

¹

©Copyright 2021

²

Sean Gasiorowski

$HH \rightarrow b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

Sean Gasiorowski

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

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Anna Goussiou, Chair

Jason Detwiler

Shih-Chieh Hsu

David Kaplan

Henry Lubatti

Thomas Quinn

Gordon Watts

Program Authorized to Offer Degree:
Physics

22

University of Washington

23

Abstract

24

$HH \rightarrow b\bar{b}b\bar{b}$ or How I Learned to Stop Worrying and Love the QCD Background

25

Sean Gasiorowski

26

Chair of the Supervisory Committee:

27

Professor Anna Goussiou

28

Physics

29

Insert abstract here

TABLE OF CONTENTS

	Page
31 List of Figures	iii
32 Glossary	xv
33 Chapter 1: The Standard Model of Particle Physics	1
34 1.1 Introduction: Particles and Fields	1
35 1.2 Quantum Electrodynamics	4
36 1.3 An Aside on Group Theory	9
37 1.4 Quantum Chromodynamics	11
38 1.5 The Weak Interaction	14
39 1.6 The Higgs Potential and the SM	20
40 1.7 The Standard Model: A Summary	26
41 Chapter 2: Di-Higgs Phenomenology and Physics Beyond the Standard Model	28
42 2.1 Intro to Di-Higgs	28
43 2.2 Resonant HH Searches	29
44 2.3 Non-resonant HH Searches	31
45 Chapter 3: Experimental Apparatus	35
46 3.1 The Large Hadron Collider	35
47 3.2 The ATLAS Experiment	38
48 Chapter 4: Simulation	47
49 4.1 Event Generation	47
50 4.2 Detector Simulation	50
51 4.3 Correlated Fluctuations in FastCaloSim	52
52 4.4 Outlook	60

53	Chapter 5:	Reconstruction	62
54	5.1	Jets	62
55	5.2	Flavor Tagging	64
56	Chapter 6:	The Anatomy of an LHC Search	74
57	6.1	Object Selection and Identification	74
58	6.2	Defining a Signal Region	74
59	6.3	Background Estimation	74
60	6.4	Uncertainty Estimation	74
61	6.5	Hypothesis Testing	74
62	Chapter 7:	Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state	75
63	7.1	Data and Monte Carlo Simulation	77
64	7.2	Triggers and Object Definitions	79
65	7.3	Analysis Selection	81
66	7.4	Background Reduction and Top Veto	87
67	7.5	Kinematic Region Definition	88
68	7.6	Background Estimation	94
69	7.7	Uncertainties	155
70	7.8	Background Validation	165
71	7.9	Overview of Other $b\bar{b}b\bar{b}$ Channels	168
72	7.10	m_{HH} Distributions	170
73	7.11	Statistical Analysis	181
74	7.12	Results	182
75	Chapter 8:	Future Ideas for $HH \rightarrow b\bar{b}b\bar{b}$	187
76	8.1	pairAGraph: A New Method for Jet Pairing	187
77	8.2	Background Estimation with Mass Plane Interpolation	192
78	Chapter 9:	Conclusions	207

LIST OF FIGURES

Figure Number		Page
80	1.1 Diagram of the elementary particles described by the Standard Model [1].	3
81	2.1 Representative diagrams for the gluon-gluon fusion production of spin-0 (X) 82 and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs 83 bosons. The spin-0 resonance considered for this thesis is a generic narrow 84 width resonance which may be interpreted in the context of two Higgs doublet 85 models [18], whereas the spin-2 resonance is considered as a Kaluza-Klein 86 graviton within the bulk Randall-Sundrum (RS) model [19, 20].	29
87	2.2 Dominant contributing diagrams for non-resonant gluon-gluon fusion produc- 88 tion of HH . κ_λ and κ_t represent variations of the Higgs self-coupling and 89 coupling to top quarks respectively, relative to that predicted by the Standard 90 Model.	30
91	2.3 Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common 92 decay mode, representing 34 % of all HH events produced at the LHC.	31
93	2.4 Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demon- 94 strating the impact of the interference between the two diagrams of Figure 2.2 95 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram con- 96 tribution, demonstrating the shape of the box diagram contribution, whereas 97 for $\kappa_\lambda = 10$, the triangle diagram dominates, with a strong low mass peak. 98 The interplay between the two is quite evident for other values, resulting in, 99 e.g., the double peaked structure present for $\kappa_\lambda = 2$ (near maximal destructive 100 interference) and $\kappa_\lambda = -5$. [21]	33
101	3.1 Diagram of the ATLAS detector [41]	38
102	3.2 Cross section of the ATLAS detector showing how particles interact with 103 various detector components [43]	40
104	3.3 2D projections of the ATLAS coordinate system	41

105	4.1	Schematic diagram of the Monte Carlo simulation of a hadron-hadron collision. The incoming hadrons are the green blobs with the arrows on the left and right, with the red blob in the center representing the hard scatter event, and the purple representing a secondary hard scatter. Radiation from both incoming and outgoing particles is shown, and the light green blobs represent hadronization, with the outermost dark green circles corresponding to the final state hadrons. Yellow lines are radiated photons. [57]	49
112	4.2	The projected ATLAS computational requirements for Run 3 and the HL-LHC relative to the projected computing budget. Aggressive R&D is required to keep resources within budget [60].	51
115	4.3	Energy and variable weta2, defined as the energy weighted lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [62].	53
118	4.4	Example of photon and pion average shapes in 5×5 calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all close to 1, while the pion ratios show significant deviation from the average.	55
123	4.5	Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing α , left to right corresponds to increasing R , with core voxels numbered 1, 10, 19, Agreement is quite good across all voxels. Results are similar for the VAE method.	56
129	4.6	Correlation coefficient of ratios of voxel energy in single events to the corresponding voxel energy in the average shape, examined between the core bin from $\alpha = 0$ to $2\pi/8$ and each of the other voxels. The periodic structure represents the binning in α , and the increasing numbers in each of these periods correspond to increasing R , where the eight points with correlation coefficient 1 are the eight core bins. Both the Gaussian and VAE generated toy events are able to reproduce the major correlation structures for 65 GeV central pions in EMB2.	57
137	4.7	Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies, η points, and layers.	58

140	4.8 Comparison of the Gaussian fluctuation model to the default FCSV2 version	59
141	and to G4 simulation, using pions of 65 GeV energy and $0.2 < \eta < 0.25$.	
142	Variables shown relate to calorimeter clusters, three-dimensional spatial group-	
143	ings of cells [68] which provide powerful insight on the structure of energy	
144	deposits in the calorimeter. Variables considered include number and energy	
145	of clusters, the 2nd moment in lambda, ($\langle \lambda^2 \rangle$), which describes the square	
146	of the longitudinal extension of a cluster, where λ is the distance of a cell from	
147	the shower center along the shower axis, and a cluster moment is defined as	
148	$\langle x^n \rangle = \frac{\sum E_i x_i}{\sum E_i}$, and the distance ΔR , between the cluster and the true pion.	
149	With the correlated fluctuations, variables demonstrate improved modeling	
150	relative to default FastCaloSimV2.	
151	5.1 Illustration of an interaction producing two light jets and one b -jet in the	67
152	transverse plane. While the light jets decay “promptly”, coinciding with	
153	the primary vertex of the proton-proton interaction, the longer lifetime of B	
154	hadrons leads to a secondary decay vertex, displaced from the primary vertex	
155	by length L_{xy} . This is typically a few mm, and therefore is not directly visible	
156	in the detector, but leads to a large transverse impact parameter, d_0 , for the	
157	resulting tracks. [76]	
158	5.2 Performance of the various low and high level flavor tagging algorithms in	71
159	$t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light	
160	and c -jet rejection. The high level taggers demonstrate significantly better	
161	performance than any of the individual low level taggers, with DL1 offering	
162	slight improvements over MV2 due to the inclusion of additional input variables.	
163	5.3 Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demon-	72
164	strating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c	
165	controls the importance of c -jet rejection in the discriminating variable, and	
166	values shown have been optimized separately for each DL1 configuration. DL1r	
167	demonstrates a significant improvement in both light and c jet rejection over	
168	MV2 and DL1. [81]	
169	7.1 Comparison of m_{H1} vs m_{H2} planes for the full Run 2 $2b$ dataset with different	85
170	pairings. As evidenced, this choice significantly impacts where events fall in	
171	this plane, and therefore which events fall into the various kinematic regions	
172	defined in this plane (see Section 7.5). Respective signal regions are shown for	
173	reference, with the min ΔR signal region shifted slightly up and to the right to	
174	match the non-resonant selection. Note that the band structure around 80 GeV	
175	in both m_{H1} and m_{H2} is introduced by the top veto described in Section 7.4.	

176	7.2 Comparison of signal distributions in the respective signal regions for the min ΔR and D_{HH} pairing for various values of the Higgs trilinear coupling in the respective signal regions. The distributions are quite similar at the Standard Model point, but for other variations, min ΔR does not pick up the low mass features.	86
181	7.3 Regions used for the resonant search, shown on the $2b$ data mass plane. The outermost region (the “control region”) is used for derivation of the nominal background estimate. The innermost region is the signal region, where the signal extraction fit is performed. The region in between (the “validation region”) is used for the assessment of an uncertainty.	89
186	7.4 Regions used for the non-resonant search. The “top” and “bottom” quadrants together comprise the control region, in which the nominal background estimate is derived. The “left” and “right” quadrants together comprise the validation region, which is used to assess an uncertainty. The signal region, in the center, is where the signal extraction fit is performed.	91
191	7.5 Impact of the m_{HH} correction on a range of spin-0 resonant signals. The corrected m_{HH} distributions (solid lines) are much sharper and more centered on the corresponding resonance masses than the uncorrected m_{HH} distributions (dashed).	92
195	7.6 Resonant Search: Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Control Region.	101
198	7.7 Resonant Search: Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.	102
201	7.8 Resonant Search: Distributions of the number of jets before and after CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	103
204	7.9 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	104
207	7.10 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region.	105
210	7.11 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	106

213	7.12 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.	107
216	7.13 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.	108
219	7.14 Resonant Search: Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied	109
222	7.15 Resonant Search: Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Validation Region.	110
225	7.16 Resonant Search: Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.	111
228	7.17 Resonant Search: Distributions of the number of jets before and after CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	112
231	7.18 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	113
234	7.19 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region.	114
237	7.20 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	115
240	7.21 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.	116
243	7.22 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.	117
246	7.23 Resonant Search: Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied	118

249	7.24 Non-resonant Search (4b): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Control Region.	119
250		
251		
252	7.25 Non-resonant Search (4b): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.	120
253		
254		
255	7.26 Non-resonant Search (4b): Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	121
256		
257		
258	7.27 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	122
259		
260		
261	7.28 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Control Region.	123
262		
263		
264	7.29 Non-resonant Search (4b): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	124
265		
266		
267	7.30 Non-resonant Search (4b): Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.	125
268		
269		
270	7.31 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.	126
271		
272		
273	7.32 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.	127
274		
275		
276	7.33 Non-resonant Search (4b): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Validation Region.	128
277		
278		
279	7.34 Non-resonant Search (4b): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.	129
280		
281		
282	7.35 Non-resonant Search (4b): Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	130
283		
284		

285	7.36 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	131
288	7.37 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region.	132
291	7.38 Non-resonant Search (4b): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	133
294	7.39 Non-resonant Search (4b): Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.	134
297	7.40 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.	135
300	7.41 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.	136
303	7.42 Non-resonant Search (3b1l): Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Control Region.	137
307	7.43 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.	138
310	7.44 Non-resonant Search (3b1l): Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.	139
313	7.45 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.	140
316	7.46 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region.	141
319	7.47 Non-resonant Search (3b1l): Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.	142

322	7.48 Non-resonant Search (3b1l): Distributions of mass of the leading and 323 subleading Higgs candidates and of the di-Higgs system before and after CR 324 derived reweighting for the 2018 3b1l Control Region.	143
325	7.49 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the lead- 326 ing and subleading Higgs candidates before and after CR derived reweighting 327 for the 2018 3b1l Control Region.	144
328	7.50 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , 329 before and after CR derived reweighting for the 2018 3b1l Control Region. 330 Reweighting is done after the cut on this variable is applied.	145
331	7.51 Non-resonant Search (3b1l): Distributions of ΔR between the closest 332 Higgs Candidate jets, ΔR between the other two, and average absolute value of 333 HC jet η before and after CR derived reweighting for the 2018 3b1l Validation 334 Region.	146
335	7.52 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading 336 Higgs Candidate jets and the p_T of the di-Higgs system before and after CR 337 derived reweighting for the 2018 3b1l Validation Region.	147
338	7.53 Non-resonant Search (3b1l): Distributions of the number of jets before 339 and after CR derived reweighting for the 2018 3b1l Validation Region. A 340 minimum of 4 jets is required in each event in order to form Higgs candidates.	148
341	7.54 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading 342 Higgs Candidate jets and ΔR between Higgs candidates before and after CR 343 derived reweighting for the 2018 3b1l Validation Region.	149
344	7.55 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd 345 leading Higgs Candidate jets before and after CR derived reweighting for the 346 2018 3b1l Validation Region.	150
347	7.56 Non-resonant Search (3b1l): Distributions of η of the 4th leading Higgs 348 Candidate jet and the p_T of the leading and subleading Higgs candidates before 349 and after CR derived reweighting for the 2018 3b1l Validation Region.	151
350	7.57 Non-resonant Search (3b1l): Distributions of mass of the leading and 351 subleading Higgs candidates and of the di-Higgs system before and after CR 352 derived reweighting for the 2018 3b1l Validation Region.	152
353	7.58 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the lead- 354 ing and subleading Higgs candidates before and after CR derived reweighting 355 for the 2018 3b1l Validation Region.	153
356	7.59 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , 357 before and after CR derived reweighting for the 2018 3b1l Validation Region. 358 Reweighting is done after the cut on this variable is applied.	154

359	7.60 Illustration of the approximate bootstrap band procedure, shown as a ratio to 360 the nominal estimate for the 2017 non-resonant background estimate. Each grey 361 line is from the m_{HH} prediction for a single bootstrap training. Figure 7.60(a) 362 shows the variation histograms constructed from median weight \pm the IQR of 363 the replica weights. It can be seen that this captures the rough shape of the 364 bootstrap envelope, but is not good estimate for the overall magnitude of the 365 variation. Figure 7.60(b) demonstrates the applied normalization correction, 366 and Figure 7.60(c) shows the final band (normalized Figure 7.60(a) + Figure 367 7.60(b)). Comparing this with the IQR variation for the prediction from each 368 bootstrap in each bin in Figure 7.60(d), the approximate envelope describes a 369 very similar variation.	159
370	7.61 Resonant Search: Example of CR vs VR variation in each H_T region for 371 2018. The variation nicely factorizes into low and high mass components.	161
372	7.62 Non-resonant Search (4b): Example of CR vs VR variation in each signal 373 region quadrant for 2018. Significantly different behavior is seen between 374 quadrants, with the largest variation in quadrant 1 and the smallest in quadrant 375 4.	162
376	7.63 Resonant Search: Performance of the background estimation method in the 377 resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background 378 estimate is trained following nominal procedures entirely within the reversed 379 $\Delta\eta_{HH}$ region, and the resulting model, including uncertainties, is compared 380 with $4b$ data in the corresponding signal region. Good agreement is shown. 381 The quoted p -value uses the χ^2 test statistic, and demonstrates no evidence 382 that the data differs from the assessed background.	166
383	7.64 Non-resonant Search: Performance of the background estimation method 384 in the $3b + 1$ fail validation region. A new background estimate is trained 385 following nominal procedures but with a reweighting from $2b$ to $3b + 1$ fail 386 events. Generally good agreement is seen, though there is some deviation at 387 very low masses in the low $\Delta\eta_{HH}$ low X_{HH} category.	167
388	7.65 Resonant Search: Demonstration of the performance of the nominal reweight- 389 ing in the control region on corrected m_{HH} , with Figure 7.65(a) showing 390 $2b$ events normalized to the total $4b$ yield and Figure 7.65(b) applying the 391 reweighting procedure. Agreement is much improved with the reweighting. 392 Note that overall reweighted $2b$ yield agrees with $4b$ yield in the control region 393 by construction.	171

394	7.66 Resonant Search: Demonstration of the performance of the control region derived reweighting in the validation region on corrected m_{HH} . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined by a reweighting trained in the validation region.	172
399	7.67 Resonant Search: Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The closure is generally quite good, though there is an evident deficit in the background estimate relative to the data for higher values of corrected m_{HH}	173
403	7.68 Non-resonant Search (4b): Demonstration of the performance of the nomi- nal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with some residual mismodeling in the low $\Delta\eta_{HH}$ region near 600 GeV.	175
407	7.69 Non-resonant Search (3b1l): Demonstration of the performance of the nominal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with similar conclusions as for the 4b region.	176
410	7.70 Non-resonant Search (4b): Demonstration of the performance of the nomi- nal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. The low $\Delta\eta_{HH}$ region is consistently overestimated, but, systematic uncertainties are defined via the difference between VR and CR estimates. .	177
414	7.71 Non-resonant Search (3b1l): Demonstration of the performance of the nominal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. A deficit is present near 600 GeV, but agreement is fairly good otherwise.	178
417	7.72 Non-resonant Search (4b): Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 4b channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Modeling is generally quite good near the Standard Model peak, but disagreements are seen at very low and high masses. A deficit is present in low $\Delta\eta_{HH}$ bins near 600 GeV.	179
423	7.73 Non-resonant Search (3b1l): Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 3b1l channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Conclusions are very similar to the 4b channels, with generally good modeling near the Standard Model peak, but disagreements at very low and high masses. A deficit is present near 600 GeV.	180

429	7.74 Expected (dashed black) and observed (solid black) 95% CL upper limits 430 on the cross-section times branching ratio of resonant production for spin-0 431 ($X \rightarrow HH$) and spin-2 $G_{KK}^* \rightarrow HH$. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the 432 expected limits are shown in the colored bands. The resolved channel expected 433 limit is shown in dashed pink and covers the range from 251 and 1500 GeV. It 434 is combined with the boosted channel (dashed blue) between 900 and 1500 GeV. 435 The theoretical prediction for the bulk RS model with $k/\bar{M}_{\text{Pl}} = 1$ [20] (solid 436 red line) is shown, with the decrease below 350 GeV due to a sharp reduction 437 in the $G_{KK}^* \rightarrow HH$ branching ratio. The nominal $H \rightarrow b\bar{b}$ branching ratio is 438 taken as 0.582.	184
439	7.75 Expected (dashed black) and observed (solid black) 95% CL upper limits on 440 the cross-section times branching ratio of non-resonant production for a range 441 of values of the Higgs self-coupling, with the Standard Model value ($\kappa_\lambda = 1$) 442 illustrated with a star. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the expected limits are 443 shown in the colored bands. The cross section limit for HH production is set 444 at 140 fb (180 fb) observed (expected), corresponding to an observed (expected) 445 limit of 4.4 (5.9) times the Standard Model prediction. κ_λ is constrained to 446 be within the range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$ expected). 447 The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582. We note that the 448 excess present for $\kappa_\lambda \geq 5$ is thought to be due to a low mass background 449 mis-modeling, present due to the optimization of this analysis for the Standard 450 Model point, and is not present in more sensitive channels in this same region 451 (e.g. $HH \rightarrow bb\gamma\gamma$).	185
452	7.76 Comparison of the limits in Figure 7.75 with an equivalent set of limits that 453 drop the m_{HH} bins below 381 GeV, with the value of 381 GeV determined 454 by the optimized variable width binning. The expected limit band with this 455 mass cut is shown in dashed blue, and the observed is shown in solid blue. 456 The excess at and above $\kappa_\lambda = 5$ is significantly reduced, demonstrating that 457 this is driven by low mass. Notably, there is minimal impact on the expected 458 sensitivity with this m_{HH} cut.	186
459	8.1 Comparison of distributions of $HH p_T$ in the 2018 resonant validation region 460 before reweighting for BDT pairing (left) and pairAGraph (right). $HH p_T$ is 461 a variable with a large difference between 2 b and 4 b , but the relative shapes 462 seem to be more similar for pairAGraph than for BDT paring, corresponding 463 to the hypothesis that pairAGraph chooses more “4 b -like” jets.	192

