

A SEARCH FOR THE HIGGS BOSON PRODUCED IN ASSOCIATION WITH TOP QUARKS IN MULTILEPTON FINAL STATES AT ATLAS

Chris Lester

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of The University of Pennsylvania
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy
2014

I. Joseph Kroll, Professor, Physics
Supervisor of Dissertation

Elliot Lipeles, Associate Professor, Physics
Graduate Group Chairperson

Dissertation Committee

Justin Khoury, Associate Professor, Physics

I. Joseph Kroll, Professor, Physics

Burt Ovrut, Professor, Physics

Evelyn Thompson, Associate Professor, Physics

I. Joseph Kroll, Professor, Physics

A SEARCH FOR THE HIGGS BOSON PRODUCED IN ASSOCIATION WITH TOP QUARKS IN MULTILEPTON FINAL STATES AT ATLAS

COPYRIGHT
2014
Chris Lester

All rights reserved.

Acknowledgements

30 Acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements ac-
31 knowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowl-
32 edgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
33 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
34 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
35 acknowledgements acknowledgements.

36 Acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements ac-
37 knowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
38 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
39 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
40 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
41 acknowledgements acknowledgements.

42 Acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements ac-
43 knowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
44 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
45 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
46 acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements acknowledgements
47 acknowledgements acknowledgements.

48

ABSTRACT

49

A SEARCH FOR THE HIGGS BOSON PRODUCED IN ASSOCIATION WITH TOP
50 QUARKS IN MULTILEPTON FINAL STATES AT ATLAS

51

Chris Lester

52

Joseph Kroll

53

Abstract abstract abstract abstract abstract abstract abstract abstract abstract
54 abstract abstract abstract abstract abstract abstract abstract abstract abstract
55 abstract abstract abstract abstract abstract abstract abstract abstract abstract
56 abstract abstract abstract abstract abstract abstract abstract abstract abstract
57 abstract abstract abstract abstract abstract abstract abstract abstract abstract
58 abstract abstract abstract abstract abstract abstract abstract abstract abstract
59 abstract abstract abstract abstract abstract abstract abstract abstract abstract
60 abstract abstract abstract abstract abstract abstract abstract abstract abstract
61 abstract abstract abstract abstract abstract abstract abstract abstract abstract.

Contents

63	Acknowledgements	iii
64	Abstract	iv
65	Contents	v
66	List of Tables	ix
67	List of Figures	xi
68	1 Introduction	1
69	2 Theoretical Background	2
70	2.1 The Standard Model	2
71	2.1.1 The Standard Model Structure	2
72	2.1.2 Electroweak Symmetry Breaking and the Higgs	3
73	2.1.3 The Standard Model Parameters	5
74	2.2 Collider Physics and the Higgs	5
75	2.2.1 Higgs Discovery at the LHC	9
76	2.2.2 The Importance $t\bar{t}H$ Production	9
77	2.3 Conclusion	11
78	3 The Large Hadron Collider and the ATLAS Experiment	13
79	3.1 The Large Hadron Collider	13
80	3.1.1 The Accelerator Complex	13
81	3.1.2 Beam Parameters and Collisions	15

82	3.2 The ATLAS Experiment	15
83	3.2.1 Detector Coordinate System	17
84	3.2.2 The Inner Detector	17
85	3.2.3 The Calorimeter	19
86	3.2.4 The Muon Spectrometer	22
87	3.2.5 The Trigger System	22
88	3.2.6 Reconstruction: Jets, Muons and Electrons	24
89	3.2.6.1 Tracks and Clusters	24
90	3.2.6.2 Electrons	25
91	3.2.6.3 Muons	25
92	3.2.6.4 Jets	25
93	3.2.6.5 B-Tagged Jets	26
94	4 Electrons	28
95	4.1 Identification of Electrons at ATLAS	28
96	4.1.1 Pile-up and Electron identification	30
97	4.1.2 2011 Menu	30
98	4.1.3 2012 Menu and Pile-up	30
99	4.1.4 Electron Likelihood	31
100	4.2 Measurement of Electron Identification Efficiency at ATLAS	32
101	4.2.1 Issues	32
102	5 Analysis Summary	33
103	5.1 Signal Characteristics	33
104	5.2 Background Overview	34
105	5.3 Analysis Strategy	35
106	6 Dataset and Simulation	36
107	6.1 Data	36
108	6.1.1 The 2012 Dataset	36
109	6.2 Simulation	37
110	6.2.1 Signal Simulation	38
111	6.2.2 Background Simulation	38
112	7 Object and Event Selection	40

113	7.1	2 ℓ Same-Charge Signal Region	40
114	7.2	3 ℓ Signal Region	42
115	7.3	4 ℓ Signal Region	42
116	7.4	Electron Selection	43
117	7.5	Muon Selection	43
118	7.6	Jet and b-Tagged Jet Selection	44
119	7.7	Object Summary and Overlap	45
120	8	Background Estimation	46
121	8.1	Vector Boson (W^\pm, Z) production in association with top quarks: $t\bar{t}V, tZ$	46
122	8.1.1	$t\bar{t}Z$ Validation Region	48
123	8.2	Di-boson Background Estimation: $W^\pm Z, ZZ$	50
124	8.2.1	$W^\pm Z$ Normalization Uncertainty	51
125	8.2.2	ZZ Normalization Uncertainty	52
126	8.3	Charge-Misidentification Background	55
127	8.3.1	Likelihood Method	56
128	8.3.2	Results	57
129	8.3.3	Systematic and Statistical Uncertainties	58
130	8.4	Fake Lepton Backgrounds	60
131	8.4.1	2 ℓ SS Fakes	61
132	8.4.2	3 ℓ Fakes	63
133	8.4.3	4 ℓ Fakes	68
134	9	Summary of Systematic Uncertainties	70
135	9.1	Systematic Uncertainties on Signal Cross-section and Acceptance	70
136	9.2	Experimental and Detector Systematic Uncertainties	73
137	9.2.1	Lepton Identification, Energy Scale, and Trigger	73
138	9.2.2	Lepton Isolation and Impact Parameter	74
139	9.2.3	Jet Energy	75
140	9.2.4	B-Tagged Jet Efficiency	76
141	9.2.5	Summary	76
142	9.3	Summary of Background and Signal Normalization Uncertainties	77
143	10	Results and Statistical Model	79

144	10.1 Results in Signal Regions	79
145	10.1.1 2ℓ SS	79
146	10.1.2 3ℓ	79
147	10.1.3 4ℓ	79
148	10.2 Statistical Model	79
149	10.2.1 The Likelihood	79
150	10.2.2 Test Statistic and Profile Likelihood	80
151	10.2.3 CL_s Method	80
152	10.2.4 Exclusion Limits	81
153	10.2.5 μ Measurements	81
154	10.2.6 Nuisance Parameter Impact on the Signal Strength	82
155	11 Conclusions	83
156	11.1 Higgs Results in Review	83
157	11.2 Prospects for Future	83
158	Bibliography	84

List of Tables

160	5.1	Contributions of the main Higgs decay modes to the 3 multi-lepton $t\bar{t}H$ signatures at generation level.	34
162	6.1	Monte Carlo samples used for signal description.	38
163	6.2	Monte Carlo samples used for background description. Unless otherwise specified MadGraph samples use Pythia 6 for showering and Alpgen samples use Herwig+Jimmy. $t\bar{t}$, single top, and $Z+jets$ samples are replaced with data-driven estimates for the final result	39
166	7.1	Selections in the 2ℓ SS, 3ℓ and 4ℓ Signal Regions	41
167	7.2	Object identification and selection used to define the 5 channels of the multi-lepton $t\bar{t}H$ analysis. Some channels use a sub-sample of objects as explained in the Remarks column.	45
169	8.1	Expected number of signal and background events in 2ℓ SS, 3ℓ and 4ℓ signal regions. For data-driven backgrounds, Monte Carlo only numbers are given for reference. The expected sensitivity $\frac{s}{\sqrt{s+b}}$ is provided for data-driven, MC only and for the inclusion of ad-hoc systematic uncertainties (20% for $t\bar{V}$ and 30% for $t\bar{t}$ fake).	47
173	8.2	NLO cross section and theoretical uncertainty calculations derived from MadGraph5+aMC@NLO.	48
174	8.3	Truth origin of highest energy b-tagged jet in the $W^\pm Z + b$ VR and $3l$ SR	52
175	8.4	Truth origin of highest energy b-tagged jet in the $ZZ + b$ VR and $4l$ SR	55
176	8.5	Number of events of the main simulated background processes and of the data in the $e^\pm e^\pm$ and $\mu^\pm \mu^\pm$ channels used for the measurement of θ_e and θ_μ . VV , $V\gamma$, $t\bar{t}V$, tV and $t\bar{t}$ prompts (or charge misID) are the backgrounds which lead to prompt same-sign dileptons and are subtracted from the data to get a measured number of fakes. Uncertainties are statistical. The numbers labeled Data fakes are used to measure θ	62
181	8.6	Expected and measured values of the θ factors. NEEDS TO BE UPDATED FOR 2,3	63
182	8.7	Summary of the uncertainties (in %) in $e^\pm e^\pm$ (reverse Id + reverse isolation method), $\mu^\pm \mu^\pm$ and $e^\pm \mu^\pm$. Statistical uncertainty is split into statistical uncertainties on θ_e and θ_μ and uncertainty due to the Control Region size ($\Delta N_{\ell anti-\ell}(n jets)$). For channels with muons, uncertainties between parenthesis are obtained when anti-tight muons are defined such as $p_T < 20$ GeV (includes blinded data)	64
187	8.8	Summary of regions and inputs to the extraction of the number of $t\bar{t}$ events with a fake muon in the SR	65

189	9.1	Theoretical uncertainties of the signal event yields in the signal regions due to the impact of QCD scale uncertainties on the event selection.	71
190	9.2	Uncertainties on $t\bar{t}H$ acceptance in signal regions due to PDF variation.	72
191	9.3	Sum in quadrature of all the systematic uncertainties on the number of event yields per channel.	77
192	9.4	Summary of systematic uncertainties for processes present in the signal regions in the anal- ysis, with their type, description, name, values and uncertainties, and status of inclusion in the final results.	78
193	10.1	95%CL limits on μ for all channels and combination with cumulative uncertainties.	81

List of Figures

199	2.1	The Standard Model Particle Content	4
200	2.2	χ^2 as a function of the Higgs mass (top left), the top quark mass (top right), the W boson mass (bottom left) and the effective weak mixing angle (bottom right) for the combined SM fit from the GFitter group. The data points placed along $\chi^2 = 1$ represent direct measurements of the respective observable and their $\pm 1\sigma$ uncertainties. The grey (blue) bands show the results when excluding (including) the new M_H measurements from (in) the fits.	6
207	2.3	Proton Parton Distribution Functions (PDFs) from the MSTW Collaboration at $Q^2 = 10$ GeV 2 and $Q^2 = 10^4$ GeV 2	7
208	2.4	Dominant Higgs production modes at the LHC	8
209	2.5	8 TeV LHC Higgs production cross-sections (left) and decay branching fractions	8
211	2.6	ATLAS Higgs combination results for all SM measurement channels as ratios of the measured to SM production cross-sections (left) and extracted Higgs coupling constraint scale-factors for a combined fit to the measurement channels, where the W, Z , top-quark, b-quark, and τ couplings are allowed to float. The p-value of this particular model is 0.13 and in agreement with SM expectations	10
215	2.7	RGE for the running of the SM parameter, λ for the Higgs self-coupling term with present values and uncertainty bands for M_H and M_t (left). The two-dimensional plot colored (right) shows regions for which the SM is stable, unstable and metastable based on this RGE.	11
219	3.1	Diagram of the Large Hadron collider and location of the 4 main experiments (ATLAS, CMS, LHCb, and ALICE) around the ring. The diagram also shows the location of the SPS, the final booster ring in the accelerator complex that accelerates the protons to 450 GeV before injection into the LHC.	14
223	3.2	Diagram of the ATLAS detector and subsystems	16
224	3.3	Diagram of the ATLAS ID in the $R - \phi$ plane showing the barrel view of the Pixel Detector, SCT, and TRT.	18
226	3.4	Diagram of the ATLAS ID in the $R - z$ plane showing the endcap view of the Pixel Detector, SCT, and TRT. Only one side of the endcap is shown.	18
228	3.5	Diagram of the ATLAS calorimeters	20
229	3.6	Diagram of the ATLAS LAr EM calorimeter showing the longitudinal segmentation and the $\eta - \phi$ cells for the central barrel region	21

231	3.7	Diagram of the ATLAS muon system	23
232	3.8	$R - \phi$ schematic of the ATLAS detector and various particle signatures	27
233	5.1	Example Feynman diagrams for the 3 $t\bar{t}H$ multi-lepton categories.	34
234	6.1	Plot showing the accumulation of the integrated luminosity delivered to the ATLAS experiment over 2011 and 2012. The rough size of the usable, physics ready dataset for 2012 is 20 fb^{-1} and is the dataset used.	37
235	6.2	The average number of interactions per bunch-crossing for the 2012 and 2011 LHC proton-proton dataset. Most of these interactions are uninteresting but leave energetic signatures in particle detectors called pile-up which interfere with measurements	37
240	7.1	Number of jets vs. number of b-tagged jet plot for the fully selected multi-lepton channels. Signal regions are outlined with a dashed line. Sub-channels are defined later in the 2ℓ SS and 4ℓ SRs. The fractional background contribution to each jet and b-tagged jet bin are shown for non-prompt (red), $t\bar{t}V + tZ$ (yellow), and VV (green). The expected signal fraction is shown in white. The expected non-prompt fraction contains charge misidentifications and fakes. It is shown for MC only, although data-based methods are used for the final result.	41
241			
242			
243			
244			
245			
246			
247	8.1	Data/MC comparison plots for $t\bar{t}Z$ control region A (≥ 4 jets, ≥ 1 b-tag and 3 jets, ≥ 2 b-tag). In all plots, the rightmost bin contains any overflows. Top left: number of electrons. Top right: 10^* the number of b-tags + the total number of jets. Middle left: the invariant mass of the (0,1) lepton pair (see the text for the definition of the lepton ordering). Middle right: the invariant mass of the (0,2) lepton pair.	49
248			
249			
250			
251			
252	8.2	NJet spectrum for 2 tight-isolation leptons with 1 b-tagged jet (MV1.70)	51
253	8.3	3 lepton $W^\pm Z$ validation using the $t\bar{t}H$ lepton identification and momentum cuts: mass, number of jet and flavor variables	52
254	8.4	$W^\pm Z + b$ validation region: NJet, NElec, and Mass Variables	53
255	8.5	Jet-inclusive 4-lepton ZZ validation region using the $t\bar{t}H$ lepton identification and momentum cuts	54
256	8.6	$ZZ + b$ validation region using the $t\bar{t}H$ lepton identification and momentum cuts	54
257	8.7	Electron charge misidentification rates measured in data with the likelihood method on Z events (black points, red squares and blue triangles) as a function of $ \eta $ and parametrized in p_T . The full 2012 dataset has been used to estimate the rates below 130 GeV. Above this value, the charge misidentification rates have been estimated by extrapolating the rates in the region where the $p_T \in [90, 130]$ GeV with a p_T dependent factor extracted from simulated $t\bar{t}$ events (green triangles). Statistical and systematic uncertainties have been included in this plot.	58
258			
259			
260			
261			
262			
263			
264			
265			
266	8.8	Closure test on simulated $Z \rightarrow e^+e^-$ events (a) and data (b). The black circles show the distribution of same-sign events while the blue histograms show the distribution of the re-weighted opposite-sign events together with the statistical and systematic uncertainties. The distributions are not expected to overlay exactly, due to the loss of energy of the trident electrons for the same-sign peak.	58
267			
268			
269			
270			
271	8.9	Relative systematic uncertainty contributions on the charge misidentification rate, for different bins in p_T and $ \eta $	59
272			

273	8.10 Ratios of regions with tight and anti-tight leptons in 2-lepton and 3-lepton channels from $t\bar{t}$ MC. These ratios are the MC calculated transfer factors for each region, i.e. $\theta_e = eee/eed$, 274 ee/ed and $\theta_\mu = mmp/mmm$, mm/mp , where 'd' refers to anti-tight electrons and 'p' refers 275 to anti-tight muons. The transfer factors are seen to be similar in the 2ℓ and 3ℓ cases and 276 stable as a function of the number of jets	61
278	8.11 4,5 Jet SS 2ℓ ed (above) and μp control regions with at least one b-tagged jet. After 279 subtraction of prompt and charge misidentification backgrounds, these regions are scaled 280 by the transfer factors, θ_e and θ_μ , to obtain the final number of fake events in the CR. 281 The top MC (red) is used for reference but not in the actual calculation.	63
282	8.12 Distributions of the properties of the anti-tight muons in data (dots), compared with 283 the total simulation (red line), rescaled to the integral of the data for a shape comparison. 284 The uncertainty on the data distribution is statistical. The number of events for 285 each of them is also presented in the legend. The variables probed are, top: p_T and 286 $\Delta R(\mu, \text{closest selected jet})$; bottom: $\text{ptcone20}/p_T$ and $\text{Etcone20}/p_T$. The selection is the 287 signal region event selection with one anti-tight muon (failing at least one of the isolation, 288 muon-jet overlap, or p_T selection criteria). A ratio plot is containing the 20% area around 289 1, is displayed, demonstrating that a 20% data-MC comparison systematic is sufficient.	66
290	8.13 Distributions of anti-tight electron variables. The variables presented are, from top left 291 to bottom right, p_T , η , VERYTIGHT Likelihood value, $\text{ptcone20}/p_T$, $\text{Etcone20}/p_T$. The 292 plotted regions have the same cuts as the signal region, except the anti-tight electron must 293 fail isolation for the plot of the VERYTIGHT identification word or fail the VERYTIGHT 294 identification word for the plots of the isolation. Data (dots) are compared with a stacked 295 histogram of the various simulated samples: top in red, V+jets (blue), VV (purple) and 296 $t\bar{t}V$ (yellow). The uncertainty on the data distribution is statistical.	67
297	8.14 Muon-xxp (above) and electron-xxd (below) anti-tight control regions: jet variables. Note: 298 the $t\bar{t}$ and single top MC in the plots is used only as comparison, but is not included in 299 the fake measurement	68
300	8.15 3ℓ fake validation regions for nominal p_T selection (above) and relaxed p_T selection, > 10 301 GeV, (below). Plotted are the number of jets and the number of electrons in each event. 302 The data and MC ratio below each plot agree with 1 within the statistics of the region 303 and the overall systematic assigned for the fake component (red)	69
304	9.1 PDF systematic uncertainty on the jet multiplicities in $t\bar{t}H$ events with at least 2 leptons. 305 The dashed red lines in the lower panel indicate the systematic uncertainty on the $t\bar{t}H$ 306 production cross section.	72
307	9.2 Muon (left) and electron(right) identification efficiencies in Data and MC as a function of 308 η and p_T respectively. For electrons, the verytight likelihood operating point is used and 309 for muons the CB+ST (combined+segment tagged) operating point is used	74
310	9.3 Muon (left) and electron(right) isolation efficiency scale-factors from the Z control sample 311 as a function of the number of jets in the event. An additional systematic uncertainty of 312 1% is added to encompass the variation in the number of jets variable	75
313	9.4 JES systematic uncertainties as a function of jet η (for jets $p_T > 25$ GeV) and p_T (for jets 314 $ \eta < 0.4$). The combined systematic uncertainty is shown with contributions from the 315 largest sources	76
316	9.5 b-Tagging data-MC efficiency scale-factors versus jet p_T calculated in the $t\bar{t}$ sample from 317 2012 data. The uncertainties are combined statistical and systematic.	76

318

CHAPTER 1

319

Introduction

CHAPTER 2

Theoretical Background

322 The Standard Model of particle physics (SM) is an extraordinarily successful description of the funda-
 323 mental constituents of matter and their interactions. Many experiments have verified the extremely
 324 precise prediction of the SM. This success has culminated most recently in the discovery of the Higgs
 325 Boson. This chapter provides a brief introduction to the structure of the SM and how scientists are
 326 able to test it using hadron colliders. It focuses primarily on the physics of the Higgs boson and its
 327 decay to top quarks. I stress the importance of a measurement of the rate at which Higgs Bosons are
 328 produced in association of top quarks, as a new, rigorous test of the SM. The experimental search for
 329 this production mode in multi-lepton final states is the general subject of this thesis.

330 **2.1 The Standard Model**

331 **2.1.1 The Standard Model Structure**

332 The Standard Model (SM) [1, 2, 3, 4] is an example of a quantum field theory that describes the
 333 interactions of all of the known fundamental particles. Particles are understood to be excitations of
 334 the more fundamental object of the theory, the field. The dynamics and interactions of the fields are
 335 derived from the Standard Model Lagrangian, which is constructed to be symmetric under transfor-
 336 mations of the group $SU(3) \times SU(2)_L \times U(1)$. $SU(3)$ is the group for the color, $SU(2)_L$ is the group
 337 for weak iso spin, and $U(1)$ is the group for weak hyper-charge.

338 Demanding these symmetries be local, gauge symmetries allows the theory to be re-normalizable
 339 [5], meaning that unwanted infinities can be absorbed into observables from theory in a way that
 340 allows the theory to be able to predict physics at multiple energy scales. Gauging the symmetries

341 results in the introduction of 8 massless gluons, or the boson¹ carriers of the strong force [6] from the
 342 generators $SU(3)$ symmetry, and the 4 massless bosons, carriers for the weak and electromagnetic
 343 forces from the 3 generators of the $SU(2)$ and 1 generator of the $U(1)$ group. The weak and the
 344 electromagnetic forces are considered part of a larger single unified electroweak group $SU(2) \times U(1)$
 345 and the associated generators mix.

346 Matter particles are half-integer spin fermions and are representations of the symmetry groups.
 347 Singlets of the $SU(3)$ are called leptons, do not have a color charge, and, therefore, do not interact
 348 with the strong force. Quarks, as triplets of the $SU(3)$ group, do interact with the strong force.
 349 The SM is a chiral theory: the weak force violates parity, as it only couples to left-chiral particles
 350 or right-chiral antiparticles. This means that right-chiral and left-chiral fermions arise from different
 351 fields, which are different representations of the $SU(2)_L$ group.

352 The discovery of particles and new interactions in various experiments is intertwined with the
 353 development of the theory that spans many decades and is not discussed in detail here. But these
 354 experiments have proven the above model and symmetries to be an overwhelming success. So far, 3
 355 separate generations of both quarks and leptons have been discovered, differing only by mass. The
 356 gluons and the 4 electroweak bosons have also been discovered (W^+ , W^- , Z^0 , and γ). The reason for
 357 this 3-fold replication is not known. Figure 2.1 shows a table of the known SM particle content.

358 2.1.2 Electroweak Symmetry Breaking and the Higgs

359 Despite the simple structure of theory, the discovery of massive fundamental particles creates two
 360 sets of problems both related to $SU(2)_L \times U(1)$ symmetry. First, the force-carrying bosons must
 361 enter the theory without mass or the symmetries will be explicitly broken in the Lagrangian. Second,
 362 adding fermion masses to theory in an ad-hoc way allows the right-chiral and left-chiral fermions to
 363 mix. Since they possess different quantum numbers, as different representations of the weak-isospin
 364 group, this too breaks gauge invariance.

365 To solve these problems, spontaneous electro-weak symmetry breaking (EWSB) is introduced via
 366 the Brout-Englert-Higgs mechanism [7, 8, 9]. A massive scalar field in an electro-weak doublet is
 367 added to the theory with 4 new degrees of freedom and a potential which includes a quartic self-
 368 interaction term. Each fermion field interacts with the scalar field via a different Yukawa coupling,
 369 which unites the left and right chiral fields of a single particle type. This field explicitly preserves
 370 all of the symmetries, but the minimum of the potential does not occur when the expectation of the

¹bosons are full integer spin particles that obey Bose-Einstein statistics, while fermions are half-integer spin particles that obey Fermi-Dirac statistics

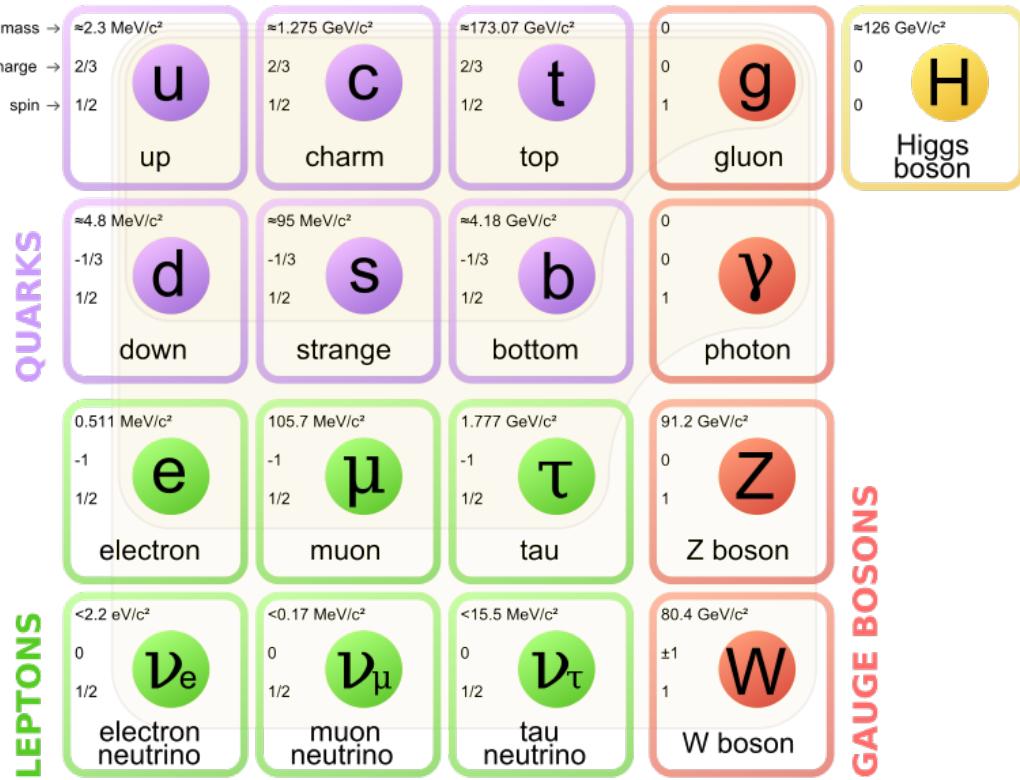


Figure 2.1: The Standard Model Particle Content

371 field is zero. The field eventually falls to a state, where it acquires a non-zero vacuum-expectation
 372 value. A non-vanishing field must point in a particular direction of weak-isospin space, breaking the
 373 symmetry.

374 The consequences of this spontaneous symmetry breaking are tremendous. The universe is filled
 375 with a field that has a non-zero expectation value. The theory can be expanded around this new value
 376 and 3 of the degrees of freedom can be interpreted as the longitudinal polarizations of the W^+ , W^- ,
 377 and Z^0 , while the 4th remains a scalar field, called the Higgs field with an associated particle called
 378 the Higgs particle or "Higgs". The weak bosons acquire a mass via their longitudinal polarizations
 379 and the Yukawa couplings of the scalar field to the fermions now behave like a mass term at this new
 380 minimum.

381 2.1.3 The Standard Model Parameters

382 Confronting the SM with experiment requires the measurement of 17^2 free parameters, which are
383 unconstrained from the theory. These free parameters include the fermion masses from the Yukawa
384 couplings, the force coupling constants, the angles and phase of the mixing between quarks, and
385 constants from the Higgs and electroweak sector³.

386 Experiments have provided a number of measurements of the parameters of the SM[10]. With
387 the discovery of the Higgs boson and the measurement of the Higgs mass, all of the parameters of
388 the SM can be estimated and statistical procedures can assess the relative agreement of overlapping
389 measurements to test the self-consistency of the SM. The GFitter collaboration assembles all relevant
390 electroweak observable measurements into a statistical model and then allows certain measurements
391 to float within their uncertainty to allow for a fit among multiple correlated measurements[11]. These
392 correlations arise for two reasons. First, measurements are made that often depend on multiple SM
393 parameters. Second, radiative corrections often cause parameters to depend on each other. For
394 instance, the Higgs mass is sensitive to both the W mass and top mass, through loop level corrections.

395 Figure 2.2 shows the fitted constraints on 4 key SM parameters (M_H , M_W , M_t , $\sin^2\theta_w$) with actual
396 measurements overlaid. The plots show both the removal and inclusion in the fit of key measurements
397 to assess their overall impact. The addition to the fit of the measured Higgs mass from the ATLAS and
398 CMS collaborations creates a small tension, as the other observables prefer the mass to be much lower
399 (~ 80 GeV). This tension in the combined electroweak fit as a result is not statistically significant
400 with a p -value of 0.07. The SM seems to be self-consistent.

401 2.2 Collider Physics and the Higgs

402 To test the theory, physicists accelerate particles to extremely high energies and force them to interact
403 through collisions. Typically, the particles accelerated are electrons or protons, since they are stable.
404 Electron-positron collider machines have a rich history of discovery and measurement in particle
405 physics. The advantage of electron accelerators is that the colliding element is itself a fundamental
406 particle. However, due to synchrotron radiation, the curvature of the beam line becomes problematic
407 for high energy beams. On the other hand, proton-proton and proton-anti-proton colliders can be
408 accelerated in rings without large losses due to synchrotron radiation, but the actual colliding objects

²There are additional parameters from neutrino mass terms and mixing but it is unclear how to include these into the Standard Model, since it does not predict right-chiral neutrinos

³ The electroweak sector includes parameters like mass of the W^\pm and Z^0 bosons, the weak mixing angle, $\sin^2\theta_w$, the fermi constant G_F , and Higgs Mass and vacuum expectation value. These parameters however are not wholly independent. As discussed above, it is only necessary theoretically to specify the two parameters relevant to the Higgs potential and the two coupling associated with the gauge groups

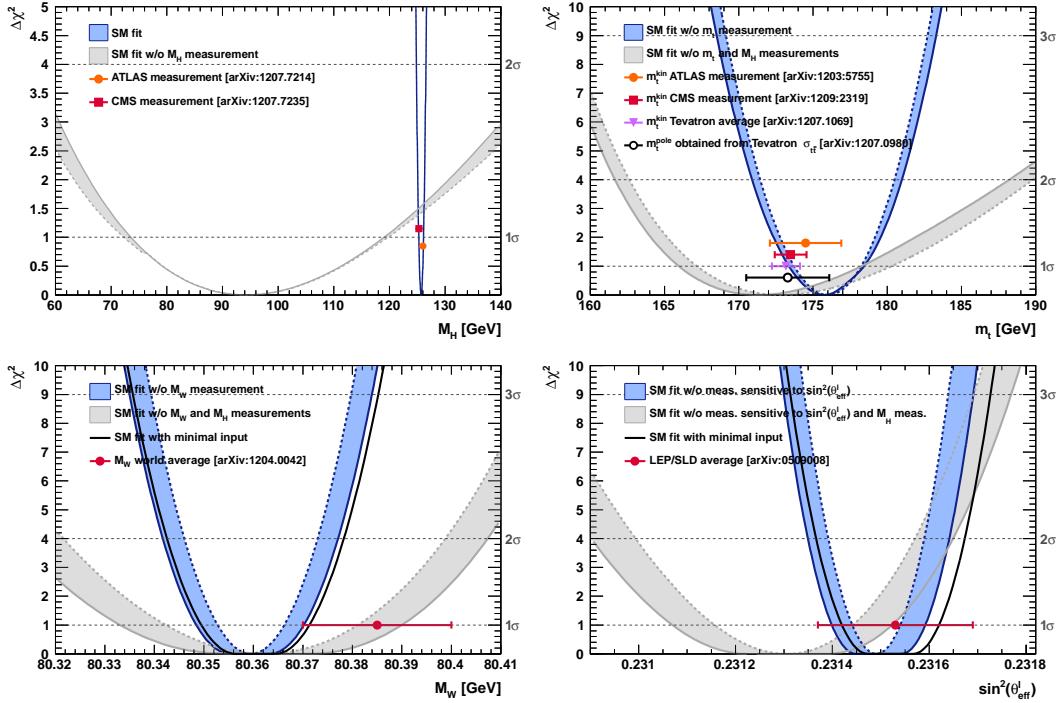


Figure 2.2: χ^2 as a function of the Higgs mass (top left), the top quark mass (top right), the W boson mass (bottom left) and the effective weak mixing angle (bottom right) for the combined SM fit from the GFitter group. The data points placed along $\chi^2 = 1$ represent direct measurements of the respective observable and their $\pm 1\sigma$ uncertainties. The grey (blue) bands show the results when excluding (including) the new M_H measurements from (in) the fits.

at high energies are the constituent quarks and gluons. This complicates analysis because the initial state of the hard-scattering system is not known on a per-collision basis and the momentum of hard-scattering system is unknown along the beam direction.

For hadron colliders, physicists must rely on form-factor descriptions of the colliding hadrons that describe the fraction of momentum carried by the hadrons constituent 'partons'. These are called parton-distribution functions, seen in Figure 2.3, and are factorized and integrated through the theoretical calculations of various collision processes [12].

At the Large Hadron Collider (LHC) protons are collided. The types of initial hard-scattering states at the LHC are quark-quark, quark-gluon, and gluon-gluon. Gluon collisions dominate overall, due to the large number of gluons inside the proton, though the relative importance of different initial states changes with the energy scale of the collision and the type of final state selected.

