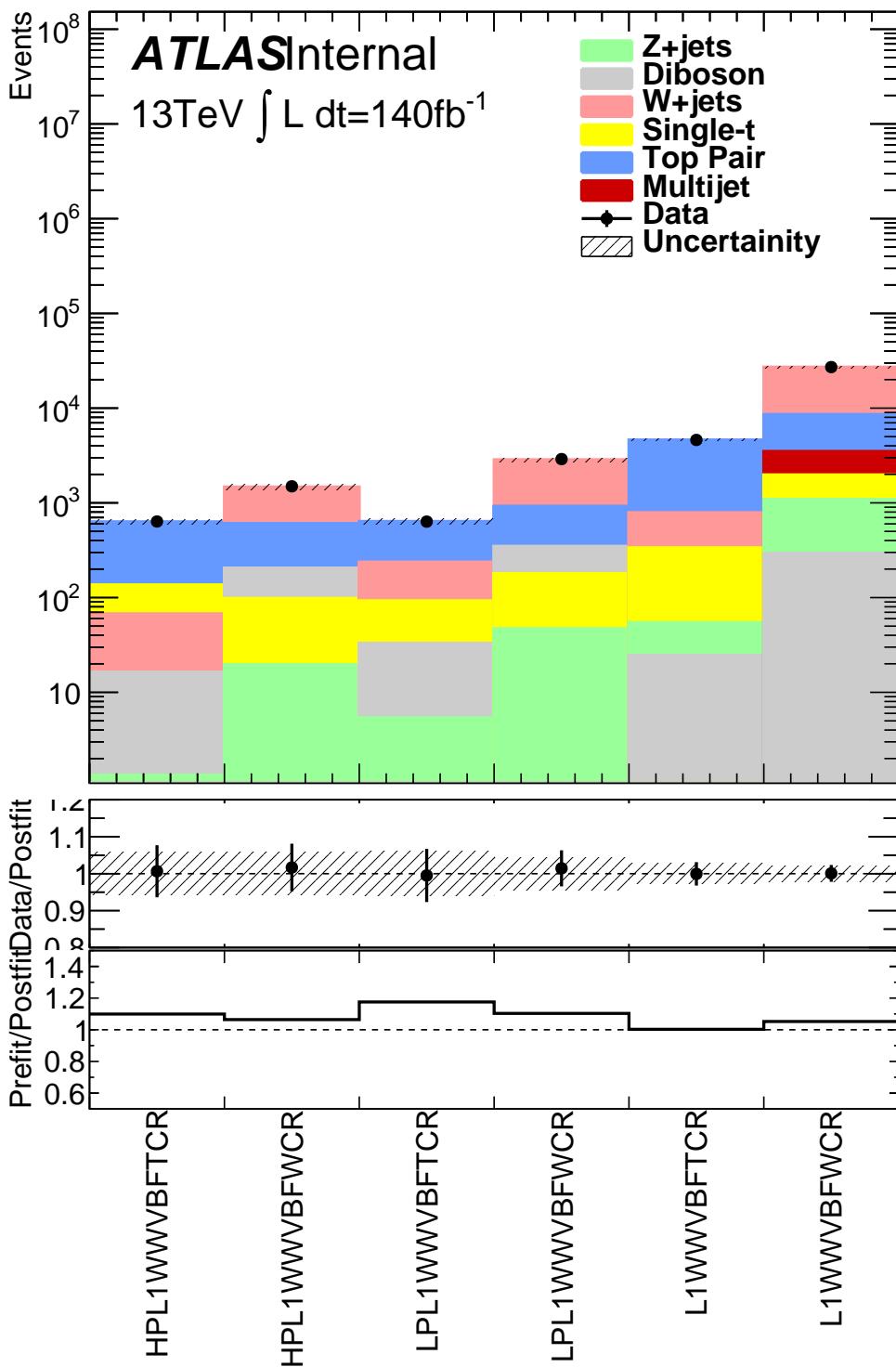
**Table 10.9:** Fitted background normalizations for  $t\bar{t}$  and  $W+jets$  backgrounds for the VBF  $WZ$  analysis region.



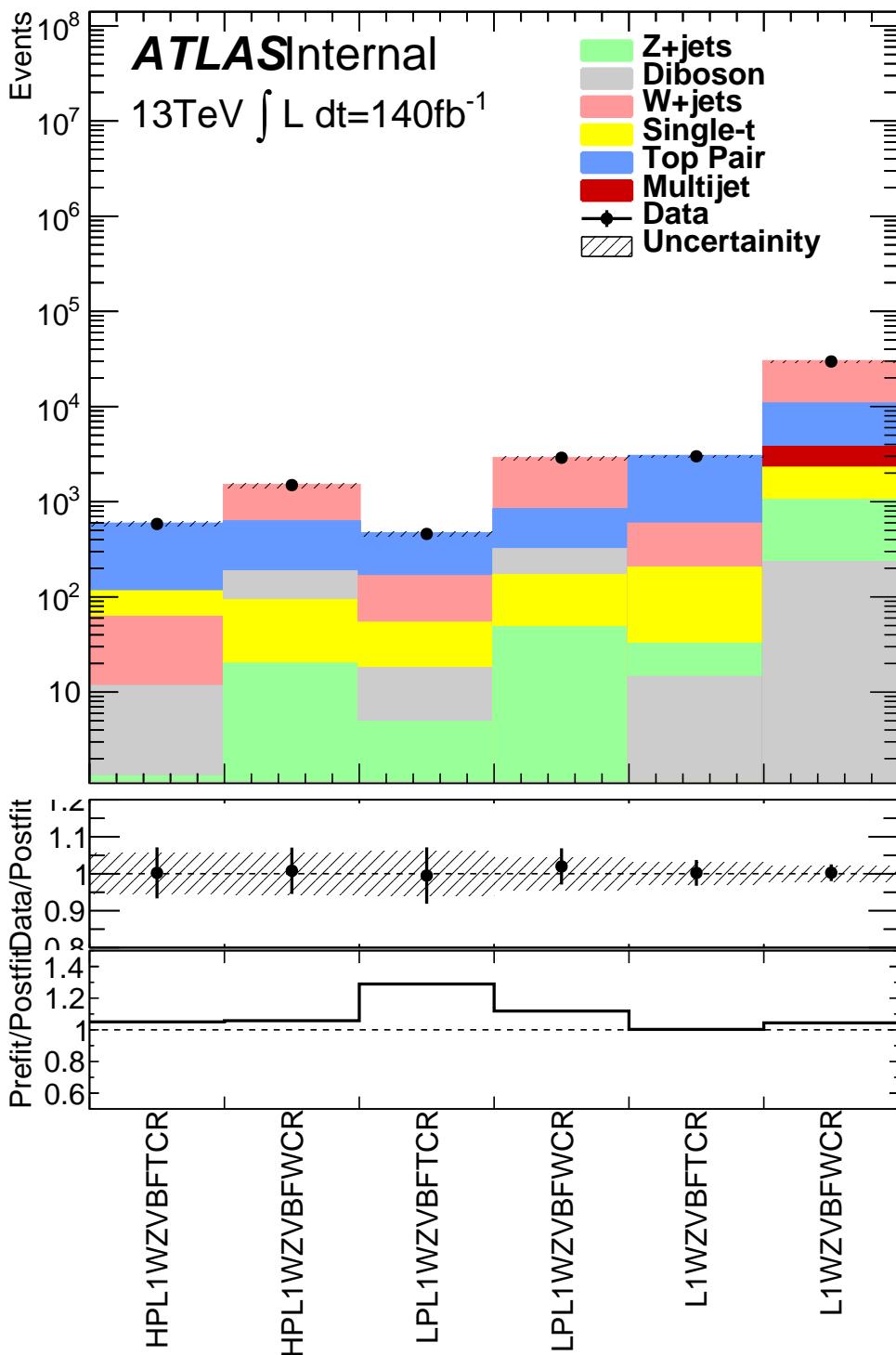
**Figure 10.9:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the DY WW control regions.