464	8.2 Comparison of distributions of $HH\ p_T$ in the 2018 resonant validation region after reweighting for BDT pairing (left) and pairAGraph (right). $HH\ p_T$ is a variable with a large difference between $2b$ and $4b$, and the reweighted agreement in the high p_T tail is significantly improved with pairAGraph, with a corresponding reduction in the assigned bootstrap uncertainty in that region.	193
465		
466		
467		
468		
469	8.3 Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for $2b$ signal region events compared to real $2b$ signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics.	200
470		
471		
472		
473	8.4 Gaussian process sampling prediction for the mass plane compared to the real $2b$ dataset for 2018. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics. Good agreement is seen.	202
474		
475		
476		
477	8.5 Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for $4b$ signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.	203
478		
479		
480	8.6 Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible. .	204
481		
482	8.7 Comparison of the interpolation background estimate with real $2b$ data in the signal region. Only 1 % of $2b$ data is used in order to mimic $4b$ statistics, and results are presented here summed across years. The “Flow Only” prediction uses samples of actual $2b$ signal region data for the input values of m_{H1} and m_{H2} , whereas the “Flow + GP” prediction uses samples following the Gaussian process procedure above, more closely mimicking a the full background estima- tion procedure. The two predictions are quite comparable, demonstrating the closure of the Gaussian process estimate, and the predicted m_{HH} shape agrees well with $2b$ data. Only $2b$ statistical uncertainty is shown.	205
483		
484		
485		
486		
487		
488		
489		
490		
491	8.8 Comparison of the interpolation background estimate in the $4b$ signal region with the control region derived reweighted $2b$ estimate, shown for each year individually. Results are generally similar, within around 10 %.	206
492		
493		

GLOSSARY

495 ARGUMENT: replacement text which customizes a L^AT_EX macro for each particular usage.

ACKNOWLEDGMENTS

497 Five years is both a short time and a long time – many things have happened and many
498 have stayed the same. I certainly know much more physics than I did at the outset, but also
499 have learned to ski, discovered a love for hiking, eaten large amounts of cheese, and survived
500 a pandemic by making sourdough and cinnamon rolls. Of course, the most important part of
501 any journey is the friends we made along the way – the utmost gratitude and appreciation
502 goes to the Seattle friends, for the many nights both hard at work and at College Inn, to
503 the CERN friends, for adventures around Europe and days by the lake, the Chicago friends,
504 for my position as European correspondent for the Dum Dum Donut Intellectuals, and the
505 friends from home, for New Years Eve parties and visits in between. Though we’re scattered
506 across the world, I hope to see you all soon.

507 A thank you, of course, to my family for their continued and constant support, for trying
508 their best to learn physics along with me, and going along on this Ph.D. adventure every
509 step of the way.

510 I am only writing this thesis because of the incredible support of my research group, so a
511 massive thank you to Anna, for your guidance and compassion, for advising me as both a
512 physicist and a person, and for always being available, and Jana, for shaping who I am as a
513 physicist and providing feedback on almost literally every talk that I’ve written – I knew I
514 was getting ready to graduate when you started running out of comments.

515 A special set of thank yous to the SLAC group, namely Nicole, Michael, and Rafael, for
516 being an excellent set of collaborators with an excellent set of ideas to push the analysis
517 forward, many of which are in this thesis, and for being my second research family.

518 And last but not least, a thank you to the entire $HH \rightarrow 4b$ group, of course to Max

519 and Rafael for their guidance and direction, Bejan for being an excellent collaborator and
520 co-editor for the resonant search, Dale and Alex for excelling on the boosted side of things,
521 Lucas, for great work on triggers and for diving headfirst into all of my code, the ggF push
522 squad, for a huge amount of hard work on the non-resonant, and of course the $HH \rightarrow 4\text{beers}$
523 team, many of whom have already been mentioned, but who deserve an extra shout out for
524 keeping things fun even during stressful times.

525 The physics is done, the rest is paperwork. Let us begin.

526

DEDICATION

527

To family, both given and found

528

Chapter 1

529

THE STANDARD MODEL OF PARTICLE PHYSICS

530

The Standard Model of Particle Physics (SM) is a monumental historical achievement, providing a formalism with which one may describe everything from the physics of everyday experience to the physics that is studied at very high energies at the Large Hadron Collider (Chapter 3). In this chapter, we will provide a brief overview of the pieces that go into the construction of such a model. The primary focus of this thesis is searches for pair production of Higgs bosons decaying to four b -quarks. Consequently, we will pay particular attention to the relevant pieces of the Higgs Mechanism, as well as the theory behind searches at a hadronic collider.

538

1.1 Introduction: Particles and Fields

539

What is a particle? The Standard Model describes a set of fundamental, point-like, objects shown in Figure 1.1. These objects have distinguishing characteristics (e.g., mass and spin). These objects interact in very specific ways. The set of objects and their interactions result in a set of observable effects, and these effects are the basis of a field of experimental physics.

543

The effects of these objects and their interactions are familiar as fundamental forces: electromagnetism (photons, electrons), the strong interaction (quarks, gluons), the weak interaction (neutrinos, W and Z bosons). Gravity is not described in this model, as the weakest, with effects most relevant on much larger distance scales than the rest. However, the description of these other three is powerful – verifying and searching for cracks in this description is a large effort, and the topic of this thesis.

549

The formalism for describing these particles and their interactions is that of quantum field theory. Classical field theory is most familiar in the context of, e.g., electromagnetism – an

551 electric field exists in some region of space, and a charged point-particle experiences a force
552 characterized by the charge of the point-particle and the magnitude of the field at the location
553 of the point-particle in spacetime. The same language translates to quantum field theory.
554 Here, particles are described in terms of quantum fields in some region of spacetime. These
555 fields have associated charges which describe the forces they experience when interacting
556 with other quantum fields. Most familiar is electric charge – however this applies to e.g., the
557 strong interaction as well, where quantum fields have an associated *color charge* describing
558 behavior under the strong force.

559 Particles are observed to behave in different ways under different forces. These behaviors
560 respect certain *symmetries*, which are most naturally described in the language of group
561 theory. The respective fields, charges, and generators of these symmetry groups are the basic
562 pieces of the SM Lagrangian, which describes the full dynamics of the theory. In the following,
563 we will build up the basic components of this Lagrangian. The treatment presented here relies
564 heavily on Jackson's Classical Electrodynamics [2] for the build-up, and Thomson's Modern
565 Particle Physics [3] for the rest, with reference to Srednicki's Quantum Field Theory [4], and
566 some personal biases and interjections.

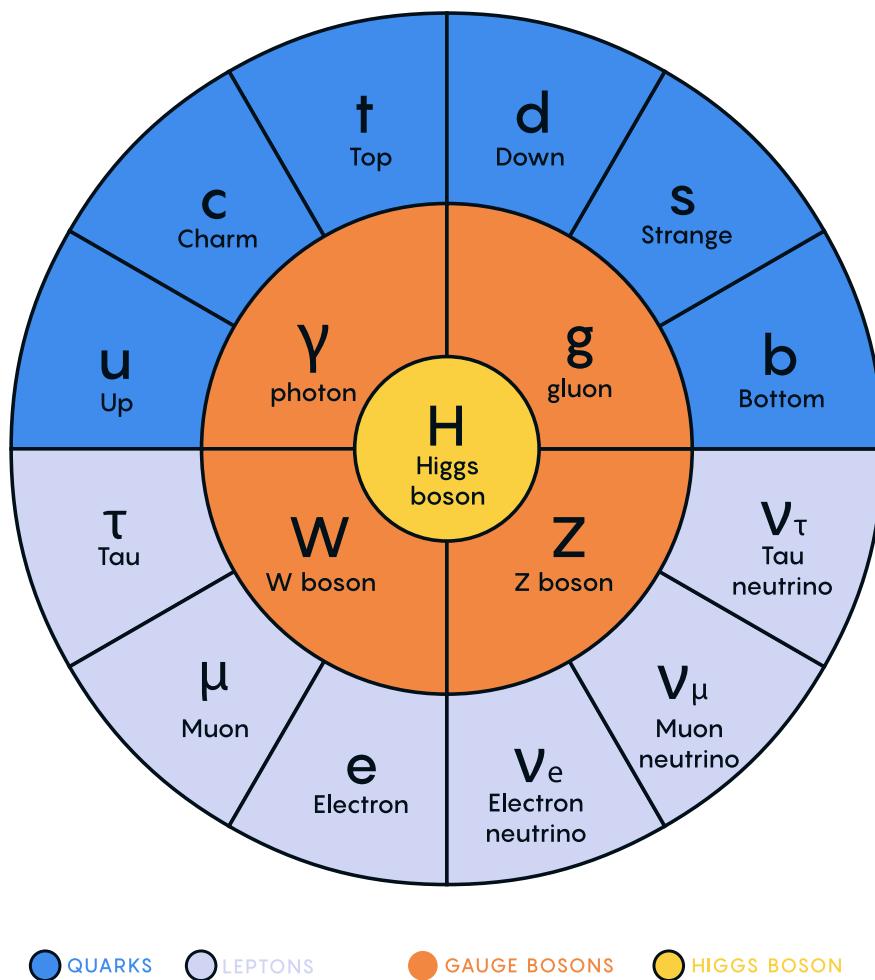


Figure 1.1: Diagram of the elementary particles described by the Standard Model [1].

⁵⁶⁷ **1.2 Quantum Electrodynamics**

Classical electrodynamics is familiar to the general physics audience: electric (\vec{E}) and magnetic (\vec{B}) fields are used to describe behavior of particles with charge q moving with velocity \vec{v} , with forces described as $\vec{F} = q\vec{E} + q\vec{v} \times \vec{B}$. Hints at some more fundamental properties of electric and magnetic fields come via a simple thought experiment: in a frame of reference moving along with the particle at velocity \vec{v} , the particle would appear to be standing still, and therefore have no magnetic force exerted. Therefore a *relativistic* formulation of the theory is required. This is most easily accomplished with a repackaging: the fundamental objects are no longer classical fields but the electric and magnetic *potentials*: ϕ and \vec{A} respectively, with

$$\vec{E} = -\nabla\phi - \frac{\partial\vec{A}}{\partial t} \quad (1.1)$$

$$\vec{B} = \nabla \times \vec{A} \quad (1.2)$$

It is then natural to fully repackage into a relativistic *four-vector*: $A^\mu = (\phi, \vec{A})$. Considering $\partial^\mu = (\frac{\partial}{\partial t}, \nabla)$, the x components of these above two equations become:

$$E_x = -\frac{\partial\phi}{\partial x} - \frac{\partial A_x}{\partial t} = -(\partial^0 A^1 - \partial^1 A^0) \quad (1.3)$$

$$B_x = \frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} = -(\partial^2 A^3 - \partial^3 A^2) \quad (1.4)$$

⁵⁶⁸ where we have used the sign convention $(+, -, -, -)$, such that $\partial^\mu = (\frac{\partial}{\partial x_0}, -\nabla)$.

This is naturally suggestive of a second rank, antisymmetric tensor to describe both the electric and magnetic fields (the *field strength tensor*), defined as:

$$F^{\alpha\beta} = \partial^\alpha A^\beta - \partial^\beta A^\alpha \quad (1.5)$$

Defining a four-current as $J_\mu = (q, \vec{J})$, with q standard electric charge, \vec{J} standard electric current, conservation of charge may be expressed via the continuity equation

$$\partial_\mu J^\mu = 0 \quad (1.6)$$

and all of classical electromagnetism may be packaged into the Lagrangian density:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} - J^\mu A_\mu. \quad (1.7)$$

569 This gets us partway to our goal, but is entirely classical - the description is of classical
 570 fields and point charges, not of quantum fields and particles. To reframe this, let us go back
 571 to the zoomed out view of the particles of the Standard Model. Two of the most familiar
 572 objects associated with electromagnetism are electrons: spin-1/2 particles with charge e , mass
 573 m , and photons: massless spin-1 particles which are the "pieces" of electromagnetic radiation.

574 We know that electrons experience electromagnetic interactions with other objects. Given
 575 this, and the fact that such interactions must be transmitted *somewhat* between e.g. two
 576 electrons, it seems natural that these interactions are facilitated by electromagnetic radiation.
 577 More specifically, we may think of photons as *mediators* of the electromagnetic force. It
 578 follows, then, that a description of electromagnetism on the level of particles must involve a
 579 description of both the "source" particles (e.g. electrons), the mediators (photons), and their
 580 interactions. Further, this description must be (1) relativistic and (2) consistent with the
 581 classically derived dynamics described above.

The beginnings of a relativistic description of spin-1/2 particles is due to Paul Dirac, with the famous Dirac equation:

$$(i\gamma^\mu \partial_\mu - m)\psi = 0 \quad (1.8)$$

where ∂_μ is as defined above, ψ is a Dirac *spinor*, i.e. a four-component wavefunction, m is the mass of the particle, and γ^μ are the Dirac gamma matrices, which define the algebraic structure of the theory. For the following, we also define a conjugate spinor,

$$\bar{\psi} = \psi^\dagger \gamma^0 \quad (1.9)$$

which satisfies the conjugate Dirac equation

$$\bar{\psi}(i\gamma^\mu \partial_\mu - m) = 0 \quad (1.10)$$

582 where the derivative acts to the left.

The Dirac equation is the dynamical equation for spin-1/2, but we'd like to express these dynamics via a Lagrangian density. Further, to have a relativistic description, we'd like to

have this be density be Lorentz invariant. These constraints lead to a Lagrangian of the form

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.11)$$

583 where the Euler-Lagrange equation exactly recovers the Dirac equation.

The question now becomes how to marry the two Lagrangian descriptions that we have developed. Returning for a moment to classical electrodynamics, we know that the Hamiltonian for a charged particle in an electromagnetic field is described by

$$H = \frac{1}{2m}(\vec{p} - q\vec{A})^2 + q\phi. \quad (1.12)$$

Comparing this to the Hamiltonian for a free particle, we see that the modifications required are $\vec{p} \rightarrow \vec{p} - q\vec{A}$ and $E \rightarrow E - q\phi$. Using the canonical quantization trick of identifying \vec{p} with operator $-i\nabla$ and E with operator $i\frac{\partial}{\partial t}$, this identification becomes

$$i\partial_\mu \rightarrow i\partial_\mu - qA_\mu \quad (1.13)$$

Allowing for the naive substitution in the Dirac Lagrangian:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu + iqA_\mu) - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.14)$$

584 where the source term may be interpreted as coming from the Dirac fields themselves, namely,

585 $-q\bar{\psi}\gamma^\mu\psi A_\mu$.

Setting $q = e$ here (as appropriate for the case of an electron), and defining $D_\mu \equiv \partial_\mu + ieA_\mu$, this may then be written in the form

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}. \quad (1.15)$$

586 which is exactly the quantum electrodynamics Lagrangian.

587 We have swept a few things under the rug here, however. Recall that the general form
588 of a Lagrangian is conventionally $\mathcal{L} = T - V$, where T is the kinetic term, and thus ought
589 to contain a derivative with respect to time (c.f. the standard $\frac{1}{2}m\frac{\partial x}{\partial t}$ familiar from basic
590 kinematics). More particularly, given the definition of conjugate momentum as $\partial\mathcal{L}/\partial\dot{q}$ for

591 $\mathcal{L}(q, \dot{q}, t)$ and $\dot{q} = \frac{\partial q}{\partial t}$, any field q which has no time derivative in the Lagrangian has 0
592 conjugate momentum, and thus no dynamics.

593 Looking at this final form, there is an easily identifiable kinetic term for the spinor fields
594 (just applying the D_μ operator). However trying to identify something similar for the A fields,
595 one comes up short – the antisymmetric nature of $F^{\mu\nu}$ term means that there is no time
596 derivative applied to A^0 .

597 What does this mean? A^μ is a four component object, but it would appear that only three
598 of the components have dynamics: we have too many degrees of freedom in the theory. This
599 is the principle behind *gauge symmetry* – an extra constraint on A^μ (a *gauge condition*) must
600 be defined such that a unique A^μ defines the theory and satisfies the condition. However,
601 we are free to choose this extra condition – the physics content of the theory should be
602 independent of this choice (that is, it should be *gauge invariant*).

To ground this a bit, let us return to basic electric and magnetic fields. These are physical quantities that can be measured, and are defined in terms of potentials as

$$\vec{E} = -\nabla\phi - \frac{\partial\vec{A}}{\partial t} \quad (1.16)$$

$$\vec{B} = \nabla \times \vec{A}. \quad (1.17)$$

603 It is easy to show, for any scalar function λ , that $\nabla \times \nabla\lambda = 0$. This implies that the physical
604 \vec{B} field is invariant under the transformation $\vec{A} \rightarrow \vec{A} + \nabla\lambda$ for any scalar function λ .

605 Under the same transformation of \vec{A} , the electric field \vec{E} becomes $-\nabla\phi - \frac{\partial\vec{A}}{\partial t} - \frac{\partial\nabla\lambda}{\partial t} =$
606 $-\nabla(\phi + \frac{\partial\lambda}{\partial t}) - \frac{\partial\vec{A}}{\partial t}$, such that, for the \vec{E} field to be unchanged, we must additionally apply
607 the transformation $\phi \rightarrow \phi - \frac{\partial\lambda}{\partial t}$.

This set of transformations to the potentials that leave the physical degrees of freedom invariant is expressed in our four vector notation naturally as

$$A_\mu \rightarrow A_\mu - \partial_\mu \lambda \quad (1.18)$$

608 where $A_\mu = (\phi, -\vec{A})$ with our sign convention. It should be noted that this function λ is an
609 arbitrary function of *local* spacetime, and thus expresses invariance of the physics content

610 under a local transformation.

Let us return to the Lagrangian for QED. In particular, focusing on the free Dirac piece

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi \quad (1.19)$$

we note that if we apply a local transformation of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$ (and correspondingly $\bar{\psi} \rightarrow \bar{\psi}e^{-iq\lambda(x)}$, by definition), the Lagrangian becomes

$$\bar{\psi}e^{-iq\lambda(x)}(i\gamma^\mu \partial_\mu - m)e^{iq\lambda(x)}\psi = \bar{\psi}e^{-iq\lambda(x)}(i\gamma^\mu \partial_\mu)e^{iq\lambda(x)}\psi - m\bar{\psi}\psi. \quad (1.20)$$

As $\partial_\mu(e^{iq\lambda(x)}\psi) = iq e^{iq\lambda(x)}(\partial_\mu \lambda(x))\psi + e^{iq\lambda(x)}\partial_\mu \psi$, this becomes

$$\bar{\psi}(i\gamma^\mu(\partial_\mu + iq\partial_\mu \lambda(x)) - m)\psi. \quad (1.21)$$

Thus, the free Dirac Lagrangian on its own is not invariant under this transformation. We may note, however, that on interaction with an electromagnetic field, as described above, this transformed Lagrangian may be packaged as:

$$\bar{\psi}(i\gamma^\mu(\partial_\mu + iq\partial_\mu \lambda(x) + iqA_\mu) - m)\psi = \bar{\psi}(i\gamma^\mu(\partial_\mu + iq(A_\mu + \partial_\mu \lambda(x))) - m)\psi. \quad (1.22)$$

611 since by the arguments above, the physics content of the Lagrangian is invariant under the
612 transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, we may directly make this transformation, and remove this
613 extra $\partial_\mu \lambda(x)$ term. It is straightforward to verify that the $\frac{1}{4}F_{\mu\nu}F^{\mu\nu}$ term is invariant under
614 this same transformation of A_μ , so we may say that the QED Lagrangian is invariant under
615 local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$.