A prime motivation for the construction of the Large Hadron Collider was the discovery or exclusion of the Higgs boson[13]. LEP and the Tevatron excluded large swaths of possible Higgs boson masses,

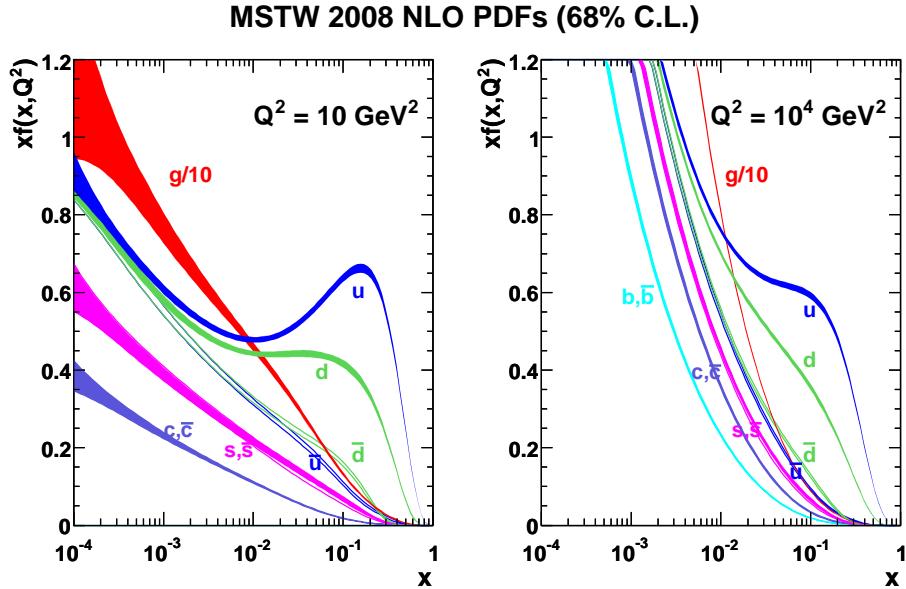


Figure 2.3: Proton Parton Distribution Functions (PDFs) from the MSTW Collaboration at $Q^2 = 10 \text{ GeV}^2$ and $Q^2 = 10^4 \text{ GeV}^2$

especially below 114 GeV. The Higgs mass was also known to have a theoretically motivated upper bound. The unitarity of diagrams including the $WWWW$ vertex required the Higgs mass to be below about 1 TeV. The LHC was designed to be able to eventually find or exclude a Higgs particle in this range [10].

Reaching this discovery or exclusion required an enormous dataset with collisions at high energies. Despite the fact that the Higgs couples to nearly every particle, Higgs boson production at the LHC is a low rate process. Because it couples to fermions proportional to mass and because the colliding particles must be stable and therefore light, production of the Higgs must occur through virtual states.

The Higgs boson can be produced through collision at the LHC via 4 mechanisms: gluon-fusion (ggF), vector-boson fusion (VBF), Higgsstrahlung (VH), and production in association with top quarks ($t\bar{t}H$). The diagrams are shown in Figure 2.4 and the production cross-sections as a function of Higgs mass for the 8 TeV LHC proton-proton running are shown in Figure 2.5 [14]. The largest production cross-section is via the gluon fusion channel at 20 pb, which proceeds through a fermion loop that is dominated by the top quark, because of its large Yukawa coupling to the Higgs. Because the Higgs couples to every massive particle, it has a rich set of decays also seen in Figure 2.5, especially for $m_H = 125 \text{ GeV}$. Studies of Higgs properties at hadron colliders offers many tests of the Standard Model and ample room for new physics searches. These tests specifically can verify

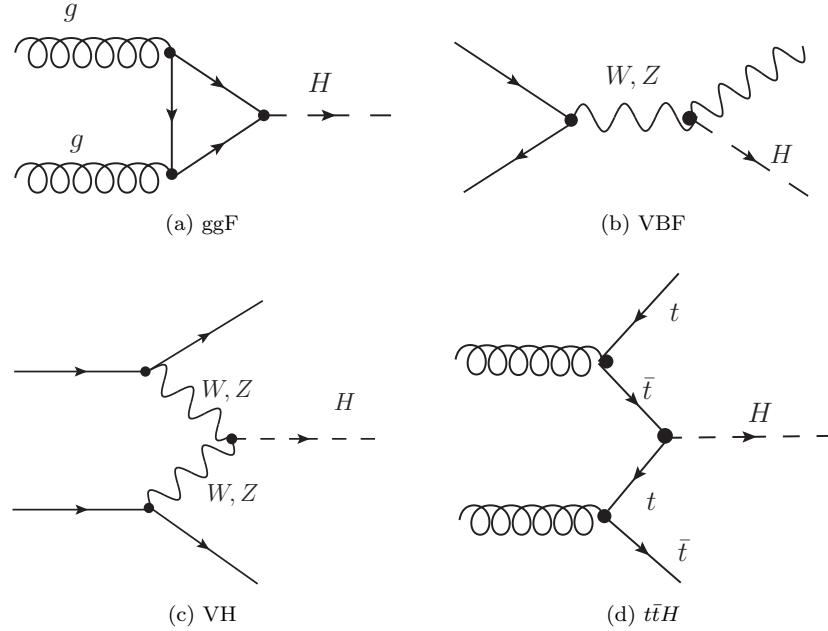


Figure 2.4: Dominant Higgs production modes at the LHC

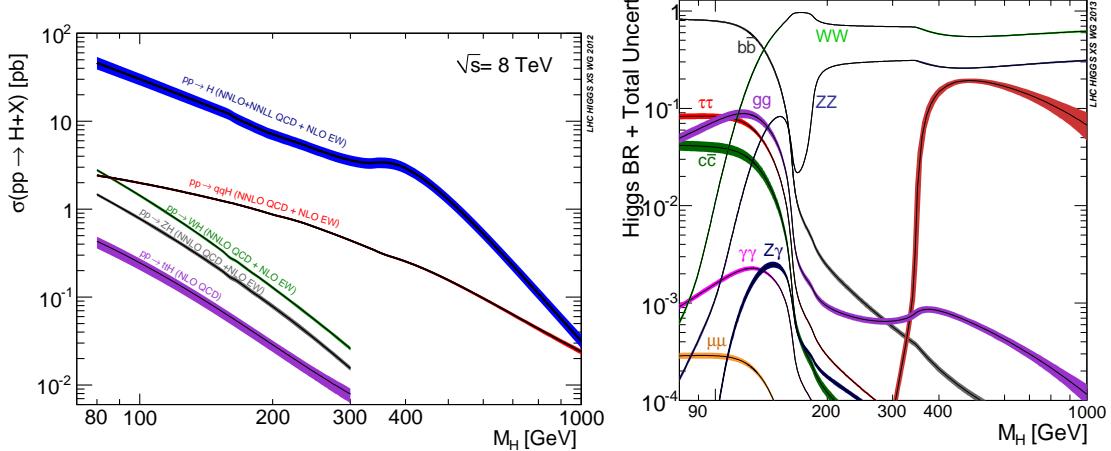


Figure 2.5: 8 TeV LHC Higgs production cross-sections (left) and decay branching fractions

439 the link between Yukawa coupling and the particles mass and further constrain details of EWSB by
 440 examining Higgs coupling to the weak bosons.

⁴⁴¹ **2.2.1 Higgs Discovery at the LHC**

⁴⁴² In 2012 both ATLAS and CMS announced the discovery of a new boson consistent with the Higgs
⁴⁴³ by examining the results of Higgs searches in a number of decay channels ($H \rightarrow W^+W^-$, $H \rightarrow Z^0Z^0$,
⁴⁴⁴ and $H \rightarrow \gamma\gamma$) in the 2011 dataset at $\sqrt{s} = 7$ TeV and part of the 2012 dataset at $\sqrt{s} = 8$ TeV. By
⁴⁴⁵ 2013 and 2014, both experiments have updated and/or finalized their results for the full 2011 and
⁴⁴⁶ 2012 datasets [15, 16]. I will focus on the ATLAS results, which measured both the Higgs mass[17]
⁴⁴⁷ and spin[18], as well as provided initial constraints of the Higgs couplings to different particles.

⁴⁴⁸ Figure 2.6 show the results of the searches in all of the measurement channels as well as constraints
⁴⁴⁹ on the SM Higgs coupling parameters in an example fit, where the couplings to the top-quark, bottom-
⁴⁵⁰ quark, W,Z, and τ are allowed to fluctuate independently. These rely on measurements binned in
⁴⁵¹ different production and decay channels. They are dominated by higher statistics results in the gluon-
⁴⁵² fusion production modes, but measurements in the VH and VBF modes are close to SM sensitivity.

⁴⁵³ The combined results show basic agreement with the SM with much room for improvement with
⁴⁵⁴ the addition of new production and decay modes and higher statistics. The coupling constraints are
⁴⁵⁵ particularly strong for the W and Z, which are the most sensitive decay channels, and top quark due
⁴⁵⁶ to the dominance of the top Yukawa in the ggF loop.

⁴⁵⁷ **2.2.2 The Importance $t\bar{t}H$ Production**

⁴⁵⁸ Notably absent thus far in the SM are searches for the Higgs in the $t\bar{t}H$ production channel, due to
⁴⁵⁹ the low production rate and lack of statistics. Searches are underway and initial results are close to
⁴⁶⁰ SM sensitivity for ATLAS and CMS.

⁴⁶¹ Measuring the $t\bar{t}H$ production rate is important, because $t\bar{t}H$ production depends on the top
⁴⁶² Yukawa coupling at tree level. Comparing the predicted Yukawa coupling from top mass measurements
⁴⁶³ to the coupling from the wholly independent Higgs production measurements is a very direct test of
⁴⁶⁴ the Higgs' involvement in providing mass for the fermions in the SM.

⁴⁶⁵ The top Yukawa coupling is already constrained from current measurements of the ggF production
⁴⁶⁶ process, since the ggF loop is dominated by top quarks. However, new, colored particles could be
⁴⁶⁷ present in the loop. Comparison of the gluon-fusion and the $t\bar{t}H$ modes would allow for disentangling
⁴⁶⁸ the effects of these possible new particles[19]. The simplest of new physics models, allowing for the
⁴⁶⁹ modification of the ggF loop, introduce a new generation of quarks. However, fourth generation
⁴⁷⁰ quarks, which obtain mass from a Higgs Yukawa coupling, are already largely excluded due to their
⁴⁷¹ enormous effects on the Higgs production cross-section[20]. Other exotic scenarios allow for new

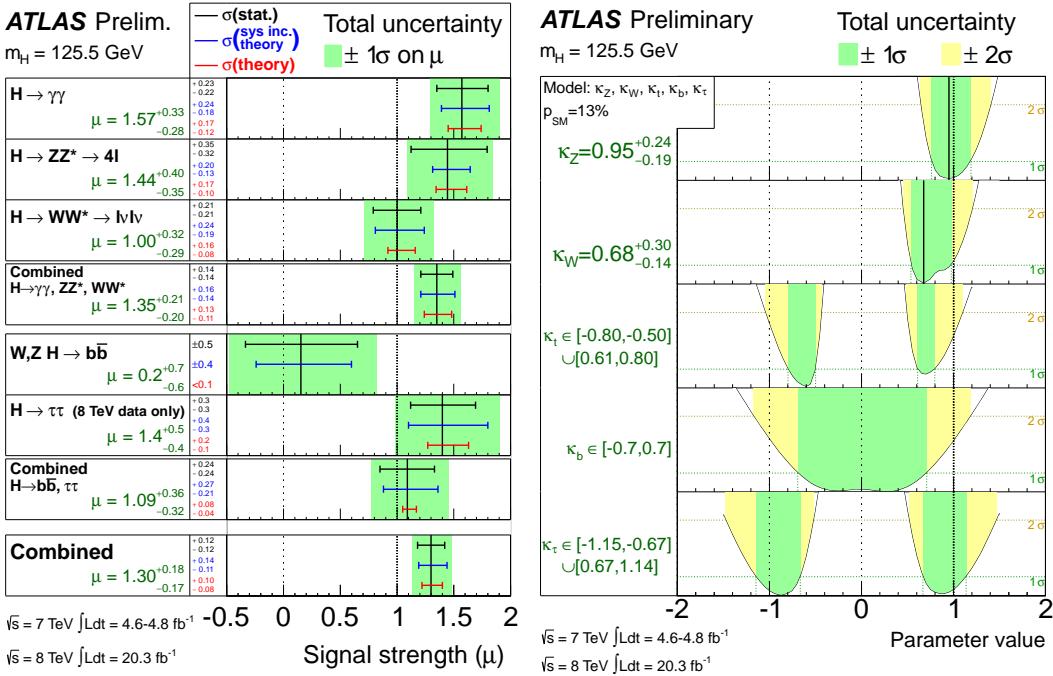


Figure 2.6: ATLAS Higgs combination results for all SM measurement channels as ratios of the measured to SM production cross-sections (left) and extracted Higgs coupling constraint scale-factors for a combined fit to the measurement channels, where the W, Z , top-quark, b -quark, and τ couplings are allowed to float. The p-value of this particular model is 0.13 and in agreement with SM expectations

472 colored particles, which are not entirely constrained by present measurements[21, 22, 23]. These
473 include, for instance, supersymmetric models involving the stop quark.

474 With the level of statistics available in Run I dataset, very strict constraints on the top Yukawa
475 coupling are simply not possible and the measurement presented in this thesis is a first step. Future,
476 high-statistics datasets will have the ability to provide better measurements and $t\bar{t}H$ production will
477 become very important. Despite similar uncertainties on the overall production cross-sections for $t\bar{t}H$
478 and the ggF, $t\bar{t}H$ has the advantage that most of these uncertainties would cancel for $t\bar{t}H$ if normalized
479 to the topologically similar $t\bar{t}Z$. Finally, the uniqueness of the experimental signature means that
480 searches for $t\bar{t}$ signatures can be performed for a variety of Higgs decays ($\gamma\gamma, b\bar{b}, WW, ZZ$, and $\tau\bar{\tau}$)
481 with roughly similar degrees of sensitivity (within a factor of 10)[19].

482 It is important to note the importance of the top Yukawa coupling due to its enormous size
483 compared to other couplings. For instance, the top Yukawa is 350000x as large as the electron
484 Yukawa coupling. The top Yukawa coupling, along with the Higgs mass, is one of the most important
485 pieces of the renormalization group equations (RGE) responsible for the running of the parameter that
486 determines the Higgs self-coupling λ . If this parameter runs negative, then the potential responsible

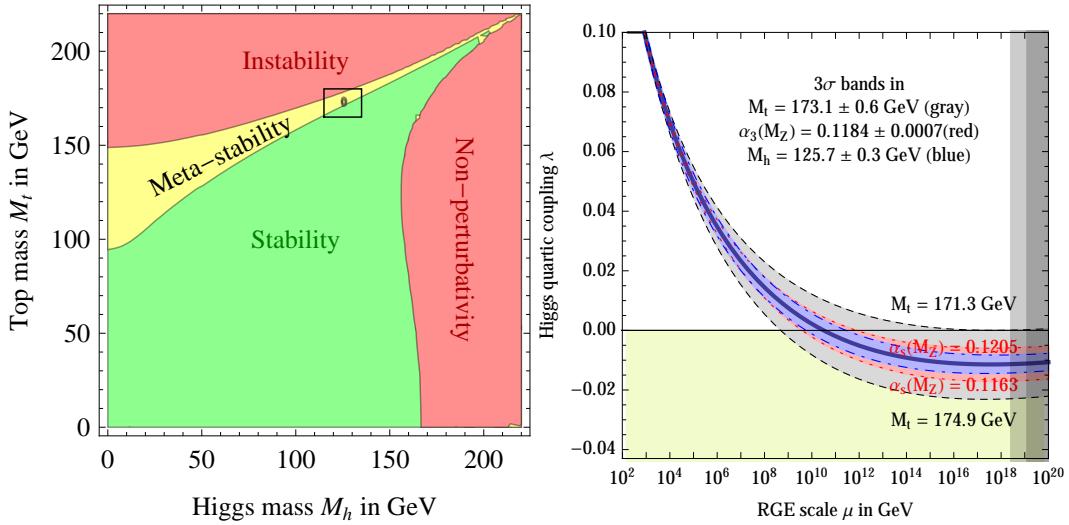


Figure 2.7: RGE for the running of the SM parameter, λ for the Higgs self-coupling term with present values and uncertainty bands for M_H and M_t (left). The two-dimensional plot colored (right) shows regions for which the SM is stable, unstable and metastable based on this RGE.

for the entire mechanism of EWSB no longer has a minimum and becomes unbounded, resulting in instability in the universe [24]. Metastability occurs when the shape of the potential allows for a false local minimum. Figure 2.7 shows the running of this parameter, the regions for which the universe is stable, unstable and metastable. Current measurements suggest that universe lies in a metastable island⁴. This is a sort of fanciful aside, intended only to highlight the importance of the top Yukawa coupling and to suggest that new discoveries in the top-Higgs sector have far reaching consequences.

2.3 Conclusion

The Standard Model, despite its success in providing a unified description of fundamental particles and interactions into single theory, has its flaws. These have been discussed in depth elsewhere, but include issues like the description of massive neutrinos, the failure to include gravity, and the unnaturalness of large quantum corrections to Higgs parameters. For these reasons, it seems the SM might be a lower energy approximation to a more fundamental theory. The discovery of the Higgs boson at the LHC provided a stunning verification of one of the fundamental aspects of the theory but at the same time offers new area to search for glimpses of something more fundamental. The production of samples of Higgs bosons allows for a rich array of new tests of the Standard Model, which is now finally over-constrained by experiment. Searches for the $t\bar{t}H$ production, one category of

⁴The RGE assumed that there is no new physics at all energy scales

503 which is the topic of this thesis, provide tree-level access to a central parameter of the theory, the top
504 Yukawa coupling, as well as access a variety of Higgs decays, which will eventually provide a rigorous
505 new test of the SM.

CHAPTER 3

The Large Hadron Collider and the ATLAS Experiment

509

3.1 The Large Hadron Collider

510 Production of a sufficient number of high energy collisions to adequately explore particle physics at
 511 the electro-weak scale required the development of one of the most complex machines ever built, the
 512 Large Hadron Collider or LHC.

513 The LHC is the world's highest energy particle accelerator and is located 100m underneath the
 514 Franco-Swiss border at the European Organization for Nuclear Research (CERN) in a 26.7 km tunnel.
 515 The technology involved in the development of the LHC is an enormous achievement in its own right
 516 and is documented in detail here [25, 26, 27]. The LHC is a circular machine capable of accelerating
 517 beams of protons and colliding them at center of mass energies up to $\sqrt{s} = 14$ TeV at 4 collision sites
 518 around the ring, where 4 experiments are housed (ATLAS[28], CMS[29], LHCb[30], and ALICE[31]).
 519 Figure 3.1 is a diagram of the layout of the LHC and its experiments[32]. The LHC also operates in
 520 a mode with beams of heavy ions. The LHC is composed of thousands of super-conducting Niobium-
 521 Titanium magnets, cooled to 1.9° K with liquid Helium, which steer and focus the particle beams,
 522 and a superconducting resonant-frequency (RF) cavity, which boosts the beam to higher energies.

523

3.1.1 The Accelerator Complex

524 The accelerator complex is a progressive series of machines with the LHC as the final stage. Protons
 525 are obtained from hydrogen atoms and are accelerated to 50 MeV using the Linac2, a linear acceler-
 526 ator, before being injected into the Proton-Synchrotron Booster (PSB). In the PSB the protons are
 527 accelerated to energies of 1.4 GeV for injection into the Proton-Synchrotron (PS). The PS accelerates

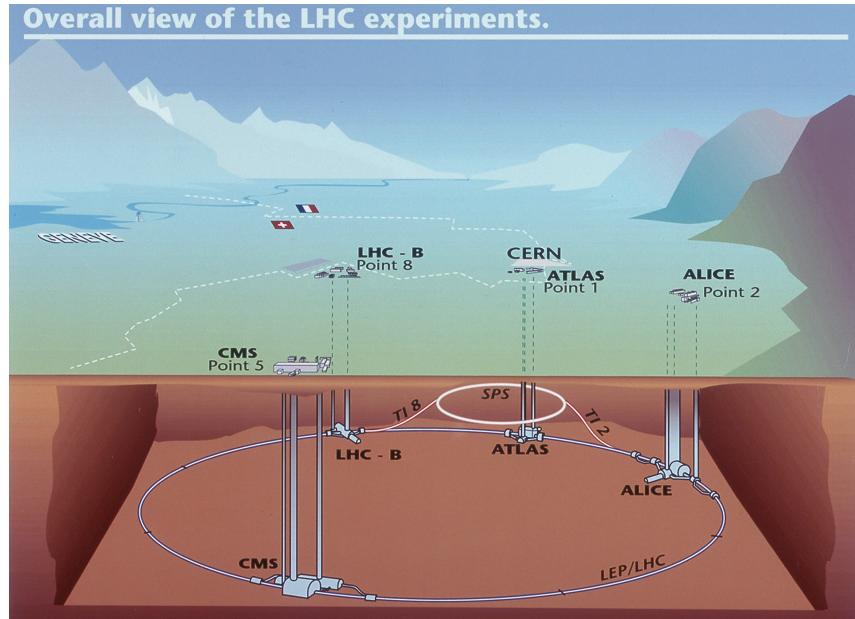


Figure 3.1: Diagram of the Large Hadron collider and location of the 4 main experiments (ATLAS, CMS, LHCb, and ALICE) around the ring. The diagram also shows the location of the SPS, the final booster ring in the accelerator complex that accelerates the protons to 450 GeV before injection into the LHC.

528 the protons to 25 GeV and dumps bunches into the Super Proton Synchrotron (SPS), where they are
 529 accelerated to 450 GeV and finally dumped into the LHC for full acceleration. The PS and SPS are
 530 circular accelerators that were important in past physics discoveries and have been re-purposed for
 531 use in the LHC complex.

532 **3.1.2 Beam Parameters and Collisions**

533 For the physics studied at the ATLAS experiment, the two most important parameters of the collisions
534 are the center of mass energy (CME) and instantaneous luminosity (\mathcal{L}). High center of mass energies
535 are necessary for the production of new high mass particles, and, because the constituents of the
536 actual collisions are the partons of the proton, the CME of the collisions must in general be much
537 higher than the mass of the particles produced.

538 The instantaneous luminosity of the collisions is a measure of the collision rate. The integrated
539 luminosity over time is a measure of the size of the dataset and when multiplied by the cross-section of a
540 particular process gives the total number of expected events produced for that process. Instantaneous
541 luminosity depends on the number of colliding bunches of protons, the intensity of those bunches, the
542 revolution frequency, and the normalized transverse spread of the beam in momentum and position
543 phase space, called the emittance, and the transverse beam size. The LHC has the option for colliding
544 beams with 2808 bunches of protons, each with around 10^{11} protons, at a rate of one bunch collision
545 every 25 ns, or 40 MHz. These parameters correspond to a design luminosity of around $10^{34} \text{ cm}^2 \text{ s}^{-1}$
546 or $10 \text{ nb}^{-1} \text{ s}^{-1}$, equivalent to 1 Higgs every 5 seconds.

547 **3.2 The ATLAS Experiment**

548 This section provides a brief overview of the ATLAS experiment. The ATLAS detector is centered on
549 one of the LHC collisions points, located 100m underground. Through the combination of a number
550 of subsystems, it designed to identify the particles arising from these collisions, measure the energy
551 and momentum of these particles, and make fast decisions about the content of each collision, in order
552 to save a small fraction of measured collision events for offline study.

553 ATLAS, shown in Figure 3.2, possesses cylindrical symmetry around the beam pipe. It weights
554 7000 tons, has a diameter of roughly 25m and length of 46m. ATLAS was designed to be a multi-
555 purpose hermetic, particle detector, able to identify many types of particles, and designed to provide
556 a snapshot of the entire collision event. The detector sub-systems form concentric rings around the
557 beam-line at increasing distance. From closest to the beam outward, they are:

- 558 • **Inner Detector:** The inner detector (ID)[33, 34] is immersed in a solenoidal magnetic field[35]
559 and provides measurements of charge particle tracks, through three subsystems: the Pixel
560 Detector[36, 37], the Semi-Conductor Tracker (SCT)[38, 39], and Transition Radiation Tracker(TRT)
561 [40, 41, 42].

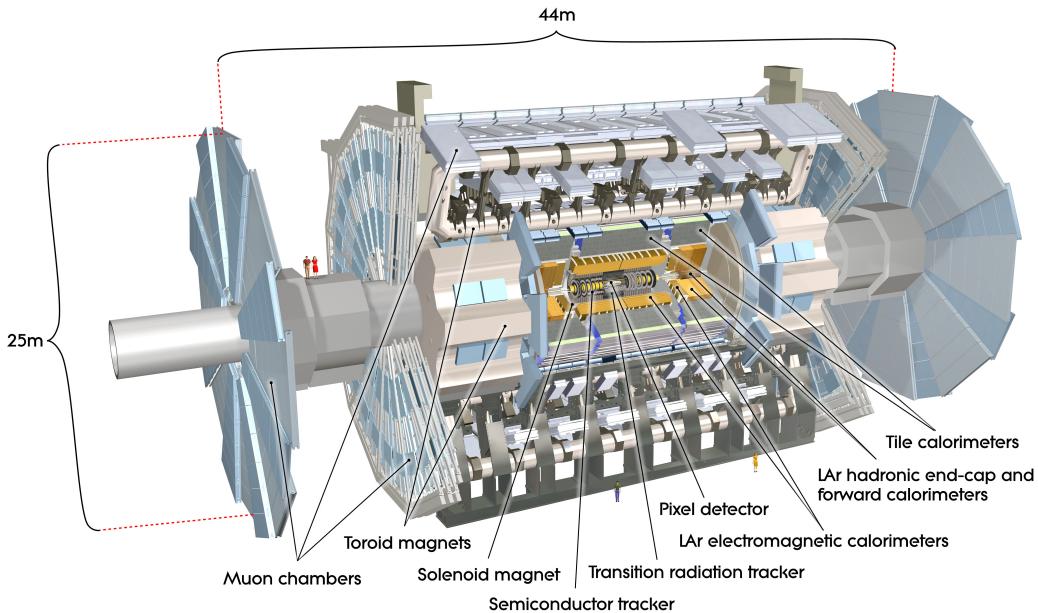


Figure 3.2: Diagram of the ATLAS detector and subsystems

- **Calorimeter:** The calorimeters measure the energy of particles that participate in the electromagnetic (photons, electrons) and hadronic interactions (pions, protons, neutrons, etc.), by forcing them to shower in dense material. The hermeticity of the calorimeters allows for missing transverse energy measurements. The calorimeter is composed of the liquid argon electromagnetic calorimeter (LAr)[43], the hadronic tile calorimeter[44], the liquid argon hadronic endcap calorimeter, and the forward calorimeters.
- **Muon Spectrometer:** The muon spectrometer (MS) sub-systems[45] form the outermost detector systems and measure the momentum of minimum ionizing muon tracks, as all other particles are stopped by the calorimeters. The muon systems are immersed in a toroidal magnetic field [35] and are composed of 4 different sub-systems for triggering and tracking measurements [46, 47, 48].
- **Triggering Systems:** The trigger and data acquisition systems[49, 50] read out data from the detector through a three-tiered hardware and software decision making framework to record the most interesting physical processes for a broad physics analysis program.

These systems are discussed in depth in the following sections.

577 **3.2.1 Detector Coordinate System**

578 ATLAS uses a right-handed coordinate system centered at the nominal proton interaction point. The
579 beam line defines the z -axis. The $x - y$ plane is perpendicular to the beam line and is referred to as the
580 transverse plane. The transverse plane holds special significance in reporting measurements, because
581 the initial momentum of the hard collision system is 0 along the transverse plane in the laboratory
582 rest frame. Particle momenta measured along the transverse plane is called transverse momenta, and
583 labeled p_T . The momentum of the colliding proton-proton system is also 0 along the z -axis but the
584 colliding partons may have vastly different momenta. Thus, momentum of the hard colliding system
585 along the z -axis differs collision to collision.

586 Because ATLAS possesses a rough cylindrical symmetry, cylindrical and polar coordinates are used
587 to describe particle trajectories and detector positions. The radial coordinate, R , describes transverse
588 distances from the beam line. An azimuthal angle, ϕ , describes angles around the z -axis, and a polar
589 coordinate θ describes angles away from the z -axis. The polar angle is often expressed in terms of
590 pseudo-rapidity, defined as $\eta = -\ln(\tan(\theta/2))$. Distances in $\eta - \phi$ space are often used to describe
591 the proximity of objects in the detector, $\Delta R = \sqrt{\eta^2 + \phi^2}$.

592 The ‘barrel’ and ‘endcap’ are classifications that are used to label the position of sub-detectors.
593 Barrel sub-detectors occupy positions more central to the detector at $|\eta|$ values roughly less than
594 1-2, while the endcap calorimeters extend farther in $|\eta|$. The barrel-endcap transition region contains
595 detector services. Also, the orientation of the detector elements are often different in the barrel and
596 endcap to have optimal particle flux.

597 **3.2.2 The Inner Detector**

598 The ID makes measurements of the position of charged particles as they move through the detectors
599 3 sub-systems (Pixel Detector, SCT, TRT). The individual position measurements can be strung
600 together to form a particle track. The entire ID is immersed in a 2T solenoidal magnetic field allowing
601 for measurements of particle momenta through the curvature of the tracks. The ID is contained with
602 a radius of 1.15 m and has a total length of 7m, allowing for particle tracking out to $|\eta| < 2.5$.
603 Figures 3.3 and 3.4 show the placement of the ID sub-systems in the $R - \phi$ and $R - z$ planes.

604 The Pixel Detector has 80 million silicon read out channels (pixels) and is closest to the interaction
605 point with the finest granularity. As charged particles traverse the silicon, they create electron-hole
606 pairs, which subsequently drift in an electric field and can be captured and registered as a current
607 pulse. The detector has three concentric layers of pixels in the barrel (to $|\eta| < 1.9$) and three endcap
608 disks on each side of the barrel (to $|\eta| < 2.5$). The closest barrel layer to the beam pipe is called the

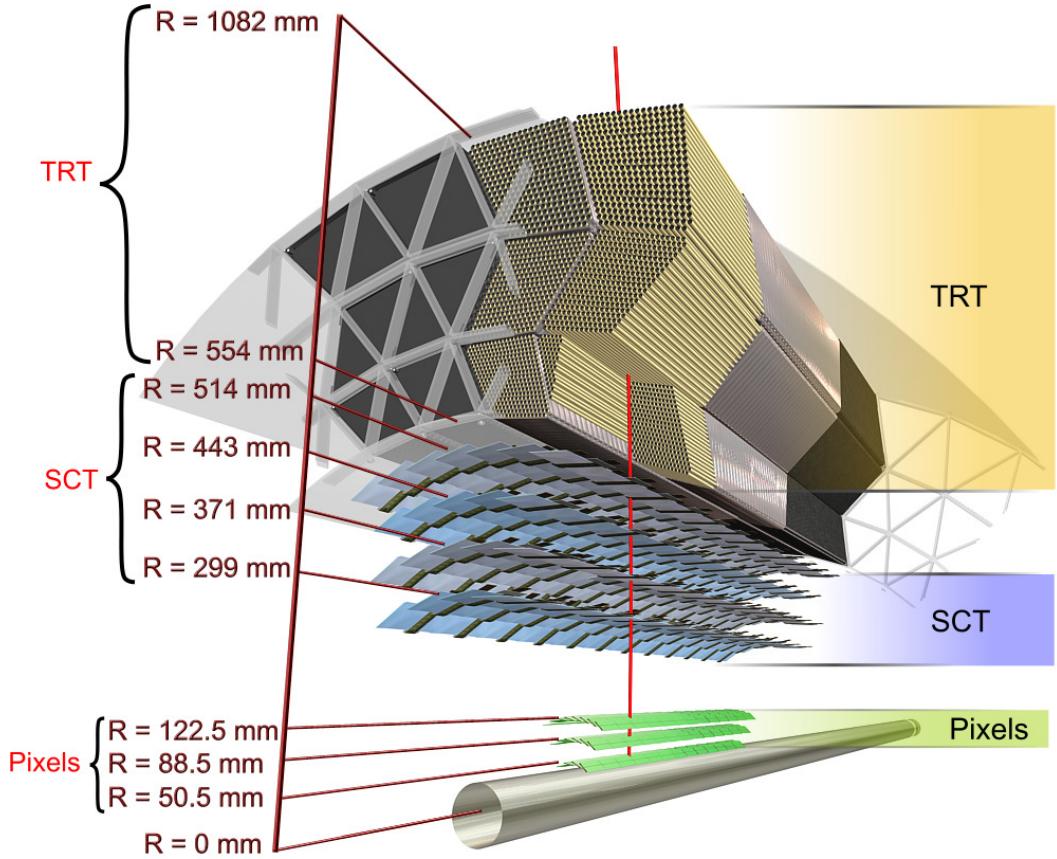


Figure 3.3: Diagram of the ATLAS ID in the $R - \phi$ plane showing the barrel view of the Pixel Detector, SCT, and TRT.