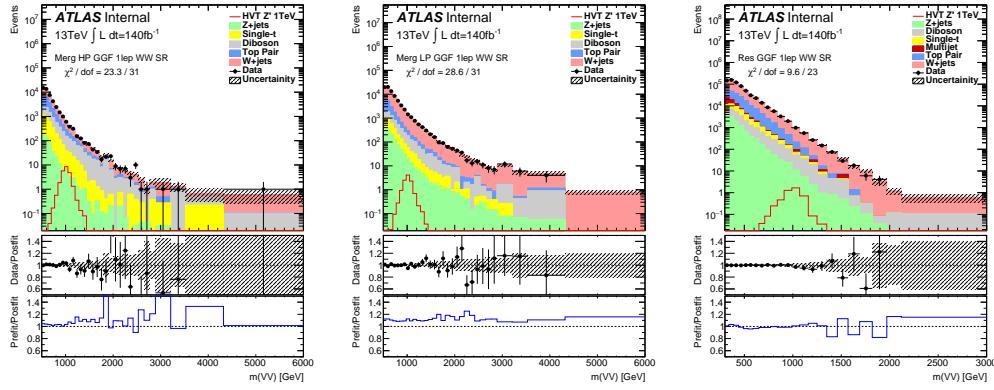
**Figure 10.10:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the DY  $WZ$  control regions.



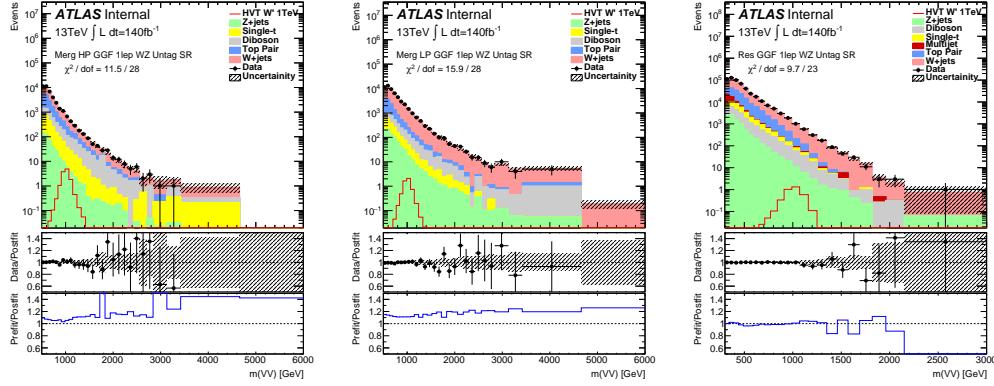
**Figure 10.11:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WW$  control regions.



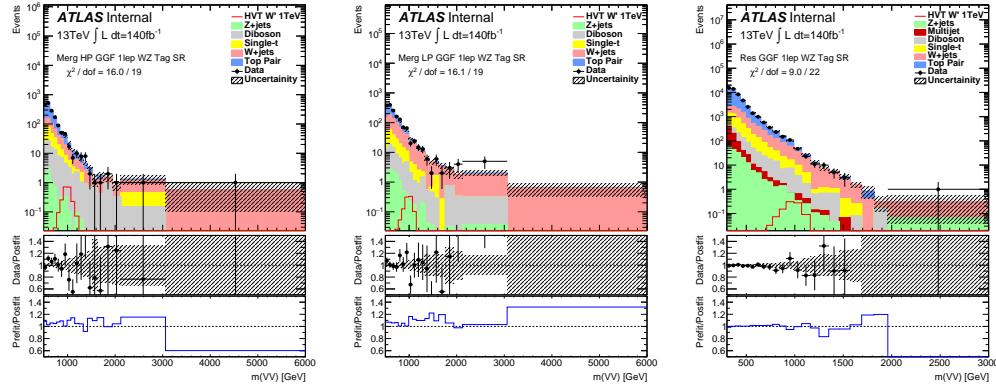
**Figure 10.12:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  control regions.



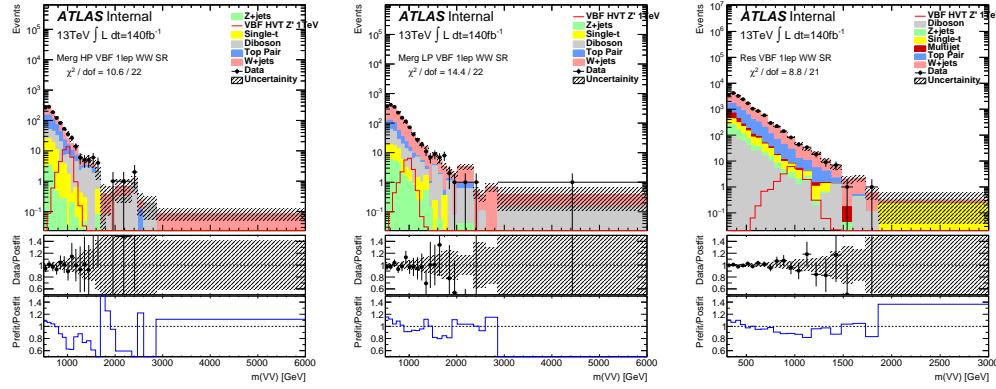
**Figure 10.13:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WW$  signal regions.



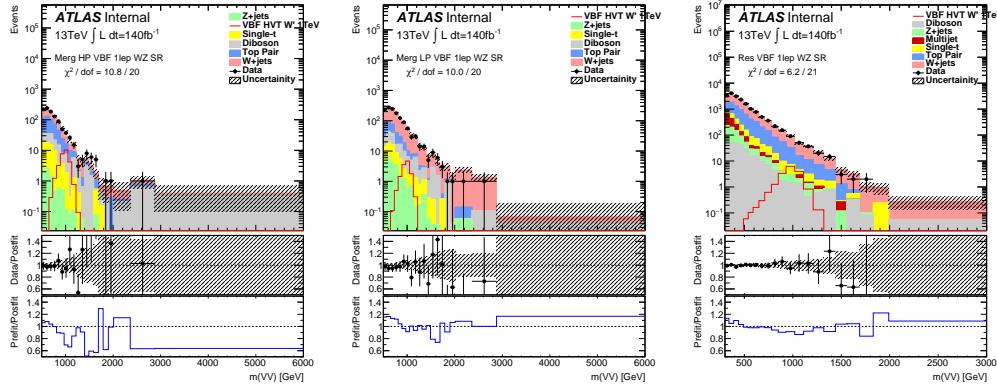
**Figure 10.14:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WZ$  Untag signal regions.



**Figure 10.15:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the GGF  $WZ$  Tag signal regions.



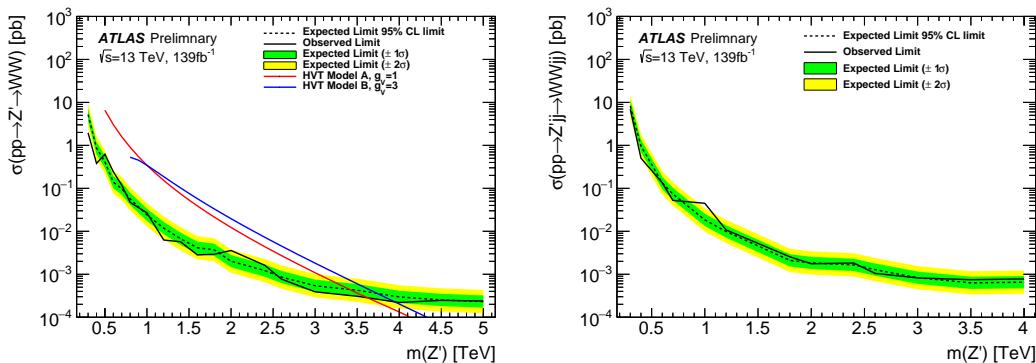
**Figure 10.16:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  Tag signal regions.