616 These arguments illuminate some important concepts which will serve us well going forward.
617 First, while we have remained grounded in the “familiar” physics of electromagnetism for the
618 above, arguments of the “top down” variety would lead us to the exact same conclusions.
619 That is, suppose we wanted to construct a theory of spin-1/2 particles that was invariant
620 under local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$. More broadly, we could say that we
621 desire this theory to be invariant under local $U(1)$ transformations, where $U(1)$ is exactly
622 this group, under multiplication, of complex numbers with absolute value 1. By very similar

arguments as above, we would see that, to achieve invariance, this theory would necessitate an additional degree of freedom, A_μ , with the exact properties that are familiar to us from electrodynamics. These arguments based on symmetries are extremely powerful in building theories with a less familiar grounding, as we will see in the following.

Second, we defined this quantity $D_\mu \equiv \partial_\mu + ieA_\mu$ above, seemingly as a matter of notational convenience. However, from the latter set of arguments, such a packaging takes on a new power: by explicitly including this gauge field A_μ which transforms in such a way as to keep invariance under a given transformation, the invariance is immediately more manifest. That is, to pose the $U(1)$ invariance in a more zoomed out way, under the transformation $\psi \rightarrow e^{iq\lambda(x)}\psi$, while

$$\bar{\psi}\partial_\mu\psi \rightarrow \bar{\psi}(\partial_\mu + iq\partial_\mu\lambda(x))\psi \quad (1.23)$$

with the extra term that gets canceled out by the gauge transformation of A_μ ,

$$\bar{\psi}D_\mu\psi \rightarrow \bar{\psi}D_\mu\psi \quad (1.24)$$

where this transformation is already folded in. This repackaging, called a *gauge covariant derivative* is much more immediately expressive of the symmetries of the theory.

Finally, to emphasize how fundamental these gauge symmetries are to the corresponding theory, let us examine the additional term needed for $U(1)$ invariance, $q\bar{\psi}\gamma^\mu A_\mu\psi$. While a first principles examination of Feynman rules is beyond the scope of this thesis, it is powerful to note that this is expressive of a QED vertex: the $U(1)$ invariance of the theory and the interaction between photons and electrons are inextricably tied together.

1.3 An Aside on Group Theory

Quantum electrodynamics is very familiar and well covered, and provides (both historically and in this thesis) a nice bridge between “standard” physics and the language of symmetries and quantum field theory. However, now that we are acquainted with the language, we may set up to dive a bit deeper. To begin, let us look again at the $U(1)$ group that is so fundamental to QED. We have expressed this via a set of transformations on our Dirac spinor

640 objects, ψ , of the form $e^{iq\lambda(x)}$. Note that such transformations, though they are local (i.e. a
 641 function of spacetime) are purely *phase* transformations. Relatedly, $U(1)$ is an Abelian group,
 642 meaning that group elements commute.

643 To set up language to generalize beyond $U(1)$, note that we may equivalently write $U(1)$
 644 elements as $e^{ig\vec{\alpha}(x)\cdot\vec{T}}$, $\vec{\alpha}(x)$ and \vec{T} and are vectors in the space of *generators* of the group,
 645 with each $\alpha^a(x)$ an associated scalar function to generator t^a , and g is some scalar strength
 646 parameter. Of course this is a bit silly for $U(1)$, which has a single generator, and thus
 647 reduces to the transformation we discussed above. However, this becomes much more useful
 648 for groups of higher degree, with more generators and degrees of freedom.

649 To discuss these groups in a bit more detail, note that $U(n)$ is the unitary group of degree
 650 n , and corresponds to the group of $n \times n$ unitary matrices (that is, $U^\dagger U = UU^\dagger = 1$). Given
 651 that group elements are $n \times n$, this means that there are n^2 degrees of freedom: n^2 generators
 652 are needed to characterize the group.

653 For $U(1)$, this is all consistent with what we have said above – the group of 1×1 unitary
 654 matrices have a single generator, and the phases we identify above clearly satisfy unitarity.
 655 Note that these degrees of freedom for the gauge group also characterize the number of gauge
 656 bosons we need to satisfy the local symmetry: for $U(1)$, we need one gauge boson, the photon.

657 Of relevance for the Standard Model are also the special unitary groups $SU(n)$. These
 658 are defined similarly to the unitary groups, with the additional requirement that group
 659 elements have determinant 1. This extra constraint removes 1 degree of freedom: groups are
 660 characterized by $n^2 - 1$ generators.

661 In particular, we will examine the groups $SU(2)$ in the context of the weak interaction,
 662 with an associated $2^2 - 1 = 3$ gauge bosons (cf. the W^\pm and Z bosons), and $SU(3)$, with an
 663 associated $3^2 - 1 = 8$ gauge bosons (cf. gluons of different flavors). Note that these groups
 664 are non-Abelian (2×2 or 3×3 matrices do not, in general, commute), leading to a variety of
 665 complications. However, both of these theories feature interactions with spin-1/2 particles,
 666 with transformations of a very similar form: $\psi \rightarrow e^{ig\vec{\alpha}(x)\cdot\vec{T}}\psi$, and the general framing of the
 667 arguments for QED will serve us well in the following.

668 **1.4 Quantum Chromodynamics**

669 In some sense, the simplest extension the development of QED is quantum chromodynamics
 670 (QCD). QCD is a theory in which, once the basic dynamics are framed (a non-trivial task!)
 671 the group structure becomes apparent. The quark model, developed by Murray Gell-Mann [5]
 672 and George Zweig [6], provided the fundamental particles involved in the theory, and had
 673 great success in explaining the expanding zoo of experimentally observed hadronic states.

674 Some puzzles were still apparent – the Δ^{++} baryon, e.g., is composed of three up quarks,
 675 u , with aligned spins. As quarks are fermions, such a state should not be allowed by the
 676 Pauli exclusion principle. The existence of such a state in nature implies the existence of
 677 another quantum number, and a triplet of values, called *color charge* was proposed by Oscar
 678 Greenberg [7]. With these pieces in place, the structure becomes more apparent, as elucidated
 679 by Han and Nambu [8].

680 Let us reason our way to the symmetries using color charge. Experimentally, we know
 681 that there is this triplet of color charge values r, g, b (the “plus” values, cf. electric charge)
 682 and correspondingly anti-color charge $\bar{r}, \bar{g}, \bar{b}$ (the “minus” values). Supposing that the force
 683 behind QCD (the *strong force*) is, similar to QED, interactions between fermions mediated
 684 by gauge bosons (quarks and gluons respectively), we can start to line up the pieces.

685 What color charge does a gluon have? Similarly to electric charge, we may associate
 686 particles with color charge, anti-particles with anti-color charge. Notably, free particles
 687 observed experimentally are colorless (have no color charge). Thus, in order for charge to
 688 be conserved throughout such processes, this already implies that there are charged gluons.
 689 Further, examining color flow diagrams such as *TODO: insert*, it is apparent first that a
 690 gluon has not one but two associated color charges and second that these two must be one
 691 color charge and one anti-color charge.

692 Counting up the available types of gluons, then, we come up with nine. Six of mixed
 693 color type: $r\bar{b}, r\bar{g}, b\bar{r}, b\bar{g}, g\bar{b}$, and $g\bar{r}$, and three of same color type: $r\bar{r}, g\bar{g}$, and $b\bar{b}$. In practice,
 694 however, these latter three are a bit redundant: all express a colorless gluon, which, if we

could observe this as a free particle, would be indistinguishable from each other. The *color singlet* state is then a mix of these, $\frac{1}{\sqrt{3}}(r\bar{r} + g\bar{g} + b\bar{b})$, leaving two unclaimed degrees of freedom, which may be satisfied by the linearly independent combinations $\frac{1}{\sqrt{2}}(r\bar{r} - g\bar{g})$ and $\frac{1}{\sqrt{6}}(r\bar{r} + g\bar{g} - 2b\bar{b})$.

We thus have an octet of color states plus a colorless singlet state. If this colorless singlet state existed, however, we would be able to observe it, not only via interactions with quarks, but as a free particle. Since do not observe this in nature, this restricts us to 8 gluons. The simplest group with a corresponding 8 generators is $SU(3)$. Under the assumption that $SU(3)$ is the local gauge symmetry of the strong interaction, we may proceed in a similar way as we did for QED. The gauge transformation is $\psi \rightarrow e^{ig_S \vec{\alpha}(x) \cdot \vec{T}} \psi$, where \vec{T} is an eight component vector of the generators of $SU(3)$, often expressed via the Gell-Mann matrices, λ^a , as $t^a = \frac{1}{2}\lambda^a$, and the spinor ψ represents the fields corresponding to quarks.

This $SU(3)$ symmetry exactly expresses the color structure elucidated above – the Gell-Mann matrices are an equivalent presentation of the color combinations described above. Proceeding by analogy to QED, gauge invariance is achieved by introducing eight new degrees of freedom, G_μ^a , which are the gauge fields corresponding to the gluons, with the gauge covariant derivative then analogously taking the form $D_\mu \equiv \partial_\mu + ig_S G_\mu^a t^a$.

Recall from the QED derivation that the field strength tensor, $F^{\mu\nu}$ is a rank two antisymmetric tensor which is manifestly gauge invariant and which describes the physical dynamics of the A_μ field. We would like to analogously define a term for the gluon fields. Repackaging this QED tensor, it is apparent that

$$[D_\mu, D_\nu] = D_\mu D_\nu - D_\nu D_\mu \quad (1.25)$$

$$= (\partial_\mu + iqA_\mu)(\partial_\nu + iqA_\nu) - (\partial_\nu + iqA_\nu)(\partial_\mu + iqA_\mu) \quad (1.26)$$

$$= \partial_\mu \partial_\nu + iq\partial_\mu A_\nu + iqA_\mu \partial_\nu + (iq)^2 A_\mu A_\nu - (\partial_\nu \partial_\mu + iq\partial_\nu A_\mu + iqA_\nu \partial_\mu + (iq)^2 A_\nu A_\mu) \quad (1.27)$$

$$= iq(\partial_\mu A_\nu - \partial_\nu A_\mu) + (iq)^2 (A_\mu A_\nu - A_\nu A_\mu) \quad (1.28)$$

$$= iq(\partial_\mu A_\nu - \partial_\nu A_\mu) + (iq)^2 [A_\mu, A_\nu]. \quad (1.29)$$

We proceed through this derivation to highlight that, in the specific case of QED, with its Abelian $U(1)$ gauge symmetry, the field commutator vanishes, leaving exactly the definition of $F_{\mu\nu}$ as described above, i.e.,

$$F_{\mu\nu} = \frac{1}{iq}[D_\mu, D_\nu]. \quad (1.30)$$

We may proceed to define an analogous field strength term for G_μ^a in a similar way:

$$G_{\mu\nu} = \frac{1}{ig_S}[D_\mu, D_\nu] \quad (1.31)$$

This has an extremely nice correspondence, but is complicated by the non-Abelian nature of $SU(3)$, with

$$G_{\mu\nu} = \partial_\mu(G_\nu^a t^a) - \partial_\nu(G_\mu^a t^a) + ig_s[G_\mu^a t^a, G_\nu^a t^a]. \quad (1.32)$$

in which the field commutator term is non-zero. In particular (since each term is summing over a , so we may relabel) as

$$[G_\mu^a t^a, G_\nu^b t^b] = [t^a, t^b]G_\mu^a G_\nu^b \quad (1.33)$$

and as $[t^a, t^b] = if^{abc}t^c$ for the Gell-Mann matrices, where f^{abc} are the structure constants of $SU(3)$, we have

$$G_{\mu\nu} = \partial_\mu(G_\nu^a t^a) - \partial_\nu(G_\mu^a t^a) - g_s f^{abc} t^c G_\mu^a G_\nu^b \quad (1.34)$$

$$= t^a(\partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{bca}G_\mu^b G_\nu^c) \quad (1.35)$$

$$= t^a G_{\mu\nu}^a \quad (1.36)$$

⁷¹² for $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{abc}G_\mu^b G_\nu^c$.

⁷¹³ This gives the component of the field strength corresponding to a particular gauge field a ,
⁷¹⁴ where the first two terms have the familiar form of the QED field strength, while the last
⁷¹⁵ term is new, and explicitly related to the group structure via the f^{abc} constants. In terms
⁷¹⁶ of the physics content of the theory, this latter term gives rise to a gluon *self-interaction*, a
⁷¹⁷ distinguishing feature of QCD.

⁷¹⁸ Similarly as in QED, a Lorentz invariant combination of field strength tensors may be made
⁷¹⁹ as $G_{\mu\nu}G^{\mu\nu}$. However, this is not manifestly gauge invariant. Under a gauge transformation

⁷²⁰ U , the covariant derivative behaves as $D^\mu \rightarrow UD^\mu U^{-1}$, corresponding to $G^{\mu\nu} \rightarrow UG^{\mu\nu}U^{-1}$.
⁷²¹ The cyclic property of the trace thus ensures the gauge invariance of $\text{tr}(G_{\mu\nu}G^{\mu\nu})$, which we
⁷²² will write as $G_{\mu\nu}^a G_a^{\mu\nu}$ with the implied sum over generators a .

Packaging up the theory, it is tempting to copy the form of the QED Lagrangian, with the identifications we have made above:

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.37)$$

However this is not quite correct due to the $SU(3)$ nature of the theory. In terms of the physics, the Dirac fields ψ have associated color charge, which must interact appropriately with the G_μ fields. Mathematically, the generators t^a are 3×3 matrices, while the ψ are four component spinors. Adding a color index to the Dirac fields, i.e., ψ_i where i runs over the three color charges, and similarly indexing the generators t_{ij}^a , we may then express the $SU(3)$ gauge covariant derivative component-wise as

$$(D_\mu)_{ij} = \partial_\mu \delta_{ij} + ig_S G_\mu^a t_{ij}^a \quad (1.38)$$

⁷²³ where δ_{ij} is the Kronecker delta, as ∂_μ does not participate in the $SU(3)$ structure.

The Lagrangian then becomes

$$\mathcal{L} = \bar{\psi}_i(i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij})\psi_j - \frac{1}{4}G_{\mu\nu}^a G_a^{\mu\nu}. \quad (1.39)$$

⁷²⁴ and we have constructed QCD.

⁷²⁵ 1.5 The Weak Interaction

⁷²⁶ One of the first theories of the weak interaction was from Enrico Fermi [9], in an effort to
⁷²⁷ explain beta decay, a process in which an electron or positron is emitted from an atomic
⁷²⁸ nucleus, resulting in the conversion of a neutron to a proton or proton to a neutron respectively.
⁷²⁹ Fermi's hypothesis was of a direct interaction between four fermions. However, in the advent of
⁷³⁰ QED, it is natural to wonder if a theory based on mediator particles and gauge symmetries
⁷³¹ applies to the weak force as well. The modern formulation of such a theory is due to Sheldon

732 Glashow, Steven Weinberg, and Abdus Salam [10], and is what we will describe in the
733 following.

734 Considering emission of an electron, Fermi's theory involves an initial state neutron that
735 transitions to a proton with the emission of an electron and a neutrino. This transition
736 gives a hint that something slightly more complicated is happening than in QED: there is an
737 apparent mixing between particle types.

738 Now, with the assumption there are mediators for such an interaction, we further know
739 from beta decay and charge conservation that there must be at least two such degrees of
740 freedom: e.g. one that decays to an electron and neutrino (W^-) and one that decays to a
741 positron and neutrino (W^+). From consideration of the process $e^+e^- \rightarrow W^+W^-$, it turns
742 out that with just these two degrees of freedom, the cross section for this process increases
743 without limit as a function of center-of-mass energy, ultimately violating unitarity (more
744 W^+W^- pairs come out than e^+e^- pairs go in). This is resolved with a third, neutral degree
745 of freedom, the Z boson, whose contribution interferes negatively, regulating this process.

746 This leads to three degrees of freedom for the gauge symmetry of the weak interactions, so
747 we thus need a theory which is locally invariant under transformations of a group with three
748 generators. The simplest such choice is $SU(2)$. We may follow a very similar prescription as
749 for QED and QCD: $SU(2)$ has three generators, which implies the existence of three gauge
750 bosons, call them W_μ^k . The gauge transformation may be expressed as $\psi \rightarrow e^{ig_W \vec{\alpha}(x) \cdot \vec{T}} \psi$, where
751 in this case the generators are for $SU(2)$, which may be written in terms of the familiar Pauli
752 matrices: $\vec{T} = \frac{1}{2}\vec{\sigma}$. The structure constants for $SU(2)$ are the antisymmetric Levi-Civita
753 tensor, so the corresponding gauge covariant derivative is $D_\mu \equiv \partial_\mu + ig_W W_\mu^k t^k$, and the field
754 strength tensor is $W_{\mu\nu}^k = \partial_\mu W_\nu^k - \partial_\nu W_\mu^k - \epsilon^{ijk} W_\mu^k W_\nu^k$.

The corresponding Lagrangian would thus be

$$\mathcal{L} = \bar{\psi}_i (i(\gamma^\mu D_\mu)_{ij} - m\delta_{ij}) \psi_j - \frac{1}{4} W_{\mu\nu}^k W_k^{\mu\nu} \quad (1.40)$$

755 where indices i and j run over $SU(2)$ charges.

756 On considering some of the details, the universe unfortunately turns out to be a bit

more complicated. However, this still provides a useful starting place for elucidating the theory of weak interactions. First off, let us consider the particle content, namely, what do the Dirac fields correspond to? This is still a theory of fermionic interactions with gauge bosons. However, we might notice that the fermion content of this theory is both a) broader than QCD, as we know experimentally (cf. beta decay) that both quarks and leptons (e.g. electrons) participate in the weak interaction and b) this fermion content seemingly has a large overlap with QED. In terms of the gauge bosons, we know that at both W^+ and W^- are electrically charged – this means that we expect some interaction of the weak theory with electromagnetism.

However, before diving deeper into this apparent connection between the weak interaction and QED, let us focus on the gauge symmetry. In QCD, the $SU(3)$ content of the theory is expressed via a contraction of color indices – the theory allows for transitions between quarks of one color and quarks of another. Thinking similarly in terms of $SU(2)$ transitions, the beta decay example is already fruitful – there is a transition between an electron and its corresponding neutrino, as well as between two types of quark. In particular, for the case of neutron (with quark content udd) and proton (with quark content udu), the weak interaction provides for a transition from down to up quark.

Such $SU(2)$ dynamics are described via a quantity called *weak isospin*, denoted I_W with third component $I_W^{(3)}$, and can be thought of in a very similar way as color charge in QCD (i.e. as the charge corresponding to the weak interaction). Since $SU(2)$ is 2×2 , there are two such charge states for the fermions, denoted as $I_W^{(3)} = \pm\frac{1}{2}$. This means that the bosons must have $I_W = 1$ such that, by sign convention corresponding to electric charge, the W^+ boson has $I_W^{(3)} = +1$, the Z boson has $I_W^{(3)} = 0$, and the W^- boson has $I_W^{(3)} = -1$.