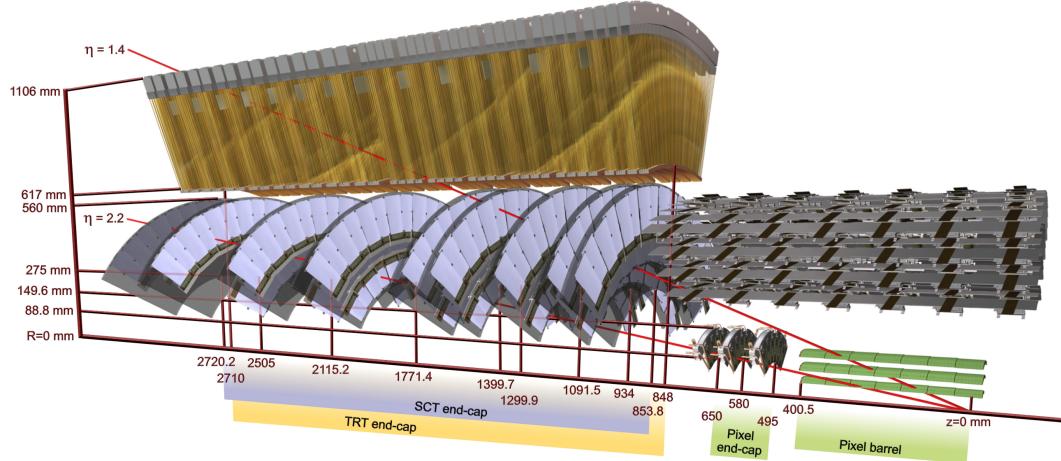


Figure 3.4: Diagram of the ATLAS ID in the $R - z$ plane showing the endcap view of the Pixel Detector, SCT, and TRT. Only one side of the endcap is shown.

609 b-layer. The pixels provide excellent hit resolution ($R - \phi$ accuracy of $10\ \mu\text{m}$ and $z(R)$ accuracy of
 610 $115\ \mu\text{m}$ in the barrel (endcap)).

611 The SCT uses similar silicon technology to the pixels, but the each SCT layer contains a double
 612 layer of silicon strips, which are much longer in length than width. The SCT has 4 million read out
 613 channels and is comprises 4 barrel layers and 9 endcap layers with coverage to $|\eta| < 2.5$. The double
 614 layers are inclined slightly with respect to each other so that these 1D sensors have 2D resolution for
 615 coincident hits. The resolutions are $580\ \mu\text{m}$ in $z(R)$ for the barrel(endcap) and $17\ \mu\text{m}$ in $R - \phi$.

616 The TRT is comprised of straw drift tubes filled with a gas mixture, containing mostly Xenon
 617 gas. Particles traversing the straws ionize the gas, and the liberated electrons drift to a wire at the
 618 center of the straw, which has an applied voltage, and induce an signal on the wire. The TRT has
 619 $\sim 300,000$ straws. The barrel straws are arranged cylindrically along the z direction out to $\sim \eta < 1$
 620 and the endcap straws point radially outward in the R direction. For this reason, the barrel(endcap)
 621 straws provide no measurement in the $R(z)$ directions. The drift tubes provide individual position
 622 measurements with resolutions of $\sim 130\ \mu\text{m}$. Each particle track has on average 35 hits, which is
 623 large compared to the Pixel and SCT tracks, which have on average 7 hits.

624 The TRT is unique in that it also provides particle identification measurements via transition
 625 radiation. Charged particles emit transition radiation when traversing a boundary between materials
 626 of different dielectric constants. The volume between the straws is filled with a radiator material, a
 627 polymer foil or foam, to provide this boundary condition. Transition radiation photons are emitted
 628 in the direction of the particle trajectory in the keV range and cause a much larger signal amplitude
 629 within the straw. Hits that cause a signal at a higher threshold are thus indicative of transition
 630 radiation. The probability for emission transition radiation depends on the relativistic γ of the
 631 traversing particle. Because electrons are much lighter than any other charged particle, their γ -
 632 factors tend to be high enough to induce transition radiations, as opposed to pions, muons and other
 633 particles.

634 Combined tracking of particles through the 3 sub-detectors results in track momentum measure-
 635 ments from 500 MeV, the minimum energy need to leave the ID due to the magnetic field, to a few
 636 TeV. The track p_{T} resolution is roughly $0.05\% \cdot p_{\text{T}} \oplus 1\%$.

637 3.2.3 The Calorimeter

638 The ATLAS calorimeters measure the energy of electrons, photons and hadrons with $|\eta| < 4.5$. They
 639 induce a particle shower via electro-magnetic and nuclear interactions with the detector material and
 640 are deep enough to ensure that all or most of the shower energy remains contained. Exceptioanlly,

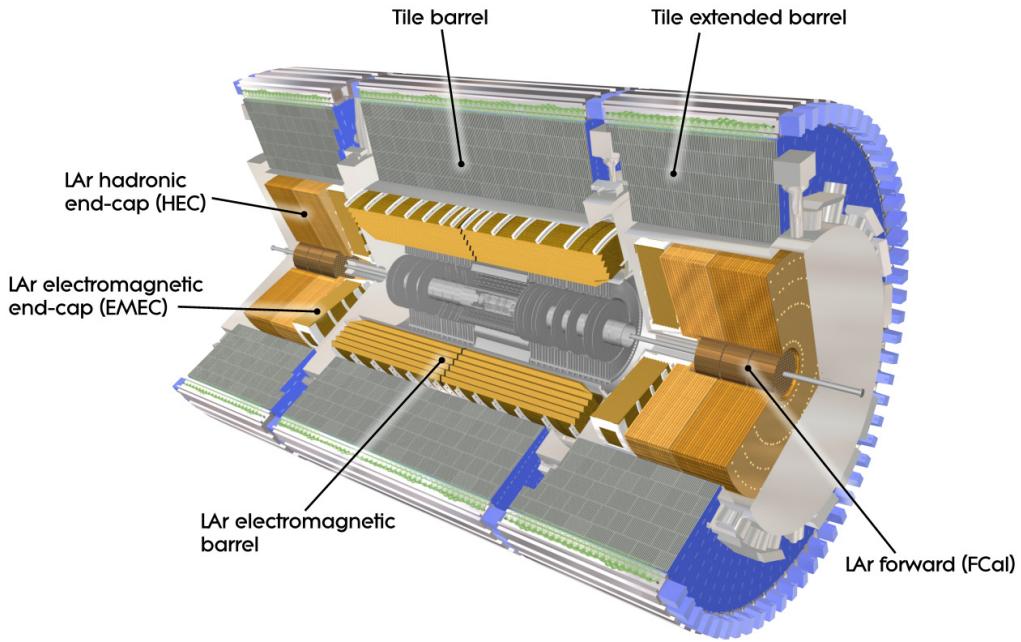


Figure 3.5: Diagram of the ATLAS calorimeters

641 muons pass through the ATLAS calorimeters leaving relatively little energy behind. ATLAS calorime-
 642 ters are sampling calorimeters meaning that the active material of the detector only measures a small
 643 fraction of the energy produced by the shower. The overall shower energy is inferred from this frac-
 644 tional measurement. The rest of the material is inactive, dense material, designed to induce showers.
 645 The calorimetry system is grossly divided longitudinally (radially) into electro-magnetic (EM) and
 646 then hadronic segments, operated with different technologies. Figure 3.5 diagrams the layout of the
 647 calorimeter system.

648 The EM calorimeter (LAr), which is located directly outside of the solenoid magnet but within
 649 the same cryostat, has an accordion design with lead absorber and liquid argon active material. The
 650 accordion design ensures uniform coverage in ϕ . The barrel and endcap LAr extend to $|\eta| < 2.47$. The
 651 LAr provides highly granular measurements in $\eta - \phi$ with 4 longitudinal segments, totaling $\sim 25\text{-}35$
 652 radiation lengths with the exception of the barrel/endcap transition region ($1.37 < |\eta| < 1.52$). The
 653 geometry of the barrel LAr calorimeter can be seen in Figure 3.6. The first longitudinal segment
 654 is called the pre-sampler, composed of a thin layer of active liquid argon, designed to detect early
 655 particle showers. The second segment is the most highly granular segment called the ‘strips’, as it
 656 is composed of thin liquid argon cells. The strips have a size of $0.025/8 \times 0.1$ in $\eta - \phi$ in the barrel

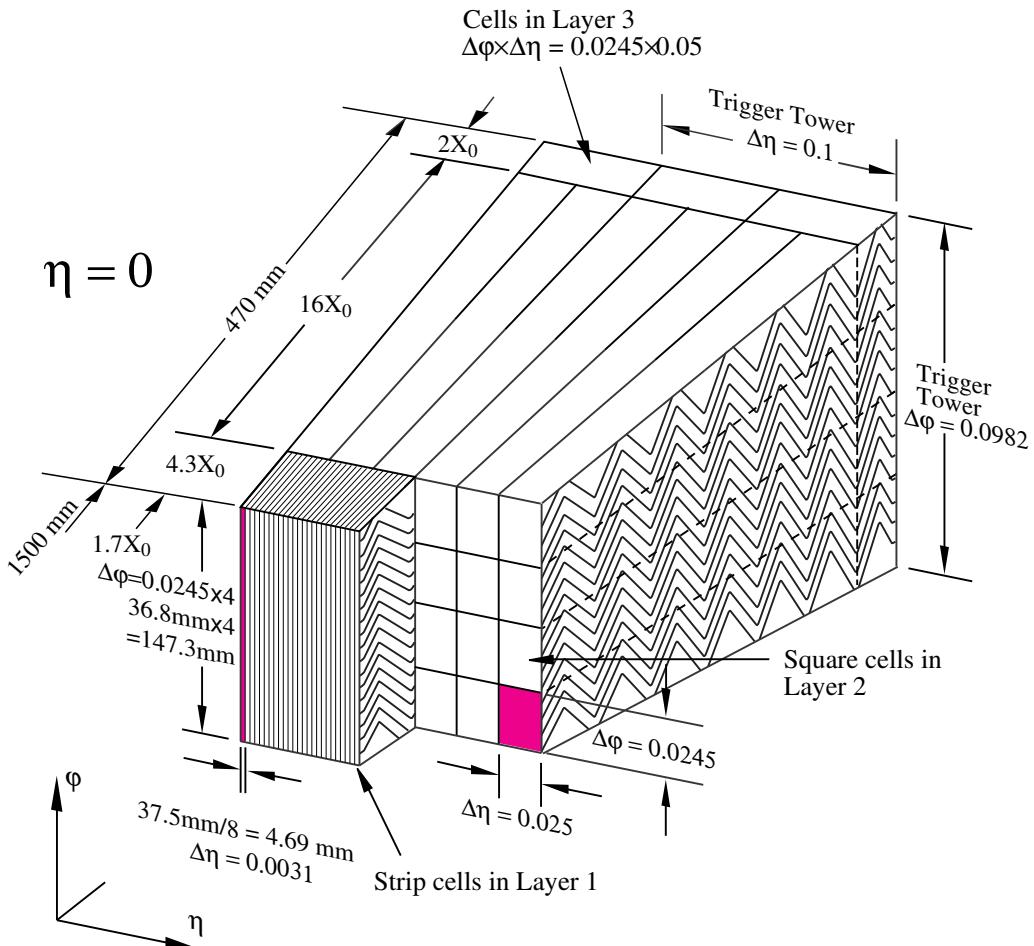


Figure 3.6: Diagram of the ATLAS LAr EM calorimeter showing the longitudinal segmentation and the $\eta - \phi$ cells for the central barrel region

with similar sizes in the endcap and are designed to be able to resolve single and double particle showers. This resolution is particularly useful in distinguishing $\pi^0 \rightarrow \gamma\gamma$ signatures from electron and photon signatures. The bulk of the radiation lengths and therefore the primary energy measurement come from the the third layer⁵. Each cell in this layer is 0.025×0.025 in $\eta - \phi$. The final layer is coarser, thinner and designed to estimate energy leaking out of the EM calorimeter. The forward EM calorimeters extend the η range and use the same technology, but are not used in this analysis. The energy resolution of the EM calorimeters is $\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$, measured in test beam data and confirmed in collision data.

The hadronic calorimeter is located directly behind the EM calorimeter. It is composed of tiles

⁵this layer is actually called 'layer 2', since the pre-sampler is referred to as 'layer 0'

666 of iron absorber and plastic scintillator in the barrel ($|\eta| < 1.6$), called the TileCal, and copper-
667 liquid argon in the endcap ($1.5 < |\eta| < 3.2$), called the HEC. The calorimeters contain $\sim 10\text{-}19$
668 hadronic interactions lengths with multiple longitudinal segments to contain showers induced by
669 the nuclear interaction of hadronic particles. The energy resolution of the hadronic calorimeters is
670 $\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$. The intrinsic resolution of hadron calorimeters is much worse than electro-
671 magnetic calorimeters, because much of the energy is lost to the inelasticity of nuclear break-up.

672 3.2.4 The Muon Spectrometer

673 The MS measures the trajectory of particles outside of the calorimeters, using multiple different
674 technologies. Generally, all charged particles except for muons are stopped by the calorimeter, and
675 therefore the majority of particles in the MS are muons, with the exception of rare cases of hadronic
676 punch-through. Particle momentum spectroscopy is made possible by an air-core toroidal magnet
677 system, embedded in the barrel MS ($|\eta| < 1.4$), and two smaller end cap toroids that provide fields
678 out to $|\eta| < 2.7$.

679 In the barrel region, the muon chambers are arranged in three cylindrical layers around the beam,
680 while in the endcap-regions the the layers are arranged perpendicular to the beam in wheels. The
681 arrangement is depicted in Figure 3.7.

682 The chambers in the barrel and most of the endcap are constructed from Monitored Drift Tubes
683 (MDTs) with an Argon gas mixture. Each chamber contains 3-7 drift tubes and provide hit resolutions
684 of $80\text{ }\mu\text{m}$ per tube and $35\text{ }\mu\text{m}$ per chamber in the bending plane. For $|\eta| > 2.0$, Cathode Strip
685 Chambers (CSCs) are used, primarily to handle the higher incident particle flux. They are composed
686 of cathode strips crossed with anode wires in the gas mixture, but use similar drift technology as the
687 MDTs and have resolutions in the bending plane $40\text{ }\mu\text{m}$ per chamber.

688 Separate chambers, called Thin Gap Chambers (TGCs), used in the endcaps, and Resistive Plate
689 Chambers (RPCs), used in the barrel, provide less precise hit information but within a much quicker
690 time window, and are therefore used for triggering, as the CSCs and MDTs are too slow.

691 3.2.5 The Trigger System

692 The ATLAS trigger system is designed to make quick decisions about individual particle collisions to
693 reduce the enormous collision rate of 20 MHz to a much more manageable 400 Hz to be stored for
694 offline analysis. Saving the full ATLAS data-stream would require space for 40 TB of raw data per
695 second, but, more importantly, most of these collisions result in the uninteresting inelastic break-up of
696 the colliding protons. To select out collisions to allow for a diverse physics program, ATLAS devotes a

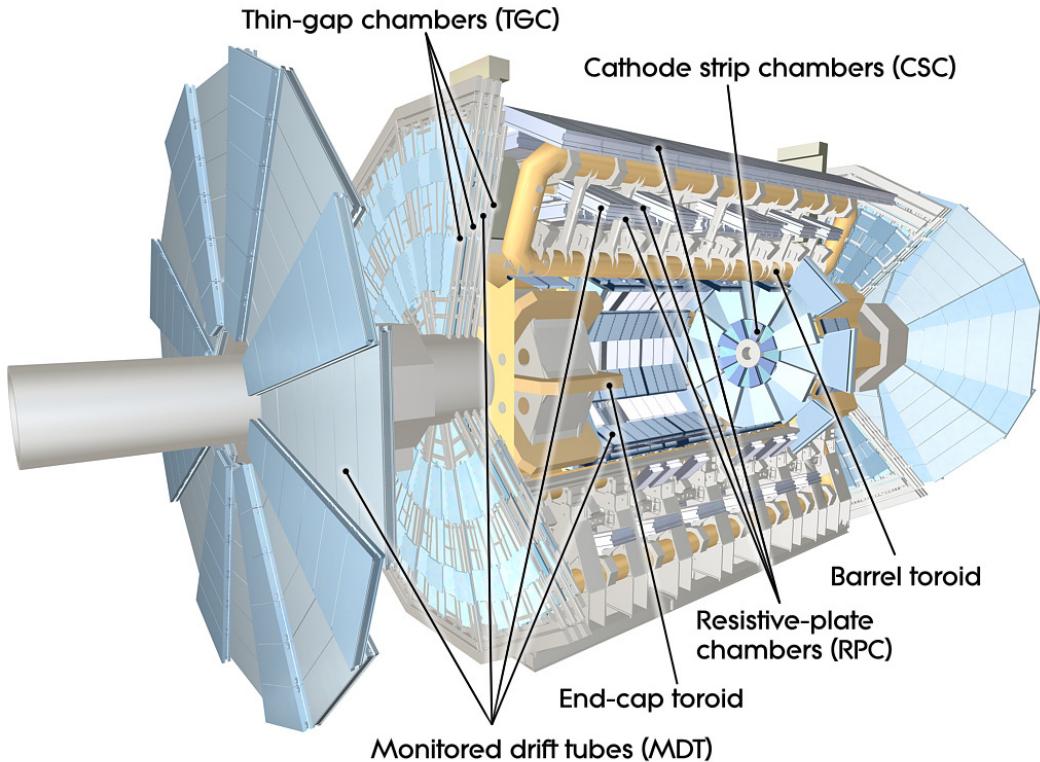


Figure 3.7: Diagram of the ATLAS muon system

697 large portion of the bandwidth to general purpose single lepton triggers (~ 250 Hz). The presence of
 698 leptons in the event indicates the presence of the weak or electro-magnetic interaction and therefore
 699 occurs at many order of magnitude less frequently than interactions involving the strong interaction.
 700 Moreover, many interesting physics signatures that are analyzable by ATLAS involve leptonic final
 701 states. The remaining bandwidth is allocated to jet, missing energy, tau, and unbiased supporting
 702 triggers.

703 The ATLAS trigger system is composed of 3 levels: level-1 (L1), level (L2), and the event filter
 704 (EF). The first level is hardware only trigger that reduces the input 20 MHz rate to ~ 75 kHz, selecting
 705 1 out of every 250 collisions. The available buffering on the FPGA chips means that the decisions
 706 need to be made within $2.5 \mu\text{s}$. The L1 selection is based on calorimeter clustering and tracking
 707 finding in the MS for small areas of the detector called regions-of-interest (ROIs). It selects ROIs
 708 with significant energy.

709 The second and third stages L2 and EF are software based. The L2 algorithms perform more
 710 detailed object reconstruction for leptons, jets and photons inside of the ROIs provided by L1, by

711 performing tracking and in depth calorimeter clustering algorithms. The decisions are made within 50
712 ms per event and pass 1 out of every 15 events to the EF. At the EF, events undergo full reconstruction
713 using similar but faster versions of the algorithms used offline. The EF makes decisions on the presence
714 of fully id-objects in the event and event topological quantities within 4s to reduce the L2 output by
715 a factor of 10. The events that pass this stage are then written to tape for offline study.

716 **3.2.6 Reconstruction: Jets, Muons and Electrons**

717 Physicists analyze the collision event as a collection of identified objects, expressed as momentum
718 4-vectors. The process of converting the disparate detector signatures and signals into a unified 4-
719 momentum description of individual objects is called reconstruction. These objects arise from the
720 final state particles in the event, which can be combined and counted to infer properties of the hard
721 scatter. The particles that make detectable signatures are those that are stable enough to pass through
722 the detector: muons, electrons, photons, and quarks and gluons. Jets and b-tagged jets, muons and
723 electrons are used in the $t\bar{t}H$ analysis to define our search regions and to separate the Higgs signal
724 from backgrounds. Other analyses may use photons, taus and missing energy⁶, but these are not
725 discussed in depth here. Figure 3.8 shows an $R - \phi$ schematic of the interaction of various particle
726 signatures in the ATLAS detector.

727 **3.2.6.1 Tracks and Clusters**

728 The basic components of reconstruction are sensor measurements, or hits, in trackers (ID, MS) and
729 energy measurements in the calorimeter. Hits in the ID and MS undergo pattern recognition, which
730 identifies hits that belong to a single track, and fitting, which fits a curve to the track to assess
731 the particle trajectory. Charged particle trajectories are generally helical in a magnetic field, but
732 the fitting algorithm takes into more detailed information about energy loss to material along the
733 tracks length. The result of the fitting is an estimation of particle momentum 3-vector. Electrons,
734 photons and hadronic particles leave clustered deposits of energy in the EM and hadronic calorimeters
735 from their showers. Electron and photon showers are primarily contained with the EM calorimeter,
736 while hadronic showers are deposit most of their energy in the hadronic calorimeters. The process
737 of associating individual read-out cells of energy in the calorimeter to clusters of energy from the
738 showers of individual particles is called clustering. From the basic pieces of tracks and clusters, more
739 complex objects can be created.

⁶missing energy is the presence of momentum imbalance in the transverse plane of the calorimeter due to escaping neutrinos

740 **3.2.6.2 Electrons**

741 Electrons leave both a track in the ID and a narrow and isolated cluster of energy in the EM calorime-
742 ter, $\Delta R < 0.1$. Electron reconstruction proceeds using a sliding window algorithm, which scans a
743 fixed size rectangle in $\eta - \phi$ space over the EM calorimeter cells to find relative maxima of energy in
744 the window[51]. These maxima seed the clustering algorithms. Because electrons are light, they both
745 lose energy to the material gradually through scattering and more catastrophically through the emis-
746 sion of a high energy photon, through interaction with nuclei. This process is called bremsstrahlung.
747 Tracks for electrons are reconstructed differently because they must include the hypothesis that the
748 electron loses significant energy through bremsstrahlung. Generally, the emitted photon is contained
749 within the same energy cluster and therefore the sliding window algorithm is always wider in the di-
750 rection of bending, ϕ . A single track is then matched to the cluster within certain minimum matching
751 requirements in η, ϕ , and p_T . Electrons are distinguished from photon conversions, which also have a
752 track, by lack of association with conversion vertices, found with a dedicated algorithm.

753 Electron have many lever arms for further identification to suppress backgrounds from fake sources.
754 The narrowness of the shower shape, quality of track, and presence of transition radiation are used
755 by cut-based and multivariate identification algorithms. This is discussed in depth in Chapter 4.
756 Electrons are reliably reconstructed and identified with energies above 7 GeV.

757 **3.2.6.3 Muons**

758 Muons are reconstructed from a combination of ID and MS tracks, when possible. The two tracks
759 must meet matching criteria to ensure they are from the same particle. The muon momentum 3-vector
760 comes from the combined ID/MS fit. Muons leave little energy in the calorimeters and are generally
761 isolated from other particles, when produced from electro-weak bosons. Identification algorithms
762 make requirements on the number of tracking hits in the ID and MS and the quality of the matching
763 of the two tracks. Muons are reliably reconstructed and identified with energies above 5 GeV. More
764 about muon reconstruction and identification can be found here [52].

765 **3.2.6.4 Jets**

766 Quarks and gluons are colored objects that cannot exist alone on the time scales of detector measure-
767 ments, due to confinement, a property of the strong force . When emitted, they undergo a process
768 called hadronization, in which the convert into 'jets' of colorless hadrons that emerge collimated from
769 the interaction point. The majority of these hadrons are charged and neutral pions, though other
770 hadrons are often present. Jets are reconstructed using conglomerations of calorimeter energy clusters

771 chosen via an anti- k_t algorithm, with a radius of $\Delta R < 0.4$ [53]. The algorithm has been shown to be
772 infrared safe, meaning the jet quantities are not sensitive to low energy, small angle radiative diver-
773 gences. Jets at ATLAS are reconstructed from 10 GeV, calibration of the energy scale and resolution
774 are only available for energies greater than 20-25 GeV.

775 **3.2.6.5 B-Tagged Jets**

776 Generally, the flavor of the initiating quark is not known from the reconstructed jet, although gluon
777 initiated jets and quark initiated jets have slightly different properties. B-quark jets, however, are
778 unique in that the long life-time of the produced b-mesons allow for measurable decays in flight. This
779 property is used to tag b-quark initiated jets. This analysis uses the MV1 tagging algorithm [54],
780 which is a neural network based algorithm that looks for secondary displaced decay vertices inside
781 the event and takes into account jet track parameters and energy flow with respect to these vertices.
782 B-quark jets often involve b-meson decays to leptons, especially muons, which can be used to tag an
783 orthogonal b-jet sample for studying tagging efficiencies.

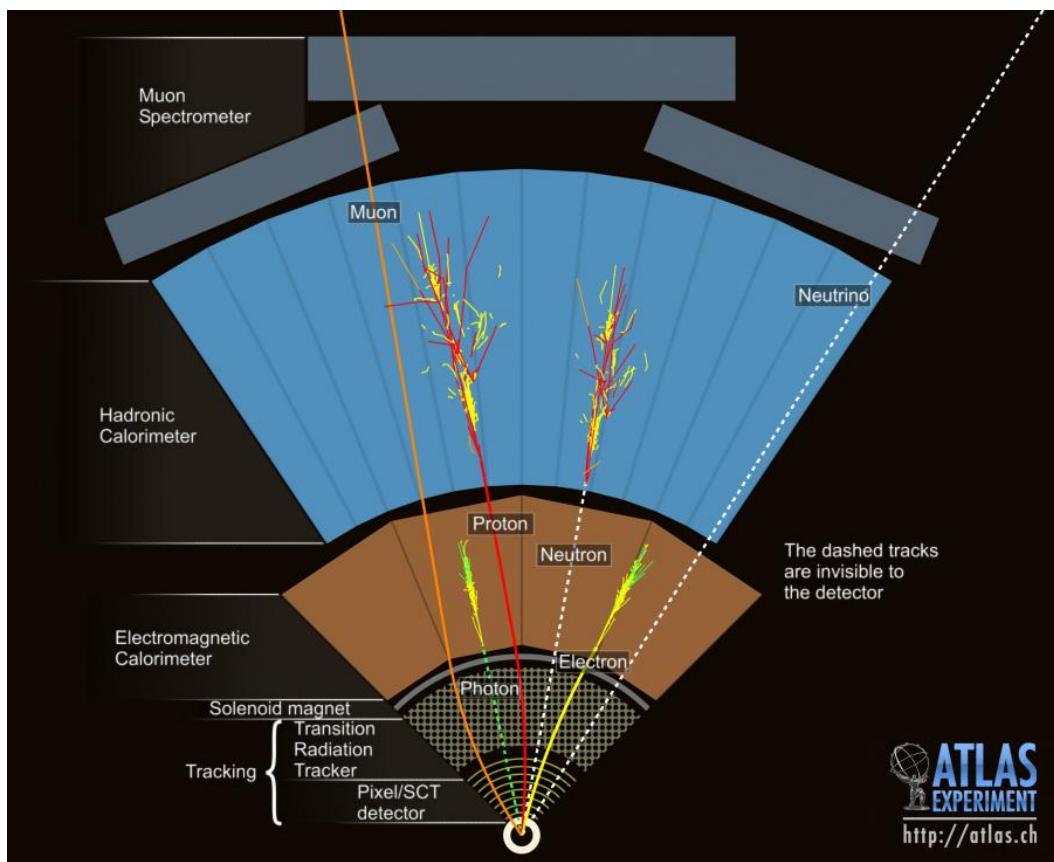


Figure 3.8: $R - \phi$ schematic of the ATLAS detector and various particle signatures

784

CHAPTER 4

785

Electrons

786 High energy electron signatures are important elements of searches and measurements at hadron col-
787 liders, because they signal the presence of important electro-weak processes in the event. Requiring
788 well-identified electrons in collision events quickly suppresses the overwhelming rate of strong-force
789 mediated scattering and allows for the collection of a manageable-sized dataset with interesting physics
790 for study. For this reason, electron signatures form one of the two pillars of the HLT trigger at ATLAS,
791 as discussed in Chapter ?? and rigorous electron identification is an important piece of many ATLAS
792 analyses. This section summarizes the development of ATLAS electron identification for the high
793 luminosity 2011 and 2012 datasets and discusses the techniques involved in measuring the electron
794 identification efficiency.

795 4.1 Identification of Electrons at ATLAS

796 Electron reconstruction is discussed briefly in Chapter ?? and, in depth here []. The result of electron
797 reconstruction is called an electron candidate, which is comprised of a narrow calorimeter energy
798 cluster with $|\eta| < 2.47$ and inner detector track that matches loosely in η and ϕ . If the electron
799 has $|\eta| < 2.01$, the inner detector track is fiducial to the TRT has the possibility of having
800 high-threshold hits, indicative of transition radiation (TR). Electron cluster reconstruction is extremely
801 efficient. The track-matching requirement is less efficient, because the presence of hard bremsstrahlung
802 may in certain cases cause the electron cluster and emitted photon cluster to have a wide separation
803 in the calorimeter[].

804 Objects that are not isolated electrons are often reconstructed as electrons, as the reconstruction
805 requirements are quite loose. Objects that often ‘fake’ isolated electrons are light quark and gluon
806 jets, heavy flavor jets that include real decays to electrons and converted photons. Light quark and

807 gluon jets fragment into a number of collimated hadronic particles. In rare cases, the jet may fragment
 808 most of its energy into a single charged pion, which showers early in the EM calorimeter. In other
 809 rare cases, the jet may fragment mostly into a neutral pion, which subsequently decays into a pair of
 810 photons. If one of these photons converts, a track will point to the EM energy cluster. These cases
 811 would result in a reconstructed electron candidate. Although the probability for this to happen is
 812 small, the enormous jet production rate means that it is a significant background. In general, light
 813 quark and gluon jet ‘fakes’ have larger transverse shower profiles and more energy leakage into the
 814 hadronic calorimeter. For the neutral pion case, there are generally two separated showers for lower
 815 energy decays. For both cases, there are often other particle signatures nearby. On the other hand,
 816 heavy-flavor jet decays and photon conversions contain real electrons. However, heavy flavor decays
 817 also involve the production of additional hadronic particles within the jet and both photon conversion
 818 and heavy flavor decays involved secondary vertices displaced from the primary interaction point.

819 In order to distinguish these fake signatures from real isolated electrons, electron identification
 820 algorithms use a number of reconstructed variables describing the electron shower in the detector
 821 and the electron track. The details of the calculated variables can be found here []. In general, the
 822 calorimeter variables take advantage of the narrowness of isolated electron shower in the transverse
 823 plane and lack of energy deposition in the hadronic calorimeter. The transverse variables include
 824 measurements of the shower width in both layer 2 and the strips, where more refined measurements
 825 are possible. In fact, the strips were designed to separate single photon and eletron showers from
 826 multiple showers from neutral pion decays, shown in Figure ???. The shower width variables are
 827 generally measured mostly in η as bremsstrahlung tends to smear the electron energy in ϕ . Electron
 828 tracks are required to have an adequate number of hits in the Pixel Detector, SCT and TRT. These
 829 hit requirements, especially the b-layer requirement suppress electron conversions which occur in the
 830 detector material. Track-cluster matching and geometric impact parameter variables require ID tracks
 831 to match the calorimeter energy well and to arise from the primary interaction point, respectively.
 832 Electrons with tracks explicitly associated with a conversion vertex may be rejected. Finally, the high
 833 threshold fraction of hits on the track, made by transition radiation is an uncorrelated discrimator of
 834 pion and electron tracks.

835 Electron identification algorithms make selections in 9 bins of $|\eta|$, [0.10, 0.60, 0.80, 1.10, 1.37, 1.52,
 836 1.81, 2.01, 2.37, 2.47] and bins of p_T , [7, 10, 15, 20, 30, 40, 50, 60, 70, 80+]. The $|\eta|$ binning changes
 837 with the calorimeter geometry, which in turn affect the shower shape distributions. The shape of
 838 most of the identification variable distributions, tracking and calorimeter, are p_T dependent.

839 **4.1.1 Pile-up and Electron identification**

840 **4.1.2 2011 Menu**

841 Electron identification in 2011 was accomplished through rectangular cuts on the identification vari-
842 ables at 3 operating points: Loose, Medium and Tight. The medium operating point was used online
843 as the primary electron trigger. At the beginning of the 2011 run, the 3 operating points possessed
844 the same cuts, but tighter operating points had cuts on more variables. The Loose operating point
845 only cut on shower shape variables in layer 2 and hadronic leakage, the medium operating point added
846 cuts on shower width variables in the strips, and tight added TR cuts, strict-track cluster matching,
847 conversion rejection and a b-layer requirement. This menu, called the ‘IsEM’ menu, was the first fully
848 data-optimized cut menu for electrons.

849 The demands of increasing luminosity demanded a tightening of the medium operating point
850 midway through the data-taking, in order to maintain a EF trigger rate of around 20-25 Hz on the
851 primary electron trigger. To accomplish this, variables cut on at the tight operating point were added
852 to the medium operating point, and the entire set of cuts was optimized to provide the targeted
853 fake rejection and reduction in the rates at the highest possible efficiency. The same procedure was
854 applied to the loose operating point, where the target was to provide efficiencies of 95% and the
855 highest possible fake rejection. The re-inventing of the menu in this way allowed for not only better
856 performance, due to the inclusion of more variables, but a more stable tightening of the backgrounds
857 from loose to medium to tight, where the same backgrounds types were targeted at each level. The new
858 menu was called the ‘IsEM++’ menu and the operating points were renamed ‘Loose++’, ‘Medium++’
859 and ‘Tight++’. Figure ?? shows the comparison of the operating points for the new menu and old
860 menu.

861 **4.1.3 2012 Menu and Pile-up**

862 Improvements in the running conditions for 2012, in particular narrowing the transverse beam emit-
863 tance and size, resulted in large increase in number of proton-proton interactions during every 50
864 ns bunch crossing. In 2011 the average number of reconstructed primary vertices in each event, an
865 indicator of the number of interaction per bunch crossing, was around 7, while in 2012 the average
866 grew to 25. Some events during 2012 running had 40 reconstructed primary vertices.