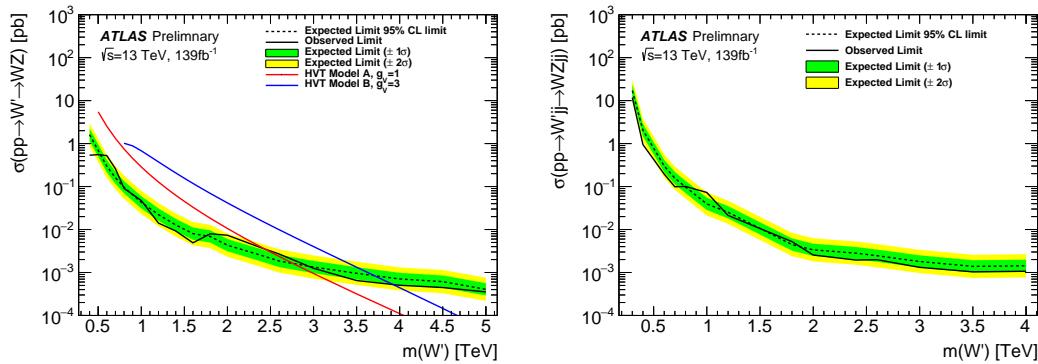
**Figure 10.17:** This figure shows the distribution of  $m_{\ell\nu qq}$  in the VBF  $WZ$  Tag signal regions.

## 1189 10.4 Limits

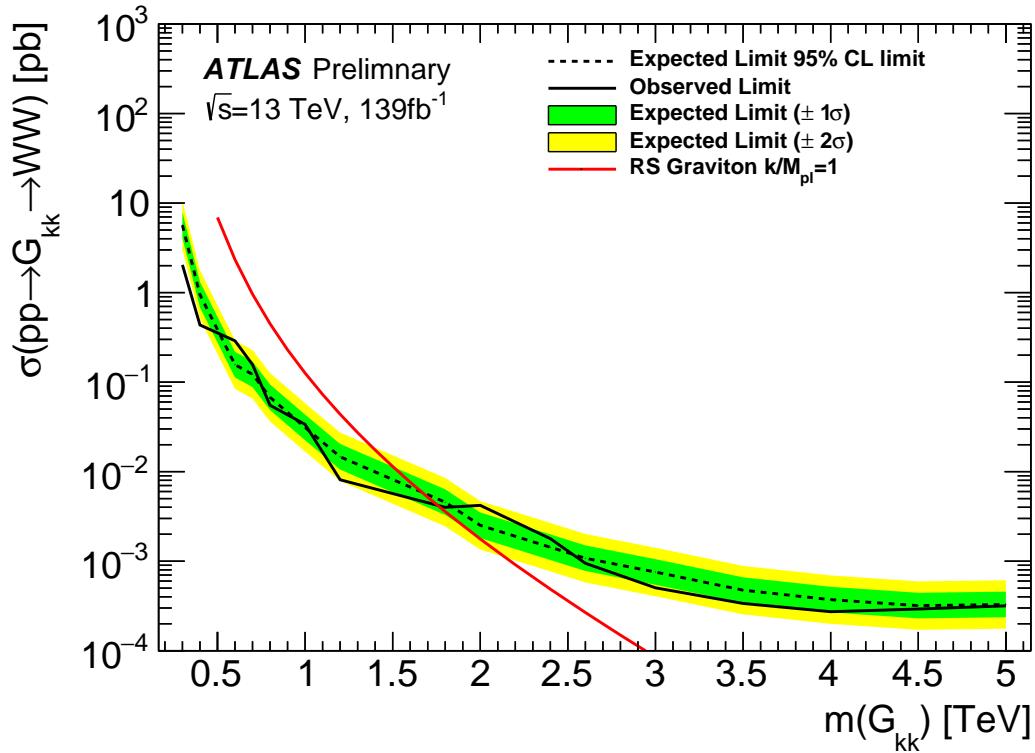
1190 Using the exclusion limits tests discussed previously, exclusion limits are set  
 1191 on  $\mu$  and consequently cross-sections for different signal models. Exclusion limits  
 1192 for the models considered are shown in Figure 10.18 - 10.20. These limits exclude  
 1193 HVT Model A  $W' < \text{blah}$  and  $Z' < \text{blah}$  and Model B  $W' < \text{blah}$  and  $Z' < \text{blah}$ .  
 1194 Randall Sundrum Gravitons are excluded for masses below blah.



**Figure 10.18:** This figure shows theory, expected and observed limits for HVT  $W'$  DY (left) and VBF (right) production.



**Figure 10.19:** This figure shows theory, expected and observed limits for HVT  $Z'$  DY (left) and VBF (right) production.



**Figure 10.20:** This figure shows theory, expected and observed limits for RS Gravitons via DY production.

1195

## Part V

1196

# Quark and Gluon Tagging

<sub>1197</sub> **Chapter 11**

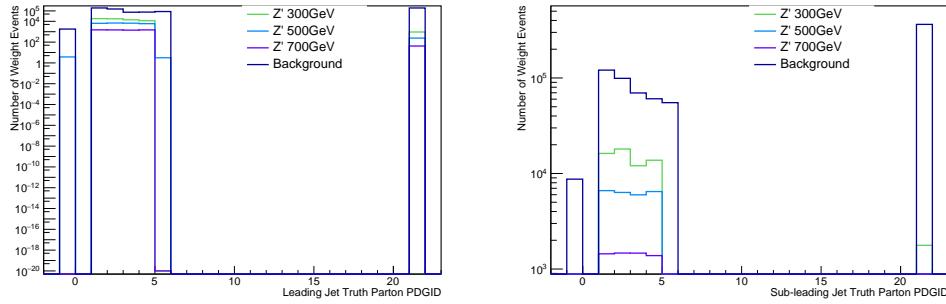
<sub>1198</sub> **Prospects**

<sub>1199</sub> For the resolved analysis, signal jets are quark enriched and background jets are  
<sub>1200</sub> gluon dominated. By classifying jets in the event as quark or gluon initiated, less  
<sub>1201</sub> background would contaminate the signal region. Figure 11.1 shows the PDGID  
<sub>1202</sub> for the truth parton matched to the jet (meaning the highest energy parton in  
<sub>1203</sub> the jet catchment area) in events passing the resolved signal region selections.  
<sub>1204</sub> PDGID = -1 corresponds to pileup jets,  $0 < \text{PDGID} < 6$  correspond to quarks  
<sub>1205</sub> and  $\text{PDGID} = 21$  corresponds to gluons. From this Figure, it is evident that a  
<sub>1206</sub> notable fraction of the background that contaminates the signal region contains  
<sub>1207</sub> gluon jets, especially for the sub-leading jet.

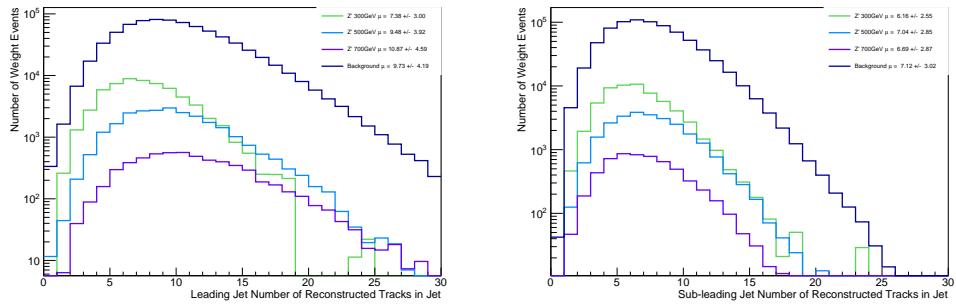
<sub>1208</sub> As gluons jets have more constituents and therefore more tracks ( $n_{trk}$ ), the  
<sub>1209</sub> background jets have more tracks than the signal jets. This is shown in Fig-  
<sub>1210</sub> ure 11.2. Therefore, by cutting on the number of tracks in a jet, quark and gluon  
<sub>1211</sub> jets may be distinguished (i.e. jets with less than a given number of tracks are  
<sub>1212</sub> classified as a quark, otherwise the jet is classified as a gluon.) Moreover, as the  
<sub>1213</sub> momentum of the jet increases the number of tracks also increases logarithmically  
<sub>1214</sub> [Cite nachman thesis Natasha]. Therefore by applying a cut on the number of  
<sub>1215</sub> tracks that scales with the  $\ln(p_T)$  is more powerful than a threshold cut on the

1216 number of tracks. Figure 11.3-Figure 11.6 show normalized heat maps of  $\ln(p_T)$   
1217 vs the number of reconstructed tracks for the background and a 300 GeV Z' signal.  
1218 In these plots it is evident that the number of tracks in the background jets grows  
1219 more quickly with  $\ln(p_T)$  than for the signal jets. This is expected given that the  
1220 signal is quark dominated and the background is gluon dominated.