From conservation of electric charge, this means that transitions involving a W^\pm are between particles that differ by ± 1 in both weak isospin $I_W^{(3)}$ and electric charge. We may thus line up all such doublets as:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}, \begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \quad (1.41)$$

780 with the top corresponding to the lower weak isospin and electric charge particles, and the
781 lower quark entries (d' , etc) corresponding to the weak quark eigenstates (which are related
782 to the mass eigenstates by the CKM matrix *TODO: more detail*). Similar doublets may be
783 constructed for the corresponding anti-particles.

The fundamental structuring of these transitions around both electric and weak charge is again indicative of a natural connection. However, nature is again a bit more complicated than we have described. This is because the weak interaction is a *chiral* theory. For massless particles, chirality is the same as the perhaps more intuitive *helicity*. This describes the relationship between a particle's spin and momentum: if the spin vector points in the same direction as the momentum vector, helicity is positive (the particle is “right-handed”), and if the two point in opposite directions, the helicity is negative (the particle is “left-handed”). More concretely:

$$H = \frac{\vec{s} \cdot \vec{p}}{|\vec{s} \cdot \vec{p}|}. \quad (1.42)$$

For massive particles, this generalizes a bit – in the language of Dirac fermions that we have developed, we define projection operators

$$P_R = \frac{1}{2}(1 + \gamma^5) \quad \text{and} \quad P_L = \frac{1}{2}(1 - \gamma^5) \quad (1.43)$$

784 for right and left-handed chiralities respectively – acting on a Dirac field with such operators
785 projects the field onto the corresponding chiral state.

Experimentally, this pops up via parity violation and the famous $V - A$ theory. For the scope of this thesis, it is sufficient to say that the weak interaction is only observed to take place for left-handed particles (and correspondingly, right-handed anti-particles). We therefore modify the theory stated above by projecting all fermions participating in the weak interaction onto respective chiral states – in particular, the $SU(2)$ gauge symmetry only acts on left-handed particles and right-handed anti-particles. We therefore modify the theory appropriately, denoting the chiral projected gauge symmetry as $SU(2)_L$, and similarly for the

Dirac fields. In particular, the weak isospin doublets listed above must now be left-handed:

$$\begin{pmatrix} \nu_e \\ e^- \end{pmatrix}_L, \begin{pmatrix} \nu_\mu \\ \mu^- \end{pmatrix}_L, \begin{pmatrix} \nu_\tau \\ \tau^- \end{pmatrix}_L, \begin{pmatrix} u \\ d' \end{pmatrix}_L, \begin{pmatrix} c \\ s' \end{pmatrix}_L, \begin{pmatrix} t \\ b' \end{pmatrix}_L \quad (1.44)$$

⁷⁸⁶ and right-handed particle states are placed in singlets and assigned 0 charge under $SU(2)_L$
⁷⁸⁷ ($I_W = I_W^{(3)} = 0$).

With all of these assignments, let us revisit our guess at the form of the weak interaction Lagrangian. First, dwelling on the kinetic term $\bar{\psi}_i(i(\gamma^\mu D_\mu)_{ij}\psi_j)$, we note that the assigning of left-handed fermions to isospin doublets and right-handed fermions to isospin singlets allows us to remove explicit $SU(2)$ indices by treating these as the fundamental objects, that is, for a single *generation* of fermions, we may write:

$$\bar{Q}i\gamma^\mu D_\mu Q + \bar{u}i\gamma^\mu D_\mu u + \bar{d}i\gamma^\mu D_\mu d + \bar{L}i\gamma^\mu D_\mu L + \bar{e}i\gamma^\mu D_\mu e \quad (1.45)$$

⁷⁸⁸ for left-handed doublets Q and L for quarks and electron fields respectively and right handed
⁷⁸⁹ singlets u and d for up and down quark fields and e for electrons.

More concisely, and summing over the three generations of fermions, we may write

$$\sum_f \bar{f}i\gamma^\mu D_\mu f \quad (1.46)$$

⁷⁹⁰ where the f are understood to run over the fermion chiral doublets and singlets as above.

This then leaves our Lagrangian as

$$\mathcal{L} = \sum_f \bar{f}i\gamma^\mu D_\mu f - \frac{1}{4}W_{\mu\nu}^k W_k^{\mu\nu} \quad (1.47)$$

$$= \sum_f \bar{f}\gamma^\mu(i\partial_\mu - \frac{1}{2}g_W W_\mu^k \sigma_k)f - \frac{1}{4}W_{\mu\nu}^k W_k^{\mu\nu}, \quad (1.48)$$

⁷⁹¹ where we have expanded the covariant derivative for clarity. You may note that we have
⁷⁹² dropped the mass term in the equation above – we will discuss this in detail in just a moment.

First, however, we return to the above comment about fermion content – we neglected to include the sum over fermions in our QED derivation for simplicity. However, all of the

fermions considered in the discussion of the weak interaction have an electric charge (except for the neutrinos). It would be nice to repackage the theory into a coherent *electroweak* theory. This is fairly straightforward when considering the gauge approach – from the discussion above we should expect the electroweak gauge group to be something like $SU(2) \times U(1)$, with four corresponding gauge bosons. Consider a gauge theory with group $SU(2)_L \times U(1)_Y$ – that is, the same weak interaction as discussed previously, but a new $U(1)_Y$ gauge group for electromagnetism, with transformations defined as

$$\psi \rightarrow e^{ig' \frac{Y}{2} \lambda(x)} \psi \quad (1.49)$$

⁷⁹³ with *weak hypercharge* Y .

Similarly to our discussion of QED, we may write the $U(1)_Y$ gauge field as B_μ , and interactions with the Dirac fields take the form $g' \frac{Y}{2} \gamma^\mu B_\mu \psi$. The relationship between this hypercharge and new B_μ field and classical electrodynamics is not so obvious – however it is convenient to parametrize as

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \quad (1.50)$$

⁷⁹⁴ where A_μ and Z_μ are the physical fields, and we pick W_μ^3 as the neutral weak boson.

⁷⁹⁵ Note that in the $SU(2)_L \times U(1)_Y$ theory, the Lagrangian must be invariant under all of
⁷⁹⁶ the local gauge transformations. In particular, this means that the hypercharge must be the
⁷⁹⁷ same for fermion fields in each weak doublet to preserve $U(1)_Y$ invariance. This gives insight
⁷⁹⁸ into the relation between the charges of $SU(2)_L \times U(1)_Y$ and electric charge. In particular
⁷⁹⁹ we know that the hypercharge, Y , of e^- ($I_W^{(3)} = -\frac{1}{2}$) and ν_e ($I_W^{(3)} = +\frac{1}{2}$) is the same.

Supposing that $Y = \alpha I_W^{(3)} + \beta Q$, we must have $-\alpha \frac{1}{2} - \beta = \alpha \frac{1}{2} \implies \beta = -\alpha$. Therefore, choosing an overall scaling from convention,

$$Y = 2(Q - I_W^{(3)}). \quad (1.51)$$

⁸⁰⁰ Some of these particular forms are best understood in the context of the Higgs mechanism
⁸⁰¹ – we will return to this discussion below.

802 **1.6 The Higgs Potential and the SM**

803 In the above, we have neglected a discussion of masses. However there are several things to
804 sort out here. In the first place, we know experimentally that the weak interactions occur
805 over very short ranges at low energies (e.g., why Fermi's effective four fermion interaction was
806 such a good description). This is consistent with massive W^\pm and Z bosons (and indeed, this
807 is seen experimentally). However, requiring local gauge invariance forbids mass terms in the
808 Lagrangian. In the simple $U(1)$ QED example, such a term would have the form $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$,
809 which is not invariant under the transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, and similar arguments hold
810 for gauge bosons in the electroweak theory and QCD.

Similar issues are encountered with fermions – in the electroweak theory above, the gauge symmetries are separated into left and right handed chirality via doublet and singlet states. This means that a mass term would need to be separated as well. Such a term would have the form:

$$m\bar{f}f = m(\bar{f}_L + \bar{f}_R)(f_L + f_R) \quad (1.52)$$

$$= m(\bar{f}_L f_L + \bar{f}_L f_R + \bar{f}_R f_L + \bar{f}_R f_R) \quad (1.53)$$

$$= m(\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.54)$$

811 where we have used that $f_{L,R} = P_{L,R}f$, $\bar{f}_{L,R} = \bar{f}P_{R,L}$, and $P_R P_L = P_L P_R = 0$. As left
812 and right-handed particles transform differently under $SU(2)_L$, this is manifestly not gauge
813 invariant.

814 The question then becomes: how do we include particle masses while preserving the
815 gauge properties of our theory? The answer, due to Robert Brout and François Englert [11],
816 Peter Higgs [12], and Gerald Guralnik, Richard Hagen, and Tom Kibble [13] comes via the
817 Higgs mechanism, which we will describe in the following. Importantly for this thesis, this
818 mechanism predicts the existence of a physical particle, the Higgs boson, and a particle
819 consistent with the Higgs boson was seen by both ATLAS [14] and CMS [15] in 2012.

To explain the Higgs, we focus first on generating masses for the electroweak gauge bosons.

Consider adding two complex scalar fields ϕ^+ and ϕ^0 to the Standard Model embedded in a weak isospin doublet ϕ . We may write the doublet as

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.55)$$

820 where we explicitly note the four available degrees of freedom.

The Lagrangian for such a doublet takes the form

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \quad (1.56)$$

where V is the corresponding potential. Considering the particular form

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.57)$$

821 we may notice that this has some interesting properties. Considering, as illustration, a similar
822 potential for a real scalar field, $\mu^2 \chi^2 + \lambda \chi^4$, taking the derivative and setting it equal to 0
823 yields extrema when $\chi = 0$ and $(\mu^2 + 2\lambda\chi^2) = 0 \implies \chi^2 = -\frac{\mu^2}{2\lambda}$. For $\mu^2 > 0$, there is a
824 unique minimum at $\chi = 0$, and for $\mu^2 < 0$ there are degenerate minima at $\chi = \pm\sqrt{-\frac{\mu^2}{2\lambda}}$.
825 Note that we take $\lambda > 0$, otherwise the only minima in the theory are trivial.

The same simple calculus for the complex Higgs doublet above yields degenerate minima for $\mu^2 < 0$ at

$$\phi^\dagger \phi = \frac{1}{2} (\phi_1^2 + \phi_2^2 + \phi_3^2 + \phi_4^2) = \frac{v}{2} = -\frac{\mu^2}{2\lambda} \quad (1.58)$$

However, though there is this degenerate set of minima, there can only be a single *physical* vacuum state (we say that the symmetry is *spontaneously broken*). Without loss of generality, we may align our axes such that the physical vacuum state is at

$$\langle 0 | \phi | 0 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (1.59)$$

826 where we have explicitly chosen a real, non-zero vacuum expectation value for the neutral
827 component of the Higgs doublet to maintain a massless photon, as we shall see. Physically,
828 however, this makes sense - the vacuum is not electrically charged.

The vacuum is a classical state – we want a quantum one. We may express fluctuations about this nonzero expectation value via an expansion as $v + \eta(x) + i\xi(x)$. However, renaming of fields is only meaningful for the non-zero vacuum component - we thus have:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + \eta(x) + i\phi_4 \end{pmatrix}. \quad (1.60)$$

where we may expand the Lagrangian listed above:

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda(\phi^\dagger \phi)^2. \quad (1.61)$$

It is an exercise in algebra to plug in the expansion about v into this Lagrangian: first expanding the potential

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \quad (1.62)$$

$$= \mu^2 \left(\sum_i \phi_i(x)^2 + (v + \eta(x))^2 \right) + \lambda \left(\sum_i \phi_i(x)^2 + (v + \eta(x))^2 \right) \quad (1.63)$$

$$= -\frac{1}{4} \lambda v^4 + \lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{1}{4} \lambda \eta^4 \quad (1.64)$$

$$+ \frac{1}{2} \lambda \sum_{i \neq j} \phi_i^2 \phi_j^2 + \lambda v \eta \sum_i \phi_i(x)^2 + \frac{1}{2} \lambda \eta^2 \sum_i \phi_i(x)^2 + \frac{1}{4} \sum_i \phi_i(x)^4 \quad (1.65)$$

where the sums are over the $i \in 1, 2, 4$, that is, the fields with 0 vacuum expectation, and we have used the definition $\mu^2 = -\lambda v^2$.

Within this potential, we note a quadratic term in $\eta(x)$ which we may identify with a mass, namely $m_\eta = \sqrt{2\lambda v^2}$, whereas the ϕ_i are massless. These ϕ_i are known as *Goldstone bosons*, and correspond to quantum fluctuations along the minimum of the potential. Of particular note for this thesis are the interaction terms $\lambda v \eta^3$ and $\frac{1}{4} \lambda \eta^4$, expressing trilinear and quartic self-interactions of the η field.

Expanding the kinetic term

$$(\partial_\mu \phi)^\dagger (\partial^\mu \phi) = \frac{1}{2} \sum_i (\partial_\mu \phi_i)(\partial^\mu \phi_i) + \frac{1}{2} (\partial_\mu(v + \eta(x)))(\partial^\mu(v + \eta(x))) \quad (1.66)$$

$$= \frac{1}{2} \sum_i (\partial_\mu \phi_i)(\partial^\mu \phi_i) + \frac{1}{2} (\partial_\mu \eta)(\partial^\mu \eta) \quad (1.67)$$

⁸³⁶ in a similar way, completing the story of three massless degrees of freedom (Goldstone bosons)
⁸³⁷ and one massive one.

Now, this doublet is embedded in an $SU(2)_L \times U(1)$ theory, so we would like to preserve that gauge invariance. This is achieved in the same way as for the Dirac fields, with the introduction of the electroweak gauge covariant derivative such that the Lagrangian for the Higgs doublet and the electroweak bosons is just

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 - \frac{1}{4} W_{\mu\nu}^k W_k^{\mu\nu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (1.68)$$

⁸³⁸ with $D_\mu = \partial_\mu + ig_W W_\mu^k t^k + ig' \frac{Y}{2} B_\mu$.

We note that it is convenient to pick a gauge such that the Goldstone fields do not appear in the Lagrangian, upon which we may identify the field $\eta(x)$ with the physical Higgs field, $h(x)$. The field mass terms then very apparently come via the covariant derivative, namely, as

$$W_\mu^k \sigma^k + B_\mu = \begin{pmatrix} W_\mu^3 + B_\mu & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 + B_\mu \end{pmatrix} \quad (1.69)$$

we may then write

$$D_\mu \phi = \frac{1}{2\sqrt{2}} \begin{pmatrix} 2\partial_\mu + ig_W W_\mu^3 + ig' Y B_\mu & ig_W W_\mu^1 + \frac{1}{2} g_W W_\mu^2 \\ ig_W W_\mu^1 - g_W W_\mu^2 & 2\partial_\mu - ig_W W_\mu^3 + ig' Y B_\mu \end{pmatrix} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \quad (1.70)$$

$$= \frac{1}{2\sqrt{2}} \begin{pmatrix} ig_W (W_\mu^1 - iW_\mu^2)(v + h) \\ (2\partial_\mu - ig_W W_\mu^3 + ig' Y B_\mu)(v + h) \end{pmatrix} \quad (1.71)$$

⁸³⁹ As identified above, $Y = 2(Q - I_W^{(3)})$. The Higgs has 0 electric charge, and the lower doublet
⁸⁴⁰ component has $I_W^{(3)} = -\frac{1}{2}$, yielding $Y = 1$.

Computing $(D_\mu \phi)^\dagger (D^\mu \phi)$, then, yields

$$\frac{1}{8} g_W^2 (W_\mu^1 + iW_\mu^2)(W^{\mu 1} - iW^{\mu 2})(v + h)^2 + \frac{1}{8} (2\partial_\mu + ig_W W_\mu^3 - ig' B_\mu)(2\partial^\mu - ig_W W^{\mu 3} + ig' B^\mu)(v + h)^2 \quad (1.72)$$

and extracting terms quadratic in the fields gives

$$\frac{1}{8} g_W^2 v^2 (W_{\mu 1} W^{\mu 1} + W_{\mu 2} W^{\mu 2}) + \frac{1}{8} v^2 (g_W W_\mu^3 - g' B_\mu)(g_W W^{\mu 3} - g' B^\mu) \quad (1.73)$$

meaning that W_μ^1 and W_μ^2 have masses $m_W = \frac{1}{2}g_W v$. The neutral boson case is a bit more complicated. Writing the corresponding term as

$$\frac{1}{8}v^2 \begin{pmatrix} W_\mu^3 & B_\mu \end{pmatrix} \begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix} \begin{pmatrix} W^{\mu 3} \\ B^\mu \end{pmatrix} \quad (1.74)$$

we note that we must diagonalize this mass matrix to get the physical mass eigenstates. Doing so in the usual way yields eigenvalues 0 , $g'^2 + g_W^2$, thus corresponding to $m_\gamma = 0$ and $m_Z = \frac{1}{2}v\sqrt{g'^2 + g_W^2}$, with physical fields as the (normalized) eigenvectors

$$A_\mu = \frac{g'W_\mu^3 + g_W B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.75)$$

$$Z_\mu = \frac{g_W W_\mu^3 - g' B_\mu}{\sqrt{g_W^2 + g'^2}} \quad (1.76)$$

From this form, the angular parametrization of the physical fields is very apparent, namely, defining

$$\tan \theta_W = \frac{g'}{g_W}, \quad (1.77)$$

these equations may be written in terms of the single parameter θ_W as

$$A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3 \quad (1.78)$$

$$Z_\mu = -\sin \theta_W B_\mu + \cos \theta_W W_\mu^3 \quad (1.79)$$

and, notably, from the above equations,

$$\frac{m_W}{m_Z} = \cos \theta_W. \quad (1.80)$$

To get the mass terms from Equation 1.72, we extracted those terms quadratic in fields, i.e., the v^2 terms within $(v + h)^2$. However there are also terms of the form VVh and $VVhh$ that arise, which describe the Higgs interactions with the corresponding vector bosons $V = W^\pm, Z$. Namely, identifying physical W bosons as

$$W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2) \quad (1.81)$$

we may express the first term of Equation 1.72 as

$$\frac{1}{4}g_W^2 W_\mu^- W^{+\mu} (v + h)^2 = \frac{1}{4}g_W^2 v^2 W_\mu^- W^{+\mu} + \frac{1}{2}g_W^2 v W_\mu^- W^{+\mu} h + \frac{1}{4}g_W^2 W_\mu^- W^{+\mu} h^2 \quad (1.82)$$

with the first term corresponding to the mass term $m_W = \frac{1}{2}g_W v$, and the second two terms corresponding to hW^+W^- and hhW^+W^- vertices. Of particular note is the coupling strength

$$g_{HWW} = \frac{1}{2}g_W^2 v = g_W m_W \quad (1.83)$$

841 which is proportional to the W mass – an analysis with the form of the physical Z boson
842 finds that the coupling g_{HZZ} is also proportional to the Z mass.

The Higgs coupling to fermions (in particular to quarks) is of particular interest for this thesis. We showed above that a naive introduction of a mass term

$$m\bar{f}f = m(\bar{f}_L f_R + \bar{f}_R f_L) \quad (1.84)$$

843 is manifestly not gauge invariant because right and left handed particles transform differently
844 under $SU(2)_L$. However, because the Higgs is constructed via an $SU(2)_L$ doublet, ϕ , writing
845 a fermion doublet as L and conjugate \bar{L} , it is apparent that $\bar{L}\phi$ is invariant under $SU(2)_L$.