867 The increase in energy in the calorimeters from these additional collisions, called pile-up, caused
868 a worsening of the resolution of electron identification variables, particularly the shower shapes and
869 hadronic leakage. The presumed cause was the increase in the number of showers of low energy hadronic

870 particles nearby electrons. Figure ?? shows an example distribution, the hadronic leakage under
871 different pile-up conditions. The distributions shows a clear widening which results in a loss of
872 efficiency.

873 In order to combat this loss, the ‘IsEM++’ menu was once again optimized to have a flatter
874 efficiency profile as a function of the amount of pile-up in the event with similar performance to the
875 2011 menu. The strategy for this menu was to loosen selections on variables sensitive to pile-up
876 energy. It was expected and confirmed that relying more on the strip variables for the shower shape
877 selection and the energy in layer 3 of the EM calorimeter for the hadronic leakage, would use a smaller
878 volume of the calorimeter and thus be less sensitive to additional energy in the neighborhood of the
879 electron. The strategy is outlined pictorially in Figure ??.

880 The efficiency of the 2011 operating points compared to the 2012 operating points is shown in
881 Figure ??, demonstrating a clear improvement in efficiency of the selections for higher pile-up conditions.
882 Finally, the 2012 primary electron trigger to include highly efficient tracking isolation criteria in order
883 to

884 4.1.4 Electron Likelihood

885 A natural step forward for the electron identification is the use of multi-variate algorithms. Multi-
886 variate identification algorithms use many identification variables at once in a multi-dimensional space
887 and judge where signals and backgrounds are found in that space to make decisions

888 For the case of electron identification, it was found that using a likelihood function, trained with
889 electron identification variables, provided clear performance gains with respect to rectangular cuts, while
890 also providing stable and easily understandable results. The likelihood scores each electron based on
891 how signal-like or background it is for each individual identification variable and then multiplies
892 these identification variables together into a final score. Figure ?? shows

893 XXX equations?

894 There are many advantages to a likelihood based approach. First, variables that show significant
895 shape differences between real and fake electrons but do not have a clear cut point can still be used
896 in a likelihood. Second, the likelihood score takes into account the entire shape of the distribution and
897 not simply an efficiency and fake rejection at a single cut point. Finally, the final cut on the likelihood
898 output score.

899 The likelihood menu for ATLAS was developed at the end of the 2012 Run to be used on advanced
900 2012 analyses. The menu uses similar variables to the cut-based menus but adds a few additional
901 variables, including a measurement of the amount of energy the electron track lost as it traversed the

902 ID. The likelihood menu makes cuts on the likelihood output score at 4 different operating points with
903 the same binning as the cut menu but tunes the cuts based on the amount of pile-up in the event.
904 The likelihood menu greatly outperforms the rejection of the ‘IsEM++’ menu for similar efficiencies.
905 Figure XXX shows. The likelihood menu, specifically the VERYTIGHT operating point, is used in the
906 $t\bar{t}H$ analysis.

907 **4.2 Measurement of Electron Identification Efficiency at ATLAS**

908 Precise measurements of the electron identification efficiency are important pieces of many ATLAS
909 analyses, including the $t\bar{t}H$ multi-leptons analysis. For analyses with low p_T leptons, systematic
910 uncertainties on the electron identification efficiency can be some of the largest systematic effects. The
911 methods used to measure the electron identification efficiency are described in depth here [].

912 Electron identifications efficiencies are measured using a method called tag-and-probe for J/Ψ
913 and Z boson decays to electrons. One object from the decay is tagged, while the other electrons is
914 left unidentified. Reasonable confidence that the second object, thought not identified, is an electron
915 based on the kinematic properties of the event, specifically the di-electron invariant mass is near the
916 Z of J/Ψ pole. The tag-and-probe method leaves a sample of unidentified and unbiased probes, where
917 the efficiency can be measured.

918 As opposed to muons, contamination from fake electron make the tag-and-probe method difficult.
919 This is especially true for electrons in with p_T of 10-20 GeV. The energy scale of Z and J/Ψ decays
920 disfavor electrons of this momentum, but low energy electrons are important in a number of Higgs
921 searches. Backgrounds from fake electrons are subtracted using fits to the Z and J/Ψ invariant
922 mass distributions. For Z electrons, fits to the electron isolation distribution are also used. The
923 final efficiencies reported are the result of statistical fit among all methods. The uncertainties are
924 at low momenta are around $\sim 5\%$ and are dominated by systematics effects from large background
925 subtractions. They are less than 1% at high momenta and dominated by tag-and-probe selection
926 effects.

927 **4.2.1 Issues**

928 For the measurement of the 2012 electron identification efficiencies, an inconsistency between the
929 isolation-based background subtraction method and

930

CHAPTER 5

931

Analysis Summary

932 This chapter provides an overview the of analysis searching for SM production of the Higgs boson in
 933 association with top quarks in multi-lepton final states. The analysis searches in signal regions (SRs)
 934 with 2 same-sign, 3 and 4 light leptons (e, μ), which are sensitive to Higgs decays to vector bosons,
 935 $H \rightarrow W^\pm W^\pm$ and $H \rightarrow Z^\pm Z^\pm$. We refer to these channels as 2ℓ SS, 3ℓ , and 4ℓ through the rest of
 936 this document.

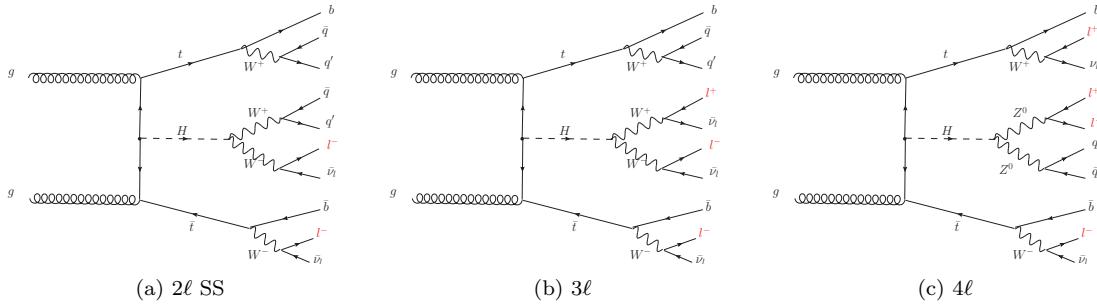
937 The multi-lepton channels form a complement to already completed $t\bar{t}H$ searches in final states
 938 targeting the $H \rightarrow b\bar{b}$ [55], $H \rightarrow \gamma\gamma$ [56]. The $t\bar{t}H$ searches in the $H \rightarrow \tau\tau$ decay modes were
 939 developed concurrently with the multi-lepton searches, but we do not discuss these here. Of this set
 940 of complementary searches, the multi-lepton and $b\bar{b}$ are the most sensitve.

941 Based on SM production cross-sections, observation lies just outside the sensitivity of the Run I
 942 dataset, even when combining all searches. Instead, the analyses provide an opportunity to constrain
 943 for the first time the $t\bar{t}H$ production mode with limits reasonably close to the actual production rate.
 944 The multi-lepton analysis is therefore optimized to overall sensitivity to the $t\bar{t}H$ production rather
 945 than individual decay modes, which would be more useful for constraining Higgs couplings.

946 Detailed description of the event and objection section are provided in Chapter 7, background
 947 modeling in Chapter 9, the effect of systematic errors and the statistical analysis in Chapter 9 and
 948 final results in Chapter 10.

949 5.1 Signal Characteristics

950 The signal is expected to be characterized by the presence of 2 b-quark jets from the top quark
 951 decays, isolated leptons from vector boson and tau decays, a high jet multiplicity and missing energy
 952 from neutrinos. Three Higgs boson decays are relevant for this analysis: W^+W^- , $\tau^+\tau^-$ and ZZ .

Figure 5.1: Example Feynman diagrams for the 3 $t\bar{t}H$ multi-lepton categories.

All modes are generally dominated by the WW signature, though the 3ℓ and 4ℓ channels possess some contribution from the $\tau\tau$ and ZZ decays. Table 5.1 provides the fractional contribution of the main Higgs decay modes at the generator level to $t\bar{t}H$ search channels and Figure 5.1 shows example diagrams for each channel. In general, the number of leptons is anti-correlated with the number of jets, since a vector boson can either decay leptonically or hadronically, such that:

- in the 2ℓ SS channel, the $t\bar{t}H$ final state contains 6 quarks⁷. These events are then characterized by the largest jet multiplicity.
- In the 3ℓ , the $t\bar{t}H$ final state contains 4 quarks
- In the 4ℓ channel, the $t\bar{t}H$ final state contains a small number of light quarks, 0 ($H \rightarrow W^+W^-$ case), 2 or 4 ($H \rightarrow ZZ$ case).

Table 5.1: Contributions of the main Higgs decay modes to the 3 multi-lepton $t\bar{t}H$ signatures at generation level.

Signature	$H \rightarrow WW$	$H \rightarrow \tau\tau$	$H \rightarrow ZZ$
Same-sign	100%	—	—
3 leptons	71%	20%	9%
4 leptons	53%	30%	17%

5.2 Background Overview

For all channels after selection, the size of the signal is of similar order to the expected size of background. Background processes can be sorted into two categories:

⁷this does not include additional quarks from radiation

- 966 • **Reducible:** These processes cannot lead to a final state compatible with the signal signature
967 without a mis-reconstructed object. This category includes events with a prompt lepton but
968 with mis-reconstructed charge and events with jets that "fake" leptons. The main backgrounds
969 of this sort are $t\bar{t}$ and $Z+jets$. Data-driven techniques are used to measure the rate of these
970 processes and strict object selection and isolation requirements are used to reduce their rate.
- 971 • **Irreducible:** Events which can lead to the same final state as the signal. The main background
972 of this category are: vector boson production (V) associated with top quarks ($t\bar{t}V$), a Z boson
973 produced in association with a top quark (tZ), $W^\pm Z$, and ZZ . They are modeled using the
974 Monte Carlo simulations. In general, these backgrounds are combatted with jet and b-tagged
975 jet requirements. Although the jet multiplicity of $t\bar{t}V$ is high, the multiplicity of $t\bar{t}H$ events is
976 still higher.

977

5.3 Analysis Strategy

978 The analysis search is conducted in 3 channels, based on counting of fully identified leptons: 2ℓ SS,
979 3ℓ , and 4ℓ , with cuts optimized separately for each. We further divide the 2ℓ SS into sub channels
980 based on the number of jets and flavor of the leptons and the 4ℓ channel into sub-channels enriched
981 and depleted in opposite-sign (OS) leptons arising from Z decays.

982 This analysis is a counting experiment, meaning that the only quantities significant to the result
983 are the event counts in the signal regions and not the event shapes. The measured background rates,
984 expected signal rates and systematic uncertainties are fed into a Poisson model and fit to the observed
985 data. The parameter of interest in the fit and the result of this measurement is, μ , the ratio of the
986 fitted number of $t\bar{t}H$ events in the signal regions to expected number of $t\bar{t}H$ events in the signal
987 regions. Since we assume SM branching ratios, μ can be considered the ratio of the measured $t\bar{t}H$
988 cross-section to the observed $t\bar{t}H$ cross-section, and we expect the fitted μ to be close to 1 with large
989 statistical errors.

990 We express the final result as a measurement of μ with uncertainties and 95% upper limit on the
991 value of μ : μ -values higher than this value will be considered excluded. We provide these results for
992 each channel individually and combined.

993

CHAPTER 6

994

Dataset and Simulation

995

6.1 Data

996

6.1.1 The 2012 Dataset

997 The $t\bar{t}H$ analysis uses the entire 2012 ATLAS dataset only, collected from April to December. The
 998 size of the dataset corresponds to 20.3 fb^{-1} , after passing data quality requirements, ensuring the
 999 proper operation of the tracking, calorimeter and muon subsystems. The LHC successfully produced
 1000 datasets for physics studies in 2010, 2011 and 2012. The 2012 proton-proton dataset was delivered
 1001 with collisions with a CME of 8 TeV with bunch collisions every 50 ns[57].

1002 Figure 6.1 shows the accumulation of the 2012 dataset over time. Despite doubling the bunch
 1003 spacing above the design of 25 ns, the luminosity neared the design luminosity due to unexpected
 1004 improvements in the transverse beam profile[58]. This increased the amount of pile-up, or number
 1005 of collisions per bunch crossing and in general collision events were busier due to these multiple
 1006 interactions. Figure 6.2 shows the average number of interaction per bunch crossing for the 2011 and
 1007 2012 datasets. The 2012 dataset shows an average of 20-25 interactions.

1008 The dataset must contain either a primary muon or primary electron trigger (`EF_e24vhi_medium1`
 1009 OR `EF_e60_medium1` OR `EF_24i_tight` OR `EF_36_tight`). The electron triggers require a electron with
 1010 at least 25 GeV of calorimeter energy, passing the medium identification requirement and loose tracking
 1011 isolation. Above 60GeV, the isolation requirement is dropped and the identification is loosened slightly.
 1012 The muon trigger requires a good inner detector track and matching hits in the muon spectrometer,
 1013 as well as loose tracking isolation, which is also dropped about 36 GeV.

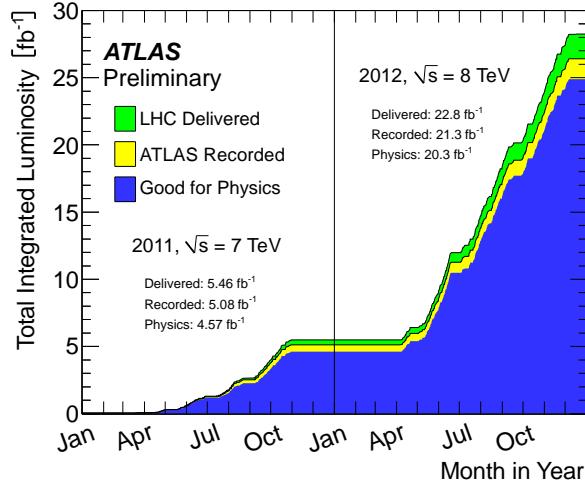


Figure 6.1: Plot showing the accumulation of the integrated luminosity delivered to the ATLAS experiment over 2011 and 2012. The rough size of the usable, physics ready dataset for 2012 is 20 fb^{-1} and is the dataset used.

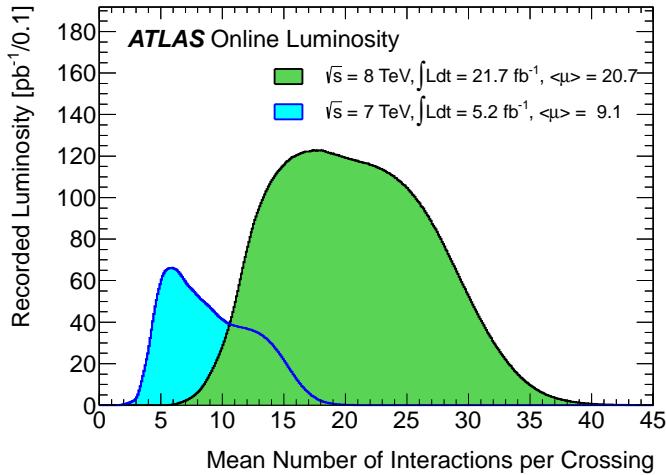


Figure 6.2: The average number of interactions per bunch-crossing for the 2012 and 2011 LHC proton-proton dataset. Most of these interactions are uninteresting but leave energetic signatures in particle detectors called pile-up which interfere with measurements

6.2 Simulation

1014 Simulation samples are used to determine the overall event selection acceptance and efficiency and
 1015 model the number of events in the signal regions for prompt backgrounds and signal. The simulated
 1016 samples are created using parton distribution function (PDF) and use Monte Carlo (MC) techniques
 1017 to model the hard parton scatter, underlying event activity and parton showering and hadronization.
 1018

Table 6.1: Monte Carlo samples used for signal description.

Process	Generator	Cross-section [fb]	\mathcal{L} [fb $^{-1}$]	Detector simulation
$t\bar{t}H \rightarrow \text{allhad} + H$	PowHel+Pythia8	59.09	2146.5	Full
$t\bar{t}H \rightarrow \text{ljets} + H$	PowHel+Pythia8	56.63	2238.9	Full
$t\bar{t}H \rightarrow ll + H$	PowHel+Pythia8	13.58	9332.0	Full

1019 The samples are then passed through a full ATLAS detector simulation[59] based on GEANT4 [60].
 1020 Small corrections are then applied to re-scale object identification efficiencies, energy scales, and the
 1021 pile-up based on control regions from data. These corrections are discussed in Chapter 9.

1022 6.2.1 Signal Simulation

1023 The $t\bar{t}H$ production is modeled using matrix elements obtained from the HELAC-Oneloop package [61]
 1024 that corresponds to the next-to-leading order (NLO) QCD accuracy. Powheg BOX [62, 63, 64] serves
 1025 as an interface to the parton shower Monte Carlo programs. The samples created using this approach
 1026 are referred to as PowHel samples. CT10NLO PDF sets are used and the factorization (μ_F) and
 1027 renormalization (μ_R) scales are set to $\mu_0 = \mu_F = \mu_R = m_t + m_H/2$. Pile-up and the underlying
 1028 events are simulated by Pythia 8.1 [65] with the CTEQ61L set of parton distribution functions and
 1029 AU2 underlying event tune. The Higgs boson mass is set to 125 GeV and the top quark mass is set
 1030 to 172.5 GeV.

1031 The signal Monte Carlo samples are summarized in Table 6.1. These large samples are generated with
 1032 inclusive Higgs boson decays with branching fractions set to the LHC Higgs Cross Section Working
 1033 Group (Yellow Report) recommendation for $m_H = 125$ GeV [66]. The inclusive cross section (129.3
 1034 fb at $m_H = 125$ GeV) is also obtained from the Yellow Report [66].

1035 6.2.2 Background Simulation

1036 The background simulations used for this analysis are listed in Table 6.2. In general, the Alpgen[67],
 1037 MadGraph[68], and AcerMC[69] samples use the CTEQ6L1[70] parton distribution function, while
 1038 the Powheg[71], Sherpa[72], are generated with the CT10 PDF. The exception is the MadGraph $t\bar{t}t\bar{t}$
 1039 sample, which is generated with the MSTW2008 PDF[73]. The highest order calculations available
 1040 are used for the cross sections.

Table 6.2: Monte Carlo samples used for background description. Unless otherwise specified MadGraph samples use Pythia 6 for showering and Alpgen samples use Herwig+Jimmy. $t\bar{t}$, single top, and $Z+jets$ samples are replaced with data-driven estimates for the final result

Process	Generator	Detector simulation
$t\bar{t}W^\pm, t\bar{t}Z$	MadGraph	Full
tZ	MadGraph	AF2
$t\bar{t}t\bar{t}$	MadGraph	Full
$t\bar{t}W^\pm W^\pm$	Madgraph+Pythia8	AF2
$t\bar{t}$	Powheg+Pythia6	Full/AF2
single top tchan	AcerMC+Pythia6	Full
single top schan $\rightarrow l$	Powheg+Pythia6	Full
single top $W^\pm t$	Powheg+Pythia6	Full
$W\gamma^*$	MadGraph	Full
$W\gamma+4p$	Alpgen	Full
W^+W^-	Sherpa	Full
$W^\pm Z$	Sherpa	Full
Same-sign WW	Madgraph+Pythia8	AF2
ZZ	Powheg+Pythia8,gg2ZZ+Herwig	Full
$Z\gamma^*$	Sherpa	Full
$Z+jets$	Sherpa	Full
ggF Higgs	Powheg+Pythia8	Full

1041

CHAPTER 7

1042

Object and Event Selection

1043 The analysis is divided into 3 signal regions based on lepton counting: 2 same-sign leptons, 3 leptons
 1044 and 4 leptons. The lepton counting occurs for fully identified leptons with full overlap removal with
 1045 transverse momenta over 10 GeV to ensure orthogonality. Lepton selections are tightened afterward
 1046 within each region.

1047 The cuts for each signal region are provided in Table 7.1 and the object selections are detailed in the
 1048 following selections. The selections are based on optimizations of the region sensitivity performed using
 1049 MC (event for data driven backgrounds) and ad-hoc values for normalization systematic uncertainties⁸
 1050 i All signal regions are comprised of three basic requirements: the presence of b-tagged jets, the
 1051 presence of additional light jets, and a veto of same flavor opposite sign leptons with an invariant
 1052 mass within the Z window. Additional requirements on the invariant mass of the leptons, the missing
 1053 transverse energy in the event, and the total object energy (H_T) proved to have negligible additional
 1054 benefit at our level of statistics. Figure 7.1 shows the background and signal fractions as a function
 1055 number of jets and number of b-tagged jets for otherwise fully selected events.

1056 **7.1 2ℓ Same-Charge Signal Region**

1057 The 2 lepton signal region requires two leptons of similar charge (2ℓ SS). The signal is symmetric
 1058 in charge but the background from opposite-sign $t\bar{t}$ di-lepton production would be overwhelming.
 1059 Requiring only two leptons allows the extra 2 W bosons in the event to decay hadronically, resulting
 1060 in on average 4 additional light jets plus 2 additional b-quark jets from the top decays.

⁸the sensitivity was approximated using the $\frac{s}{\sqrt{b+\Delta b}}$ formula. The systematic errors considered were 20% for $t\bar{t}V$ and VV and 30% for fakes. These ended up being close the final systematic errors assessed in Chapter 9. The objects of optimization were the lepton momenta, identification operating points, isolation and event kinematic variables

Table 7.1: Selections in the 2ℓ SS, 3ℓ and 4ℓ Signal Regions

Signal Region	2ℓ SS	3ℓ	4ℓ
Trigger Matched Lepton	Yes	Yes	Yes
N_l^9	=2	=3	=4
Lepton Charge Sum	+2 or -2	+1 or -1	0
Lepton Momentum (GeV) ¹⁰	$p_{T0} > 25$ $p_{T1} > 20$	$p_{T0} > 10$ $p_{T1,2} > 25, 20$	$p_{T0} > 25$ $p_{T1} > 20$ $p_{T2,3} > 10$
Jet Counting	$N_b \geq 1, N_{jet} = 4$	$N_b \geq 1, N_{jet} \geq 4$ or $N_b \geq 2, N_{jet} = 3$	$N_b \geq 1, N_{jet} \geq 2$
Mass Variables (GeV)	$ M_{ee} - M_Z < 10$	$ M_{SFOS} - M_Z < 10$	$M_{SFOS} > 10$ $150 < M_{4\ell} < 500$ $ M_{SFOS} - M_Z < 10$
Sub-channels	$2 (N_{jet} = 4, N_{jet} \geq 5)$ $\times 3(ee, e\mu, \mu\mu)$	none	2 (No SFOS leps, SFOS leps)

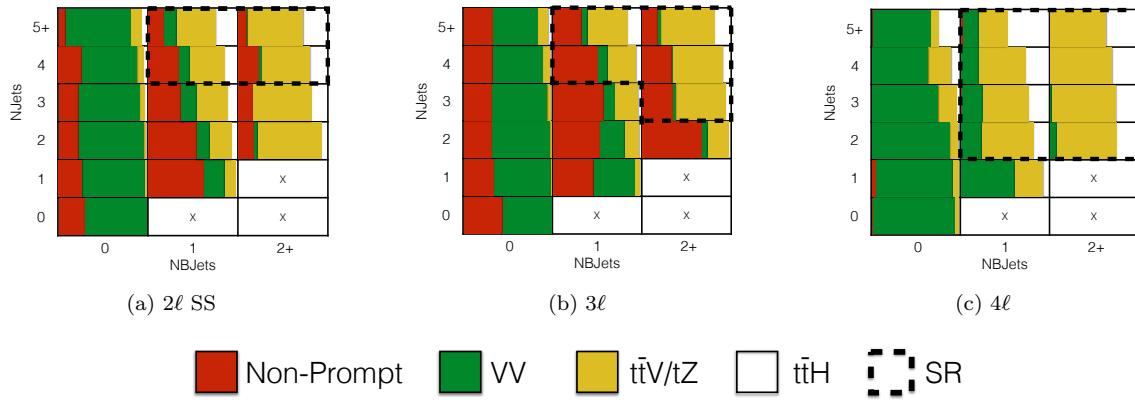


Figure 7.1: Number of jets vs. number of b-tagged jet plot for the fully selected multi-lepton channels. Signal regions are outlined with a dashed line. Sub-channels are defined later in the 2ℓ SS and 4ℓ SRs. The fractional background contribution to each jet and b-tagged jet bin are shown for non-prompt (red), $ttV + tZ$ (yellow), and VV (green). The expected signal fraction is shown in white. The expected non-prompt fraction contains charge misidentifications and fakes. It is shown for MC only, although data-based methods are used for the final result.

1061 We require a leading lepton with transverse momentum of at least 25 GeV that matches to a
 1062 trigger and a sub-leading lepton of at least 20 GeV, a b-tagged jet, and at least 4 jets in total.

1063 In order to suppress non-prompt backgrounds, the lepton isolation criteria for tracking and
 1064 calorimeter are tightened from less than 10% of the lepton momentum to 5%. To suppress charge
 1065 misidentification, the electron is required to be extremely central ($|\eta| < 1.37$) to avoid the material-
 1066 rich regions of the detector. Additionally, ee events with a lepton pair invariant mass within 10 GeV

1067 of the Z pole are removed. To maintain orthogonality with the τ analyses, events with fully identified
 1068 taus are vetoed.

1069 For the statistical combination the channel is divided into 6 sub-channels: 2 jets counting bins
 1070 ($N_{Jet} = 4, N_{Jet} \geq 5$) \times 3 lepton flavor bins (ee, $\mu\mu$,e μ).

1071 7.2 3ℓ Signal Region

1072 The 3 lepton channel requires 3 leptons, whose summed charge is either -1 or $+1$. The leptons are
 1073 ordered in this way:

- 1074 • **lep0**: the lepton that is opposite in charge to the other two leptons
- 1075 • **lep1**: the lepton that is closer in ΔR to lep0
- 1076 • **lep2**: the lepton that is farther in ΔR from lep1

1077 Since events with a fake lepton arise exclusively from opposite sign di-lepton processes, $t\bar{t}$ and
 1078 Z+jets, where additional jets are misidentified as the third lepton, lep0 is never the fake lepton. As
 1079 a result, the transverse momentum requirement of lep0 (> 10 GeV) is lower than the other two, > 20
 1080 GeV. One lepton must match a trigger and have $p_T > 25$ GeV.

1081 The 3ℓ channel further requires at least one b-tagged jet and at least 4 jets in total, or two b-
 1082 tagged jets and exactly 3 jets in total. Additionally, to suppress $W^\pm Z$ and Z+jets events, events with
 1083 same-flavor opposite-sign (SFOS) pairs within 10 GeV of the Z pole are vetoed.

1084 Additional cuts, including a di-lepton mass cut, and splittings were investigated but low statistics
 1085 proved to wash out any advantages. The di-lepton mass cut will be a useful discriminator in future
 1086 analyses since the spin statistics of Higgs decay in W bosons often causes the two emitted opposite-sign
 1087 leptons to point in the same direction, resulting in a small measured invariant mass.

1088 7.3 4ℓ Signal Region

1089 In the four lepton signal region, selected events must have exactly four leptons with a total charge
 1090 of zero. At least one lepton must be matched to one of the applied single lepton trigger and have
 1091 a transverse momentum above 25 GeV. The leading and sub-leading leptons are required to have
 1092 transverse momentum of 25 and 15 GeV respectively. In order to suppress background contributions
 1093 from low-mass resonances and Drell-Yan radiation, all SFOS lepton pairs are required to have a
 1094 dilepton invariant mass of at least 10 GeV.

1095 The four-lepton invariant mass is required to be between 100 and 500 GeV. This choice of mass
 1096 window suppresses background from the on-shell $Z \rightarrow 4\ell$ peak and exploits the high-mass differences
 1097 between the signal and the dominant $t\bar{t}Z$ background. Events containing an OS-SF lepton pair
 1098 within 10 GeV of the Z boson mass are discarded. This Z-veto procedure greatly reduces background
 1099 contributions from ZZ and $t\bar{t}Z$. Finally, selected events are required to have at least two jets, at
 1100 least one of which must be tagged as a b-quark jet.

1101 The contribution from $t\bar{t}Z$ comprises approximately 75% of the total background in the inclusive
 1102 signal region. A signal region categorization which factorizes $t\bar{t}Z$ from the remaining backgrounds is
 1103 thus beneficial. The signal region is accordingly divided into two categories based on the presence of
 1104 OS-SF lepton pairs in the final state.

1105 7.4 Electron Selection

1106 The electrons are reconstructed by a standard algorithm of the experiment [51] and the electron
 1107 cluster is required to be fiducial to the barrel or endcap calorimeters: $|\eta_{cluster}| < 2.47$. Electrons in
 1108 the transition region, $1.37 < |\eta_{cluster}| < 1.52$, are vetoed. Electrons must have $p_T > 10$ GeV and pass
 1109 the **VERYTIGHT** likelihood identification criteria.

1110 In order to reject jets misidentified as electrons, electron candidates must also be well isolated
 1111 from additional tracks and calorimeter energy around the electron cluster. Both the tracking and
 1112 calorimeter energy within $\Delta R = 0.2$ of the electron cluster must be less than 5% of the electron trans-
 1113 verse momentum: $\text{ptcone20}/P_T < 0.05$ and $\text{Etcone20}/E_T < 0.05$. All quality tracks with momentum
 1114 greater than 400 MeV contribute to the isolation energy. Calorimeter isolation energy is calculated
 1115 using topological clusters with corrections for energy leaked from the electron cluster. Pile-up and
 1116 underlying event corrections are applied using a median ambient energy density correction.

1117 The electron track must also match the primary vertex. The longitudinal projection of the track
 1118 along the beam line, $z0 \sin \theta$, must be less than 1 cm) and the transverse projection divided by the
 1119 parameter error, $d0$ significance, must be less than 4. These cuts are used in particular to suppress
 1120 backgrounds from conversions, heavy-flavor jets and electron charge-misidentifications.

1121 The electron selection is summarized in Table 7.2.

1122 7.5 Muon Selection

1123 Muons used in the analysis are formed by matching reconstructed inner detector tracks with either
 1124 a complete track or a track-segment reconstructed in the muon spectrometer (MS), called Chain 2

1125 muons. The muons have $p_T > 10$ GeV and satisfy $|\eta| < 2.5$. The muon track are required to be a good
 1126 quality combined fit of inner detector hits and muon spectrometer segments, unless the muon is not
 1127 fiducial to the inner detector, $|\eta| > 2.47$. Muons with inner detector tracks are further required to
 1128 pass standard inner detector track hit requirements [52].

1129 As with electrons, muons are required to be isolated from additional tracking or calorimeter energy:
 1130 $\text{ptcone20}/P_T < 0.1$, $\text{Etcone20}/E_T < 0.1$. A cell-based Etcone20/ P_T relative isolation variable is used.
 1131 A pile-up energy subtraction based on the number of reconstructed vertices in the event is applied.
 1132 The subtraction is derived from a Z boson control sample.

1133 The muons must also originate from the primary vertex and have impact parameter requirements,
 1134 d_0 significance < 3 , and $z_0 \sin \theta < 0.1$ cm, similar to the electrons.

1135 The muon selection is summarized in Table 7.2.

1136 7.6 Jet and b-Tagged Jet Selection

1137 Jets are reconstructed in the calorimeter using the anti- k_t [53] algorithm with a distance parameter of
 1138 0.4 using locally calibrated topologically clusters as input (LC Jets). Since the jets in the $t\bar{t}H$ signal
 1139 mostly arise from the decay massive resonances and not radiation, they are expected to be central
 1140 and high energy. Jets must have $p_T > 25$ GeV and $|\eta| < 2.5$.

1141 Jets must also pass loose quality requirement, ensuring the proper functioning of the calorimeter
 1142 at the time of data taking. Jets near a hot calorimeter cell in data periods B1/B2 are rejected. The
 1143 local hadronic calibration is used for the jet energy scale, and ambient energy corrections are applied
 1144 to account for energy due to pileup.

1145 Jets within $|\eta| < 2.4$ and $p_T < 50$ GeV are further required to be associated with the primary
 1146 vertex. The the fraction of track p_T associated with the jet that comes from the primary vertex, must
 1147 exceed 0.5 (or there must be no track associated to the jet). This requirement rejects jets that arise
 1148 from pile-up vertices.

1149 B-jets are tagged using a Multi-Variate Analysis (MVA) method called MV1 and relying on infor-
 1150 mation of the impact parameter and the reconstruction of the displaced vertex of the hadron decay
 1151 inside the jet[54]. The output of the tagger is required to be above 0.8119 which corresponds to a
 1152 70% efficient Working Point (WP).

¹¹⁵³ **7.7 Object Summary and Overlap**

¹¹⁵⁴ Since many fully identified objects may be reconstructed as two different objects, an overlap removal
¹¹⁵⁵ procedure is applied. Electrons within $\Delta R < 0.1$ of muons are rejected in favor of the muon. Jets
¹¹⁵⁶ within $\Delta R < 0.3$ of electrons are then removed. Finally, muons within $\Delta R < 0.04 + 10\text{GeV}/p_T$ of
¹¹⁵⁷ jets are rejected, as these muons are thought to arise from jet fragmentation.