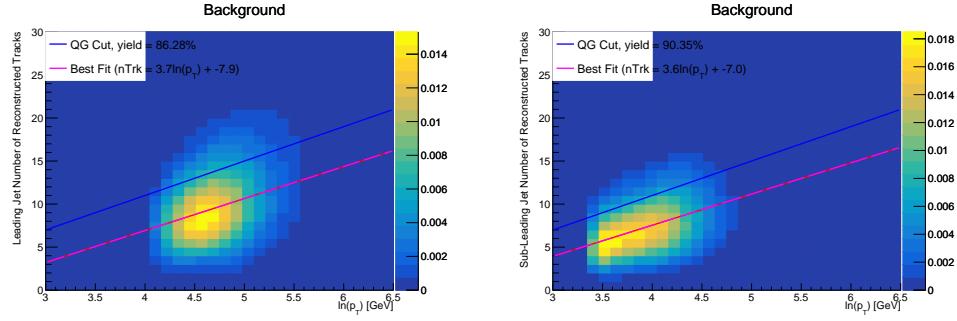
1221 In Figure 11.8 is the ROC Curve for quark gluon tagging with cut on the  
1222 number of tracks in a jet that depends on  $\ln(p_T)$ . The sum of the backgrounds in  
1223 the signal region were used for this curve. Here the quark tagging efficiency is the  
1224 ratio of quarks tagged as quarks to the total number of quarks in the signal region.  
1225 The gluon rejection is calculated as the reciprocal of the gluon tagging efficiency.  
1226 Choosing a 90% efficient working point with a rejection of 1.4 corresponds to a  
1227 slope of 4 and intercept of -5. Tagging both jets in this analysis would yield an  
1228 efficiency of  $90\%^{n_{jets}}$ . Focusing on the background in Figure 11.9, this cut helps  
1229 minimize gluon contamination in the signal region. Also, from these heat maps it  
1230 is obvious that the number of tracks in gluon jets grows more quickly than those  
1231 in quark jets.



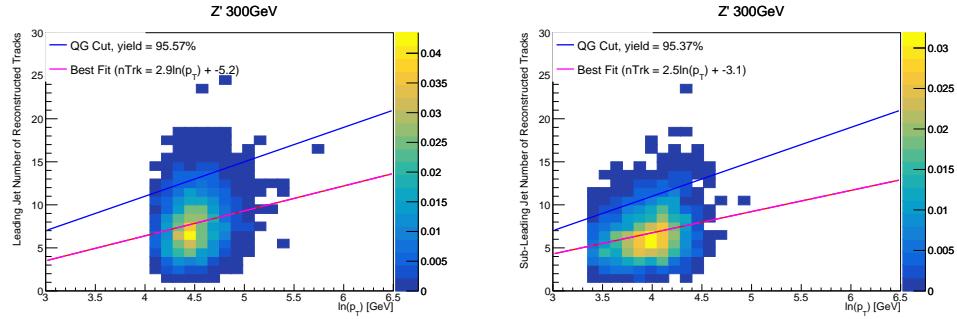
**Figure 11.1:** PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV Z' signals and the background.



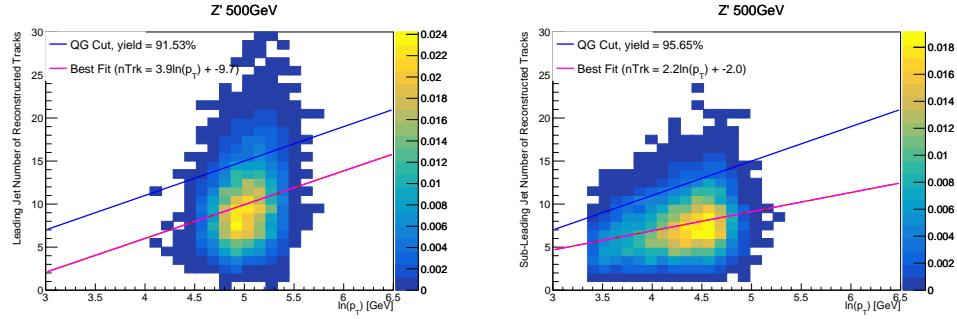
**Figure 11.2:** The number of tracks in small-R jets in events passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets. These distributions are shown for 300, 500, and 700GeV Z' signals and the background.



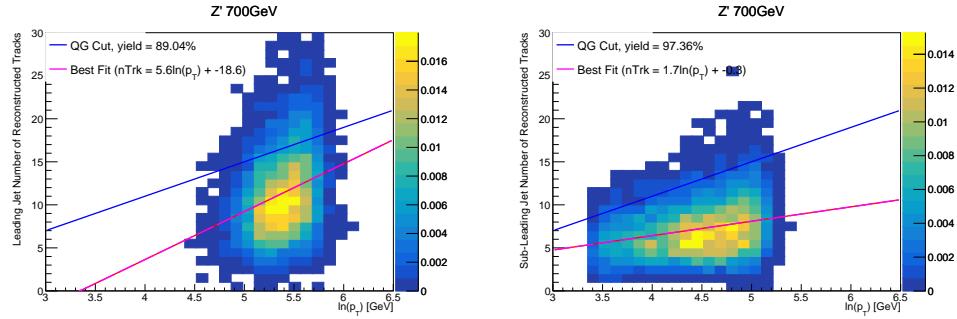
**Figure 11.3:** The number of tracks in background small-R jets in events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



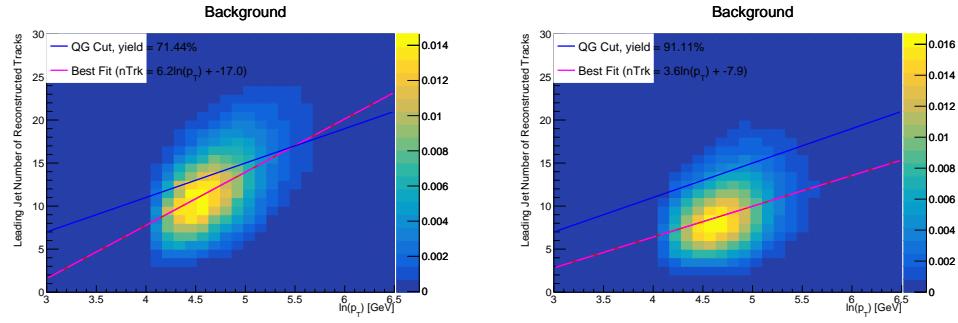
**Figure 11.4:** The number of tracks in small-R jets in 300GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