Combining with the right handed singlet, R , creates a term invariant under $SU(2)_L \times U(1)_Y$, $\bar{L}\phi R$ (and correspondingly $(\bar{L}\phi R)^\dagger$), such that we may include Yukawa [16] terms

$$\mathcal{L}_{Yukawa} = -g_f \left[\begin{pmatrix} \bar{f}_1 & \bar{f}_2 \end{pmatrix}_L \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} f_R + \bar{f}_R \begin{pmatrix} \phi^{+*} & \phi^{0*} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}_L \right] \quad (1.85)$$

846 where g_f is a corresponding Yukawa coupling, f_1 and f_2 have been used to denote components
847 of the left-handed doublet and f_R the corresponding right-handed singlet.

After spontaneous symmetry breaking, with the gauge as described above to remove the Goldstone fields, the Higgs doublet becomes

$$\phi(x) = \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (1.86)$$

giving rise to terms such as

$$-\frac{1}{\sqrt{2}}g_f v(\bar{f}_{2L}\bar{f}_R + \bar{f}_R f_{2L}) - \frac{1}{\sqrt{2}}g_f h(\bar{f}_{2L}\bar{f}_R + \bar{f}_R f_{2L}) \quad (1.87)$$

where we have kept the subscript f_{2L} to emphasize that these terms *only* impact the lower component of the left-handed doublet because of the 0 in the upper component of the Higgs doublet. Leaving this aside for a second, we note that the first term has the form of the desired mass term above (identifying f_{2L} to f_L) while the second term describes the coupling of the fermion to the physical Higgs field. The corresponding Yukawa coupling may be chosen to be consistent with the observed fermion mass, namely

$$g_f = \sqrt{2} \frac{m_f}{v} \quad (1.88)$$

such that

$$\mathcal{L}_f = -m_f \bar{f}f - \frac{m_f}{v} \bar{f}fh. \quad (1.89)$$

⁸⁴⁸ Notably here, the fermion coupling to the Higgs boson scales with the mass of the fermion, a
⁸⁴⁹ fact that is extremely relevant for this thesis analysis.

As we said above, these terms *only* impact the lower component of the left-handed doublet. The inclusion of terms for the upper component is accomplished via the introduction of a Higgs conjugate doublet, defined as

$$\phi_c = -i\sigma_2\phi^* = \begin{pmatrix} -\phi^{0*} \\ \phi^- \end{pmatrix}. \quad (1.90)$$

⁸⁵⁰ The argument proceeds similarly to the above, with similar results for couplings and masses
⁸⁵¹ of upper components.

⁸⁵² 1.7 The Standard Model: A Summary

After all of the above, we may write the Standard Model as a theory with a local $SU(3) \times SU(2)_L \times U(1)_Y$ gauge symmetry, described by the Lagrangian

$$\mathcal{L} = \sum_f \bar{f}i\gamma^\mu D_\mu f - \frac{1}{4} \sum_{gauges} F_{\mu\nu}F^{\mu\nu} + (D_\mu\phi)^\dagger(D^\mu\phi) - \mu^2\phi^\dagger\phi - \lambda(\phi^\dagger\phi)^2 \quad (1.91)$$

where $D_\mu = \partial_\mu + ig_W W_\mu^k t^k + ig' \frac{Y}{2} B_\mu + ig_S G_\mu^a t^a$, in addition to the Yukawa terms, which we write generally as

$$\mathcal{L}_{Yukawa} = - \sum_{f,\phi=\phi_c,-\phi_c} y_f (\bar{f}\phi f + (\bar{f}\phi f)^\dagger) \quad (1.92)$$

with the sum running over appropriate chiral fermion and Higgs doublets.

The $SU(2)_L \times U(1)_Y$ subgroup is spontaneously broken to a $U(1)$ symmetry, lending mass to the associated gauge bosons and fermions. Of relevance for this thesis is the resulting physical Higgs field, with a predicted trilinear self-interaction and associated coupling λv , related to the experimentally observed Higgs boson mass by $m_H = \sqrt{2\lambda v^2}$, as well as the fact that the strength of the Higgs coupling to fermions scales proportionally with the fermion mass.

The Standard Model has been monumentally successful, with predictions consistent across many varied experimental cross-checks. This thesis participates in one such cross check. However, the Standard Model is notably not a complete theory of the universe – there is no inclusion of gravity, for instance, though a consistent description may be provided with the introduction of a spin-2 particle. Neutrino oscillations demonstrate that neutrinos have mass, but right-handed neutrinos have not been observed, leading to questions about whether there is a different mechanism to provide neutrinos with mass than that described above. Cosmology tells us that dark matter exists, but there is no corresponding particle within the Standard Model. This thesis therefore also participates in searches for physics beyond the Standard Model. We will provide a sketch of the relevant theories in the following chapter, though a detailed theoretical discussion is beyond the scope of this work.

871

Chapter 2

872

DI-HIGGS PHENOMENOLOGY AND PHYSICS BEYOND THE STANDARD MODEL

873

874 This thesis focuses on searches for di-Higgs production in the $b\bar{b}b\bar{b}$ final state. In this
 875 chapter, we will provide a brief overview of the practical theoretical information motivating
 876 such searches. Though the searches test for physics beyond the Standard Model, particularly
 877 in the search for resonances, the goal of the experimental results is to be somewhat agnostic
 878 to particular theoretical frameworks. An in depth treatment of such models is therefore
 879 beyond the scope of this thesis, though we will attempt to provide a grounding for the models
 880 that we consider.

881 **2.1 Intro to Di-Higgs**

882 Di-Higgs searches can be split into two major theoretical categories: *resonant searches*, in
 883 which a physical resonance is produced that subsequently decays into two Higgs bosons, and
 884 a *non-resonant searches* in which no physical resonance is produced, but where the HH
 885 production cross section has a contribution from an exchange of a *virtual* or *off-shell* particle.

886 The focus of this thesis is gluon initiated processes – in the case of di-Higgs this is termed
 887 gluon-gluon fusion (ggF). HH production may also occur via vector boson fusion [17]. However
 888 the cross section for such production is significantly smaller. Representative Feynman diagrams
 889 are shown for gluon-gluon fusion resonant production in Figure 2.1 and for non-resonant
 890 production in Figure 2.2.

891 As shown in Chapter 1, the Higgs coupling to fermions scales with particle mass. As the
 892 top quark has a mass of 173 GeV, whereas the H has a mass of 125 GeV, such that $H \rightarrow t\bar{t}$ is
 893 kinematically disfavored, $H \rightarrow b\bar{b}$ is the dominant fermionic Higgs decay mode, and, in fact,

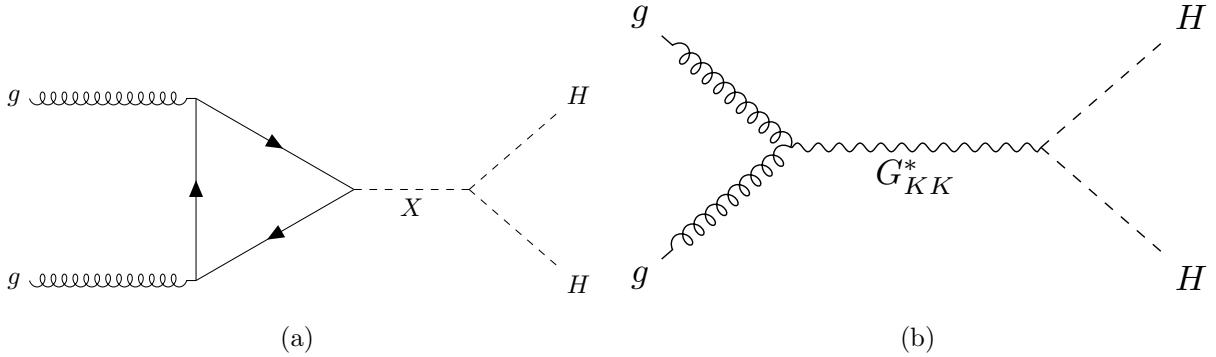


Figure 2.1: Representative diagrams for the gluon-gluon fusion production of spin-0 (X) and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs bosons. The spin-0 resonance considered for this thesis is a generic narrow width resonance which may be interpreted in the context of two Higgs doublet models [18], whereas the spin-2 resonance is considered as a Kaluza-Klein graviton within the bulk Randall-Sundrum (RS) model [19, 20].

the dominant overall decay mode, with a branching fraction of around 58 %. The dominant top quark Yukawa coupling to the H does play a role in H production, however – gluon-gluon fusion is dominated by processes including a top loop.

The single H properties translate to HH production, with $HH \rightarrow b\bar{b}b\bar{b}$ accounting for around 34 % of all HH decays. The H H branching fractions are shown in Figure 2.3.

2.2 Resonant HH Searches

Resonant di-Higgs production is predicted in a variety of extensions to the Standard Model. In particular, this thesis presents searches for both spin-0 and spin-2 resonances. The decay of spin-1 resonances to two identical spin-0 bosons is prohibited, as the final state must correspondingly be symmetric under particle exchange, but this process would require orbital angular momentum $\ell = 1$, and thus an anti-symmetric final state. Each model considered here is implemented in a particular theoretical context, but set up experimental results for generic searches.

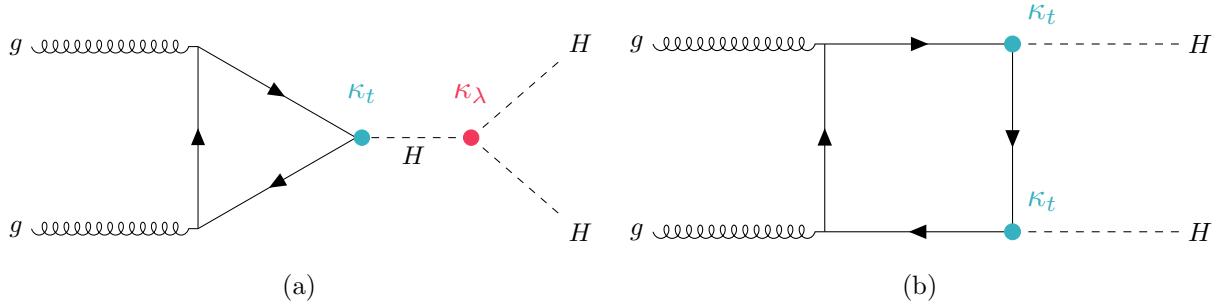


Figure 2.2: Dominant contributing diagrams for non-resonant gluon-gluon fusion production of HH . κ_λ and κ_t represent variations of the Higgs self-coupling and coupling to top quarks respectively, relative to that predicted by the Standard Model.

The spin-2 signal considered is implemented within the bulk Randall-Sundrum (RS) model [19, 20], which features spin-2 Kaluza-Klein gravitons, G_{KK}^* , that are produced via gluon-fusion and which may decay to a pair of Higgs bosons. The model predicts such gravitons as a consequence of warped extra dimensions, and is correspondingly parametrized by a value $c = k/\overline{M}_{\text{Pl}} = 1$, where k describes a curvature scale for the extra dimension and \overline{M}_{Pl} is the Planck mass. The model considered here has $c = 1.0$. However, this model was considered in the early Run 2 HH analyses [21], and was excluded across much of the relevant mass range.

The primary theoretical focus of this work is therefore the spin-0 result, which is implemented as a generic resonance with width below detector resolution. Scalar resonances are interesting, for instance, in the context of two Higgs doublet models [18], which posit the existence of a second Higgs doublet. This leads to the existence of five scalar particles in the Higgs sector – roughly, two complex doublets provide eight degrees of freedom, three of which are “eaten” by the electroweak bosons, leaving five degrees of freedom which may correspond to physical fields.

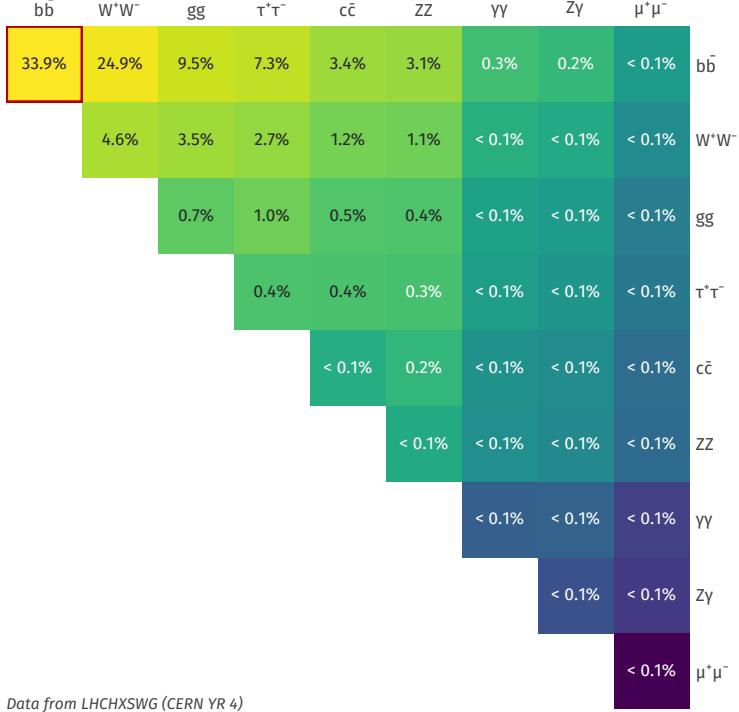


Figure 2.3: Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common decay mode, representing 34 % of all HH events produced at the LHC.

922 2.3 Non-resonant HH Searches

Non-resonant HH production is predicted by the Standard Model via the trilinear coupling discussed above, as well as via production in a fermion loop. More explicitly, after electroweak symmetry breaking, we have

$$\mathcal{L}_{SM} \supset -\lambda v^2 h^2 - \lambda v h^3 - \frac{1}{4} \lambda h^4 \quad (2.1)$$

$$= -\frac{1}{2} m_H^2 - \lambda_{HHH}^{SM} v h^3 - \lambda_{HHHH}^{SM} h^4 \quad (2.2)$$

where $m_H = \sqrt{2\lambda v^2}$ so that

$$\lambda_{HHH}^{SM} = \frac{m_H^2}{2v^2}. \quad (2.3)$$

923 The mass of the SM Higgs boson has been experimentally measured to be 125 GeV [22],
 924 and the vacuum expectation value $v = 246$ GeV has a precise determination from the muon
 925 lifetime [23]. This coupling is therefore precisely predicted in the Standard Model, such that
 926 an observed deviation from this prediction would be a clear sign of new physics.

927 The relevant diagrams for non-resonant HH production are shown in Figure 2.2. Notably,
 928 the diagrams *interfere* with each other, which can be easily seen by counting the fermion
 929 lines. A detailed theoretical discussion is provided by, e.g. [24].

930 For the searches presented here, the quark couplings to the Higgs are considered to be
 931 consistent with the Standard Model value, with measurements of the dominant top Yukawa
 932 coupling left to more sensitive direct measurements, e.g. from $t\bar{t}$ final states [25]. Variations of
 933 the trilinear coupling away from the Standard Model are considered, however. Such variations
 934 are parametrized via

$$\kappa_\lambda = \frac{\lambda_{HHH}}{\lambda_{HHH}^{SM}} \quad (2.4)$$

935 where λ_{HHH} is a varied coupling and λ_{HHH}^{SM} is the Standard Model prediction. As this
 936 variation only comes as a prefactor only with the *triangle* diagram, significant and interesting
 937 effects are observed due to the interference. Examples of the impact of this tradeoff on the
 938 di-Higgs invariant mass are shown in Figure 2.4. Generally speaking, the triangle diagram
 939 contributes more at low mass, while the box diagram contributes more at high mass.

From a quick analysis of Figure 2.2, one may see that, at leading order, the box diagram, B has amplitude proportional to κ_t^2 , defined as the ratio of the top Yukawa coupling to the value predicted by the Standard Model, whereas the triangle diagram, T has amplitude proportional to $\kappa_t \kappa_\lambda$. Therefore, the cross section is proportional to

$$\sigma(\kappa_t, \kappa_\lambda) = |A(\kappa_t, \kappa_\lambda)|^2 \quad (2.5)$$

$$\sim |\kappa_t^2 B + \kappa_t \kappa_\lambda T|^2 \quad (2.6)$$

$$= \kappa_t^4 |B|^2 + \kappa_t^3 \kappa_\lambda (BT + TB) + \kappa_t^2 \kappa_\lambda^2 |T|^2, \quad (2.7)$$

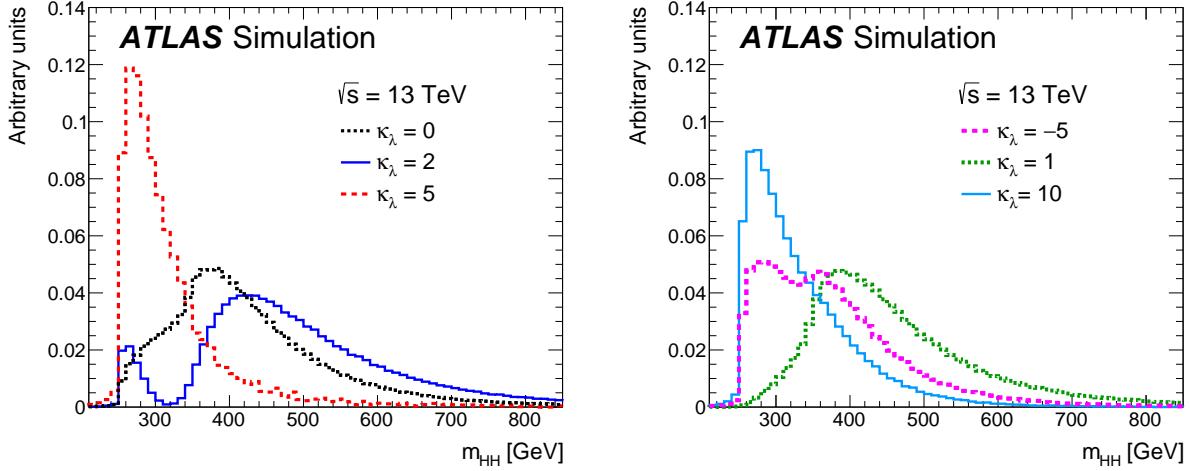


Figure 2.4: Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demonstrating the impact of the interference between the two diagrams of Figure 2.2 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram contribution, demonstrating the shape of the box diagram contribution, whereas for $\kappa_\lambda = 10$, the triangle diagram dominates, with a strong low mass peak. The interplay between the two is quite evident for other values, resulting in, e.g., the double peaked structure present for $\kappa_\lambda = 2$ (near maximal destructive interference) and $\kappa_\lambda = -5$. [21]

and thus non-resonant HH production cross section may be parametrized as a second order polynomial in κ_λ .

For positive values of κ_λ , due to the relative minus sign between the triangle and box diagrams, the interference between the two diagrams is *destructive*, with a maximum interference near $\kappa_\lambda = 2.3$, corresponding to the minimum cross section prediction. One may note that the Standard Model value of $\kappa_\lambda = 1$ is not far away from this minimum – correspondingly the Standard Model cross section for HH production is quite small, namely 31.05 fb at $\sqrt{s} = 13$ TeV for production via gluon-gluon fusion [26–33] compared to, e.g. single Higgs production, with a gluon-gluon fusion production cross section of 46.86 pb at

949 $\sqrt{s} = 13 \text{ TeV}$ [34] roughly 1500 times larger! For negative values of κ_λ , the interference is
950 constructive.

951 ATLAS projections [35] of $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ predict an expected signal strength
952 for Standard Model HH of 3.5σ with no systematic uncertainties and 3.0σ with systematic
953 uncertainties using the 3000 fb^{-1} of data from the HL-LHC (around $20\times$ the full Run 2
954 dataset considered in this thesis), constituting an *observation* of HH . As the cross section
955 for Standard Model HHH production, corresponding to the quartic Higgs interaction, is
956 much smaller (around 0.1 fb at $\sqrt{s} = 14 \text{ TeV}$ [36]), observation of triple Higgs production is
957 even farther in the future, and so is not considered here. However this may be interesting for
958 future work in a variety of Beyond the Standard Model scenarios (e.g. [37–39]).