Parameter	Values	Remarks
Electrons		
p_T	$> 10 \text{ GeV}$	
$ \eta $	< 2.47 veto crack	
ID	Very Tight Likelihood	< 1.37 for 2ℓ SS channel
Isolation	$E_{\text{T}}\text{Cone}20/p_T, p_{\text{T}}\text{Cone}20/p_T < 0.05$	
Jet overlap removal	$\Delta R > 0.3$	
$ d_0^{sig} $	$< 4\sigma$	
$z_0 \sin\theta$	$< 1 \text{ cm}$	
Muons		
p_T	$> 10 \text{ GeV}$	
$ \eta $	< 2.5	
ID	Tight	
Jet overlap removal	$\Delta R > 0.04 + 10 \text{ GeV}/p_T$	
Isolation	$E_{\text{T}}\text{Cone}20/p_T, p_{\text{T}}\text{Cone}20/p_T < 0.1$	< 0.05 for 2 leptons
$ d_0^{sig} $	$< 3\sigma$	
z_0	$< 1 \text{ cm}$	
Jets		
p_T	$> 25 \text{ GeV}$	
$ \eta $	< 2.5	
JVF	$\text{JVF} > 0.5$ or no associated track or $p_T > 50 \text{ GeV}$	
b-Tag	MV1 70% operating point	

Table 7.2: Object identification and selection used to define the 5 channels of the multi-lepton $t\bar{t}H$ analysis. Some channels use a sub-sample of objects as explained in the Remarks column.

1158

CHAPTER 8

1159

Background Estimation

1160 The $t\bar{t}H$ multi-lepton signal regions discussed in Chapter 5 are contaminated by background con-
1161 tributions at a similar order of magnitude to the signal. The dominant background for each region
1162 is $t\bar{t}V$. Sub-dominant but important backgrounds include the production of vector boson pairs in
1163 associated with jets and b-quark jets (VV) and $t\bar{t}$ production with a jet misidentified as a lepton
1164 (fakes). The 2ℓ SS regions possesses a unique background of charge misidentification from Z and top
1165 events. The methods for estimating these backgrounds are discussed in this chapter. Monte Carlo
1166 simulation is used for the prompt $t\bar{t}V$ and VV contributions. The non-prompt backgrounds from $t\bar{t}$
1167 jet-misidentification and charge-misidentification are estimated using data-driven methods. Table 8.1
1168 provides a summary of the $t\bar{t}H$ signal and background expectation for each of the signal regions,
1169 including the data-driven estimates discussed in this section.

1170 **8.1 Vector Boson (W^\pm , Z) production in association with top quarks:**

1171 $t\bar{t}V$, tZ

1172 Production of top quarks plus vector boson is an important background in all multi-lepton channels.
1173 A large part of the $t\bar{t}V$ component, arising from on-shell $Z \rightarrow \ell\ell$, can be removed via a Z mass veto
1174 on like-flavor, opposite sign leptons. However the $Z \rightarrow \tau\tau$ and γ^* components remain. The $t\bar{t}W^\pm$ and
1175 tZ processes generally require extra jets to reach the multiplicity of our signal regions, as such it is
1176 important to ascertain uncertainties associated with QCD radiation. We consider uncertainties on both
1177 the $t\bar{t}W^\pm$ and $t\bar{t}Z$ production cross-sections of these two processes and event selection efficiencies in
1178 the signal regions. The latter is sensitive to the NJet modelling in the MC. We assess the size of these
1179 uncertainties by investigating the effects of the choice of the factorization (μ_F) and renormalisation
1180 μ_R scales and PDF sets.

Table 8.1: Expected number of signal and background events in 2ℓ SS, 3ℓ and 4ℓ signal regions. For data-driven backgrounds, Monte Carlo only numbers are given for reference. The expected sensitivity $\frac{s}{\sqrt{s+b}}$ is provided for data-driven, MC only and for the inclusion of ad-hoc systematic uncertainties (20% for $t\bar{t}V$ and 30% for $t\bar{t}$ fake).

	Same-sign				4 leptons	
	$e^\pm e^\pm$	≥ 5 jets	$e^\pm e^\pm$	4 jets	Z enriched	Z depleted
	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$	$e^\pm \mu^\pm$	$\mu^\pm \mu^\pm$		
tH	0.73 ± 0.03	2.13 ± 0.05	1.41 ± 0.04	0.44 ± 0.02	0.16 ± 0.03	0.74 ± 0.03
tV	2.60 ± 0.13	7.42 ± 0.17	5.01 ± 0.16	3.05 ± 0.13	8.39 ± 0.24	5.79 ± 0.20
tZ						
VV	0.48 ± 0.25	0.37 ± 0.23	0.68 ± 0.30	0.77 ± 0.27	1.93 ± 0.80	0.54 ± 0.30
t, tX (MC)	1.31 ± 0.67	2.55 ± 0.84	1.76 ± 0.67	4.99 ± 1.19	8.19 ± 1.41	3.70 ± 1.03
$Z+jets$ (MC)	0.16 ± 0.16	0.28 ± 0.20	0.12 ± 0.12	1.37 ± 0.78	0	0.23 ± 0.23
fake leptons (DD)	2.31 ± 0.97	3.87 ± 1.01	1.24 ± 0.41	3.43 ± 1.38	6.82 ± 1.63	2.38 ± 0.78
Q misid (DD)	1.10 ± 0.09	0.85 ± 0.08	—	1.82 ± 0.11	1.39 ± 0.08	—
Tot Background (fake MC)	4.56 ± 1.17	10.62 ± 1.54	7.57 ± 1.31	10.18 ± 2.43	18.51 ± 2.54	10.26 ± 1.82
Tot Background (fake DD)	6.49 ± 1.04	12.51 ± 1.04	6.93 ± 0.52	9.07 ± 1.42	18.53 ± 1.83	8.71 ± 0.88
s/\sqrt{b} (fake MC)	0.34	0.65	0.51	0.14	0.27	0.23
s/\sqrt{b} (fake DD)	0.29	0.60	0.54	0.15	0.27	0.25
$s/\sqrt{b} \oplus 0.3fake(MC) \oplus 0.2t t V$	0.33	0.58	0.47	0.12	0.22	0.21
$s/\sqrt{b} \oplus 0.3fake(DD) \oplus 0.2t t V$	0.27	0.53	0.50	0.14	0.23	0.23

Table 8.2: NLO cross section and theoretical uncertainty calculations derived from MadGraph5+aMC@NLO.

Process	$\sigma_{NLO} [fb]$	Scale Uncertainty [%]		PDF Uncertainty [%]		Total symmetrized uncertainty [%]
$t\bar{t}W^+$	144.9	+10	-11	+7.7	-8.7	13.3
$t\bar{t}W^-$	61.4	+11	-12	+6.3	-8.4	13.6
$t\bar{t}Z$	206.7	+9	-13	+8.0	-9.2	14.0
tZ	160.0	+4	-4	+7	-7	8.0
tZ	76.0	+5	-4	+7	-7	8.6

1181 Monte Carlo events for these processes are generated with MadGraph 5 and showered with Pythia
 1182 6. $t\bar{t}W^\pm$ events are generated with up to two extra partons at matrix element level, while for $t\bar{t}Z$ up
 1183 to one extra parton at matrix-element level is produced. The tZ process is simulated without extra
 1184 partons. The next-to-leading-order (NLO) cross sections are implemented by applying a uniform
 1185 k -factor to the leading-order (LO) events for each process. For $t\bar{t}Z$, the k -factor is determined by
 1186 comparing LO and NLO cross sections for on-shell Z production only and then applied to the off-shell
 1187 signal regions.

1188 The $t\bar{t}V$ uncertainties are calculated using the internal QCD scale and PDF re-weighting that is
 1189 available with MadGraph5+aMC@NLO. The prescription for the scale envelope is taken from [74]:
 1190 the central value $\mu = \mu_R = \mu_F = m_t + m_V/2$ and the uncertainty envelope is $[\mu_0/2, 2\mu_0]$. The
 1191 PDF uncertainty prescription used is the recipe from [75]: calculate the PDF uncertainty using the
 1192 MSTW2008nlo [73] PDF for the central value and then the final PDF uncertainty envelope is derived
 1193 from three PDF error sets each with different α_S values (the central value and the upper and lower
 1194 90% CL values). The final NLO cross section central values and uncertainties are given in Table 8.2.

1195 The tZ process is normalized to NLO based on the calculation in Ref. [76]. Here the scales are set
 1196 to $\mu_0 = m_t$ and the scale variations are by a factor of four; the scale dependence is found to be quite
 1197 small.

1198 8.1.1 $t\bar{t}Z$ Validation Region

1199 Unlike $t\bar{t}W^\pm$, a $t\bar{t}Z$ validation region can be obtained by simply inverting the veto on same-flavor
 1200 opposite-sign (SFOS) lepton pairs near the Z pole in the 3 lepton signal region. This region thus
 1201 requires 3 leptons (with momentum and identification cuts discussed in Chapter 7, at least one
 1202 opposite sign, same-flavor pair of leptons within 10 GeV of the Z mass, and either 4 jets and at least 1
 1203 b-tagged jet or exactly 3 jet and 2 or more b-tagged jets. The resulting region has low statistics and is

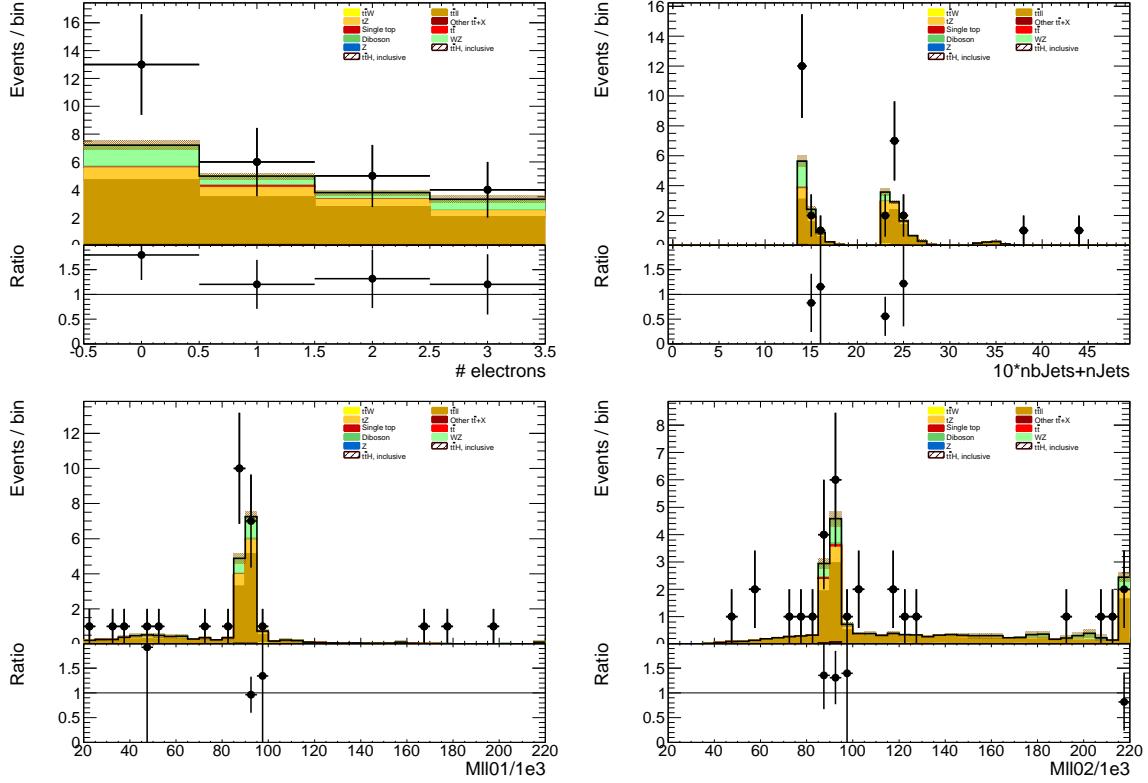


Figure 8.1: Data/MC comparison plots for $t\bar{t}Z$ control region A (≥ 4 jets, ≥ 1 b -tag and 3 jets, ≥ 2 b -tag). In all plots, the rightmost bin contains any overflows. Top left: number of electrons. Top right: 10^* the number of b -tags + the total number of jets. Middle left: the invariant mass of the (0,1) lepton pair (see the text for the definition of the lepton ordering). Middle right: the invariant mass of the (0,2) lepton pair.

not used as a control region but is instead used as a validation to demonstrate that the normalization uncertainty, discussed above, is properly evaluated.

The region defined by this is predicted to be 67% $t\bar{t}Z$, 17% $W^\pm Z$, and 13% tZ . We predict 19.3 \pm 0.5 events and observe 28, giving a observed-to-predicted ratio of $1.45 \pm 0.27 \pm 0.03$, where the errors are from data and simulation statistics, respectively. Given the large errors, the region is still in agreement with the predictions to within 1-1.5 σ . Distributions of various variables are shown in Fig. 8.1.

1211 8.2 Di-boson Background Estimation: $W^\pm Z, ZZ$

1212 $W^\pm Z$ and ZZ di-boson production with additional and b-tagged jets constitute small contributions
 1213 to the 3ℓ and 4ℓ channels. For the 3ℓ case $W^\pm Z$ comprises $\sim 10\%$ of the total background, while for
 1214 the 4ℓ case ZZ contribution accounts comprises $\sim 10\%$ of the total background. Because of the small
 1215 size of these contributions, each of the above processes can be assigned a non-aggressive uncertainty
 1216 based on similar previous analyses with ATLAS and cross-checked with data validation regions and
 1217 MC truth studies.

1218 Both $W^\pm Z$ and ZZ production have been studied by ATLAS [77][78], but neither process has
 1219 been investigated thoroughly in association with multiple jets and b-quark jets. However, single
 1220 boson production with b-quark jets has been investiaged. Both $W + b$ [79] and $Z + b$ [80] production
 1221 in 7 TeV data have been shown to agree with MC models to within 20-30%.

1222 A single W produced in association with b-tagged jets possesses a similar topology to the $W^\pm Z + b$
 1223 process at a different energy scale and has been shown to be dominated by c mis-tags and b-jets from
 1224 gluon splitting and multiple parton interaction. The $W + b$ analysis, referenced above, uses Alpgen
 1225 MC with Herwig PS modeling, only provides results to 1 additional jet, and uses the CombNN tagger
 1226 (we use MV1). Its results are therefore not directly comparable to this $t\bar{t}H$ analysis (where $W^\pm Z$ is
 1227 modeled using Sherpa with massive c and b quarks). $Z + b$ production originates from slightly different
 1228 diagrams than $ZZ + b$, but the sources of the b-tags are similar. The 7 TeV analysis, referenced above,
 1229 provides results with Sherpa MC with an agreement of $\sim 30\%$. However, it also used the CombNN
 1230 tagger instead of MV1. Beause of the differences of the 2011 single boson analyses (type of tagger
 1231 used, type of MC and tunes used), we would like to verify the general 20-30% level of agreement in
 1232 2012 data with the simulation and tagger used in the $t\bar{t}H$ analysis: Sherpa MC, 2012 tunes, MV1.
 1233 With the data skims available to use we are able to do this in the $Z + b$ region but not the $W + b$.

1234 Figure 8.2 shows the spectrum of the number of reconstructed and selected jets (NJet) in a $Z + b$
 1235 validation region, defined by 2 tight-isolated leptons within 10 GeV of the Z mass and with at least
 1236 one b-tagged jet, using the $t\bar{t}H$ analysis definitions. The level of agreement in this region confirms at
 1237 the 30% level seen in the 7 TeV analysis, discussed above.

1238 In the following two sections, we assess the truth origin of jets in the $W^\pm Z + b$ and $ZZ + b$ regions
 1239 and leverage data/MC agreement where we can. We see that the data allows us to constrain the
 1240 $W^\pm Z$ to 50%. We claim this 50% as a systematic. The 20-30% agreement in the single boson regions
 1241 above bolsters our confidence in this number.

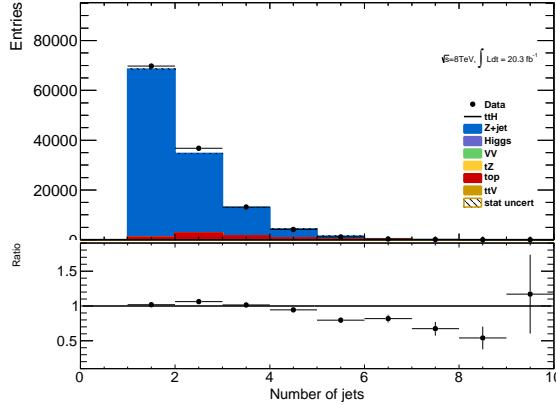


Figure 8.2: NJet spectrum for 2 tight-isolation leptons with 1 b-tagged jet (MV1_70)

8.2.1 $W^\pm Z$ Normalization Uncertainty

The $t\bar{t}H$ analyses has two validation regions to test the Sherpa agreement with data for $W^\pm Z$: one inclusive 3 lepton region, using the three-lepton channel object and p_T cuts and a $W^\pm Z + b$ region with 1 b-tagged jet and a requirement that at least one SFOS pair have an invariant mass within 10 GeV of the Z mass. The region with fewer than 4 jets is $W^\pm Z$ dominated. Figure 8.3 shows kinematic variables for the inclusive region. The overall data normalization is $\sim 10\%$ higher than MC, but this will be well within our systematic uncertainty. The NJet shape shows good agreement across the full spectrum, giving confidence about the Sherpa high NJet SR extrapolation. Figure 8.4 shows NJet spectrum for the $W^\pm Z + b$ validation region with agreement with in statistical uncertainties. The region has low statistics and around $\sim 60\%$ purity and statistical analysis of the region suggests that a 50% normalization error on the $W^\pm Z$ component is enough to cover any possible mismodelings, especially in higher NJet bins, which are closer to the signal regions.

We also examine the $W^\pm Z$ truth origins of the b-jet in the $W^\pm Z + b$ validation region (VR) and the signal region using MC to assess the validity of the extrapolation from the VR to the SR and to confirm the similarity in jet origin to the single boson analyses, references above. The flavour of the closest matching truth particle ($p_T > 5$ GeV, after FSR) in ΔR determines the true-jet flavor. If there are no quarks, taus or gluons within ΔR of 0.3, the label defaults to light. Table 8.3 shows the origin fraction of b-tagged jets in the various $W^\pm Z + b$ VRs and the SR. If there are two b-tagged jets, the highest p_T is used, but this is a small fraction of the number of b-tags. As expected the c and b contributions dominate, as was the case with the 2011 single boson analyses referenced above. It is important also that the VR has similar composition to the SR. There is a small dependence on

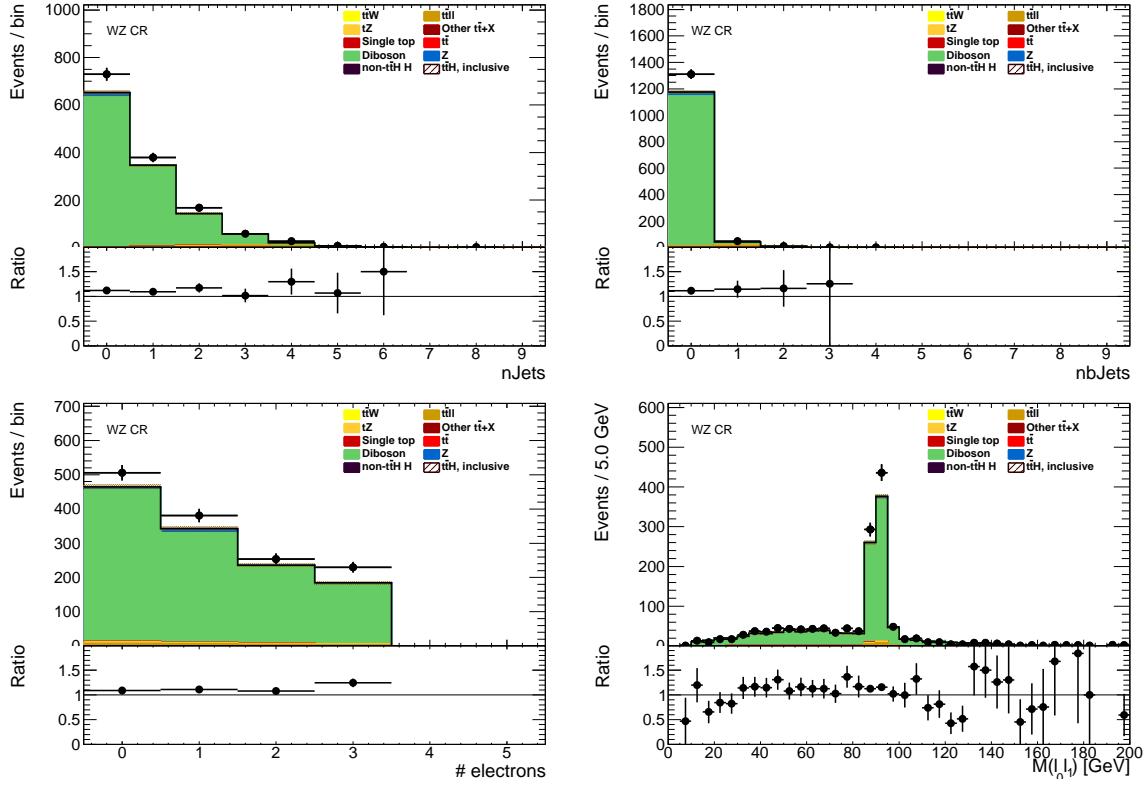


Figure 8.3: 3 lepton $W^\pm Z$ validation using the $t\bar{t}H$ lepton identification and momentum cuts: mass, number of jet and flavor variables

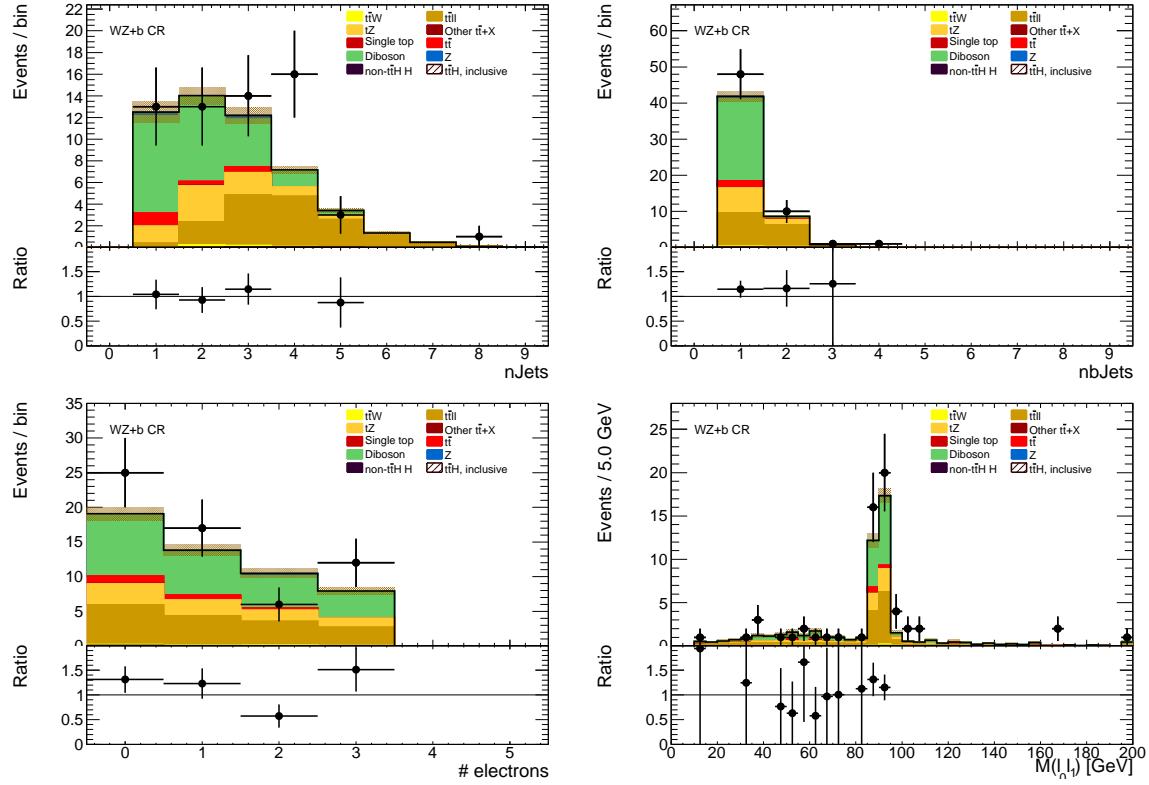
1263 the number of jets.

	Bottom	Charm	Light
$W^\pm Z + b$ VR 1 Jet	0.25 ± 0.03	0.54 ± 0.04	0.20 ± 0.03
$W^\pm Z + b$ VR 2 Jet	0.34 ± 0.04	0.52 ± 0.06	0.13 ± 0.03
$W^\pm Z + b$ VR 3 Jet	0.40 ± 0.07	0.41 ± 0.07	0.18 ± 0.04
$3l$ SR	0.43 ± 0.14	0.38 ± 0.17	0.18 ± 0.11

Table 8.3: Truth origin of highest energy b-tagged jet in the $W^\pm Z + b$ VR and $3l$ SR

1264 8.2.2 ZZ Normalization Uncertainty

In order to investigate the MC agreement with data in the ZZ case, two validation regions similar to the $W^\pm Z$ case are defined. First, a 4 lepton ZZ region is constructed using the object selections for the 4-lepton channel and requiring exactly two pairs of opposite sign-same flavor leptons with a di-lepton invariant mass within 10 GeV of the Z mass. Additionally, the $ZZ + b$ process is investigated directly using a similar validation region which again requires exactly two Z -candidate lepton pairs

Figure 8.4: $W^\pm Z + b$ validation region: NJet, NElec, and Mass Variables

as well as at least 1 b-tagged jet. Some kinematic distributions are shown in Figures 8.5 and 8.6, and particular attention should be paid to the NJet spectrum, which shows good data-MC agreement in the high-jet bins, with a slight discrepancy in the 1-jet bin. The agreement for the region with at least 2 jets yields confidence in the NJet MC modeling in this region which lies close to the 4-lepton signal region.

Based on the study of the ZZ and $ZZ + b$ validation regions and the overall agreement noted with the $Z + b$ analysis, we expect a similar error to $W^\pm Z$ to be appropriate in the ZZ case. A truth origin study is undertaken in MC to demonstrate a similar b-jet origin to the $W^\pm Z$ case. The true origin of the leading (highest energy) b-tagged jet is shown in Table 8.4 for the 4-lepton signal region as well as the $ZZ + b$ validation region described above divided into jet bins. As it was in the $W^\pm Z$ case above, the true origin of the b-jet in $ZZ + b$ is dominated by c and b. Taking this study in tandem with the results from the $W^\pm Z$ investigation, it is appropriate to take the central value of the $ZZ + b$ background contribution in the 4-lepton SR from MC and to assign an overall systematic of 50% in order to account for the MC modeling limitations.

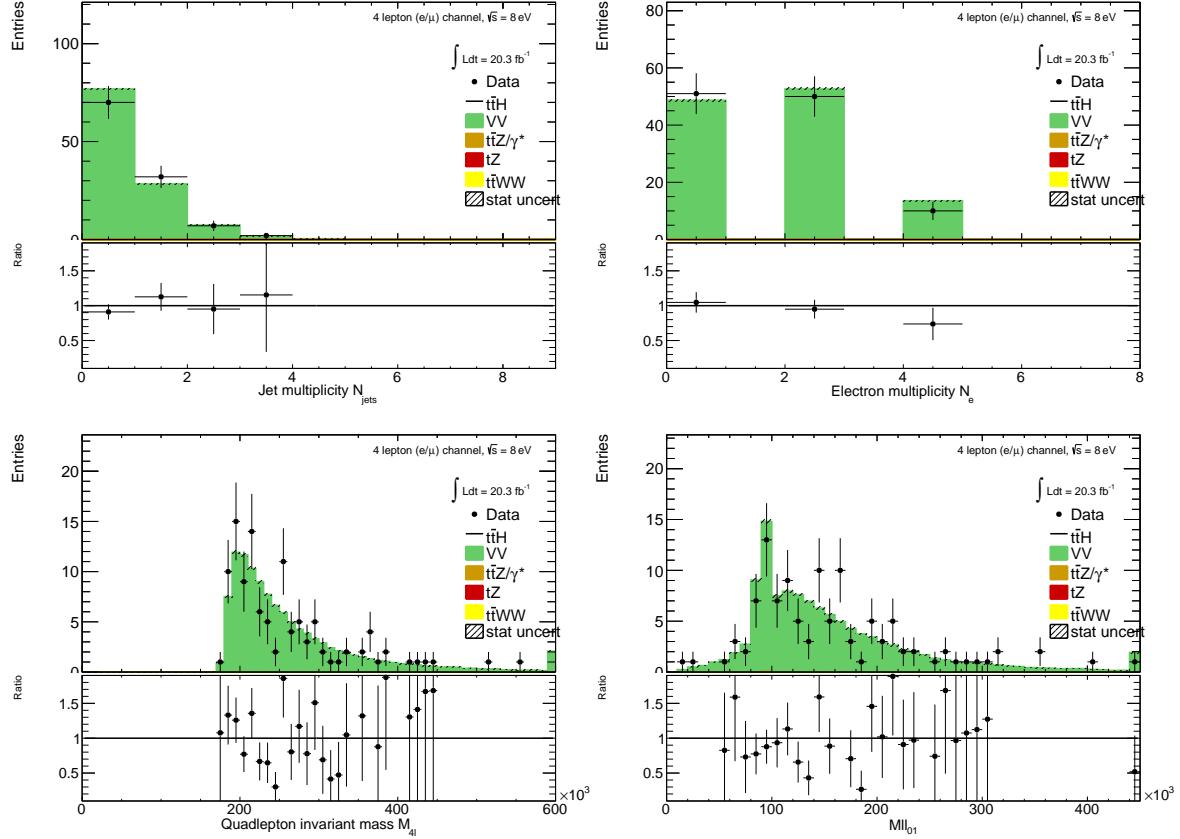


Figure 8.5: Jet-inclusive 4-lepton ZZ validation region using the $t\bar{t}H$ lepton identification and momentum cuts

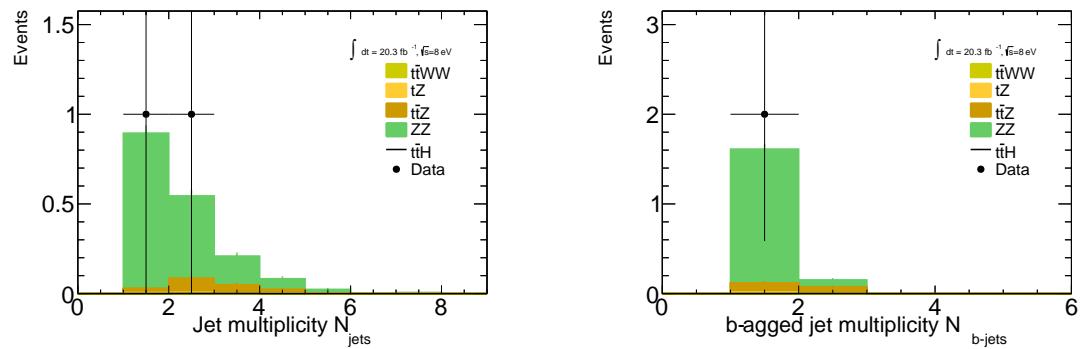


Figure 8.6: $ZZ + b$ validation region using the $t\bar{t}H$ lepton identification and momentum cuts

	Bottom	Charm	Light
$ZZ + b$ VR 1 Jet	0.56 ± 0.03	0.24 ± 0.01	0.20 ± 0.01
$ZZ + b$ VR 2 Jet	0.52 ± 0.05	0.25 ± 0.02	0.23 ± 0.02
$ZZ + b$ VR 3 Jet	0.53 ± 0.11	0.25 ± 0.08	0.22 ± 0.07
$4l$ SR	0.34 ± 0.15	0.42 ± 0.16	0.24 ± 0.10

Table 8.4: Truth origin of highest energy b-tagged jet in the $ZZ + b$ VR and $4l$ SR

8.3 Charge-Misidentification Background

Charge-misidentification contributes to the background for 2ℓ SS case and only for flavor channels, which include electrons. The same-sign requirement is essential in removing large SM opposite sign backgrounds, but because of their size even small charge misidentification rates result in contamination in same-sign regions. For the 2ℓ SS signal regions, charge-misidentification background arise primarily from $t\bar{t}$ di-lepton events with a smaller contribution from leptonic Z decays.

In general, charge-misidentification can arise in two ways. The first occurs for ultra-high energy electrons and muons, which leave tracks in the detector that are too straight for the fit to determine the direction of curvature with high confidence. This type of charge misidentification is not a concern to the $t\bar{t}H$ multi-lepton analysis, as most of the leptons have transverse momentum < 150 GeV. The second source of charge misidentification is from 'tridents', which only occurs for electrons, because their low mass allows for high rate bremsstrahlung in the detector material. In some cases, after an electron releases a photon through bremsstrahlung, the photon may convert nearby resulting in three electron tracks. The reconstruction algorithms may sometimes match the wrong track to the calorimeter energy deposit, resulting in a possible charge misidentification. As discussed in the selection, tight track-cluster geometric and energy matching requirements are applied on the electron candidates to reduce the overall rate and the electron acceptance is narrowed to ($|\eta| < 1.37$), since most of the material is concentrated more forward in the detector.