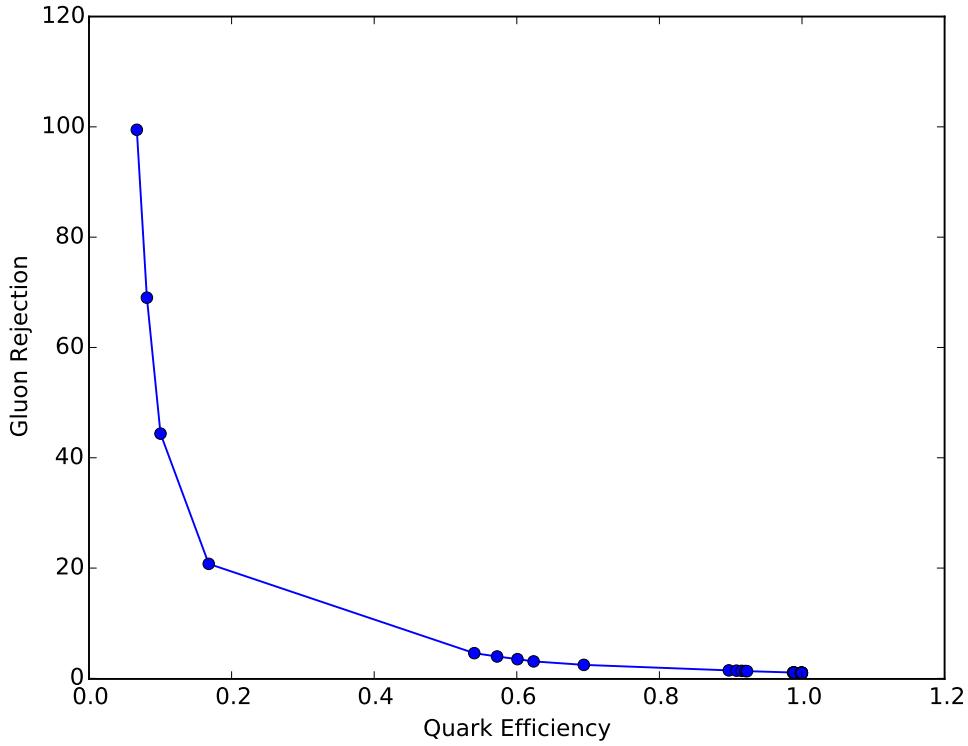
**Figure 11.5:** The number of tracks in small-R jets in 500GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



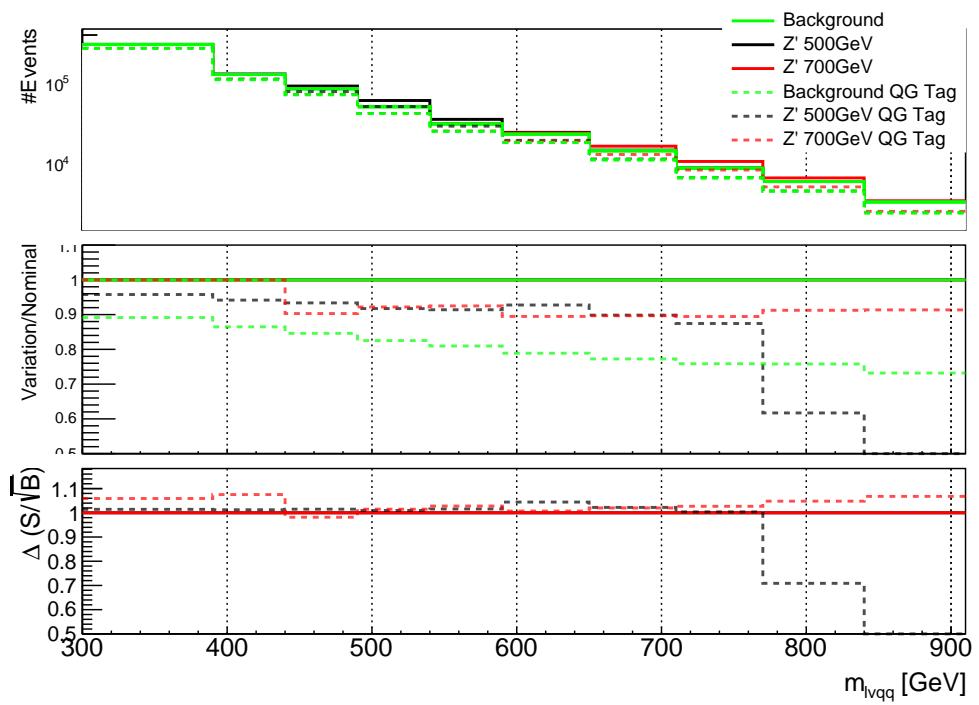
**Figure 11.6:** The number of tracks in small-R jets in 700GeV  $Z'$  events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Leading (b) Sub-Leading jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ . Note the number of total entries in these plots has been normalized to one.



**Figure 11.7:** The number of tracks in leading small-R jets in background events passing the Resolved GGF WW Signal region selection vs.  $\ln(p_T)$  for (a)Gluons (b) Quarks jets. The best fit line for the distribution is also shown, as well as the percentage of jets that pass a cut of number of tracks  $< 4 \times \ln(p_T) - 5$ .Note the number of total entries in these plots has been normalized to one.



**Figure 11.8:** ROC Curve for Quark and Gluon Tagging with a cut on the number of tracks that depends on the  $\ln(p_T)$ .



**Figure 11.9:** The top panel shows the distribution of  $m_{lvqq}$  with and without quark gluon tagging. The middle panel shows the ratio of the signals and backgrounds with and without quark gluon tagging. The bottom panel shows the change in  $S/\sqrt{B}$  with quark gluon tagging.

# 1232 Chapter 12

## 1233 $n_{trk}$ Calibration

1234 As tagger based on nTrk cuts on the number of tracks in jets, a quantity that  
1235 is not known with infinite precision, means that relevant systematic uncertainties  
1236 must be evaluated. The sources of uncertainty in  $n_{trk}$  may be split into modeling  
1237 and experimental uncertainties.

1238 Modeling uncertainties are obtained by assessing PDF and ME uncertainties  
1239 on the number of charged particles in particle-level jets in dijet events. The  
1240 number of charged particles as a function of jet  $p_T$  is calculated using an Iterative  
1241 Bayesian (IB) technique [cite paper].

1242 This measurement ([7]) uses the ATLAS 2012 pp collision dataset, correspond-  
1243 ing to  $20.3/\text{fb}$  at center-of-mass energy  $\sqrt{s} = 8\text{TeV}$ . The number of charged con-  
1244 stituents depends on fragmentation modeling and matrix elements, which do not  
1245 depend on  $s$ . For this reason, it is safe to use these uncertainties for  $\text{sqrt}(s)=13\text{TeV}$ .  
1246 Monte Carlo (MC) samples are used to determine the response matrix. The MC  
1247 sample is a dijet sample generated with Pythia 8.175 using CT10 PDF and AU2  
1248 tune. The anti- $k_T$  algorithm is used to cluster jets with a radius parameter  $R$   
1249  $= 0.4$ . Jets are required to have  $|\eta| < 2.1$ . Tracks in jets are required to have  
1250  $p_T > 500\text{MeV}$ ,  $|\eta| < 2.5$ , track-fit  $\chi^2 < 3.0$  and originate from the primary ver-

tex. Matching tracks to jets is accomplished using ghost-association [cite]. In this technique, jets are re-clustered with the track collection augmented with "ghost" versions of tracks. These "ghosts" tracks have the same direction as their parent track, but infinitesimal track  $p_T$ . This insures meta-jet properties (e.g.  $\eta$ ,  $p_T$ , etc) are unchanged. A track is matched to a jet if it's ghost version remains in the jet after re-clustering. Further details of the data, object, and event selection may be found in [cite 35].

To select dijet topologies events are required to have at least two jets with  $p_T > 50GeV$  that are relatively well-balanced ( $p_T^{lead}/p_T^{sub-lead} < 1.5$ ).

In the IB technique, the prior distribution and number of iterations are the inputs [cite Bayesian paper]. The IB response matrix connects number of charged particles to the number of tracks in jets determined using the simulated samples. This response matrix is used to unfold data to extract the  $n_c$ . Before applying the response matrix a fake factor is applied. This accounts for jets that pass detector level selections, but not particle level selections. Following this, the IB method iteratively applies the response matrix using the nominal Pythia 8.175 sample as a prior. The number of IB iterations is chosen to minimize unfolding bias and statistical fluctuations. For this measurement four iterations was found to be optimal by minimizing the unfolding bias from pseudodata simulated with Herwig++ with a prior from Pythia 8 AU2. Finally, the inefficiency factor is applied to account for events passing particle level selection but not detector level, yielding the unfolded nCharged distribution.