959

Chapter 3

960

EXPERIMENTAL APPARATUS

961 What machines must we build to examine the smallest pieces of the universe? The famous
 962 equation $E = m$ provides that to create massive particles, we need to provide enough energy.
 963 In order to give kinematic phase space to the types of processes that are examined in this
 964 thesis (and many others besides), a system must be created in which there is enough energy
 965 to (at bare minimum), overcome kinematic thresholds: if you want to search for HH decays,
 966 you should have at least 250 GeV ($= 2 \times m_H$) to work with. It is not enough to simply induce
 967 such processes, however. These processes need to be captured in some way, emitted energy
 968 and particles must be characterized and identified, and in the end all of this information must
 969 be put into a useful and useable form such that selections can be made, statistics can be run,
 970 and a meaningful statement can be made about the universe. In this chapter, we describe the
 971 machines behind the physics, namely the Large Hadron Collider and the ATLAS experiment.

972 **3.1 The Large Hadron Collider**

973 The Large Hadron Collider is a particle accelerator near Geneva, Switzerland. In broad scope,
 974 it is a ring with a 27 kilometer circumference. Hadrons (usually protons or heavy ions) move
 975 in two counter-circulating beams, which are made to collide at four collision points at various
 976 points on the ring. These four collision points correspond to the four detectors placed around
 977 the ring: two “general purpose” experiments: ATLAS and CMS; LHCb, focused primarily on
 978 flavor physics; and ALICE, focused primarily on heavy ions.

979 The focus of this thesis is proton-proton collisions at center of mass energy $\sqrt{s} = 13$ TeV.
 980 The process to achieve such collisions proceeds as follows: first, an electric field strips hydrogen
 981 of its electrons, creating protons. A linear accelerator, LINAC 2, accelerates protons to

982 50 MeV. The resulting beam is injected into the Proton Synchrotron Booster (PSB), which
 983 pushes the protons to 1.4 GeV, and then the Proton Synchrotron, which brings the beam to
 984 25 GeV.

985 Protons are then transferred to the Super Proton Synchrotron (SPS), which ramps up
 986 the energy to 450 GeV. Finally, the protons enter the LHC itself, bringing the beam up to
 987 6.5 TeV [40].

988 While there is, of course, much that goes into the Large Hadron Collider development and
 989 operation, perhaps two of the most fundamental ideas are (1) how are the beams directed
 990 and manipulated and (2) what do we mean when we say “protons are accelerated”. These
 991 questions both are directly answered by pieces of hardware, namely (1) magnets and (2)
 992 radiofrequency (RF) cavities.

993 One of fundamental components of the LHC is a large set of superconducting niobium-
 994 titanium magnets. These are cooled by liquid helium to achieve superconducting temperatures,
 995 and there are several types with very specific purposes. The obvious first question with a
 996 circular accelerator is how to keep the particle beam moving around in that circle. This job
 997 is done via a set of dipole magnets placed around the *beam pipes*: the tubes containing the
 998 beam. These are designed such that the magnetic field in the center of the beam pipe runs
 999 perpendicular to the velocity of the charged particles, providing the necessary centripetal
 1000 force for the synchrotron motion.

1001 A proton beam is not made of a single proton, however, but of many protons, grouped
 1002 into a series of *bunches*. As all of these are positively charged, if unchecked, these bunches
 1003 would become diffuse and break apart. What we want is a stable beam with tightly clustered
 1004 protons to maximize the chance of a high energy collision. Such clustering is done via a series
 1005 of quadropole magnets, with field distributed as in *TODO: grab image from General Exam*.
 1006 Alternating sets of quadropoles provide the necessary forces for a tight, stable beam. While
 1007 these are the two major components of the LHC magnet system, it is not the full story –
 1008 higher order magnets are used to correct for small imperfections in the beam.

1009 Magnetic fields do no work, however, so the magnet system is unable to do the job of the

actual acceleration. This is accomplished via a set of radiofrequency (RF) cavities. Within these cavities, an electric field is made to oscillate (switch direction) at a precise rate. These rates interact with the beam via in RF *buckets*, with bunches corresponding to groups of protons that fill a given bucket. The timing is such that protons will always experience an accelerating voltage, corresponding to the 25 ns bunch spacing used at the LHC.

A nice property of this bucket/bunch configuration is that there is some self-correction – there is some finite spread in the grouping of particles. If a particle arrives too early, it will experience some decelerating voltage; if too late, it will experience a higher accelerating voltage.

3.1.1 The LHC Schedule

The physics program at the Large Hadron Collider is split into a variety of data taking periods called *runs*. These runs correspond to various detector/accelerator configurations, and are interspersed with *long shutdowns* – periods used for detector/accelerator upgrades in preparation for the next run. The LHC timeline is as follows

1. Run 1 (2010–2013): First run of the LHC, operating at center of mass energy $\sqrt{s} = 7 \text{ TeV}$, increased to 8 TeV in 2012. ATLAS recorded 4.57 fb^{-1} and 20.3 fb^{-1} of data usable for physics at $\sqrt{s} = 7 \text{ TeV}$ and 8 TeV respectively.
2. Long Shutdown 1 (LS1; 2013–2015): Upgrades to accelerator complex, magnet system, to allow for increase in energy. Design energy was $\sqrt{s} = 14 \text{ TeV}$, delays in “training” of superconducting magnets led to decrease to $\sqrt{s} = 13 \text{ TeV}$.
3. Run 2 (2015–2018): Second run of the LHC, operating at center of mass energy $\sqrt{s} = 13 \text{ TeV}$. Data from this run is used in this thesis, with 139 fb^{-1} of data available for physics from the ATLAS experiment.
4. Long Shutdown 2 (LS2; 2019–2021): Upgrades to ATLAS muon spectrometer (New

1034 Small Wheel), liquid argon calorimeter; upgrades in preparation for the High Luminosity
1035 LHC (HL-LHC).

1036 5. Run 3 (2021–2023?): Third run of the LHC, target center of mass energy $\sqrt{s} =$
1037 $13 - 14 \text{ TeV}$, total target luminosity 300 fb^{-1} .

1038 6. Long Shutdown 3 (LS3; 2024?–2026?): Further upgrades for the HL-LHC.

1039 7. Run 4, 5, ... (2026? onward): High Luminosity LHC – goal is to achieve instantaneous
1040 luminosities by a factor of five, massively enlarging available statistics for physics.
1041 Projected 3000 to 4000 fb^{-1} , > 20 times the full Run 2 ATLAS dataset.

1042 3.2 The ATLAS Experiment

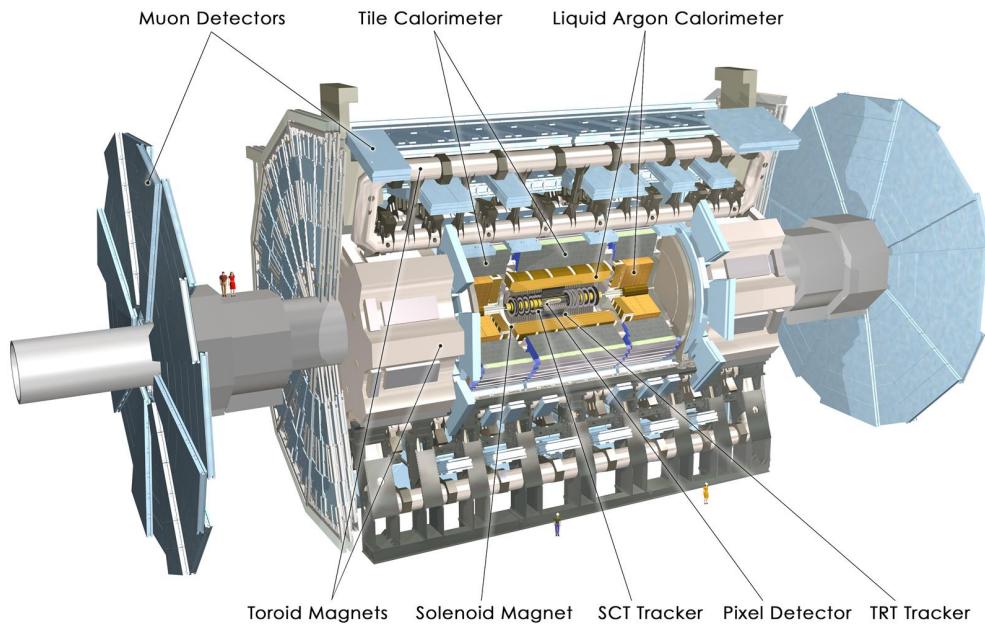


Figure 3.1: Diagram of the ATLAS detector [41]

1043 This thesis focuses on searches done with the ATLAS experiment. As mentioned, this is one

1044 of two “general purpose” experiments at the LHC, by which we mean there is a very large and
 1045 broad variety of physics done within the experimental collaboration. This broad physics focus
 1046 has a direct relation to the design of the ATLAS detector [42], pictured in Figure 3.1, which
 1047 is composed of a sophisticated set of subsystems designed to fully characterize the physics of
 1048 a given high energy particle collision. It consists of an inner tracking detector surrounded
 1049 by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon
 1050 spectrometer incorporating three large superconducting toroidal magnets. The ATLAS
 1051 detector covers nearly the entire solid angle around the collision point, fully characterizing
 1052 the “visible” components of a collision and allowing for indirect sensitivity to particles that
 1053 do not interact with the detector (e.g. neutrinos) via “missing” energy (roughly momentum
 1054 balance). We will go through the design and physics contribution of each of the detector
 1055 components in the following. A schematic of how various particles interact with the detector
 1056 is shown in Figure 3.2.

1057 3.2.1 ATLAS Coordinate System

1058 Of relevance for the following discussion, as well as for the analysis presented in Chapter 7,
 1059 is the ATLAS coordinate system. ATLAS uses a right-handed coordinate system with its
 1060 origin at the nominal interaction point (IP) in the center of the detector and the z -axis along
 1061 the beam pipe. The x -axis points from the IP to the centre of the LHC ring, and the y -axis
 1062 points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse plane, ϕ being the
 1063 azimuthal angle around the z -axis. The pseudorapidity is defined in terms of the polar angle
 1064 θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.
 1065 These coordinates are shown in Figure 3.3.

1066 3.2.2 Inner Detector

1067 The purpose of the inner detector is the reconstruction of the trajectory of charged particles,
 1068 called *tracking*. This is accomplished primarily through the collection of electrons displaced
 1069 when a charged particle passes through a tracking detector. By setting up multiple layers of

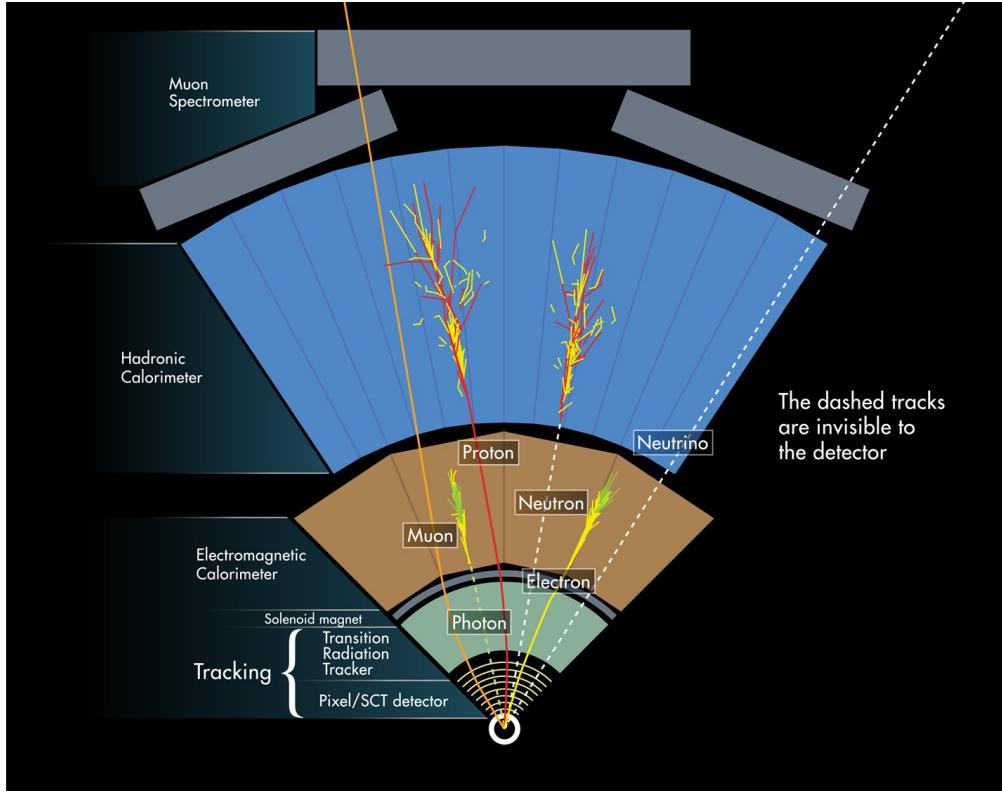


Figure 3.2: Cross section of the ATLAS detector showing how particles interact with various detector components [43]

such detectors, such that a given particle leaves a signature, known as a “hit”, in each layer, the trajectory of the particle may be inferred via “connecting the dots” between these hits.

The raw trajectory of a particle only provides positional information. However, the trajectory of a charged particle in a known magnetic field additionally provides information on particle momentum and charge via the curvature of the corresponding track (cf. $\vec{F} = q\vec{v} \times \vec{B}$). The inner detector system is therefore surrounded by a solenoid magnet, providing a 2 T magnetic field along the z -axis (yielding curvature in the transverse $x - y$ plane).

The inner detector provides charged particle tracking in the range $|\eta| < 2.5$ via a series of detector layers. The innermost of these is the high-granularity silicon pixel detector which typically provides four measurements per track, with the first hit in the insertable B-layer

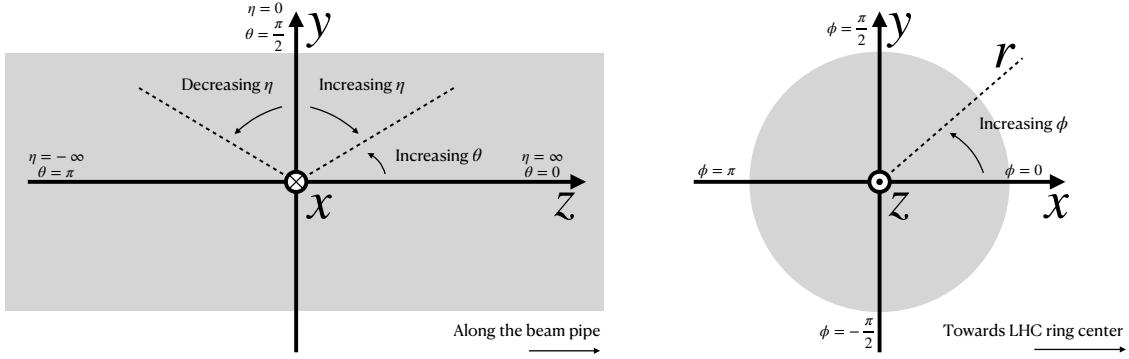


Figure 3.3: 2D projections of the ATLAS coordinate system

1080 (IBL) installed before Run 2 [44, 45]. This is very close to the interaction point with a
 1081 high degree of positional information, and is therefore very important for e.g. b -tagging (see
 1082 Chapter 5). It is followed by the silicon microstrip tracker (SCT), which usually provides
 1083 eight measurements per track. This is lower granularity, but similar in concept to the pixel
 1084 detector.

1085 Both of these silicon detectors are complemented by the transition radiation tracker
 1086 (TRT), which extends the radial track reconstruction within the range $|\eta| < 2.0$. This is
 1087 a different design, composed of *drift tubes*, i.e. straws filled with Xenon gas with a wire
 1088 in the center, but similarly collects electrons displaced by ionizing particles. In addition,
 1089 the TRT includes materials with widely varying indices of refraction, which leads to the
 1090 production of transition radiation, namely radiation produced by a charged particle passing
 1091 through an inhomogeneous medium. The energy loss on such a transition is proportional
 1092 to the Lorentz factor $\gamma = E/m$ – correspondingly, lighter particles (e.g. electrons) tend to
 1093 lose more energy and emit more photons compared to heavier particles (e.g. pions). In the
 1094 detector, this corresponds to a larger fraction of hits (typically 30 in total) above a given

1095 high energy-deposit threshold for electrons, providing particle identification information.

1096 *3.2.3 Calorimeter*

1097 Surrounding the inner detector in ATLAS is the calorimeter. The principle of the calorimeter
1098 is to completely absorb the energy of a produced particle in order to measure it. However,
1099 a pure block of absorber does not provide much information about the particle interaction
1100 with the material. The ATLAS calorimeter therefore has a *sampling calorimeter* structure,
1101 namely, layers of absorber interspersed with layers of sensitive material, giving the calorimeter
1102 “stopping power” while allowing detailed measurement of the resulting particle shower and
1103 corresponding deposited energy.

1104 The ATLAS calorimetersystem covers the pseudorapidity range $|\eta| < 4.9$, and is primarily
1105 composed of two components, an electromagnetic calorimeter, designed to measure particles
1106 which primarily interact via electromagnetism (e.g. photons and electrons), and a hadronic
1107 calorimeter, designed to measure particles which interact via the strong force (e.g. pions,
1108 other hadrons). We will return to the differences between these in a moment.

1109 In ATLAS, the electromagnetic calorimeter covers the region of $|\eta| < 3.2$, and uses
1110 lead for the absorbers and liquid-argon for the sensitive material. It is high granularity
1111 and, geometrically, has two components: the “barrel”, which covers the cylindrical body of
1112 the detector volume and the “endcap”, covering the ends. An additional thin liquid-argon
1113 presampler covers $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters.

1114 The hadronic calorimeter is composed of alternating steel and plastic scintillator tiles,
1115 segmented into three barrel structures within $|\eta| < 1.7$, in addition to two copper/liquid-argon
1116 endcap calorimeters.

1117 The solid angle coverage is completed with forward copper/liquid-argon and tungsten/liquid-
1118 argon calorimeter modules optimized for electromagnetic and hadronic energy measurements
1119 respectively.

1120 *3.2.4 Muon Spectrometer*

1121 While muons interact electromagnetically, they are around 200 times heavier than electrons
 1122 ($m_\mu = 106 \text{ MeV}$, while $m_e = 0.510 \text{ MeV}$). Therefore, electromagnetic interactions with ab-
 1123 sorbers in the calorimeter are not sufficient to stop them, and, as they do not interact via the
 1124 strong force, hard scattering with nuclei is rare. A dedicated system for muon measurements
 1125 is therefore required.

1126 The muon spectrometer (MS) is the outermost layer of ATLAS and is designed for this
 1127 purpose. It is composed of three parts: a set of triggering chambers, which detect if there is
 1128 a muon and provide a coordinate measurement, in conjunction with high-precision tracking
 1129 chambers, which measure the deflection of muons in a magnetic field to measure muon
 1130 momentum, similar to the inner detector solenoid. The magnetic field is generated by the
 1131 superconducting air-core toroidal magnets, with a field integral between 2.0 and 6.0 T m
 1132 across most of the detector. The toroid magnetic field runs roughly in a circle in the $x - y$
 1133 plane around the beam line, leading to muon curvature along the z-axis.