We estimate the contribution of charge-misidentification events in our 2ℓ SS signal regions and relevant control regions by applying a weight per electron in the OS region with otherwise identical cuts. The weight is related to the charge-misidentification rates. We measure these rates using a likelihood method in the OS and SS $Z \rightarrow ee$ control region in data. The rate measured from these control regions is binned in electron p_T and η , to account for dependencies in these variables. The method, validations and associated errors are discussed in detail in the following sub-sections.

1308 **8.3.1 Likelihood Method**

1309 The number of reconstructed same-sign (N_{ss}) $Z \rightarrow ee$ events is related to total number of produced
 1310 $Z \rightarrow ee$ (N) through factors related to the charge misidentification rate, ϵ :

$$N_{ss}^{ij} = N^{ij}(\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j) \quad (8.1)$$

1311 where ϵ_i and ϵ_j are the charge misidentification rates for each electron separately. If we drop terms
 1312 quadratic in ϵ , we have:

$$N_{ss}^{ij} = N^{ij}(\epsilon_i + \epsilon_j). \quad (8.2)$$

1313 Although it is impossible to know event-by-event which electron's charge was misidentified, we can
 1314 use a likelihood method over the whole Z sample to measure how ϵ depends on the electron p_T and
 1315 $|\eta|$. As illustration, we first consider the case, where ϵ depends on only one variable, $|\eta|$, and then
 1316 generalize to the two-dimensional case of $|\eta|$ vs p_T .

1317 N_{ss}^{ij} is described by a Poisson distribution:

$$f(k, \lambda) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (8.3)$$

1318 where k is the observed number of occurrences of the event, i.e. $k = N_{ss}^{ij}$, and λ is the expected
 1319 number, i.e. $\lambda = N^{ij}(\epsilon_i + \epsilon_j)$. Thus, the probability for an observed number of same-sign Z events
 1320 given the sample size and charge misidentification rates is expressed by:

$$P(N_{ss}^{ij}|N^{ij}, \epsilon_i, \epsilon_j) = \frac{[N^{ij}(\epsilon_i + \epsilon_j)]^{N_{ss}^{ij}} e^{-N^{ij}(\epsilon_i + \epsilon_j)}}{N_{ss}^{ij}!}. \quad (8.4)$$

1321 The likelihood L for all the events is obtained by evaluating all the $|\eta|$ combinations:

$$L(\epsilon|N_{ss}, N) = \prod_{i,j} \frac{[N^{ij}(\epsilon_i + \epsilon_j)]^{N_{ss}^{ij}} e^{-N^{ij}(\epsilon_i + \epsilon_j)}}{N_{ss}^{ij}!}. \quad (8.5)$$

1322 In this process, the $-\ln L$ is used in order to simplify and make easier the minimization. Terms which
 1323 do not depend on the rates ϵ_i and ϵ_j are removed in this step. This way, the final function to minimize
 1324 is given by the following expression:

$$-\ln L(\epsilon|N_{ss}, N) \approx \sum_{i,j} \ln[N^{ij}(\epsilon_i + \epsilon_j)]N_{ss}^{ij} - N^{ij}(\epsilon_i + \epsilon_j). \quad (8.6)$$

1325 The likelihood can be easily extended to depend on the charge misidentification rates as a function
 1326 of two parameters. The probability to find a same-sign event given the rates for each electron is
 1327 $(\epsilon_{i,k} + \epsilon_{j,l})$, where the two indices represent binned $|\eta|$ - and p_T -dependence. Thus, the Eq. 8.6

1328 transforms into

$$-\ln L(\epsilon | N_{ss}, N) \approx \sum_{i,j,k,l} \ln[N^{ij,kl}(\epsilon_{i,k} + \epsilon_{j,l})] N_{ss}^{ij,kl} - N^{ij,kl}(\epsilon_{i,k} + \epsilon_{j,l}). \quad (8.7)$$

1329 We use events selected within the Z peak using the $t\bar{t}H$ electron object cuts. The events are stored
 1330 in two matrices: one for the same-sign events $N_{ss}^{ij,kl}$, and the other one for all events $N^{ij,kl}$. Small
 1331 backgrounds need to be subtracted. The background subtraction is done using a simple side-band
 1332 method. This method consists in dividing the Z invariant mass in three regions, i.e. A , B and C ,
 1333 where B is the central region corresponding to the Z peak. The number of events is counted in the
 1334 regions on the sides of the peak, i.e. n_A and n_C , and removed from the total number of events in the
 1335 peak region B , n_B . This way, the number of signal events N_Z is given by

$$N_Z = n_B - \frac{n_A + n_C}{2}. \quad (8.8)$$

1336 Once the background has been subtracted, the likelihood is minimized for the 2D matrix of ϵ bins.
 1337 Knowing ϵ as a function of $|\eta|$ and p_T for any single electron, it is now possible to estimate the number
 1338 of same-sign events from the number of opposite sign events in any sample:

1339 • $N^{ss} = \frac{\epsilon_i + \epsilon_j - 2\epsilon_i\epsilon_j}{1 - \epsilon_i - \epsilon_j + 2\epsilon_i\epsilon_j} N^{os}$ for ee channels

1340 • $N^{ss} = \frac{\epsilon}{1 - \epsilon} N^{os}$ for the $e\mu$ channels

1341 8.3.2 Results

1342 The charge misidentification rate is calculated in 7 $|\eta|$ bins [0.0, 0.6, 1.1, 1.37, 1.52, 1.7, 2.3, 2.47] by
 1343 4 p_T bins [15, 60, 90, 130, 1000] GeV. For p_T bins above 130 GeV, the Z dataset becomes too small
 1344 and the rates are calculated using $t\bar{t}$ MC, scaled by the data-MC ratio of the rates in the lower p_T
 1345 bins, [90-130] GeV. Figure 8.7 shows the extracted rates in all bins.

1346 As a cross-check, we apply the full method to the Z MC samples (extracting rates via a likelihood
 1347 fit and applying them to opposite sign events) and compare to the MC predicted number of same-sign
 1348 events. The invariant mass of the Z from our charge misidentification and directly from the MC
 1349 can be seen on Figure 8.8. In the simulated Z samples, the number of same-sign Z events is 5 049
 1350 while the estimation is $5 031^{+375}_{-365}$. The uncertainties combine both statistical systematic uncertainties,
 1351 which are discussed in depth below. The validation gives compatible results within uncertainties.

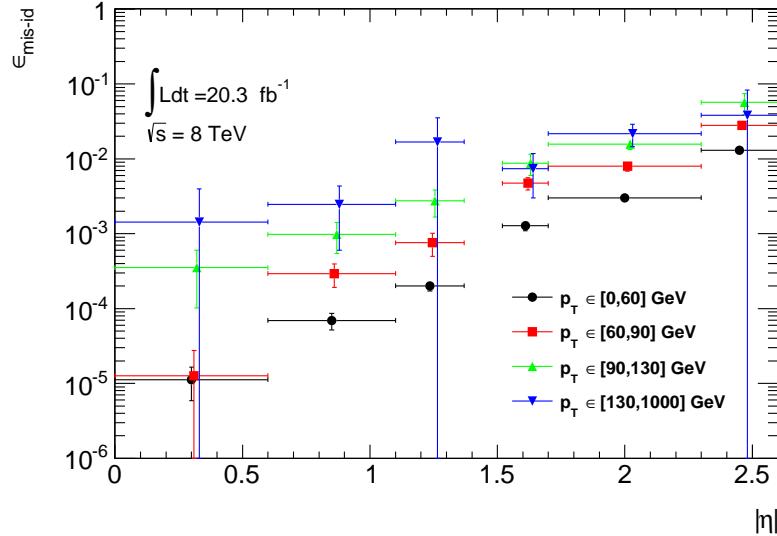


Figure 8.7: Electron charge misidentification rates measured in data with the likelihood method on Z events (black points, red squares and blue triangles) as a function of $|\eta|$ and parametrized in p_T . The full 2012 dataset has been used to estimate the rates below 130 GeV. Above this value, the charge misidentification rates have been estimated by extrapolating the rates in the region where the $p_T \in [90, 130]$ GeV with a p_T dependent factor extracted from simulated $t\bar{t}$ events (green triangles). Statistical and systematic uncertainties have been included in this plot.

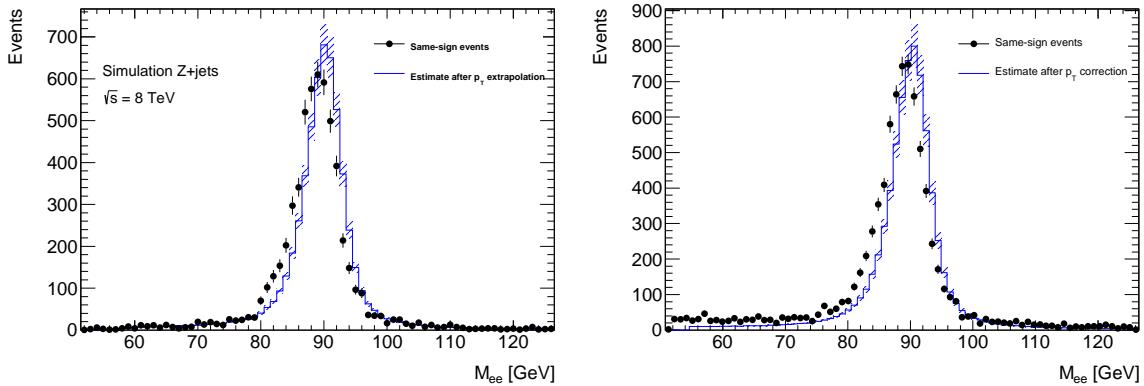


Figure 8.8: Closure test on simulated $Z \rightarrow e^+e^-$ events (a) and data (b). The black circles show the distribution of same-sign events while the blue histograms show the distribution of the re-weighted opposite-sign events together with the statistical and systematic uncertainties. The distributions are not expected to overlay exactly, due to the loss of energy of the trident electrons for the same-sign peak.

8.3.3 Systematic and Statistical Uncertainties

1353 Statistical uncertainties dominate the combined uncertainty on the charge misidentification estimate.

1354 The statistical uncertainties come primarily from the size of the Z same-sign sample in data and are

especially large for central, material-poor regions where the charge misidentification rate is extremely low. Additional systematic uncertainties are included for a comparison between the positron and electron rate, the per-bin MC closure test discussed above, and for the effect of varying the invariant mass window used for the background subtraction for three different cases. Figure 8.9 shows the relative uncertainties for all rate bins.

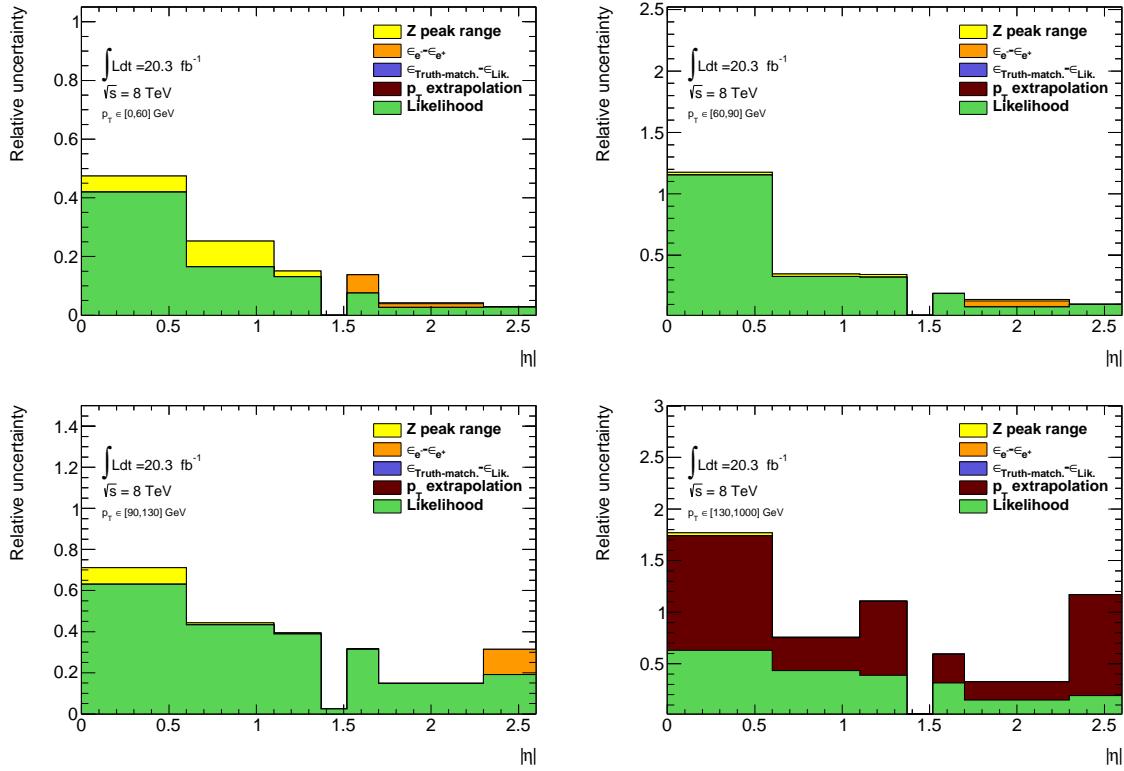


Figure 8.9: Relative systematic uncertainty contributions on the charge misidentification rate, for different bins in p_T and $|\eta|$.

We apply the rates to estimate the charge misidentification background in the 2ℓ SS signal regions, and find $\sim 25\%$ contamination in the $e^\pm e^\pm$ regions and a $\sim 10\%$ contribution to the $e^\pm \mu^\pm$ regions with a 10% systematic error overall. The low overall error can be attributed to the fact that the statistical error is lowest where the bulk of charge misidentifications occur. The charge flip contribution measured in the signal regions from this method is detailed in Table 8.1.

1365 8.4 Fake Lepton Backgrounds

1366 Fake Leptons, from the misidentification of jets as either electrons or muons, primarily arise from $t\bar{t}$
 1367 and single top processes in the 2ℓ SS, 3ℓ and 4ℓ channels. Smaller contributions come from $Z+jet$
 1368 events. Fake backgrounds are sub-dominant but important in the 2ℓ SS and 3ℓ channels. They
 1369 are extremely small in the 4ℓ channels. Truth studies suggest that these misidentified leptons arise
 1370 overwhelmingly from b-quark initiated jets. The general method for estimating fakes in all channels
 1371 is to define a reversed object selection control region (usually isolation) for each lepton flavor with
 1372 otherwise identical signal region selection (N_{CR}^e , N_{CR}^μ). This region is fake-dominated. The total
 1373 number of fake events in these regions are then scaled by transfer factors (θ) to estimate the number
 1374 of fake events of the appropriate flavor in the signal region. The simple formula for determining fakes
 1375 is defined in Equation 8.9.

$$N_{fake} = \theta_e \cdot N_{CR}^e + \theta_\mu \cdot N_{CR}^\mu \quad (8.9)$$

1376 This approach factorizes the background model into two separate measurements. N_{CR} is sensitive
 1377 the overall $t\bar{t}$ production rate, especially in the presence of additional jets from QCD ratio, as well as
 1378 the object-level misidentification of a jet as a lepton. The transfer factors are sensitive to only the
 1379 object level properties of the misidentified jet, and in particular only the variables which are reversed
 1380 in the anti-tight identification.

1381 The transfer factors are obtained obtained in a different way for each channel, due to unique issues
 1382 with statistics and contamination, but each method relies heavily on the data-based control regions
 1383 with fewer jets. Figure 8.10 shows a truth study of the stability of the transfer factor for the 2ℓ SS
 1384 and 3ℓ cases as a function of the number of jets in the event for events with one-b-tagged jet. This
 1385 suggest that the regions with fewer jets are a good model of the fakes in the signal regions with more
 1386 jets and is expected because of the homogeneity of origin of the fakes across all jet bins.

1387 The details of the methods for each channel are discussed in depth in the following sections. For
 1388 all methods, the overall systematic uncertainty on the normalization of the fake estimate is in the
 1389 range 30%-50% and arise primarily from statistics and the closure on assumptions used to obtain the
 1390 transfer factor.

1391 Because these methods do not provide a per-object transfer-factor that depends on the properties
 1392 of the faking object, we must use the MC to model the shapes of the fake kinematic distributions
 1393 in the signal regions. This is not an essential issue, since the analysis only considers only the total
 1394 number of events in each signal region in the final measurement of $t\bar{t}H$ production.

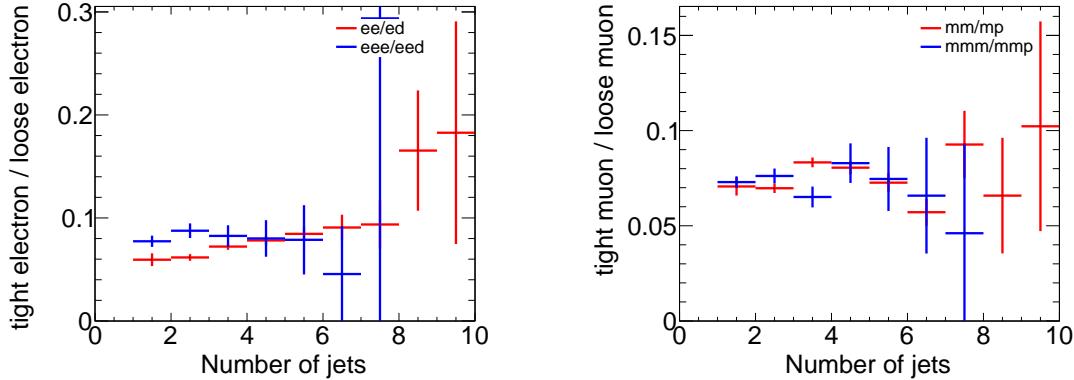


Figure 8.10: Ratios of regions with tight and anti-tight leptons in 2-lepton and 3-lepton channels from $t\bar{t}$ MC. These ratios are the MC calculated transfer factors for each region, i.e. $\theta_e = eee/eed$, ee/ed and $\theta_\mu = mmp/mmm$, mm/mp , where 'd' refers to anti-tight electrons and 'p' refers to anti-tight muons. The transfer factors are seen to be similar in the 2ℓ and 3ℓ cases and stable as a function of the number of jets

1395 8.4.1 2ℓ SS Fakes

1396 The 2ℓ SS fake method follows the procedure outlined in general above. We define anti-tight electron
 1397 and muon control regions with reversed particle identification criteria for each signal region, including
 1398 the 6 flavor and jet-counting sub regions. The anti-tight muon and electron criteria are provided
 1399 below:

- 1400 • **anti-tight electron (d):** fails the verytight likelihood operating point, but still passes the
 1401 veryloose operating point. fails relative tracking and calorimeter isolation, $E_T^{rel} > 0.05$ and
 1402 $p_T^{rel} > 0.05$.
- 1403 • **anti-tight muon (p):** $6 \text{ GeV} < p_T < 10 \text{ GeV}$

1404 The electron and muon transfer factors, θ_e and θ_μ , are calculated in the region with signal region
 1405 selection but fewer jets, $NJet == 2$ or $NJet == 3$ and are defined as the ratio of the number of
 1406 events for two fully identified leptons to the number of events with one fully identified lepton and
 1407 one anti-identified lepton, after the prompt and charge misidentification backgrounds are subtracted.
 1408 Only same-flavor channels are used to ensure that muon and electron transfer factors maybe estimated
 1409 separately: on every region, the prompt and charge-misidentification backgrounds are subtracted from
 1410 the data.

$$\theta_e = \frac{N_{ee}}{N_{ed}} = \frac{N_{ee}^{Data} - N_{ed}^{Prompt SS} - N_{ed}^{QMisId}}{N_{ed}^{Data} - N_{ed}^{Prompt SS} - N_{ed}^{QMisId MC}} \quad (8.10)$$

Process	N(events)
<i>ed</i> \leq 3 jets	
<i>VV</i>	7.13 ± 0.63
<i>V</i> γ	7.55 ± 1.27
<i>t</i> $\bar{t}V, tV$	6.68 ± 0.18
<i>V + jets</i>	59.4 ± 18.51
<i>t</i> $\bar{t}, t + X$	671.26 ± 12.76
<i>t</i> $\bar{t} prompts$	32.97 ± 2.83
Total MC	752.0 ± 22.5
Data	967
Data fakes (Data - prompts)	912.66
<i>ee</i> \leq 3 jets	
<i>VV</i>	3.30 ± 0.42
<i>V</i> γ	1.31 ± 0.65
<i>t</i> $\bar{t}V, tV$	3.96 ± 0.16
<i>V + jets</i>	8.3 ± 8.8
<i>t</i> $\bar{t}, t + X$	11.65 ± 1.67
Charge misID	8.54 ± 0.23
Total MC	28.52 ± 8.96
Data	32
Data fakes (Data - prompts)	14.26
<i>$\mu\mu$</i> \leq 3 jets	
<i>VV</i>	6.27 ± 0.56
<i>V</i> γ	0.06 ± 0.25
<i>t</i> $\bar{t}V, tV$	10.00 ± 0.27
<i>V + jets</i>	1.22 ± 11.78
<i>t</i> $\bar{t}, t + X$	15.4 ± 2.1
Total MC	32.95 ± 2.21
Data	44
Data fakes (Data - prompts)	27.65

Table 8.5: Number of events of the main simulated background processes and of the data in the $e^\pm e^\pm$ and $\mu^\pm \mu^\pm$ channels used for the measurement of θ_e and θ_μ . VV , $V\gamma$, $t\bar{t}V, tV$ and $t\bar{t}$ prompts (or charge misID) are the backgrounds which lead to prompt same-sign dileptons and are subtracted from the data to get a measured number of fakes. Uncertainties are statistical. The numbers labeled Data fakes are used to measure θ .

1411

$$\theta_\mu = \frac{N_{\mu\mu}}{N_{\mu p}} = \frac{N_{\mu\mu}^{Data} - N_{\mu\mu}^{Prompt SS}}{N_{\mu p}^{Data} - N_{\mu p}^{Prompt SS}} \quad (8.11)$$

1412

1413 The 2,3 jet anti-tight regions used in obtaining the transfer factors are shown in Table 8.5 and the
 1414 4,5 jet anti-tight regions, scaled by the transfer factors to get the fake estimates in the SR are shown
 1415 in Figure 8.11. The $t\bar{t}$ and single top MC are included in the plots and tables for reference, although
 1416 they are not used in the measurements.

1417 The transfer factors obtained from the 2 and 3 jet regions are shown in 8.6 with statistical errors
 1418 and propagated systematic errors from the subtraction of relevant backgrounds ($t\bar{t}V$ and charge
 1419 misidentification). The MC values are just for comparison. An additional systematic error is added
 1420 by comparing the transfer factors, obtained from the low jet control region, to those obtained from
 1421 the higher jet signal regions, using $t\bar{t}$ MC. The value of this systematic is about 20 % and can be seen

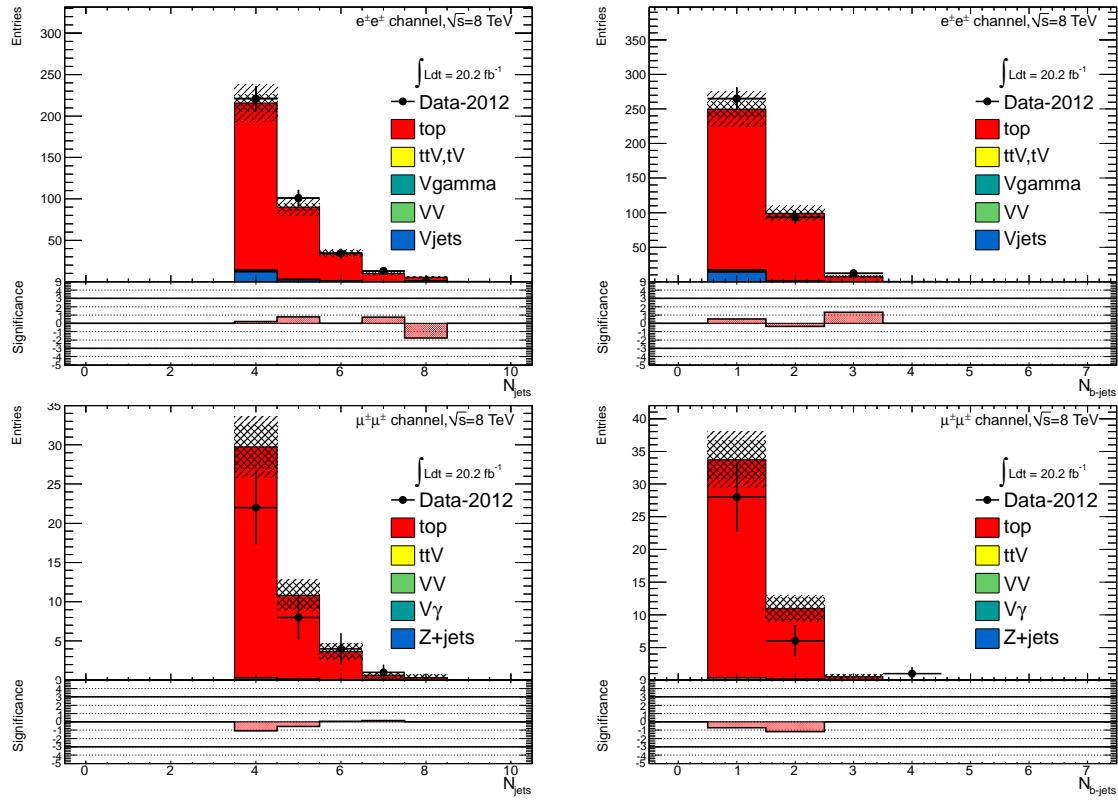


Figure 8.11: 4,5 Jet SS 2ℓ ed (above) and μp control regions with at least one b-tagged jet. After subtraction of prompt and charge misidentification backgrounds, these regions are scaled by the transfer factors, θ_e and θ_μ , to obtain the final number of fake events in the CR. The top MC (red) is used for reference but not in the actual calculation.

Factor	Expected (MC)	Measured (data)
θ_e Rev. Id.	0.0136 ± 0.0062	0.0156 ± 0.0062
θ_μ Rev. p_T	0.078 ± 0.012	0.1156 ± 0.0288

Table 8.6: Expected and measured values of the θ factors. NEEDS TO BE UPDATED FOR 2,3

in the above Figure 8.10. The overall systematic uncertainties and contribution from each source in all of the sub-channels of the signal region are shown in Table 8.6 and the final contribution of fake events to the signal region are show in Table 8.1 found at the beginning of the chapter.

8.4.2 3ℓ Fakes

The 3ℓ fake method follows the same general strategy as the 2ℓ SS case. Transfer factors are used to extrapolate from anti-tight, fake-rich control regions in data into the signal region. However, the equivalent low jet control regions are too low in statistics to provide the transfer factors from data

Uncertainties		Channels			
		4 jets		≥ 5 jets	
		$e^\pm e^\pm$	$\mu^\pm \mu^\pm$	$e^\pm e^\pm$	$\mu^\pm \mu^\pm$
Statistical	$\Delta\theta_e^{\text{stat}}$	39.6	—	14.2	39.6
	$\Delta\theta_\mu^{\text{stat}}$	—	24.7	15.8	24.7
	$\Delta N_{\ell\text{anti}-\ell}(n \text{ jets})(\text{stat})$	6.8	18.1	21.3	25.9
Systematics	$\Delta\theta_e^{\text{syst}}$ (closure)	21.8	—	7.8	26.7
	$\Delta\theta_\mu^{\text{syst}}$ (closure)	—	23.3	18.4	31.2
	Q Mis Id ($\ell\text{anti}-\ell$)	2.2	—	1.5	—
Total		45.7	38.5 (36.3)	35.7	48.5 47.8 (43.9) 39.6
Systematic	Q Mis Id ($\ell\ell$)	24.0	—	8.6	24.0 — 11.3

Table 8.7: Summary of the uncertainties (in %) in $e^\pm e^\pm$ (reverse Id + reverse isolation method), $\mu^\pm \mu^\pm$ and $e^\pm \mu^\pm$. Statistical uncertainty is split into statistical uncertainties on θ_e and θ_μ and uncertainty due to the Control Region size ($\Delta N_{\ell\text{anti}-\ell}(n \text{ jets})$). For channels with muons, uncertainties between parenthesis are obtained when anti-tight muons are defined such as $p_T < 20$ GeV (includes blinded data)

1429 directly. Instead, the transfer factors are obtained directly from the $t\bar{t}$ simulation. Data control
 1430 regions, called auxiliary regions, are used to determine the modeling of the identification and isolation
 1431 variables, used in the transfer factor extrapolation. The low jet regions are still employed in a low
 1432 statistics validation of the entire fake procedure.

1433 Anti-tight electrons and muons are defined in slightly different ways, compared to the 2ℓ SS case:

- 1434 • **anti-tight electron (d):** fails to pass the verytight likelihood operating point, but still verifies
 1435 the veryloose operating point. The isolation selection is released $E_T^{\text{rel}} > 0.05$, $p_T^{\text{rel}} > 0.05$.
- 1436 • **anti-tight muon (p):** muons must pass identification but the p_T cuts is lowered to 6 GeV.
 1437 The overlap removal with jets and isolation cuts are released.

1438 The transfer factors, θ_e and θ_μ , extracted from MC, is defined as the ratio of the number of top ($t\bar{t}$
 1439 + single-top) events in the signal region, to the number of $t\bar{t}$ events in the anti-tight regions. The
 1440 factors are calculated in separate flavor regions to ensure that the electron jet fakes and muon jet
 1441 fakes are calculated separately. The calculation follows the same form as for the 2ℓ SS case, but now
 1442 lep0, which by construction is almost never a fake is allowed to be either electron or muon in both
 1443 cases, denoted below in Equations 8.12 and 8.13.

$$\theta_e = \frac{N_{xee}^{\text{top}}}{N_{xed}^{\text{top}}} \quad (8.12)$$

$$\theta_\mu = \frac{N_{x\mu\mu}^{\text{top}}}{N_{x\mu p}^{\text{top}}} \quad (8.13)$$

1444 The MC modeling of the variables involved in the transfer factor can be verified when another
 1445 variable fails. For instance, the MC modeling of the electron isolation variable can be compared to
 1446 data when the particle identification variable fails and vice-versa. The modeling of muon-jet ΔR ,
 1447 involved in the overlap removal, can be compared when either the isolation variable or the p_T fails.
 1448 The comparison of the electron variables in this manner can be seen in Figure 8.13 and the muon
 1449 variables in Figure 8.12. The regions used have the same selection as the signal region with an added
 1450 missing transverse energy requirement, > 60 GeV, to ensure only top fakes. 20% and 30% systematic
 1451 uncertainties are assigned to the muon and electron transfer factors, respectively, to account for data-
 1452 MC discrepancies. This method for evaluating data-MC agreement for individual electron and muon
 1453 variables in turn relies on the assumption that these variables are largely uncorrelated and that the
 1454 transfer factor itself is factorizable into pieces for each variable. Factorized and fully correlated transfer
 1455 factors have been compared using MC and shown to have differences smaller than the systematic
 1456 quoted, suggesting that the uncorrelated assumption is reasonable.

1457 The electron and muon anti-tight control regions, which are scaled by the transfer factors are shown
 1458 in Figure 8.14. The prompt MC subtracted data in these regions are scaled by the transfer factors to
 1459 obtain the overall contribution of fake electron and muon events in the signal region. The systematic
 1460 uncertainties are split between the statistical error on the transfer factor and normalization of the
 1461 anti-tight control regions and the data-MC comparisons outlined above. For muon fakes the total
 1462 systematic uncertainty is 25% and for electrons it is 34%. The numbers and uncertainties involved in
 1463 the calculation are shown in Table 8.8.

Stage	Muon	Electron
Anti-Tight CR Normalization	364.62 ± 20.02 (5%)	38.2 ± 6.9 (17%)
Transfer factor θ_μ	0.0047 ± 0.0011 (23%)	0.0240 ± 0.0064 (29%)
SR Contribution	1.71 ± 0.42 (25%)	0.91 ± 0.29 (34%)

Table 8.8: Summary of regions and inputs to the extraction of the number of $t\bar{t}$ events with a fake muon in the SR

1464 The low jet region (1, 2, 3) is used as a validation for the method. The $t\bar{t}$ and single top fakes in
 1465 this region are estimated using the procedure above. Similar systematics are assessed. This region
 1466 with the fake estimate is plotted in Figure 8.15. The agreement of data and summed prediction for
 1467 the fakes and prompt backgrounds is well within the systematic and statistical uncertainties. The
 1468 figure also shows the same region with relaxed p_T cuts on all leptons to 10 GeV, which enriches the
 1469 fake contributions greatly. The data and summed fake and prompt predictions are also well within
 1470 the statistical and systematic uncertainties.