This process is prone to three main sources of bias: response matrix, correction factor, and unfolding procedure uncertainties. The response matrix is sensitive to experimental uncertainties impacting jet track reconstruction and calorimeter jet  $p_T$ . Correction factors are also sensitive to experimental uncertainties (e.g. JES)

1277 as such uncertainties modify detector level acceptance. Sensitivity to particle  
 1278 level acceptance is calculated by comparing Pythia and Herwig. Finally, the bias  
 1279 from the IB prior choice is determined by reweighting the particle-level spectrum,  
 1280 so the simulated detector level spectrum more closely matches the uncorrected  
 1281 data. Unfolding this modified detector-level simulation and comparing it to the re-  
 1282 weighted particle-level spectrum indicates bias from the prior distribution choice.  
 1283

1283 A summary of all the systematic uncertainties associated with this unfolding  
 1284 may be found in [ref paper]. Total uncertainties are < 7% for the number of  
 1285 charged particles in jets. The unfolded distribution of the nCharged in jets from  
 1286 data are further analyzed to extract the quark and gluon nCharged distributions.  
 1287 In dijet events, the jet with a larger  $\eta$  is more energetic and therefore more likely  
 1288 to be a quark. This is due to the quarks in protons generally having a larger  
 1289 fraction of the total momentum of the proton constituents. The more central jet  
 1290 is more likely to be a gluon-initiated jet. This correlation between jet  $\eta$  and flavor  
 1291 may then be used to extract nCharged in  $p_T$  bins using:  
 1292

$$\langle n_c^f \rangle = f_q^f \langle n_c^q \rangle + f_g^f \langle n_c^g \rangle \quad (12.1)$$

$$\langle n_c^c \rangle = f_q^c \langle n_c^q \rangle + f_g^c \langle n_c^g \rangle \quad (12.2)$$

1293 In this equation the f and c subscripts denote the more forward and central  
 1294 jets, respectively. The q and g subscripts denote quark and gluon. The fraction  
 1295 of more forward jets that are say gluons is denoted by  $f_g^f$ . The other relevant jet  
 1296 fractions are denoted with the same naming scheme. Finally,  $\langle n_c \rangle$  is the average  
 1297 number of charged particles in a jet in a given  $p_T$  bin. To show that Eq. (??) may  
 1298 be used to extract quark and gluon  $n_c$  distributions the extracted distributions  
 1299 are compared to  $n_c$  distributions determined using the jet flavor in simulation.  
 1300 Figure [add figure natasha] shows that the extracted and true distributions differ

1301 by < 1% over the  $p_T$  range probed for this study. Moreover, this implies that  $n_c$   
1302 depends only on the flavor of the initiating parton and jet  $p_T$ .

1303 These extracted distributions are prone to PDF and ME biases. The bias from  
1304 the choice of the CT10 PDF for the Pythia sample is accounted for by comparing  
1305 quark/gluon fractions for the nominal CT10 sample with its eigenvector variations.  
1306 Comparing the quark/gluon fractions from Pythia 8 and Herwig++ quantify the  
1307 uncertainty from the ME calculation. These uncertainties are added in quadra-  
1308 ture with the unfolding uncertainty to give the total modelling uncertainty on  
1309 the extracted  $n_c$  distribution. This is shown in Figure 13.2.

1310 To apply these uncertainties in  $n_c$  distributions in data, per-jet event weights  
1311 are associated with each uncertainty according to:

$$w_i(n_c) = \frac{P(n_c | n_c > \pm \sigma_{n_c}^i)}{P(n_c | n_c >)} \quad (12.3)$$

1312 In Eq. (??), i denotes the uncertainty considered, P is the Poisson probability,  
1313 and  $\sigma_{n_c}^i$  represents the average impact of the uncertainty on  $n_c$ .

1314 The previous uncertainties described accounted for modeling uncertainty as-  
1315 sociated with the number of charged particles in a jet. However,  $n_c$  is not a  
1316 measurable quantity. Instead the number of tracks in a jet is measured, which is  
1317 a proxy for  $n_c$ . Therefore the uncertainties associated with the measurement of  
1318 nTracks must also be considered ([9]). These uncertainties were calculated using  
1319 a Pythia 8 dijet sample with NNPDF 23 and Run 2 data. Track reconstruction  
1320 efficiency and fake rates are the dominant sources of nTrack uncertainties.

1321 The track reconstruction efficiency is affected by the uncertainty of the de-  
1322 scription of the ID material in simulation and the modeling of charged-particle  
1323 interactions with this material. These uncertainties are accounted for by varying  
1324 the ID material by 5-25% (dependent on the region of the detector considered).

1325 The difference in the tracking efficiency between the nominal and varied simula-  
1326 tion give the uncertainty on the track reconstruction efficiency. Another important  
1327 source of track reconstruction inefficiency arises in the core of jets. The high den-  
1328 sity of tracks in the jet cores can cause ID clusters to merge. The fraction of lost  
1329 tracks due to merging is given by the fraction of tracks that have a charge of two  
1330 minimum ionizing particles. This quantity is compared between data and simu-  
1331 lation resulting in an uncertainty of 0.4% on tracks with  $\Delta R < 0.1$ . Combining  
1332 these effects gives a total uncertainty as a function of  $p_T$  and  $\eta$  that is generally  
1333  $< 2\%$  [references figure 44 from [9]).

1334       Fake tracks are the other dominant source of nTrk uncertainty. Fake tracks  
1335 are tracks that cannot be associated to a single particle. Often these tracks are a  
1336 result of random combinations of hits from charged particles that overlap in space.  
1337 In dense environments, such as the core of jets or high-pileup environments, fake  
1338 tracks are more likely. Fake tracks are estimated with a 'control region method'  
1339 which is briefly summarized here [[8]]. By applying a series of track selections  
1340 to enrich the fraction of fake tracks (e.g.  $|d_0| > 0.1$ , track  $\chi^2 > 1.4$ , etc) in  
1341 simulation, templates for fake track parameters are calculated. These templates  
1342 are then fit to data to determine the fraction of fake tracks. On average the fake  
1343 rate is found to be 30% (independent of  $p_T$  and  $\eta$ ).

1344       To assess the impact of these two detector level uncertainties, tracks are ran-  
1345 domly dropped according to the rates described above. Reconstruction and fake  
1346 uncertainties both lower the number of tracks, hence these uncertainties are one-  
1347 sided. By dropping tracks in this way a varied nTrk distribution is calculated for  
1348 both uncertainties. The associated per-jet event weights are then calculated in  
1349 the same way as the modeling weights as:

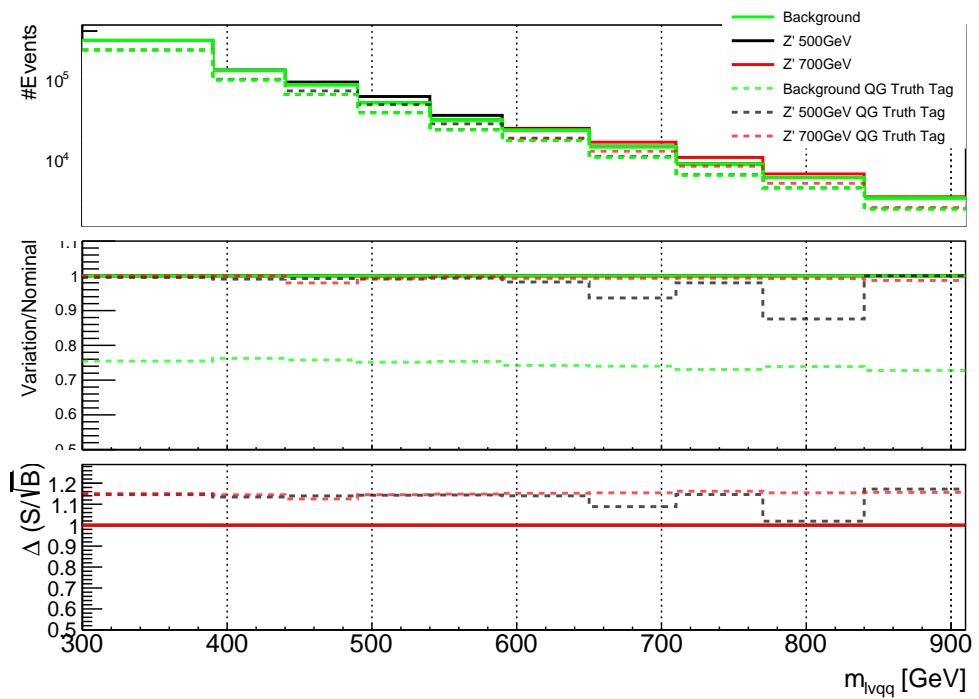
$$w_i(n_c) = \frac{P(n_{trk} | < n_{trk} > \pm \sigma_{n_{trk}}^i)}{P(n_{trk} | < n_{trk} >)} \quad (12.4)$$

1350        Adding the modeling and detector level uncertainties in quadrature gives the  
 1351        overall nTrack uncertainty. The effects of the individual uncertainties on the nTrk  
 1352        distributions can be seen in Fig 13.4. Fig 13.3 shows the  $m_{lvqq}$  and nTrk distri-  
 1353        butions for the W and Top control regions before likelihood fitting. In these plots  
 1354        the nTrk uncertainties improve agreement between data and MC. The remaining  
 1355        differences are likely covered by likelihood fitting and improving the analysis itself.