1134 The precision tracking system covers the region $|\eta| < 2.7$ via three layers of monitored
 1135 drift tubes, and is complemented by cathode-strip chambers in the forward region, where the
 1136 background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate
 1137 chambers in the barrel, and thin-gap chambers in the endcap regions.

1138 *3.2.5 Triggering*

1139 During a typical run of the LHC, there are roughly 1 billion collisions in ATLAS per second
 1140 (1 GHz), corresponding to a 40 MHz bunch crossing rate [46]. Saving the information from
 1141 all of them is not only unnecessary, but infeasible. The ATLAS trigger system provides a
 1142 sophisticated set of selections to filter the collision data and only keep those collision events
 1143 useful for downstream analysis.

1144 These events are selected by the first-level trigger system, which is implemented in custom
 1145 hardware, and accepts events at a rate below 100 kHz. Selections are then made by algorithms

1146 implemented in software in the high-level trigger [47], reducing this further, and, in the end,
1147 events are recorded to disk at much more manageable rate of about 1 kHz.

1148 An extensive set of ATLAS software [48] is open source, including the software used for
1149 real and simulated data reconstruction and analysis and that used in the trigger and data
1150 acquisition systems of the experiment.

1151 3.2.6 Particle Showers and the Calorimeter

1152 The design of the ATLAS detector is directly tied to the physics it is trying to detect. Of these,
1153 possibly the most non-trivial distinction is in the calorimeter design. It is therefore useful to
1154 discuss in more detail the various properties of electromagnetic and hadronic interactions
1155 with material, and how these correspond to the particle showers measured by the detector
1156 described above.

1157 Electromagnetic showers in ATLAS predominantly occur via bremsstrahlung, or “braking
1158 radiation”, and electron-positron pair production. This proceeds roughly as follows: an electron
1159 entering a material is deflected by the electromagnetic field of a heavy nucleus. This results in
1160 the radiation of a photon. That photon produces an electron-positron pair, and the process
1161 repeats, resulting in a shower structure. At each step, characterized by *radiation length*, X_0 ,
1162 the number of particles approximately doubles and the average particle energy decreases by
1163 approximately a factor of two. *TODO: Include nice Thomson image*

Note that bremsstrahlung and pair production only dominate in specific energy regimes, with other processes taking over depending on particle energy. For electrons, bremsstrahlung only dominates for higher energies, as low energy electrons will form ions with the atoms of the material. The point where the rates for the two processes are equal is called the *critical energy*, and is roughly

$$E_c \approx \frac{800 \text{ MeV}}{Z} \quad (3.1)$$

1164 where Z is the nuclear charge. From a similar analysis of rates, we may see that the
1165 bremsstrahlung rate is inversely proportional to the square of the mass of the particle. This

₁₁₆₆ explains why muons do not shower in a similar way, as the rate of bremsstrahlung is suppressed
₁₁₆₇ by $(m_e/m_\mu)^2$ relative to electrons.

For lead, the absorber used for the ATLAS electromagnetic calorimeter, which has $Z = 82$, this critical energy is therefore around 10 MeV. Electrons resulting from LHC collisions are of a 1.3×10^3 GeV scale. With the approximation of a reduction in particle energy by a factor of two every radiation length, the number of radiation lengths before the critical energy is reached is

$$x = \frac{\ln(E/E_c)}{\ln 2} \quad (3.2)$$

₁₁₆₈ such that for a 100 GeV shower in lead, $x \sim 13$. The radiation length for lead is around
₁₁₆₉ 0.56 cm, such that an electromagnetic shower could be expected to be captured within 10 cm
₁₁₇₀ of lead.

₁₁₇₁ Electromagnetic showers are therefore characterized by depositing much of their energy
₁₁₇₂ within a small region of space. As we show below (Chapter 4) though electromagnetic
₁₁₇₃ showering is not deterministic, the large number of particles and the restricted set of processes
₁₁₇₄ involved means that the shower development as a whole is very similar between individual
₁₁₇₅ electromagnetic showers of the same energy.

₁₁₇₆ For completeness, note as well that pair production dominates for photons of energy greater
₁₁₇₇ than around 10 MeV, whereas for lower energies (below around 1 MeV), the photoelectric
₁₁₇₈ effect, namely atomic photon absorption and electron emission, dominates.

₁₁₇₉ Hadronic showers are distinguished by the fact that they interact strongly with atomic
₁₁₈₀ nuclei. They are correspondingly more complex because (1) they involve a wider variety
₁₁₈₁ of processes than electromagnetic showers, and (2) these processes have a wide variety of
₁₁₈₂ associated length scales. Because these are heavier than electrons (e.g. protons and charged
₁₁₈₃ pions) bremsstrahlung is suppressed, but ionization interactions with the electrons will cause
₁₁₈₄ these particles to lose energy as they pass through the material. Hadronic showering occurs
₁₁₈₅ on interaction with atomic nuclei. This may lead to production of, e.g. both charged (π^\pm)
₁₁₈₆ and neutral (π^0) pions. The π^0 lifetime is much much shorter than that of the charged pions
₁₁₈₇ (around a factor of 10^8), and immediately decays to two photons, starting an electromagnetic

shower, as described above. The longer lived π^\pm travel further in the detector before experiencing another strong interaction with more particles produced, also with varying lifetimes and decay properties.

It is therefore immediately apparent that hadronic showers are more complex than electromagnetic ones (electromagnetic showers can be a subset of the hadronic!), and therefore much more variable from shower to shower. The length scales involved are also significantly larger due to the reliance on nuclear interactions, characterized by length λ_I , which is around 17 cm for iron (used in the ATLAS hadronic calorimeter). This motivates the calorimeter design, and results in the properties demonstrated in Figure 3.2.

1197

Chapter 4

1198

SIMULATION

1199

Simulated physics samples are a core piece of the physics output of the Large Hadron Collider, providing a map from a physics theory into what is observed in our detector. This is crucial for searches for new physics, where simulation is necessary to describe what a given signal model looks like, but also extremely valuable for describing the physics of the Standard Model, providing detailed predictions of background processes for use in everything from designing simple cuts to training multivariate discriminators. Broadly, simulation can be split into two stages: *event generation*, in which physics theory is used to generate a description of particles present after a proton-proton collision, and *detector simulation*, which passes this particle description through a simulation of the detector material, providing a view of the physics event as it would be seen in ATLAS data. Such simulation is often called Monte Carlo in reference to the underlying mathematical framework, which relies on random sampling.

1210

4.1 Event Generation

1211

A variety of tools are used to simulate various aspects of event generation. One such aspect is generation of the “hard scatter” event, i.e., two protons collide and some desired physics process happens. In practice, this is not quite as simple as two quarks or gluons interacting. Protons are composed of three “valence” quarks with various momenta interacting with each other via exchange of gluons, but also a sea of virtual gluons which may decay into other quarks. A hard scatter event is therefore characterized by the corresponding particle level diagrams, but additionally by a set of *parton distribution functions* (PDFs), which describe the probability to find constituent quarks or gluons at carrying various momenta at a given energy scale (often written Q^2). Such PDFs are measured experimentally *TODO: cite* and

1220 the selection of a “PDF set” and a given physics process characterizes the hard scatter.
 1221 Depending on the model being considered and the particular theoretical constraints, processes
 1222 are often simulated at either leading (LO) or next to leading order (NLO), corresponding to
 1223 the order of the perturbative expansion (i.e. tree level or 1 loop diagrams). Various additional
 1224 tools are developed for such NLO calculations, including POWHEG Box v2 [49–51], which is
 1225 used for this thesis. MADGRAPH [52] is used in this thesis for leading order simulation.

1226 The hard scatter is not the only component of a given collider event, however. Incoming
 1227 and outgoing particles are themselves very energetic and may radiate particles along their
 1228 trajectory. In particular, gluons, which have a self-interaction term as described in Chapter 1,
 1229 may be radiated, which subsequently themselves radiate gluons or decay to quarks which can
 1230 also radiate gluons, in a whole mess of QCD that both contributes to the particle content
 1231 of a collider event and is not directly described by the hard scatter. This cascade, called a
 1232 *parton shower*, has a dedicated set of simulation tools. For this thesis, HERWIG 7 [53][54] and
 1233 PYTHIA 8 [55] are used, which interface with tools such as MADGRAPH for simulation.

1234 Due to color confinement (Chapter 1), quarks and gluons cannot be observed free particles,
 1235 but rather undergo a process called hadronization, in which they are grouped into colorless
 1236 hadrons (e.g. *mesons*, consisting of one quark and one antiquark). In simulation, this is also
 1237 handled with tools such as HERWIG 7 or PYTHIA 8.

1238 The physics of b -quarks is quite important for a variety of searches for new physics and
 1239 measurements of the Standard Model, including this thesis work. Correspondingly, the decay
 1240 of “heavy flavor” particles (e.g. B and D mesons, containing b and c quarks respectively)
 1241 has been very well studied, and a dedicated simulation tool, EVTGEN [56], is used for such
 1242 processes.

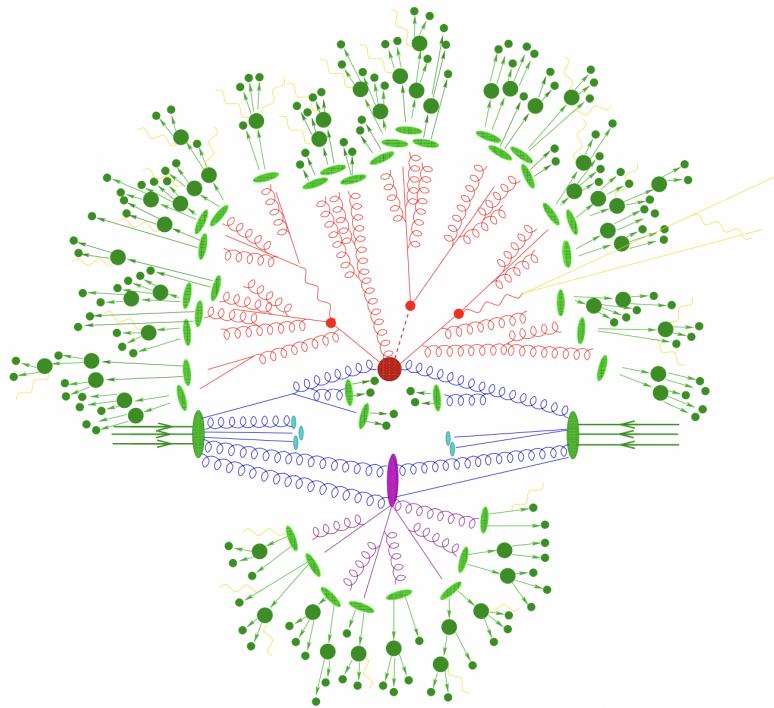


Figure 4.1: Schematic diagram of the Monte Carlo simulation of a hadron-hadron collision. The incoming hadrons are the green blobs with the arrows on the left and right, with the red blob in the center representing the hard scatter event, and the purple representing a secondary hard scatter. Radiation from both incoming and outgoing particles is shown, and the light green blobs represent hadronization, with the outermost dark green circles corresponding to the final state hadrons. Yellow lines are radiated photons. [57]

1243 **4.2 Detector Simulation**

1244 Event generation provides a full and exact description of the particle content of a given
1245 collider event. This description is useful, but is an artifact of the simulation – for real physics
1246 events, we must rely on the information collected by sophisticated detectors (Chapter 3) to
1247 make statements about the physics content of collider events. The simulation of how particles
1248 interact with the physical detector and of the corresponding information that is collected is
1249 therefore a necessary step of physics simulation at the LHC. The design and components of
1250 the ATLAS detector are described in Chapter 3. Simulation of this detector quickly becomes
1251 complicated – there are a variety of different materials and subdetectors, each with particular
1252 configurations and resolutions. Interactions of particles with the detector materials can cause
1253 showering, and such showers must be simulated and characterized.

1254 In ATLAS, the GEANT4 [58] simulation toolkit is used for detailed simulation of the
1255 ATLAS detector, often referred to as *full simulation*. The method can be thought of as
1256 proceeding step by step as a particle moves through the detector, simulating the interaction
1257 of the material at each stage, and following each branch of each resulting shower with a
1258 similarly detailed step by step simulation.

1259 This type of simulation is very computationally intensive, especially in the calorimeter,
1260 which has a high density of material, leading to an extremely large set of material interactions
1261 to simulate. There is correspondingly a large effort within ATLAS to develop techniques to
1262 decrease the computational load – these techniques will be of increasing importance for Run
1263 3 and the HL-LHC, which will have increased computational need due to the high complexity
1264 and large volume of collected physics events, along with the corresponding set of simulated
1265 physics events [59]. The divergence of the baseline computing model from the projected
1266 computing budget is shown in Figure 4.2.

1267 The fast simulation used for this thesis, AtlFast-II [61], is one such technique, which uses
1268 a parametrized simulation of the calorimeter, called FastCaloSim, in conjunction with full
1269 simulation of the inner detector, to achieve an order of magnitude speed up in simulation

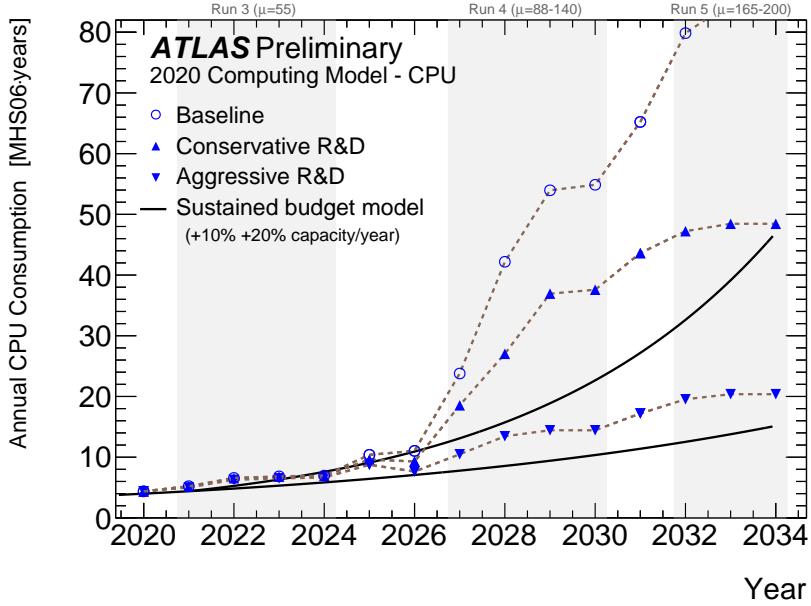


Figure 4.2: The projected ATLAS computational requirements for Run 3 and the HL-LHC relative to the projected computing budget. Aggressive R&D is required to keep resources within budget [60].

time. This parametrized simulation uses a simplified detector geometry, in conjunction with a simulation of particle shower development based on statistical sampling of distributions from fully simulated events, to massively speed up simulation time and computational load.

Such a speed up comes at a bit of a cost in performance. In particular, the modeling of jet substructure (see Chapter 5) historically has been an issue for FastCaloSim. The ATLAS authorship qualification work supporting this thesis is an effort to improve such modeling, and is part of a suite of updates being considered for a new fast simulation targeting Run 3. We briefly describe this work in the following.

1278 **4.3 Correlated Fluctuations in FastCaloSim**

1279 A variety of developments have been made to FastCaloSim, improving on the version used for
 1280 AtlFast-II. This new fast calorimeter simulation [62] is largely based on two components: one
 1281 which describes the *total energy* deposited in each calorimeter layer as a shower moves from
 1282 the interaction point outward, and one which describes the *shape*, i.e., the pattern of energy
 1283 deposits, of a shower in each respective calorimeter layer. Both methods are parametrizations
 1284 of the full simulation, and therefore are considered to be performing well if they are able
 1285 to reproduce corresponding full simulation distributions. Of course, directly sampling from
 1286 a library of showers would identically reproduce such distributions – however a statistical
 1287 sampling of various shower *properties* provides much more generality in the simulation.

1288 For the simulation of total energy in each given layer, the primary challenge is that such
 1289 energy deposits are highly correlated. The new FastCaloSim thus relies on a technique called
 1290 Principal Component Analysis (PCA) [63] to de-correlate the layers, aiding parametrization.

1291 The PCA chain transforms N energy inputs into N Gaussians and projects these Gaussians
 1292 onto the eigenvectors of the corresponding covariance matrix. This results in N de-correlated
 1293 components, as the eigenvectors are orthogonal. The component of the PCA decomposition
 1294 with the largest corresponding eigenvalue is then used to define bins, in which showers
 1295 demonstrate similar patterns of energy deposition across the calorimeter layers. To further
 1296 de-correlate the inputs, the PCA chain is repeated on the showers within each such bin. This
 1297 full process is reversed for the particle simulation. A full description of the method can be
 1298 found in [62].

1299 Modeling of the lateral shower shape makes use of 2D histograms filled with GEANT4
 1300 hit energies in each layer and PCA bin. Binned in polar $\alpha - R$ coordinates in a local plane
 1301 tangential to the surface of the calorimeter system, these histograms represent the spatial
 1302 distribution of energy deposits for a given particle shower. Such histograms are constructed
 1303 for a number of Geant4 events, and the histograms for each event are normalized to total
 1304 energy deposited in the given layer. The average of these histograms is then taken (what is

1305 called here the “average shape”).

1306 In simulation, these average shape histograms are used as probability distributions, from
 1307 which a finite number of equal energy hits are drawn. This finite drawing of hits induces
 1308 a statistical fluctuation about the average shape which is tuned to match the expected
 1309 calorimeter sampling uncertainty.

1310 As an example, the intrinsic resolution of the ATLAS Liquid Argon calorimeter has a
 1311 sampling term of $\sigma_{\text{samp}} \approx 10\%/\sqrt{E}$ [64]. The number of hits to be drawn for each layer, $N_{\text{hits}}^{\text{layer}}$,
 1312 is thus taken from a Poisson distribution with mean $1/\sigma_{\text{samp}}^2$, where the energy assigned to
 1313 each hit is then just $E_{\text{hit}} = \frac{E_{\text{layer}}}{N_{\text{hits}}^{\text{layer}}}$. This induces a fluctuation of the order of $10\%/\sqrt{E_{\text{bin}}}$ for
 1314 each bin in the average shape.

1315 Figure 4.3 shows a comparison of energy and weta2 [65], defined as the energy weighted
 1316 lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons
 1317 simulated with the new FastCaloSim and with full GEANT4 simulation. The agreement is
 1318 quite good, with FastCaloSim matching the Geant4 mean to within 0.3 and 0.03 percent
 respectively. Similar results are seen for other photon energies and η points.

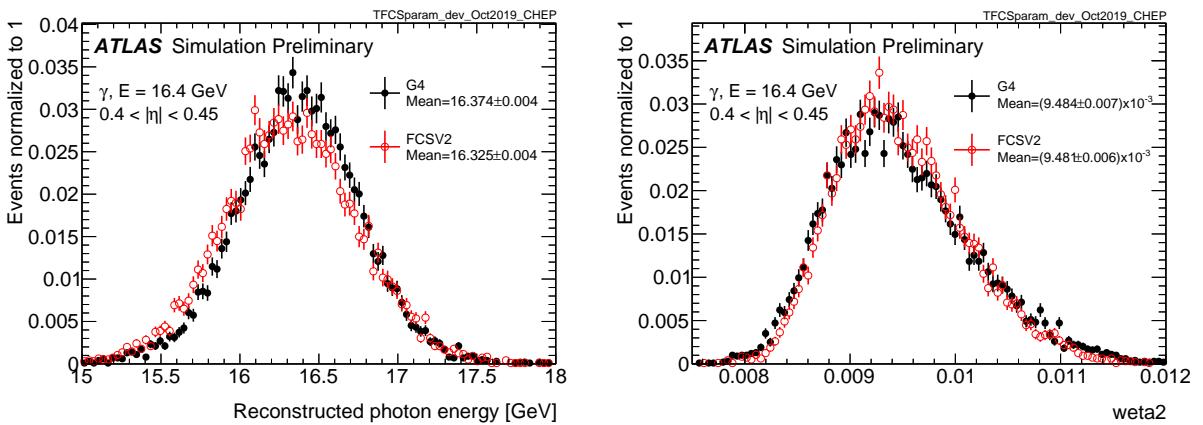


Figure 4.3: Energy and variable weta2, defined as the energy weighted lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [62].