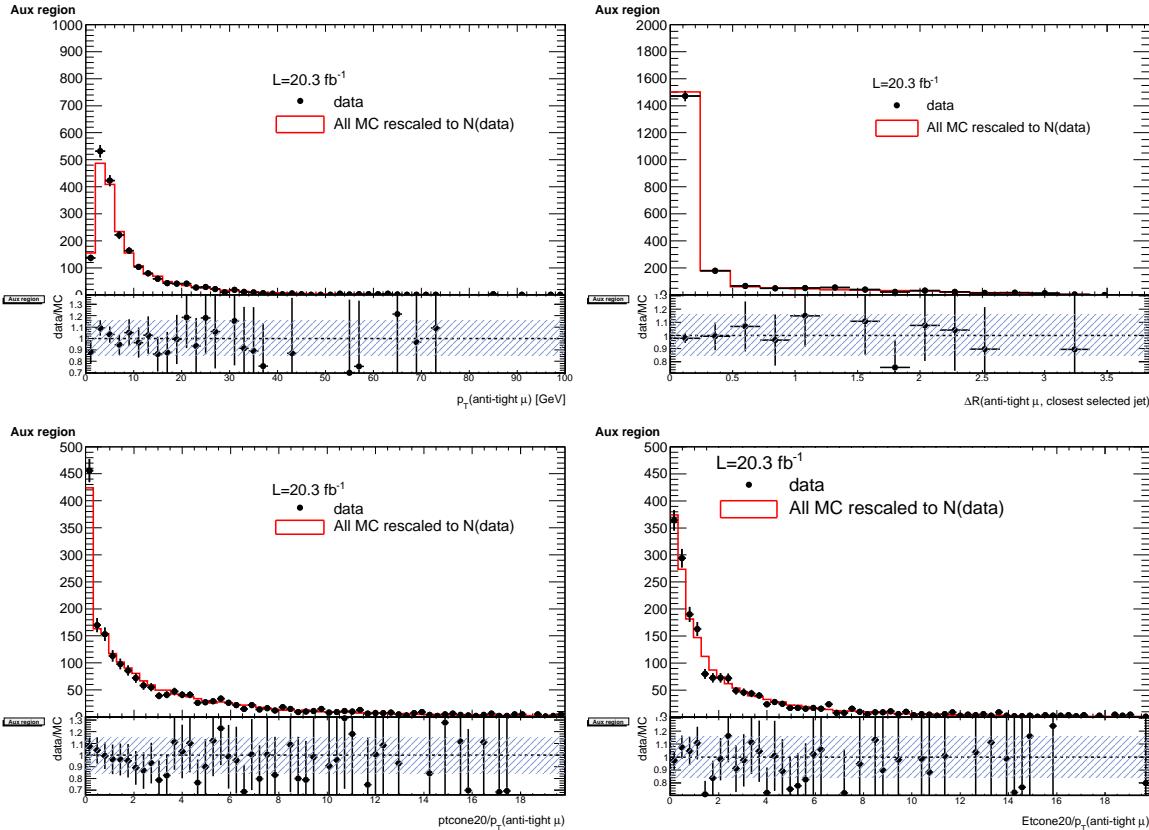


Figure 8.12: Distributions of the properties of the anti-tight muons in data (dots), compared with the total simulation (red line), rescaled to the integral of the data for a shape comparison. The uncertainty on the data distribution is statistical. The number of events for each of them is also presented in the legend. The variables probed are, top: p_T and $\Delta R(\mu, \text{closest selected jet})$; bottom: $ptcone20/p_T$ and $Etcone20/p_T$. The selection is the signal region event selection with one anti-tight muon (failing at least one of the isolation, muon-jet overlap, or p_T selection criteria). A ratio plot is containing the 20% area around 1, is displayed, demonstrating that a 20% data-MC comparison systematic is sufficient.

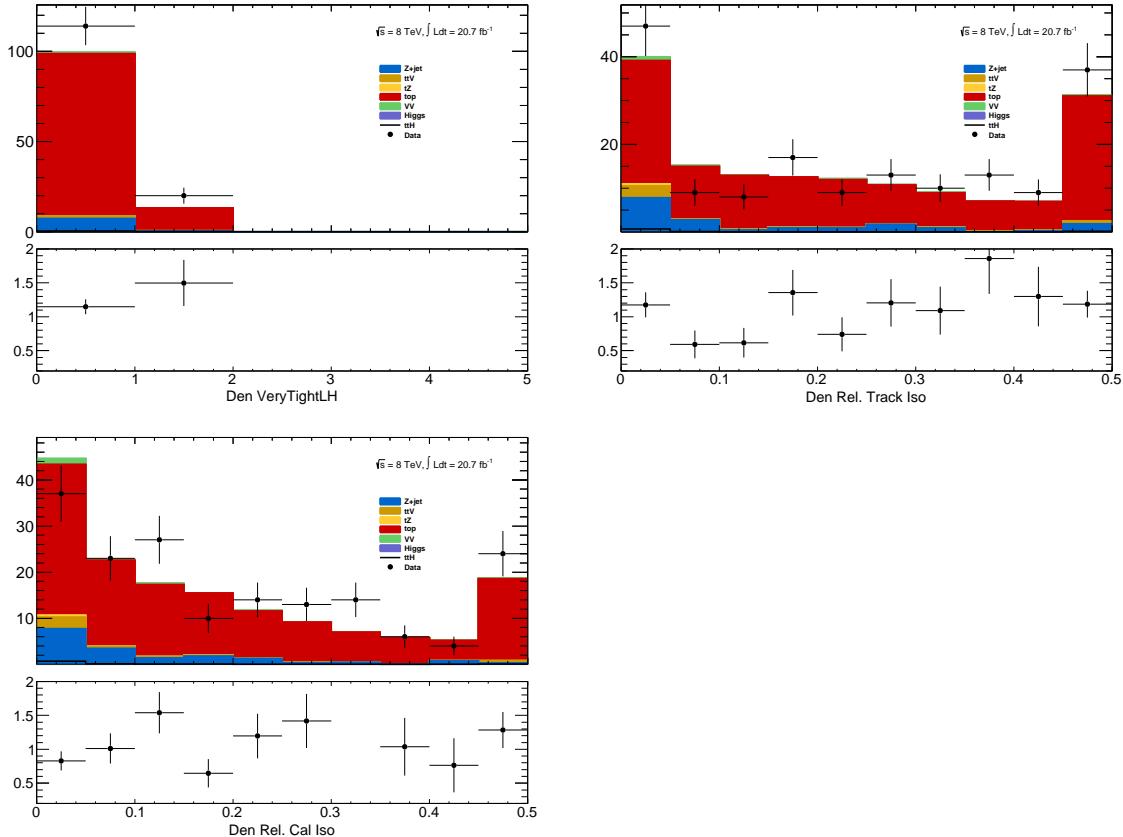


Figure 8.13: Distributions of anti-tight electron variables. The variables presented are, from top left to bottom right, p_T , η , VERYTIGHT Likelihood value, $\text{ptcone20}/p_T$, $\text{Etcone20}/p_T$. The plotted regions have the same cuts as the signal region, except the anti-tight electron must fail isolation for the plot of the VERYTIGHT identification word or fail the VERYTIGHT identification word for the plots of the isolation. Data (dots) are compared with a stacked histogram of the various simulated samples: top in red, $V+jets$ (blue), VV (purple) and $t\bar{t}V$ (yellow). The uncertainty on the data distribution is statistical.

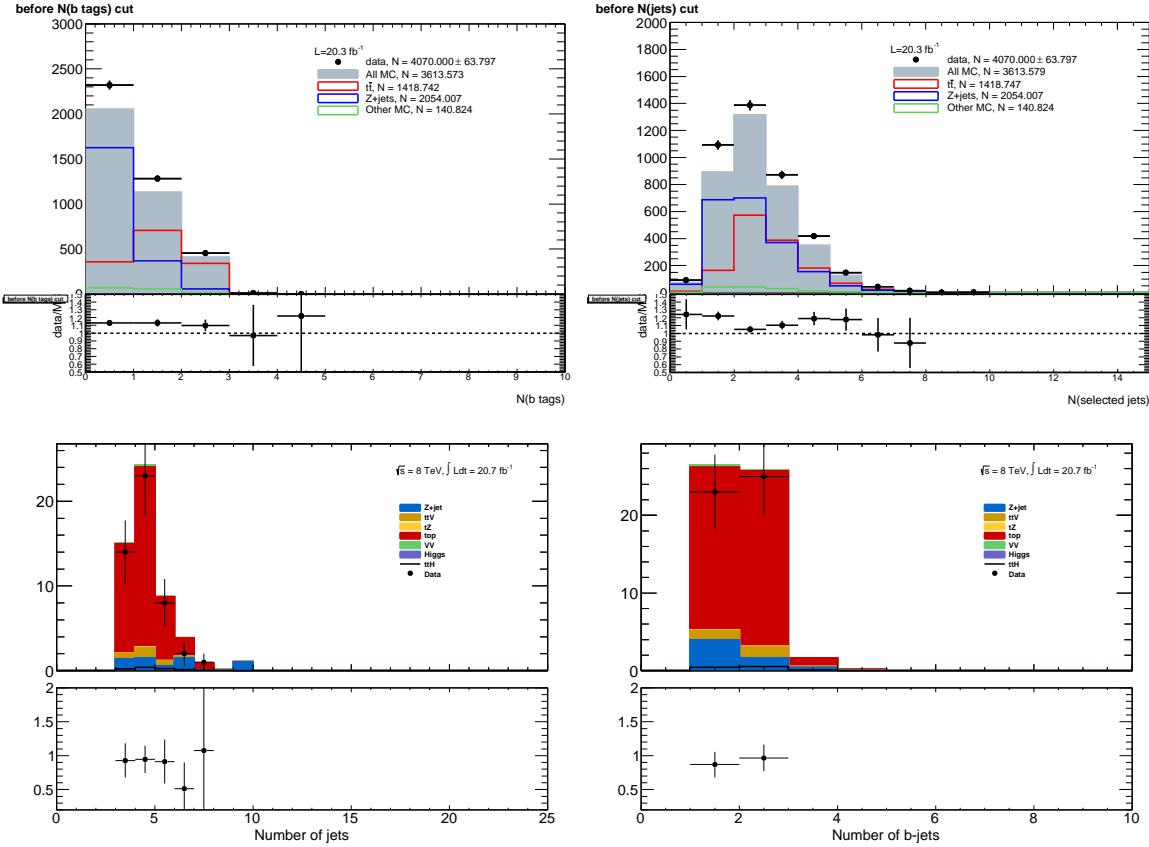


Figure 8.14: Muon-xxp (above) and electron-xxd (below) anti-tight control regions: jet variables. Note: the $t\bar{t}$ and single top MC in the plots is used only as comparison, but is not included in the fake measurement

8.4.3 4ℓ Fakes

We will not discuss the 4ℓ fakes in depth, as it is a very small background - at the % level and will have almost no impact on the final result. The fake method used in the the 4ℓ case is similar to the 2ℓ and 3ℓ cases discussed above. All fakes arise from $t\bar{t}$ and single-top events, where *two* jets are misidentified as leptons. To measure the contribution of this background, control regions with 2 fully identified and 2 anti-identified leptons are created. These control regions do not have a number of jets requirement in order to increase statistics. From these control regions, two extrapolations are made. First, a transfer factor is applied to extrapolate from the anti-tight to tight regions for electrons and muons. The regions are defined with identical object identification selection and reversal as the 3ℓ case, and the same transfer factors can be used. They must be used twice however, because there are two anti-identified leptons in each event. Second, the jet inclusive regions are extrapolated into the 2-jet signal

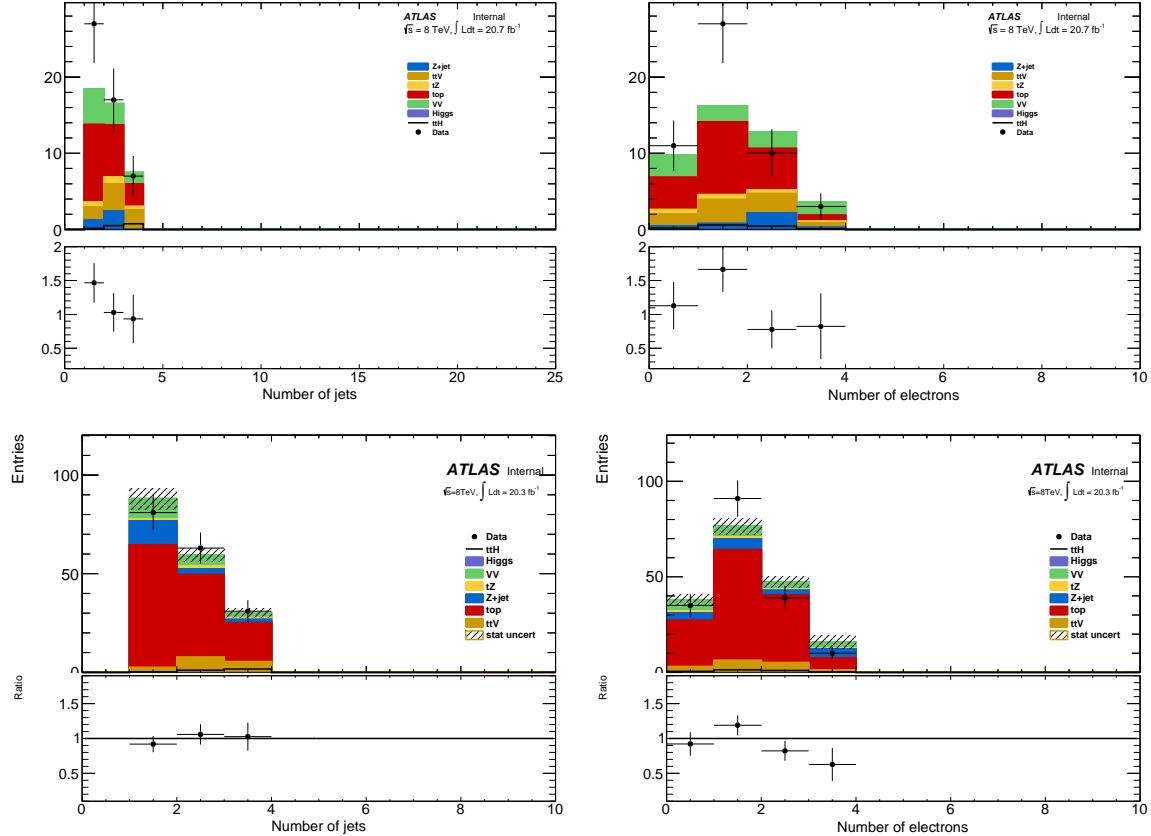


Figure 8.15: 3ℓ fake validation regions for nominal p_T selection (above) and relaxed p_T selection, > 10 GeV, (below). Plotted are the number of jets and the number of electrons in each event. The data and MC ratio below each plot agree with 1 within the statistics of the region and the overall systematic assigned for the fake component (red)

region, using a second extrapolation factor derived from $t\bar{t}$ events. Since, the majority of fake leptons arise from b-quark initiated jets, the jet spectrum $t\bar{t}$ events with the additional requirement of 2-b-tagged jets from data are used as a model for the jet extrapolation. The overall systematic uncertainty on this measurement arises from the statistics in the control regions and MC based assessments of non-closure and are 35%-50% depending on the sub-channel.

CHAPTER 9

Summary of Systematic Uncertainties

1489 This chapter summarizes the systematic uncertainties that enter the measurement of the limit of
 1490 $t\bar{t}H$ multi-lepton analysis. The systematic uncertainties arise from three main sources. The first are
 1491 the normalization uncertainties on the background process estimation methods, which are discussed
 1492 in depth in Chapter . The second source is the theoretical uncertainties on the $t\bar{t}H$ production
 1493 cross-section and acceptance. The final source are the experimental and detector related systematic
 1494 uncertainties related to event selection efficiencies and measurements and identification of the objects.
 1495 They affect only the non-data driven backgrounds and the $t\bar{t}H$ signal, as simulation is used to model
 1496 their acceptance and efficiency for the analysis selection.

1497 These systematic uncertainties, the estimated background and signal event counts in each of the
 1498 signal regions, and the observed data in each signal region are combined in a statistical fit to an
 1499 analysis model to extract the measurement of interest. We measure per-channel and combined ratios
 1500 of the observed production rate to the theoretically predicted production rate of $t\bar{t}H$, a parameter
 1501 called μ . In the absence of a statistically significant observation, this measurement is translated into
 1502 a upper confidence limit on μ . The details of this procedure are discussed in the following sections
 1503 and the results with the observed data are discussed in Chapter 10

1504 **9.1 Systematic Uncertainties on Signal Cross-section and Acceptance**

1505 The $t\bar{t}H$ signal is simulated with matrix elements at NLO in QCD with Powheg. The simulation
 1506 details are discussed in Chapter 6. The production cross section and the Higgs boson decay branching
 1507 fractions together with their theoretical uncertainties from the QCD scale and PDF choice are taken
 1508 from the NLO theoretical calculations reported in Ref. [66]. The uncertainty from the QCD scale
 1509 estimated by varying μ_0 by a factor of 2 from the nominal value is $^{+3.8\%}_{-9.3\%}$, while the uncertainty from

Table 9.1: Theoretical uncertainties of the signal event yields in the signal regions due to the impact of QCD scale uncertainties on the event selection.

QCD scale [%]	$2\ell 4\text{jets}$	$2\ell \geq 5\text{jets}$	3ℓ	4ℓ
Static	+0.6 −0.0 +1.7 −0.8	+2.7 −1.3 +2.0 −2.6	+2.3 −0.8 +1.7 −1.1	+0.9 −0.2 +0.5 −0.0
Dynamic				

1510 the PDF set and the value of α_S is $\pm 8.1\%$.

1511 The impact of the choice of the QCD scale on the simulation of the $t\bar{t}H$ event selection efficiency
1512 is estimated in two independent ways.

1513 First, the factorization and renormalisation scales μ_0 are varied by a factor of 2, as $\mu = 2\mu_0$ and
1514 $\mu = \mu_0/2$. The effects of these new scales are estimated via the application of event re-weighting
1515 procedures on the nominal simulation using kinematic distributions at parton level. The weights used
1516 are dependent on the transverse momenta of both the $t\bar{t}H$ system and of the top quark, as described
1517 in Ref. [81].

1518 Second, the choice of the factorization and renormalisation scales, dependent on fixed (“static”)
1519 parameters in the nominal simulation, is tested comparing its prediction with an alternative (“dy-
1520 namic”), but still physics motivated choice $\mu_0 = (m_T^t m_T^{\bar{t}} m_T^H)^{\frac{1}{3}}$, which depends on kinematic variables.
1521 This comparison is performed via event re-weighting of the nominal static simulation based on weights
1522 derived as a function of the $t\bar{t}H$ transverse momentum [81]. In order to take the difference between
1523 the choices of scale as systematic uncertainties, a symmetric envelope around the nominal simulation
1524 is built applying the weights and also their inverses.

1525 In order to not double-count the variations on the total cross section the predictions from the
1526 different QCD scales are normalized to the same total cross section. That means that the observed
1527 differences are only coming from the event selection. Significant variations on the jet multiplicities can
1528 be seen and these translate into different predictions on the signal event yields in the signal regions.
1529 Such differences, listed in Table 9.1, are taken as theoretical systematic uncertainties in addition to
1530 the ones affecting the total $t\bar{t}H$ production cross section. The “Static” uncertainties come from the
1531 variations by a factor of 2 from the nominal scale and they are correlated with the uncertainties on
1532 the total cross section, which are estimated with the same procedure. The “Dynamic” uncertainties
1533 come from the difference between the nominal and the alternative dynamic scale and are treated as
1534 an independent source of theoretical uncertainty.

1535 The uncertainty of the $t\bar{t}H$ event selection due to the PDF sets is estimated comparing the predic-
1536 tions with three different PDF sets, varying each set within errors and taking the width of the envelope

Table 9.2: Uncertainties on $t\bar{t}H$ acceptance in signal regions due to PDF variation.

Sample	2ℓ 4j	2ℓ 5j	3ℓ	4ℓ
$t\bar{t}H$	0.3%	1.0%	0.5%	1.4%

as systematic uncertainty. The recommended sets are CT10, MSTW2008nlo68c1 and NNPDF21_100. We determine the change in the acceptance due to the PDF sets via the formula 9.1 to disentangle it from the change in production cross section.

$$\left(\frac{\text{Reweighted yield in SR}}{\text{Re-weighted total number of events}} \right) \left(\frac{\text{Original yield in SR}}{\text{Original total number of events}} \right)^{-1} - 1 \quad (9.1)$$

Fig. 9.1 shows the estimated PDF systematic uncertainties as a function of the jet multiplicity in $t\bar{t}H$ events with at least two leptons. The uncertainties are compatible with the uncertainty on the production cross section estimated in Ref. [66] and indicated by the dashed red lines in the lower panel. Table 9.2 shows the half-width of the envelope of the acceptance under all eigenvector variations of the three PDF sets. No significant dependence on the event topology is observed, so that the PDF systematic uncertainty on the $t\bar{t}H$ event selection is neglected.

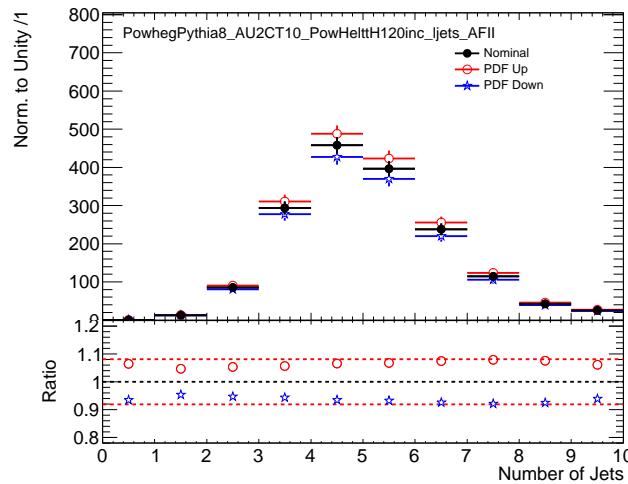


Figure 9.1: PDF systematic uncertainty on the jet multiplicities in $t\bar{t}H$ events with at least 2 leptons. The dashed red lines in the lower panel indicate the systematic uncertainty on the $t\bar{t}H$ production cross section.

1546 9.2 Experimental and Detector Systematic Uncertainties

1547 Experimental and detector systematic uncertainties affect the efficiencies of identifying objects and
1548 the efficiencies for events to pass our cuts . These uncertainites affect only MC models of physics
1549 processes, $t\bar{t}V$, $t\bar{t}H$, VV and thus alter ther number of expected events from signal and backgroun in
1550 our signal regions. Data-driven backgrounds take into account these uncertainties by construction. We
1551 consider systematic effects from a number of sources: the lepton and jet energy scale measurements,
1552 the lepton identification and isolation selections, the efficiency and mis-identification rate associated
1553 with tagging b-quark jets. Effects due to modeling the energy and objects from additional vertices
1554 were studied and found to be negligible. The vast majority of the individual detector systematics
1555 effects are small. The sum total of the systematic effects are comparable to the overall normalization
1556 and cross-section uncertainties on some of the physics processes and is shown in Table 9.3.

1557 9.2.1 Lepton Identification, Energy Scale, and Trigger

1558 The electron[51] and muon identification efficiencies[82] are measured in data using Z boson and
1559 J/Ψ control samples. They are shown in Figure 9.2. The uncertainty on the muon efficiencies are
1560 measured as functions of η and p_T and are generally less 1%. The uncertainty on the electron and
1561 muon efficiencies are also measured as functions of η and p_T and are at the 1% level for p_T above
1562 30 GeV, but become much larger 5-10% for the lower p_T regimes. These translate into sub-% level
1563 effects on the $t\bar{t}V$ and $t\bar{t}H$ event counts in the signal regions for the muons and 1% level effects for
1564 the electrons. The effects become more important with increasing numbers of leptons.

1565 The electron[83] and muon[82] energy scale and resolution are also measured using the Z -boson
1566 control samples in data. The uncertainties related to the scale and resolution for the leptons affect
1567 the overall event acceptance through the lepton momentum cuts primary and have negligible impact
1568 on the event count uncertainties in the signal regions.

1569 The efficiencies for muons and electrons to pass muon[84] and electron triggers[85] have been
1570 calculated with respect to the offline identification operating points using the Z boson control samples.
1571 They are in the range of 90% for electron triggers and 70% for muon triggers, owing to gaps in muon
1572 trigger coverage, and have 1% level uncertainties. When statistically combined for 2ℓ SS, 3ℓ and 4ℓ
1573 lepton signal regions, the overall trigger efficiency is high and the uncertainties on the number of
1574 expected events is negligible.

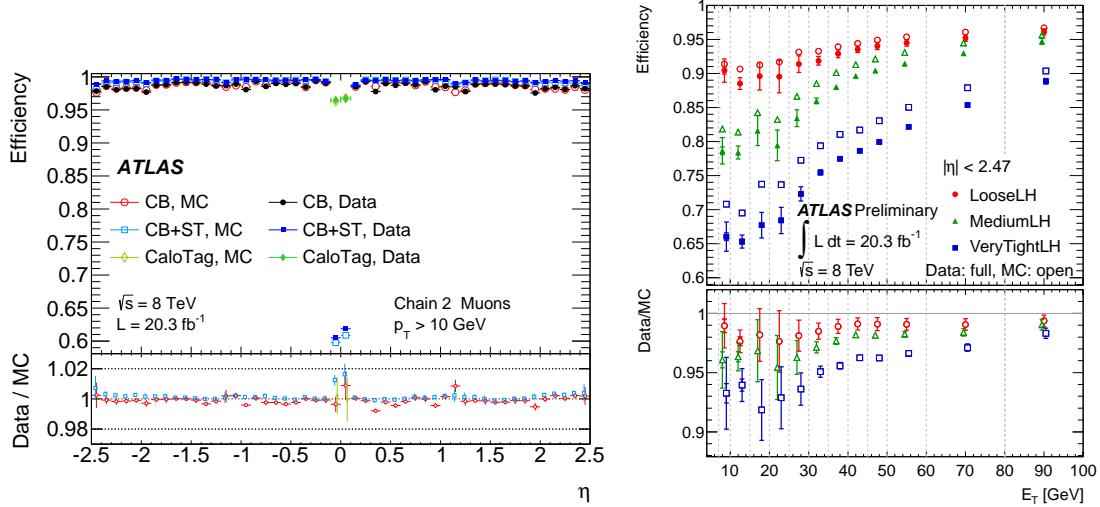


Figure 9.2: Muon (left) and electron(right) identification efficiencies in Data and MC as a function of η and p_T respectively. For electrons, the verytight likelihood operating point is used and for muons the CB+ST (combined+segment tagged) operating point is used

9.2.2 Lepton Isolation and Impact Parameter

The isolation and impact parameter selections are specific to this analysis and are discussed in depth in Chapter 7. We calculated their combined efficiency with respect to the full lepton identification selection using the Z boson control samples and define data-MC scale factors to correct the efficiency in the simulation. Background are subtracted using shape templates in the di-lepton invariant mass spectrum. The Z -event template is derived from MC, while the background template is derived from the same-sign control region. We measure the efficiency scale-factors in bins of lepton momentum. Uncertainties are assigned per-bin to account for the level of statistics and variations caused by the fit parameters. An additional 1% uncertainty envelope is added to both the electron and muon measurements to account for trends observed in the dependence of the data-MC efficiency scale-factor as a function of the number of jets. Stability of the efficiency scale-factor as a function of the number of jets is important for this analysis, because event activity in the low jet Z sample where the efficiency is measured is much different from the high jet signal regions where the efficiencies are applied. The dependence of the scale-factor on the number of jets can be seen Figure 9.3. The isolation scale-factor uncertainties are around 1-3% depending on the particle momentum, but these

1590 uncertainties propagate to 2-5 % (some of the largest) effects in the event counts in the signal regions.
 1591 The uncertainties are more important in the regions with more leptons.

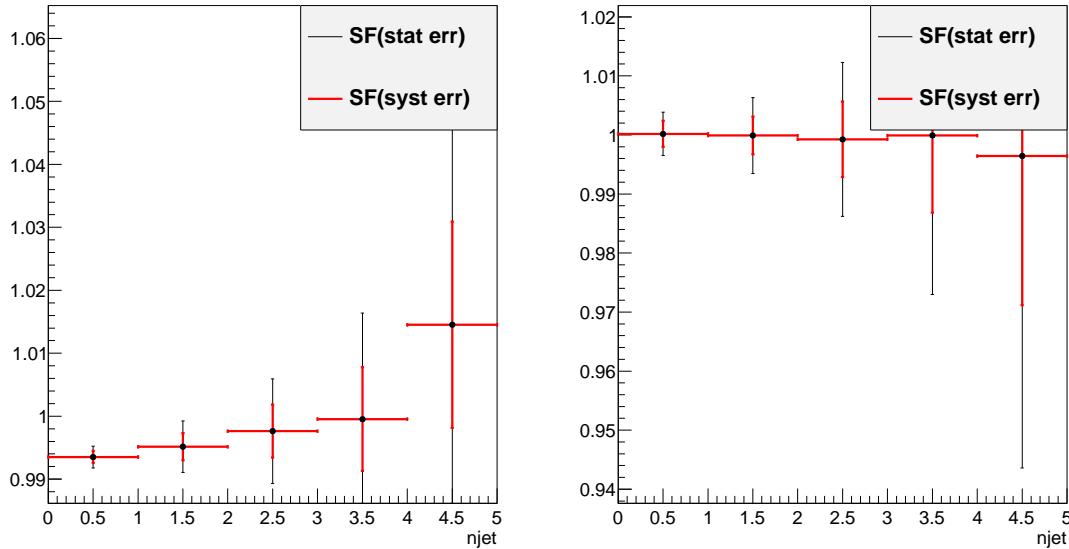


Figure 9.3: Muon (left) and electron(right) isolation efficiency scale-factors from the Z control sample as a function of the number of jets in the event. An additional systematic uncertainty of 1% is added to encompass the variation in the number of jets variable

1592 9.2.3 Jet Energy

1593 The jet energy scale (JES) is calculated using a combination of data-based in-situ techniques, where
 1594 jet transverse momentum is balanced with respect to a reference photon or a Z boson, as well as
 1595 single particle test-stand studies[86]. Additional smaller effects are taken into account including the
 1596 b-quark jet specific response, near-by jets, the effects of pile-up and an inter-calibration of similar
 1597 η regions using di-jet events. These effects are measured in 2012 data. The JES systematic errors
 1598 arises from numerous sources that are diagonalized into eigenvectors so that they can be combined
 1599 in an uncorrelated way. The combined uncertainty is plotted in Figure 9.4 as a function of jet η and
 1600 p_T and is the range 2-4% for jets used in this analysis. The jet energy resolution is calculated in a
 1601 similar way with slightly larger errors, 10% [87]. The combined scale and resolution systematics are
 1602 of non-negligible effects 6-7% on the signal and background event counts in the signal regions.

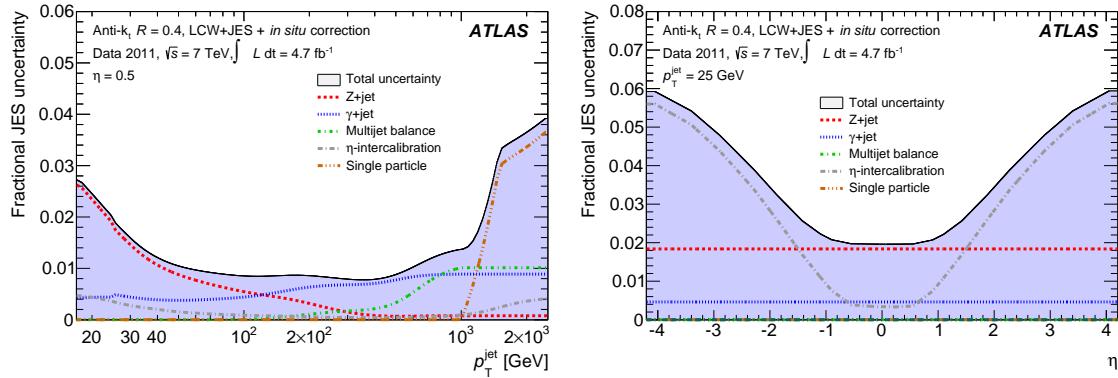


Figure 9.4: JES systematic uncertainties as a function of jet η (for jets $p_T > 25 \text{ GeV}$) and p_T (for jets $|\eta| < 0.4$). The combined systematic uncertainty is shown with contributions from the largest sources

9.2.4 B-Tagged Jet Efficiency

The b -quark tagging efficiency must be calculated separately for charm, light and b -quarks. ATLAS uses three data based control regions: an inclusive jet sample for mis-tagged light quarks[88], the $t\bar{t}$ sample for b -quarks[89], and a sample of D^* mesons for charm quarks[90]. These efficiencies and rates are well-measured in MC and the data-based corrections are small. The data-MC efficiency scale-factor shown in Figure 9.5 is close to 1 and has an overall systematic uncertainty of around 5%. The uncertainties are applied to the analysis via a number of eigenvectors. Together these uncertainties have a 4 % effect in the event expectation in the signal regions.

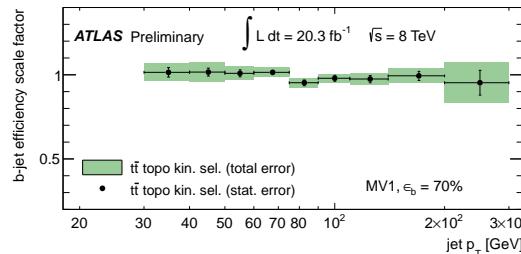


Figure 9.5: b-Tagging data-MC efficiency scale-factors versus jet p_T calculated in the $t\bar{t}$ sample from 2012 data. The uncertainties are combined statistical and systematic.

9.2.5 Summary

The combined effect of these detector and experimental systematics on the $t\bar{t}V$ and $t\bar{t}H$ is provided in Table 9.3. The effects are smaller than the normalization uncertainties on some of the backgrounds.