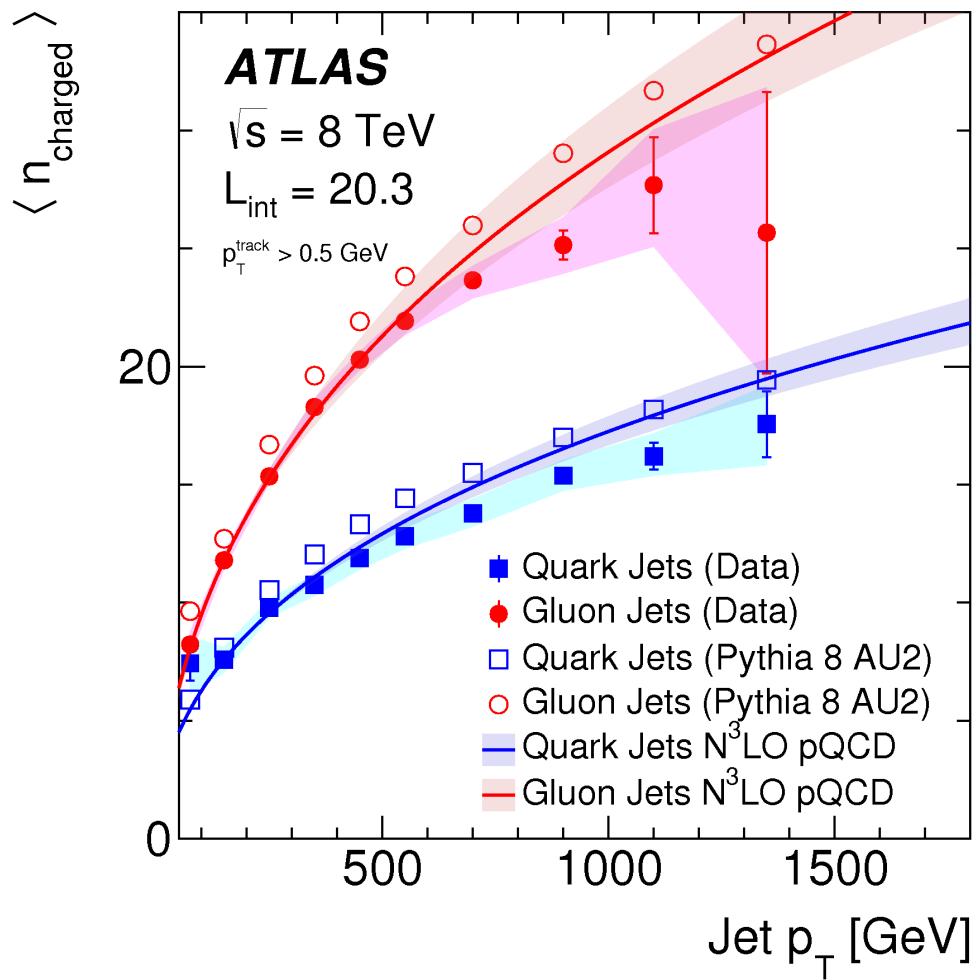
<sub>1356</sub> **Chapter 13**

<sub>1357</sub> **Application**

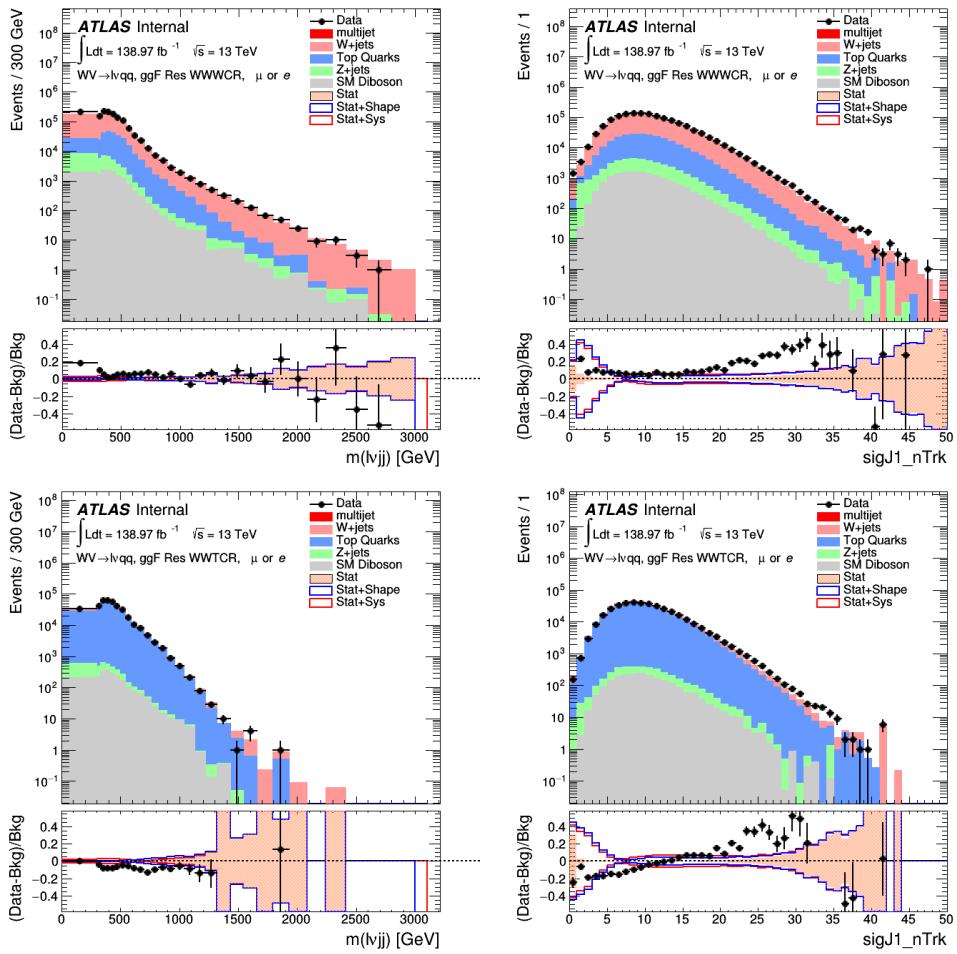
<sub>1358</sub> Using the 90% WP of the  $n_{trk}$  tagger improves  $S/\sqrt{B}$  is  $\sim 3\%$  as shown in  
<sub>1359</sub> Figure 11.9. Although,  $n_{trk}$  is the single most powerful discriminating variable  
<sub>1360</sub> for quark and gluon jets, the addition of other jet variables would improve the  
<sub>1361</sub> classification efficiency. Figure 13.1 shows the possible improvement of 10%  
<sub>1362</sub> in jet classification using the truth label of the jets to classify jets. This type of  
<sub>1363</sub> improvement is possible by using variables such as jet width, and energy correlata-  
<sub>1364</sub> tors. Figure [add BDT figure/use 1612.01551.pdf] shows for a 90% quark tagging  
<sub>1365</sub> efficiency for a 100 GeV jet, a BDT improve the gluon rejection by 0.4. Once this  
<sub>1366</sub> tagger is calibrated it would improve the analysis sensitivity of this channel.



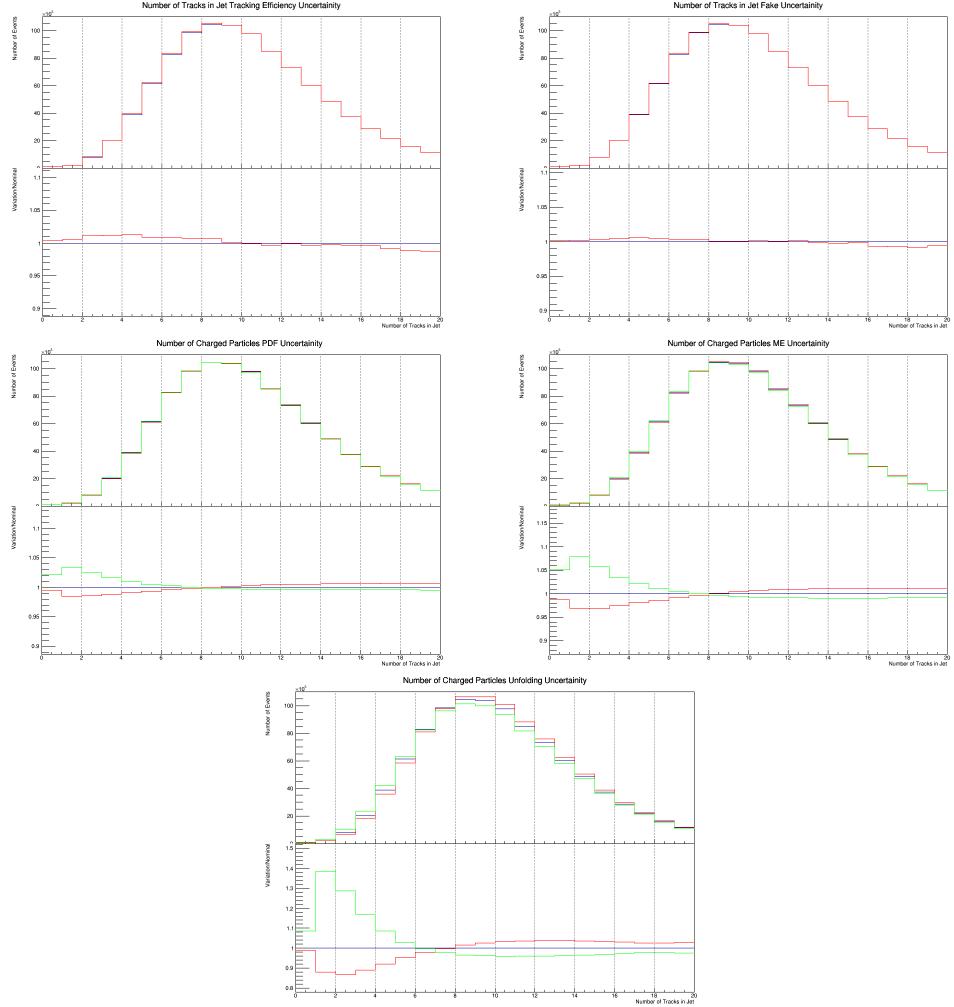
**Figure 13.1:** The top panel shows the distribution of  $m_{lvqq}$  with and without requiring jets to be true quarks. The middle panel shows the ratio of the signals and backgrounds with and without requiring jets to be true quarks. The bottom panel shows the change in  $S/\sqrt{B}$  when requiring jets to be true quarks..



**Figure 13.2:** Unfolded and extracted  $n_C$  qg dstbs..



**Figure 13.3:** PDGID of the truth-level parton matched to the small-R jets passing the Resolved GGF WW Signal Region selections for the (a) Leading (b) Sub-Leading jets . These distributions are shown for 300, 500, and 700GeV  $Z'$  signals and the background.



**Figure 13.4:** These figures show the impact of the uncertainties on the number of tracks in the leading jet in the sum of the background sample in the Resolved GGF WW SR (a) tracking efficiency (b) fake (c) PDF (d) ME (e) unfolding uncertainties.

## **Part VI**

1367

## **Conclusion**

1368

<sub>1369</sub> **Chapter 14**

<sub>1370</sub> **Conclusions**

<sub>1371</sub> This is where I am going to add my final thoughts.

# <sup>1372</sup> Bibliography

- <sup>1373</sup> [1] Lecture notes particle physics ii.
- <sup>1374</sup> [2] Kaustubh Agashe, Hooman Davoudiasl, Gilad Perez, and Amarjit Soni.  
<sup>1375</sup> Warped Gravitons at the LHC and Beyond. *Phys. Rev.*, D76:036006, 2007.
- <sup>1376</sup> [3] G. Altarelli and G. Parisi. Asymptotic freedom in parton language. *Nuclear  
1377 Physics B*, 126(2):298 – 318, 1977.
- <sup>1378</sup> [4] ATLAS Collaboration. Atlas muon reconstruction performance in lhc run 2.
- <sup>1379</sup> [5] ATLAS Collaboration. Summary plots from the atlas standard model physics  
1380 group.
- <sup>1381</sup> [6] ATLAS Collaboration. Jet energy scale measurements and their systematic  
1382 uncertainties in proton–proton collisions at  $\sqrt{s} = 13$  tev with the atlas  
1383 detector. arXiv: 1703.09665 [hep-ex].
- <sup>1384</sup> [7] ATLAS Collaboration. Measurement of the charged-particle multiplicity  
1385 inside jets from  $s=\sqrt{8}$  tev pp collisions with the atlas detector.  
1386 arXiv:1602.00988 [hep-ex].
- <sup>1387</sup> [8] ATLAS Collaboration. Performance of the atlas track reconstruction algo-  
1388 rithms in dense environments in lhc run 2. arXiv:1704.07983 [hep-ex].
- <sup>1389</sup> [9] ATLAS Collaboration. Properties of jet fragmentation using charged par-  
1390 ticles measured with the atlas detector in pp collisions at  $\sqrt{s} = 13$  tev.  
1391 arXiv:1906.09254 [hep-ex].
- <sup>1392</sup> [10] Alex Dias and V. Pleitez. Grand unification and proton stability near the  
1393 peccei-quinn scale. *Physical Review D*, 70, 07 2004.
- <sup>1394</sup> [11] Stefan Höche, Frank Krauss, Marek Schönherr, and Frank Siegert. Qcd ma-  
1395 trix elements + parton showers. the nlo case. *Journal of High Energy Physics*,  
1396 2013(4), Apr 2013.
- <sup>1397</sup> [12] Gregory Soyez Matteo Cacciari, Gavin P. Salam. The anti- $k_T$  jet clustering  
1398 algorithm. arXiv:0802.1189 [hep-ph].

- 1399 [13] Duccio Pappadopulo, Andrea Thamm, Riccardo Torre, and Andrea Wulzer.  
1400 Heavy vector triplets: bridging theory and data. *Journal of High Energy*  
1401 *Physics*, 2014(9), Sep 2014.
- 1402 [14] Antonio Pich. The Standard Model of Electroweak Interactions. In *Proceed-  
1403 ings, High-energy Physics. Proceedings, 18th European School (ESHEP 2010):  
1404 Raseborg, Finland, June 20 - July 3, 2010*, pages 1–50, 2012. [,1(2012)].
- 1405 [15] Lisa Randall and Raman Sundrum. A Large mass hierarchy from a small  
1406 extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, 1999.
- 1407 [16] Tania Robens and Tim Stefaniak. Lhc benchmark scenarios for the real higgs  
1408 singlet extension of the standard model. *The European Physical Journal C*,  
1409 76(5), May 2016.
- 1410 [17] Alex Sherstinsky. Fundamentals of recurrent neural network (RNN) and long  
1411 short-term memory (LSTM) network. *CoRR*, abs/1808.03314, 2018.