Total Systematic Uncertainty	2ee4j Down-Up	2ee5jincl Down-Up	2em4j Down-Up	2em5jincl Down-Up
ttH	-4.68 5.84	-8.24 6.14	-5.10 3.50	-5.52 6.40
ttW	-7.20 5.45	-8.72 11.30	-3.63 6.22	-9.72 7.95
ttZ	-9.68 5.07	-5.87 10.98	-4.07 6.16	-8.37 4.99
Total Systematic Uncertainty	2mm4j Down-Up	2mm5jincl Down-Up	3ℓ Down-Up	4ℓ Down-Up
ttH	-5.20 7.51	-7.28 6.75	-5.84 5.59	-6.54 6.54
ttW	-4.54 5.23	-8.63 6.88	6.36 8.16	— —
ttZ	-5.24 8.69	-9.73 8.18	-6.14 6.66	-9.58 6.94

Table 9.3: Sum in quadrature of all the systematic uncertainties on the number of event yields per channel.

1614 They are dominated by the lepton isolation scale-factor measurements and the electron identification
 1615 with smaller contributions from the JES and b-tagging efficiencies. These detector systematic uncer-
 1616 tainties enter the fit individually and their ranking of influence on the overall measurement uncertainty
 1617 can be seen in Figure ??.

1618 9.3 Summary of Background and Signal Normalization Uncertainties

1619 Tab.9.4 gives the summary of the systematic uncertainties that are included in the analysis for the
 1620 normalization and acceptance of each process. The relative importance of these uncertainties to the
 1621 final fit can be seen in Figure.

Type	Description	Name	Uncertainty
Signal (ttH)			
QCD Scale	Cross Section (Dynamic Scale) Analyses Acceptance	QCDscale_XS_ttH QCDscale_Acceptance_ttH	+3.8% –9.3% (Section 9.1) 0.-2.6%
PDF+ α_S	Cross Section Analyses Acceptance	pdf_ttH pdf_Acceptance_ttH	\pm 8.1% Negligible
ttW (Irreducible background)			
QCD Scale	Cross Section (Dynamic Scale) Analyses Acceptance	QCDscale_XS_ttW QCDscale_Acceptance_ttW	\pm 15% (Section 8.1) 0.4-3.5%
PDF+ α_S	Cross Section Analyses Acceptance	pdf_ttW PDF_Acceptance_ttW	\pm 13% 1.1-4.8%
ttZ (Irreducible background)			
QCD Scale	Cross Section (Dynamic Scale) Analyses Acceptance	QCDscale_XS_ttZ QCDscale_Acceptance_ttZ	\pm 12% (Section 8.1) 0.1-3.1%
PDF+ α_S	Cross Section Analyses Acceptance	pdf_ttZ PDF_Acceptance_ttZ	\pm 9% 0.9-2.7%
VV Backgrounds			
Normalization Uncertaintiy	WZ,ZZ Processes	WZxsec,ZZxsec	\pm 50% (Section 8.2)
Data-Driven Backgrounds			
Normalization Uncertainty		Jet Fakes	\pm 30-50% (Section 8.4)
Normalization Uncertainty		Charge MisID	\pm 20-30% (Section 8.3)

Table 9.4: Summary of systematic uncertainties for processes present in the signal regions in the analysis, with their type, description, name, values and uncertainties, and status of inclusion in the final results.

1622

CHAPTER 10

1623

Results and Statistical Model

1624

10.1 Results in Signal Regions

1625

10.1.1 2ℓ SS

1626

10.1.2 3ℓ

1627

10.1.3 4ℓ

1628

10.2 Statistical Model

1629 We use the above results to make two sets of measurements: an upper confidence limit on μ , the
 1630 signal strength parameter, and a measurement of μ . These measurements are done for each channel
 1631 individually and then combined. The interpretation of the results in the form of a statistical model
 1632 follow the procedure, discussed here [91]. We interpret the results as counting experiments in each
 1633 signal region. Therefore agreement in kinematic shapes do not affect the statistical procedure.

1634

10.2.1 The Likelihood

1635 The observed and expected event yields in the signal regions are analyzed using a binned likelihood
 1636 function (\mathcal{L}), built from product of Poisson models of expected event counts for each bin, where the
 1637 bins are the separate signal regions:

$$\mathcal{L} \propto \prod_{i=0}^{N_{SR}} P(N_{obs}^i | \mu \cdot s_{exp}^i + b_{exp}^i) \quad (10.1)$$

1638 where s_{exp}^i is the SM signal expectation in the signal region, b_{exp}^i are the background expectations, i
 1639 counts over the signal regions, and P is the Poisson distribution. The signal strength parameter is the

1640 parameter of interest in the model (POI) and acts as a simple scale-factor to the SM $t\bar{t}H$ production
 1641 rate and is common to all channels. Setting μ to 0 corresponds to the background only scenario. The
 1642 background parameter, b , is a sum over all background processes.

1643 The signal and background expectations, s and b , depend on systematic errors. These are included
 1644 in the likelihood function in the form of a vector nuisance parameters, $\vec{\theta}$, which are constrained to fluc-
 1645 tuate within Gaussian distributions. These fluctuations affect the background and signal expectations
 1646 by response functions, $\nu(\vec{\theta})$, set by systematic uncertainties measured in the previous section. For
 1647 instance, the $W^\pm Z$ normalization uncertainty is 50% from Section 8.2 and is included in the fit as its
 1648 own unit gaussian, $G(\theta|0, 1)$. The fluctuations of the gaussian, θ_{WZ} scale the background contribution
 1649 via the form, $0.5 \cdot (1 + \theta_{WZ}) \cdot b_{WZ}$. For many of the detector systematics, the uncertainties are two
 1650 sided and are included as piecewise Gaussians. We add correlations to various uncertainties by hand,
 1651 when appropriate. With these nuisance parameters, the likelihood takes this form:

$$\mathcal{L}(\mu, \vec{\theta}) = \left(\prod_{i=0}^{N_{ch}} P(N_{obs}^i; \mu \cdot \nu_s(\vec{\theta}) \cdot s_{exp}^i + \nu_b(\vec{\theta}) \cdot b_{exp}^i) \right) \times \prod_j^{N_\theta} G(\theta_j; 1, 0) \quad (10.2)$$

1652 10.2.2 Test Statistic and Profile Likelihood

1653 Values of μ are tested with the negative log quantity, $q_\mu = -2\ln(\lambda(\mu))$, where $\lambda(\mu)$ is the test statistic.
 1654 $\lambda(\mu)$ is defined as:

$$\lambda(\mu) \equiv \frac{\mathcal{L}(\mu, \hat{\vec{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\vec{\theta}})} \quad (10.3)$$

1655 where $\hat{\vec{\theta}}_\mu$ are values of the nuisance parameter vector that maximize the likelihood for a given value
 1656 of μ and $\hat{\mu}$ and $\hat{\vec{\theta}}$ are the fitted values of signal strength and nuisance parameters that maximize the
 1657 likelihood overall. μ is constrained to be positive.

1658 10.2.3 CL_s Method

1659 Exclusions limits on the signal strength are calculated with the test statistic using a modified fre-
 1660 quentist method, called the CL_s method[92]. CL_s is defined as a ratio of two frequentist quantities.
 1661 The numerator quantifies the probability of finding the observed data given the signal + background
 1662 hypothesis. The denominator quantifies the probability of the data given the background only hy-
 1663 pothesis.

1664 Using the numerator alone has the undesirable property that, if the data fluctuates below the
 1665 expectation, an exclusion limit can be reached that is far beyond the sensitivity of the experiment.
 1666 Normalizing to the background only hypothesis penalizes these low sensitivity cases.

1667 The probability of obtaining an observation as extreme as the data given a particular signal +
 1668 background hypothesis is given by the p-value,i p_μ defined as:

$$p_\mu = \int_{q_\mu^{obs}}^{\infty} f(q_\mu) dq_\mu \quad (10.4)$$

1669 and the probability of obtaining an observation as extreme as the data given the background hypoth-
 1670 esis, p_b is:

$$p_b = \int_{q_{\mu=0}^{obs}}^{\infty} f(q_{\mu=0}) dq_{\mu=0} \quad (10.5)$$

1671 where $f(q_\mu)$ is the distribution of q_μ for all possible observations for a given μ and q is defined above.
 1672 Therefore,

$$CL_s = \frac{p_\mu}{1 - p_b} \quad (10.6)$$

1673 . A μ value is considered excluded at 95% confidence when CL_s is less than 5%.

1674 10.2.4 Exclusion Limits

1675 Table ?? shows *expected* exclusion limits for all channels, including the analysis uncertainties cumula-
 1676 tively. It shows the relative importance of the statistical and systematic uncertainties to the analysis
 1677 sensitivity. The *observed* limits using observed data and predictions can be seen in Figures ??-?? for
 1678 splitting and combining the sub-channels and in Table XX by numbers. We expect a combined limit
 1679 of XXX (background only) and XXX (signal injected) and see a limit of XXX. The channel sensitivity
 1680 is dominated by the 2ℓ and 3ℓ channels.

Channels	Stat	+Fakes Unc.	+Theory	+ Experimental
2ℓ	2 ℓ ee	7.44	8.52	8.94
	2 ℓ em	3.46	3.81	4.18
	2 ℓ mm	4.03	4.14	4.57
	2 ℓ tau	8.08	8.92	10.00
	All	2.16	2.44	2.90
3ℓ	3.40	3.43	3.59	3.66
4ℓ	15.16	15.16	15.44	15.55
1l2tau	10.41	13.84	14.20	14.22
All	1.68	1.85	2.14	2.22

Table 10.1: 95%CL limits on μ for all channels and combination with cumulative uncertainties.

1681 10.2.5 μ Measurements

1682 In addition to setting a limit on the signal strength, we also fit the best value of the signal strength
 1683 for μ . We do this by minimizing the negative log likelihood value, q_μ or conversely maximizing the

1684 likelihood. The $1-\sigma$ error band is set via a profile likelihood scan, where the value q_μ is scanned as a
1685 function of μ . Values of μ that increase q_μ^{min} by 1 form the edges of the error band. The fitted values
1686 of μ with errors are provided in Table XXX for each sub-channel fit as well as the combined fit.

1687 **10.2.6 Nuisance Parameter Impact on the Signal Strength**

1688 Finally, we examine the post-fit impact of the various nuisance parameters on the final fit. We expect
1689 to have measured the various analysis uncertainties well and do not expect the fit to have much further
1690 constraint. For that reason, we expect the pulls of the nuisance parameters to be close to 0 and the
1691 measured uncertainties on those parameters to be consistent with the input uncertainties. Figures
1692 XXXX show.

1693

CHAPTER 11

1694

Conclusions

1695 **11.1 Higgs Results in Review**

1696 **11.2 Prospects for Future**

Bibliography

- 1698 [1] S. L. Glashow, *Partial-symmetries of weak interactions*, Nucl. Phys. **22** (1961) no. 4, 579. [2.1.1](#)
- 1699 [2] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264. [2.1.1](#)
- 1700 [3] A. Salam and J. C. Ward, *Gauge theory of elementary interactions*, Phys. Rev. **136** (1964) 763–768. [2.1.1](#)
- 1701 [4] S. Weinberg, *Non-abelian gauge theories of the strong interactions*, Phys. Rev. Lett. **31** (1973) 494–497. [2.1.1](#)
- 1702 [5] G. 't Hooft and M. Veltman, *Regularization and renormalization of gauge fields*, Nuclear Physics B **44** (1972) 189 – 213. [2.1.1](#)
- 1703 [6] D. J. Gross and F. Wilczek, *Ultraviolet behavior of non-abelian gauge theories*, Phys. Rev. Lett. **30** (1973) 1343–1346. [2.1.1](#)
- 1704 [7] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508. [2.1.2](#)
- 1705 [8] P. W. Higgs, *Spontaneous symmetry breakdown without massless bosons*, Phys. Rev. **145** (1966) 1156. [2.1.2](#)
- 1706 [9] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–322. [2.1.2](#)
- 1707 [10] The ALEPH, CDF, DØ, DELPHI, L3, OPAL, SLD Collaborations, the LEP Electroweak Working Group, the Tevatron Electroweak Working Group, and the SLD electroweak and heavy flavour groups, *Precision Electroweak Measurements and Constraints on the Standard Model*, CERN-PH-EP-2010-095 (2010) , [arXiv:1012.2367 \[hep-ex\]](#). [2.1.3](#), [2.2](#)
- 1708 [11] M. Baak, M. Goebel, J. Haller, A. Hoecker, D. Kennedy, R. Kogler, K. Mnig, M. Schott, and J. Stelzer, *The electroweak fit of the standard model after the discovery of a new boson at the LHC*, The European Physical Journal C **72** (2012) no. 11, .
<http://dx.doi.org/10.1140/epjc/s10052-012-2205-9>. [2.1.3](#)
- 1709 [12] J. C. Collins, D. E. Soper, and G. Sterman, *Factorization for short distance hadron-hadron scattering*, Nuclear Physics B **261** (1985) 104 – 142. [2.2](#)
- 1710 [13] CERN, . CERN, Geneva, 1984. [2.2](#)

- 1725 [14] LHC Higgs Cross Section Working Group, S. Dittmaier, C. Mariotti, G. Passarino, and
 1726 R. Tanaka (Eds.), *Handbook of LHC Higgs Cross Sections: 2. Differential Distributions*,
 1727 CERN-2012-002 (CERN, Geneva, 2012) , [arXiv:1201.3084 \[hep-ph\]](https://arxiv.org/abs/1201.3084). **2.2**
- 1728 [15] *Updated coupling measurements of the Higgs boson with the ATLAS detector using up to 25*
 1729 *fb⁻¹ of proton-proton collision data*, Tech. Rep. ATLAS-CONF-2014-009, CERN, Geneva, Mar,
 1730 2014. **2.2.1**
- 1731 [16] CMS Collaboration Collaboration, *Precise determination of the mass of the Higgs boson and*
 1732 *studies of the compatibility of its couplings with the standard model*, Tech. Rep.
 1733 CMS-PAS-HIG-14-009, CERN, Geneva, 2014. **2.2.1**
- 1734 [17] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the Higgs boson mass from*
 1735 *the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4\ell$ channels with the ATLAS detector using 25 fb⁻¹ of pp*
 1736 *collision data*, [arXiv:1406.3827 \[hep-ex\]](https://arxiv.org/abs/1406.3827). **2.2.1**
- 1737 [18] *Evidence for the spin-0 nature of the Higgs boson using {ATLAS} data*, Physics Letters B **726**
 1738 (2013) no. 13, 120 – 144.
 1739 <http://www.sciencedirect.com/science/article/pii/S0370269313006527>. **2.2.1**
- 1740 [19] S. Dawson, A. Gritsan, H. Logan, J. Qian, C. Tully, et al., *Working Group Report: Higgs*
 1741 *Boson*, [arXiv:1310.8361 \[hep-ex\]](https://arxiv.org/abs/1310.8361). **2.2.2**
- 1742 [20] O. Eberhardt, G. Herbert, H. Lacker, A. Lenz, A. Menzel, et al., *Impact of a Higgs boson at a*
 1743 *mass of 126 GeV on the standard model with three and four fermion generations*,
 1744 *Phys.Rev.Lett.* **109** (2012) 241802, [arXiv:1209.1101 \[hep-ph\]](https://arxiv.org/abs/1209.1101). **2.2.2**
- 1745 [21] M. Carena, S. Gori, N. R. Shah, C. E. Wagner, and L.-T. Wang, *Light Stops, Light Staus and*
 1746 *the 125 GeV Higgs*, *JHEP* **1308** (2013) 087, [arXiv:1303.4414](https://arxiv.org/abs/1303.4414). **2.2.2**
- 1747 [22] N. Arkani-Hamed, K. Blum, R. T. D’Agnolo, and J. Fan, *2:1 for Naturalness at the LHC?*,
 1748 *JHEP* **1301** (2013) 149, [arXiv:1207.4482 \[hep-ph\]](https://arxiv.org/abs/1207.4482). **2.2.2**
- 1749 [23] D. Carmi, A. Falkowski, E. Kuflik, and T. Volansky, *Interpreting LHC Higgs Results from*
 1750 *Natural New Physics Perspective*, *JHEP* **1207** (2012) 136, [arXiv:1202.3144 \[hep-ph\]](https://arxiv.org/abs/1202.3144). **2.2.2**
- 1751 [24] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, et al., *Higgs mass and*
 1752 *vacuum stability in the Standard Model at NNLO*, *JHEP* **1208** (2012) 098, [arXiv:1205.6497](https://arxiv.org/abs/1205.6497)
 1753 [\[hep-ph\]](https://arxiv.org/abs/1205.6497 [hep-ph]). **2.2.2**
- 1754 [25] L. Evans and P. Bryant, *LHC Machine*, *JINST* **3** (2008) no. 08, S08001. **3.1**
- 1755 [26] T. S. Pettersson and P. Lefevre, *The Large Hadron Collider: conceptual design.*, Tech. Rep.
 1756 CERN-AC-95-05 LHC, CERN, Geneva, Oct, 1995. <https://cdsweb.cern.ch/record/291782>.
 1757 **3.1**
- 1758 [27] T. Linnecar et al., *Hardware and Initial Beam Commissioning of the LHC RF Systems.*
 1759 *oai:cds.cern.ch:1176380*, Tech. Rep. LHC-PROJECT-Report-1172.
 1760 CERN-LHC-PROJECT-Report-1172, CERN, Geneva, Oct, 2008.
 1761 <https://cdsweb.cern.ch/record/1176380>. **3.1**
- 1762 [28] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, *JINST* **3**
 1763 (2008) S08003. **3.1**

- 1764 [29] The CMS Collaboration, *The CMS experiment at the CERN LHC*, Journal of Instrumentation
1765 **3** (2008) no. 08, S08004. [3.1](#)
- 1766 [30] The LHCb Collaboration, *The LHCb Detector at the LHC*, Journal of Instrumentation **3** (2008)
1767 no. 08, S08005. [3.1](#)
- 1768 [31] The ALICE Collaboration, *The ALICE experiment at the CERN LHC*, Journal of
1769 Instrumentation **3** (2008) no. 08, S08002.
1770 <http://stacks.iop.org/1748-0221/3/i=08/a=S08002>. [3.1](#)
- 1771 [32] A. Team, *The four main LHC experiments*, Jun, 1999. [3.1](#)
- 1772 [33] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report 1*. Technical Design
1773 Report ATLAS. CERN, Geneva, 1997. <https://cdsweb.cern.ch/record/331063>. [3.2](#)
- 1774 [34] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 2*. Technical Design
1775 Report ATLAS. CERN, Geneva, 1997. <https://cdsweb.cern.ch/record/331064>. [3.2](#)
- 1776 [35] ATLAS Collaboration, *ATLAS magnet system: Technical Design Report, 1*. Technical Design
1777 Report ATLAS. CERN, Geneva, 1997. <https://cdsweb.cern.ch/record/338080>. [3.2](#)
- 1778 [36] ATLAS Collaboration, *ATLAS pixel detector: Technical Design Report*. Technical Design
1779 Report ATLAS. CERN, Geneva, 1998. <https://cdsweb.cern.ch/record/381263>. [3.2](#)
- 1780 [37] ATLAS Collaboration, *ATLAS pixel detector electronics and sensors*, JINST **3** (2008) P07007.
1781 [3.2](#)
- 1782 [38] ATLAS Collaboration, *The barrel modules of the ATLAS semiconductor tracker*,
1783 Nucl.Instrum.Meth. **A568** (2006) 642–671. [3.2](#)
- 1784 [39] ATLAS Collaboration, *The ATLAS semiconductor tracker end-cap module*, Nucl.Instrum.Meth.
1785 **A575** (2007) 353–389. [3.2](#)
- 1786 [40] The ATLAS TRT Collaboration, *The ATLAS Transition Radiation Tracker (TRT) proportional
1787 drift tube: Design and performance*, JINST **3** (2008) P02013. [3.2](#)
- 1788 [41] The ATLAS TRT Collaboration, *The ATLAS TRT barrel detector*, JINST **3** (2008) P02014. [3.2](#)
- 1789 [42] E. Abat et al., *The ATLAS TRT electronics*, J. Instrum. **3** (2008) P06007. [3.2](#)
- 1790 [43] T. A. Collaboration, *ATLAS liquid argon calorimeter: Technical design report*, .
1791 CERN-LHCC-96-41. [3.2](#)
- 1792 [44] ATLAS Collaboration, *ATLAS tile calorimeter: Technical Design Report*. Technical Design
1793 Report ATLAS. CERN, Geneva, 1996. <https://cdsweb.cern.ch/record/331062>. [3.2](#)
- 1794 [45] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*. Technical Design
1795 Report ATLAS. CERN, Geneva, 1997. <https://cdsweb.cern.ch/record/331068>. [3.2](#)
- 1796 [46] G. Aielli, A. Aloisio, M. Alviggi, V. Aprodu, V. Bocci, et al., *The RPC first level muon trigger
1797 in the barrel of the ATLAS experiment*, Nucl.Phys.Proc.Suppl. **158** (2006) 11–15. [3.2](#)
- 1798 [47] F. Bauer, U. Bratzler, H. Dietl, H. Kroha, T. Lagouri, et al., *Construction and test of MDT
1799 chambers for the ATLAS muon spectrometer*, Nucl.Instrum.Meth. **A461** (2001) 17–20. [3.2](#)

- 1800 [48] T. Argyropoulos, K. A. Assamagan, B. H. Benedict, V. Chernyatin, E. Cheu, et al., *Cathode*
 1801 *strip chambers in ATLAS: Installation, commissioning and in situ performance*, IEEE
 1802 *Trans.Nucl.Sci.* **56** (2009) 1568–1574. [3.2](#)
- 1803 [49] ATLAS Collaboration, *ATLAS level-1 trigger: Technical Design Report*. Technical Design
 1804 Report ATLAS. CERN, Geneva, 1998. <https://cdsweb.cern.ch/record/381429>. [3.2](#)
- 1805 [50] P. Jenni, M. Nessi, M. Nordberg, and K. Smith, *ATLAS high-level trigger, data-acquisition and*
 1806 *controls: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 2003.
 1807 <https://cdsweb.cern.ch/record/616089>. [3.2](#)
- 1808 [51] *Electron efficiency measurements with the ATLAS detector using the 2012 LHC proton-proton*
 1809 *collision data*, Tech. Rep. ATLAS-CONF-2014-032, CERN, Geneva, Jun, 2014. [3.2.6.2](#), [7.4](#),
 1810 [9.2.1](#)
- 1811 [52] *Preliminary results on the muon reconstruction efficiency, momentum resolution, and*
 1812 *momentum scale in ATLAS 2012 pp collision data*, Tech. Rep. ATLAS-CONF-2013-088, CERN,
 1813 Geneva, February, 2013. [3.2.6.3](#), [7.5](#)
- 1814 [53] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$ jet clustering algorithm*, *JHEP* **0804**
 1815 (2008) 063, [arXiv:0802.1189 \[hep-ph\]](https://arxiv.org/abs/0802.1189). [3.2.6.4](#), [7.6](#)
- 1816 [54] ATLAS Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in*
 1817 *the 7 TeV collision data*, Tech. Rep. ATLAS-CONF-2011-102, CERN, Geneva, Jul, 2011.
 1818 <https://cdsweb.cern.ch/record/1369219>. [3.2.6.5](#), [7.6](#)
- 1819 [55] ATLAS Collaboration Collaboration, G. Aad et al., *Search for $H \rightarrow \gamma\gamma$ produced in association*
 1820 *with top quarks and constraints on the Yukawa coupling between the top quark and the Higgs*
 1821 *boson using data taken at 7 TeV and 8 TeV with the ATLAS detector*, [arXiv:1409.3122](https://arxiv.org/abs/1409.3122)
 1822 [\[hep-ex\]](#). [5](#)
- 1823 [56] *Search for the Standard Model Higgs boson produced in association with top quarks and decaying*
 1824 *to $b\bar{b}$ in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector at the LHC*, Tech. Rep.
 1825 ATLAS-CONF-2014-011, CERN, Geneva, Mar, 2014. [5](#)
- 1826 [57] ATLAS Collaboration, G. Aad et al., *Improved luminosity determination in pp collisions at \sqrt{s}*
 1827 *= 7 TeV using the ATLAS detector at the LHC*, *Eur.Phys.J.* **C73** (2013) 2518, [arXiv:1302.4393](https://arxiv.org/abs/1302.4393)
 1828 [\[hep-ex\]](#). [6.1.1](#)
- 1829 [58] CERN, . CERN, Geneva, 2012. [6.1.1](#)
- 1830 [59] ATLAS Collaboration Collaboration, G. Aad et al., *The ATLAS Simulation Infrastructure*,
 1831 *Eur.Phys.J.* **C70** (2010) 823–874, [arXiv:1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568). [6.2](#)
- 1832 [60] GEANT4 Collaboration, S. Agostinelli et al., *GEANT4: A Simulation toolkit*,
 1833 *Nucl.Instrum.Meth.* **A506** (2003) 250–303. [6.2](#)
- 1834 [61] G. Bevilacqua, M. Czakon, M. Garzelli, A. van Hameren, A. Kardos, C. Papadopoulos,
 1835 R. Pittau, and M. Worek. [6.2.1](#)
- 1836 [62] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, *11*
 1837 (2004) 040. [6.2.1](#)
- 1838 [63] S. Frixione, P. Nason, and C. Oleari **11** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](https://arxiv.org/abs/0709.2092). [6.2.1](#)

- [64] S. Alioli, P. Nason, C. Oleari, and E. Re **06** (2010) 040, [arXiv:1002.2581 \[hep-ph\]](#). 6.2.1
- [65] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to Pythia 8.1*, [arXiv:0710.3820 \[hep-ph\]](#). 6.2.1
- [66] LHC Higgs Cross Section Working Group Collaboration, S. Heinemeyer et al., *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, [arXiv:1307.1347 \[hep-ph\]](#). 6.2.1, 9.1
- [67] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, **JHEP 0307** (2003) 001, [arXiv:hep-ph/0206293 \[hep-ph\]](#). 6.2.2
- [68] F. Maltoni and T. Stelzer, *MadEvent: Automatic event generation with MadGraph*, **JHEP 0302** (2003) 027, [arXiv:hep-ph/0208156 \[hep-ph\]](#). 6.2.2
- [69] B. P. Kersevan and E. Richter-Was, *The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1*, **Comput.Phys.Commun.** **184** (2013) 919–985, [arXiv:hep-ph/0405247 \[hep-ph\]](#). 6.2.2
- [70] P. M. Nadolsky, H.-L. Lai, Q.-H. Cao, J. Huston, J. Pumplin, et al., *Implications of CTEQ global analysis for collider observables*, **Phys.Rev. D78** (2008) 013004, [arXiv:0802.0007 \[hep-ph\]](#). 6.2.2
- [71] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, **JHEP 0711** (2007) 070, [arXiv:0709.2092 \[hep-ph\]](#). 6.2.2
- [72] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, **JHEP 0902** (2009) 007, [arXiv:0811.4622 \[hep-ph\]](#). 6.2.2
- [73] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, **Eur.Phys.J. C63** (2009) 189–285, [arXiv:0901.0002 \[hep-ph\]](#). 6.2.2, 8.1
- [74] M. Garzelli, A. Kardos, C. Papadopoulos, and Z. Trocsanyi, *$t\bar{t}W^\pm$ and $t\bar{t}Z$ Hadroproduction at NLO accuracy in QCD with Parton Shower and Hadronization effects*, **JHEP 1211** (2012) 056, [arXiv:1208.2665 \[hep-ph\]](#). 8.1
- [75] J. M. Campbell and R. K. Ellis, *$t\bar{t}W^{+-}$ production and decay at NLO*, **JHEP 1207** (2012) 052, [arXiv:1204.5678 \[hep-ph\]](#). 8.1
- [76] Campbell, John and Ellis, R. Keith and Röntsch, Raoul, *Single top production in association with a Z boson at the LHC*, **Phys.Rev. D87** (2013) 114006, [arXiv:1302.3856 \[hep-ph\]](#). 8.1
- [77] ATLAS Collaboration, G. Aad et al., *Measurement of WZ production in proton-proton collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, **Eur.Phys.J. C72** (2012) 2173, [arXiv:1208.1390 \[hep-ex\]](#). 8.2
- [78] ATLAS Collaboration, G. Aad et al., *Measurement of ZZ production in pp collisions at $\sqrt{s} = 7$ TeV and limits on anomalous ZZZ and $ZZ\gamma$ couplings with the ATLAS detector*, **JHEP 1303** (2013) 128, [arXiv:1211.6096 \[hep-ex\]](#). 8.2
- [79] ATLAS Collaboration, G. Aad et al., *Measurement of the cross-section for W boson production in association with b -jets in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, **JHEP 1306** (2013) 084, [arXiv:1302.2929 \[hep-ex\]](#). 8.2

- 1877 [80] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of differential production*
 1878 *cross-sections for a Z boson in association with b-jets in 7 TeV proton-proton collisions with*
 1879 *the ATLAS detector*, arXiv:1407.3643 [hep-ex]. 8.2
- 1880 [81] S. Guindon, E. Shabalina, J. Adelman, M. Alhroob, S. Amor dos Santos, A. Basye, J. Bouffard,
 1881 M. Casolino, I. Connelly, A. Cortes Gonzalez, V. Dao, S. D'Auria, A. Doyle, P. Ferrari,
 1882 F. Filthaut, R. Goncalo, N. de Groot, S. Henkelmann, V. Jain, A. Juste, G. Kirby, D. Kar,
 1883 A. Knue, K. Kroeninger, T. Liss, E. Le Menedeu, J. Montejo Berlingen, M. Moreno Llacer,
 1884 O. Nackenhorst, T. Neep, A. Onofre, M. Owen, M. Pinamonti, Y. Qin, A. Quadt, D. Quilty,
 1885 C. Schwanenberger, L. Serkin, R. St Denis, J. Thomas-Wilsker, and T. Vazquez-Schroeder,
 1886 *Search for the Standard Model Higgs boson produced in association with top quarks and decaying*
 1887 *to bb in pp collisions at sqrt(s) = 8 TeV with the ATLAS detector at the LHC*, Tech. Rep.
 1888 ATL-COM-PHYS-2013-1659, CERN, Geneva, Dec, 2013. The note contains internal
 1889 documentation of the ttH(bb) analysis approved as a preliminary result
 1890 (ATLAS-CONF-2014-011). 9.1
- 1891 [82] ATLAS Collaboration Collaboration, G. Aad et al., *Measurement of the muon reconstruction*
 1892 *performance of the ATLAS detector using 2011 and 2012 LHC proton-proton collision data*,
 1893 arXiv:1407.3935 [hep-ex]. 9.2.1, 9.2.1
- 1894 [83] ATLAS Collaboration Collaboration, G. Aad et al., *Electron and photon energy calibration with*
 1895 *the ATLAS detector using LHC Run 1 data*, Eur.Phys.J. C74 (2014) no. 10, 3071,
 1896 arXiv:1407.5063 [hep-ex]. 9.2.1
- 1897 [84] ATLAS Collaboration Collaboration, *Performance of the ATLAS muon trigger in 2011*, Tech.
 1898 Rep. ATLAS-CONF-2012-099, CERN, Geneva, Jul, 2012. 9.2.1
- 1899 [85] ATLAS Collaboration Collaboration, *Performance of the ATLAS Electron and Photon Trigger*
 1900 *in p-p Collisions at sqrt(s) = 7 TeV in 2011*, Tech. Rep. ATLAS-CONF-2012-048, CERN,
 1901 Geneva, May, 2012. 9.2.1
- 1902 [86] ATLAS Collaboration Collaboration, G. Aad et al., *Jet energy measurement and its systematic*
 1903 *uncertainty in proton-proton collisions at sqrt(s) = 7 TeV with the ATLAS detector*,
 1904 arXiv:1406.0076 [hep-ex]. 9.2.3
- 1905 [87] ATLAS Collaboration Collaboration, G. Aad et al., *Jet energy resolution in proton-proton*
 1906 *collisions at sqrt(s) = 7 TeV recorded in 2010 with the ATLAS detector*, Eur.Phys.J. C73 (2013)
 1907 2306, arXiv:1210.6210 [hep-ex]. 9.2.3
- 1908 [88] ATLAS Collaboration Collaboration, *Measurement of the Mistag Rate with 5 fb^-1 of Data*
 1909 *Collected by the ATLAS Detector*, Tech. Rep. ATLAS-CONF-2012-040, CERN, Geneva, 2012.
 1910 9.2.4
- 1911 [89] ATLAS Collaboration Collaboration, *Measuring the b-tag efficiency in a top-pair sample with*
 1912 *4.7 fb^-1 of data from the ATLAS detector*, Tech. Rep. ATLAS-CONF-2012-097, CERN,
 1913 Geneva, 2012. 9.2.4
- 1914 [90] ATLAS Collaboration Collaboration, *b-jet tagging calibration on c-jets containing D?+ mesons*,
 1915 Tech. Rep. ATLAS-CONF-2012-039, CERN, Geneva, 2012. 9.2.4
- 1916 [91] Glen Cowan, Kyle Cranmer, Eilam Gross, Ofer Vitells, *Asymptotic formulae for likelihood-based*
 1917 *tests of new physics*, Eur.Phys.J.C 71 (2011) 1554. 10.2

- 1918 [92] A. L. Read, *Presentation of search results: the CL s technique*, Journal of Physics G: Nuclear
1919 and Particle Physics **28** (2002) no. 10, 2693.
1920 <http://stacks.iop.org/0954-3899/28/i=10/a=313>. 10.2.3