1320 *4.3.1 Fluctuation Modeling*

1321 Figure 4.4 shows the ratio of calorimeter cell energies for single GEANT4 photon and pion
 1322 events to the corresponding cell energies in their respective average shapes. While the photon
 1323 event is quite close to the corresponding average, the pion event shows a deviation from the
 1324 average which is much larger and has a non-trivial structure, reflecting the different natures
 1325 of electromagnetic and hadronic showering.

1326 While the shape parametrization described above is thus sufficient for describing electro-
 1327 magnetic showers, we will demonstrate below that it is not sufficient for describing hadronic
 1328 showers (Figures 4.7 and 4.8). We therefore present and validate methods to improve this
 1329 hadronic shower modeling. Such methods have been presented as well in [66].

1330 Two methods for modeling deviations from the average shape have been studied: (1)
 1331 a neural network based approach using a Variational Autoencoder (VAE) [67] and (2) a
 1332 map through cumulative distributions to an n -dimensional Gaussian. With both methods,
 1333 the shape simulation then proceeds as described in Section 4.3, with the drawing of hits
 1334 according to the average shape. However, these hits no longer have equal energy, but have
 1335 weights applied to increase or decrease their energy depending on their spatial position.
 1336 This application of weights is designed to mimic a realistic shower structure and to encode
 1337 correlations between energy deposits.

1338 Both methods are trained on ratios of energy in binned units called voxels. This voxelization
 1339 is performed in the same polar $\alpha - R$ coordinates as the average shape, with a 5 mm core in
 1340 R and 20 mm binning thereafter. There are a total of 8 α bins from 0 to 2π and 8 additional
 1341 R bins from 5 mm to 165 mm. The 5 mm core is filled with the average value of core voxels
 1342 across the 8 α bins when creating the parametrisation. However, during simulation, each of
 1343 these 8 core bins is treated independently. The outputs of both methods mimic these energy
 1344 ratios and are used in the shape simulation as the weights described above. In contrast to
 1345 an approach based on, e.g., calorimeter cells, using voxels allows for flexibility in tuning the
 1346 binning used in creating the parametrisation. Further, due to their relatively large size, using

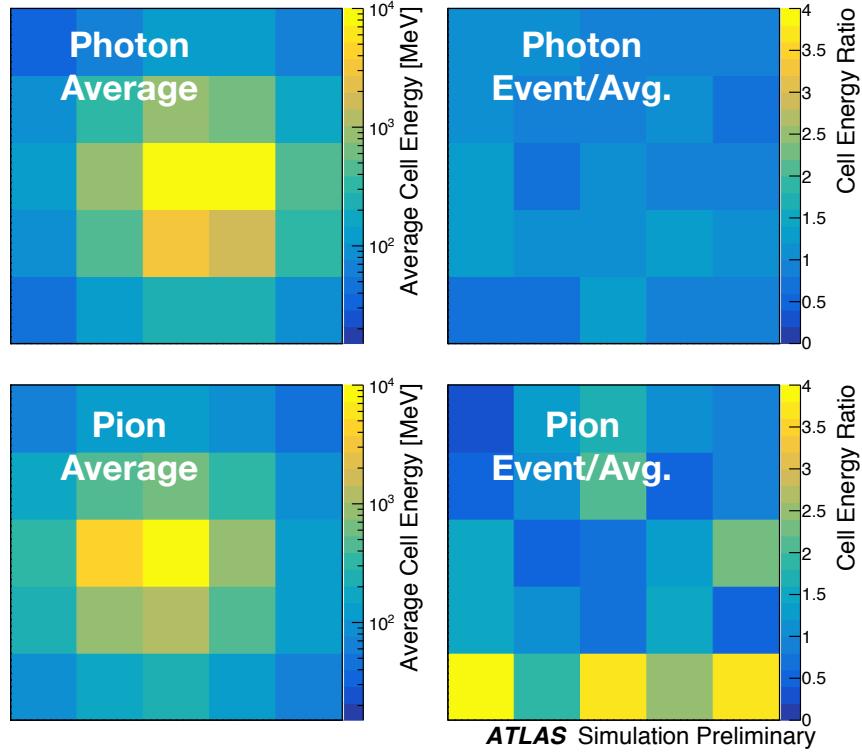


Figure 4.4: Example of photon and pion average shapes in 5×5 calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all close to 1, while the pion ratios show significant deviation from the average.

1347 calorimeter cells is subject to “edge effects”, where the splitting of energy between cells has a
 1348 non-trivial effect on the observed energy ratio. The binning used here is of the order of half
 1349 of a cell size, mitigating this effect.

1350 The Gaussian method operates by using cumulative distributions to map GEANT4 energy
 1351 ratios to a multidimensional Gaussian distribution. New events are generated by randomly
 1352 sampling from this Gaussian distribution.

1353 For the VAE method, a system of two linked neural networks is trained to generate events.

1354 The first “encoder” neural network maps input GEANT4 energy ratios to a lower dimensional
 1355 latent space. A second “decoder” neural network then samples from that latent space and
 1356 tries to reproduce the inputs. In simulation, events are generated by taking random samples
 1357 from the latent space and passing them through the trained decoder.

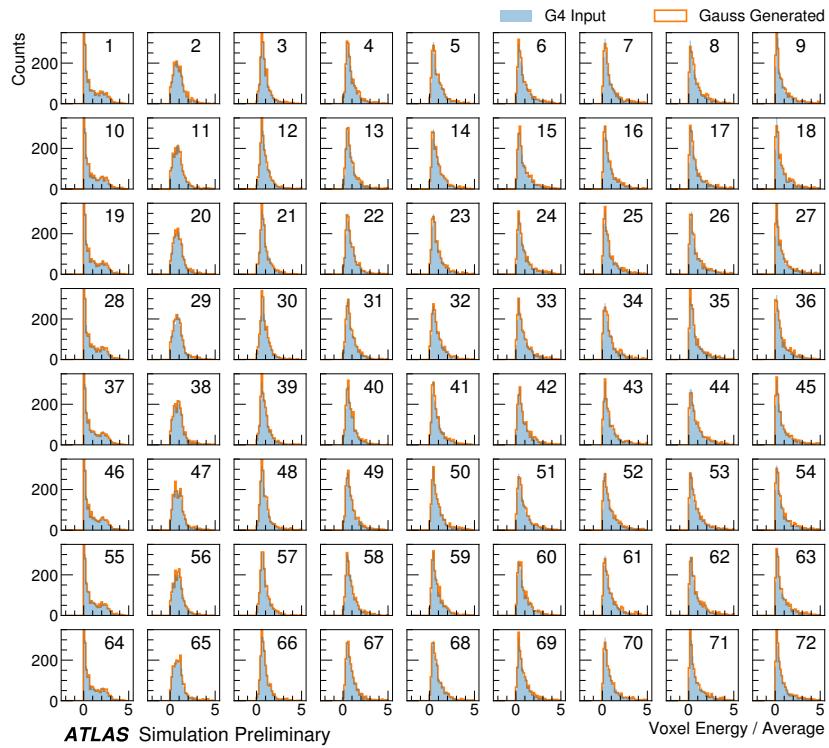


Figure 4.5: Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing α , left to right corresponds to increasing R , with core voxels numbered 1, 10, 19, Agreement is quite good across all voxels. Results are similar for the VAE method.

1358 Figure 4.5 shows the distributions of input GEANT4 and Gaussian method generated
 1359 energy ratios in the grid of voxels. Figure 4.6 shows the correlation coefficient between the
 1360 center voxel from $\alpha = 0$ to $2\pi/8$ for input GEANT4 and the Gaussian and VAE fluctuation

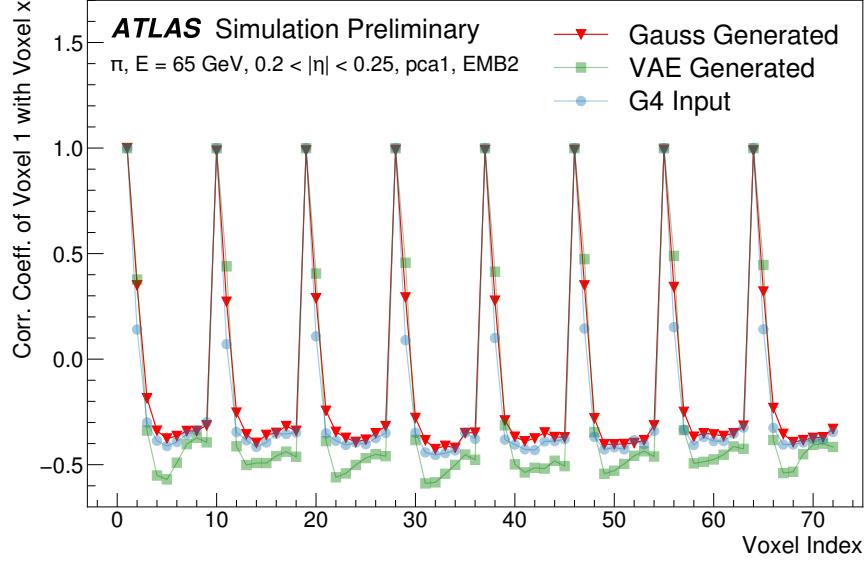


Figure 4.6: Correlation coefficient of ratios of voxel energy in single events to the corresponding voxel energy in the average shape, examined between the core bin from $\alpha = 0$ to $2\pi/8$ and each of the other voxels. The periodic structure represents the binning in α , and the increasing numbers in each of these periods correspond to increasing R , where the eight points with correlation coefficient 1 are the eight core bins. Both the Gaussian and VAE generated toy events are able to reproduce the major correlation structures for 65 GeV central pions in EMB2.

1361 methods. Agreement is good throughout.

1362 Validation of the Gaussian and VAE fluctuation methods was performed within FastCaloSimV2.

1363 Figure 4.7 shows the energy ratio of cells for a given simulation to the corresponding cells in
 1364 the average shape as a function of the distance from the shower center. The mean for all
 1365 simulation methods is expected to be around 1, with deviation from the average (the RMS
 1366 fluctuation) shown by the error bars. The Gaussian method RMS (red) and VAE method
 1367 RMS (green) both match the GEANT4 RMS (yellow) better than the case without correlated
 1368 fluctuations (blue) for a variety of energies, η points, and layers, often reproducing 80 – 100 %

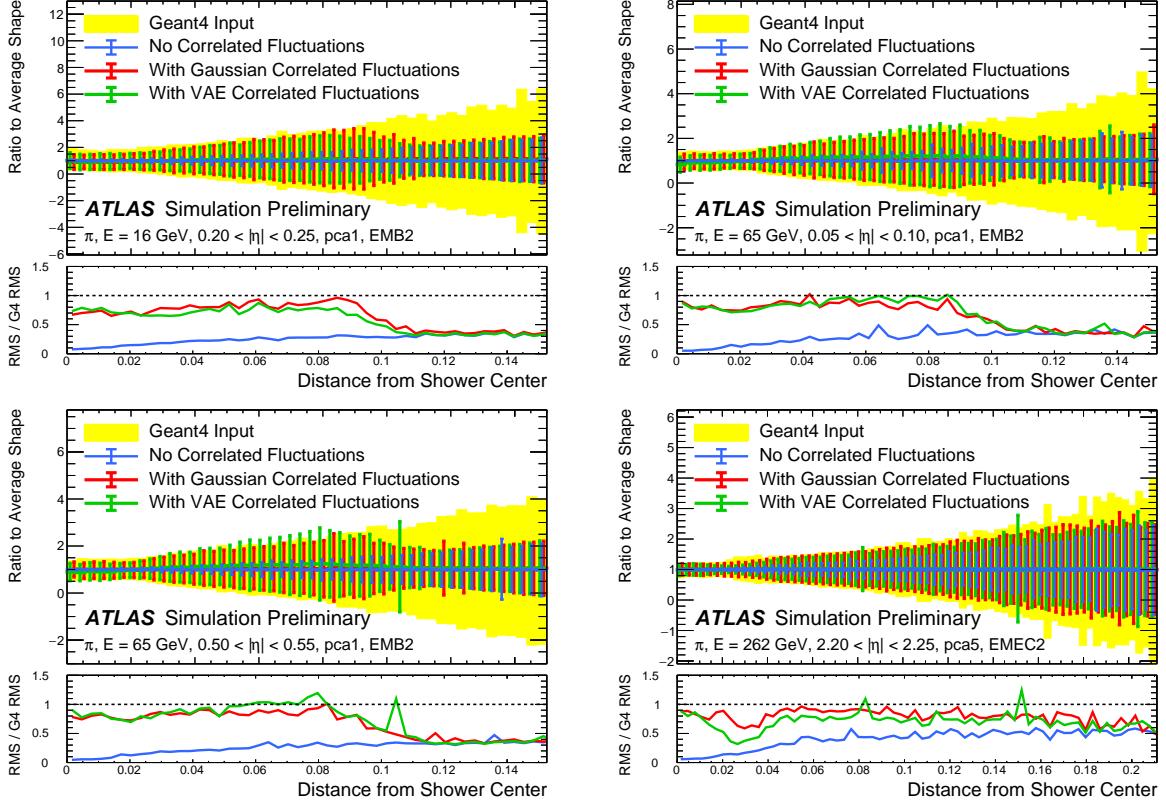


Figure 4.7: Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies, η points, and layers.

1369 of the GEANT4 RMS magnitude, compared to the 5 – 30 % observed in the no correlated
1370 fluctuations case.

1371 Figure 4.8 shows the result of a simulation with full ATLAS reconstruction for 65 GeV
1372 central pions with the Gaussian fluctuation model. Here a *cluster* [68] is defined as a three-
1373 dimensional spatial grouping of calorimeter cells which are summed based on the input signals
1374 relative to their neighboring cells. The multiplicity, shape, and spatial distribution of such
1375 clusters provides a powerful insight on the structure of energy deposits in the calorimeter,
1376 and good performance in cluster variables is a promising step towards good performance

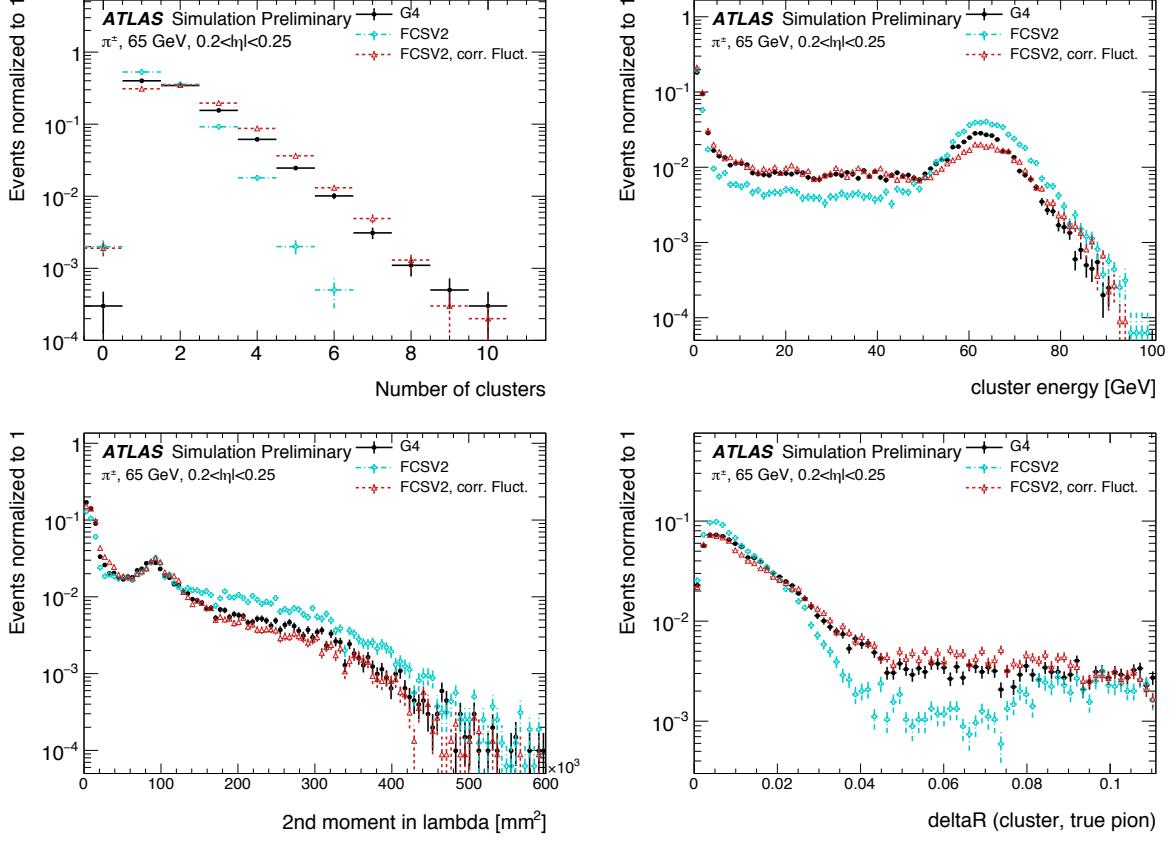


Figure 4.8: Comparison of the Gaussian fluctuation model to the default FCSV2 version and to G4 simulation, using pions of 65 GeV energy and $0.2 < |\eta| < 0.25$. Variables shown relate to calorimeter clusters, three-dimensional spatial groupings of cells [68] which provide powerful insight on the structure of energy deposits in the calorimeter. Variables considered include number and energy of clusters, the 2nd moment in lambda, ($\langle \lambda^2 \rangle$), which describes the square of the longitudinal extension of a cluster, where λ is the distance of a cell from the shower center along the shower axis, and a cluster moment is defined as $\langle x^n \rangle = \frac{\sum E_i x_i}{\sum E_i}$, and the distance ΔR , between the cluster and the true pion. With the correlated fluctuations, variables demonstrate improved modeling relative to default FastCaloSimV2.

1377 in the modeling of jet substructure, as these clusters may themselves be summed to form
 1378 jets (see Chapter 5). The simulation with the Gaussian fluctuation model demonstrates
 1379 improved modeling of several of these cluster variables relative to baseline FastCaloSimV2,
 1380 reproducing the distributions of events simulated with GEANT4. These include number and
 1381 energy of clusters, the 2nd moment in lambda, ($\langle \lambda^2 \rangle$), which describes the square of the
 1382 longitudinal extension of a cluster, where λ is the distance of a cell from the shower center
 1383 along the shower axis, and a cluster moment is defined as $\langle x^n \rangle = \frac{\sum E_i x_i}{\sum E_i}$, and the distance
 1384 ΔR , between the cluster and the true pion.

1385 The new fast calorimeter simulation is a crucial part of the future of simulation for the
 1386 ATLAS Experiment at the LHC. The per event simulation time of the full detector with
 1387 GEANT4, calculated over 100 $t\bar{t}$ events, is 228.9 s. Using FastCaloSim for the calorimeter
 1388 simulation reduces this to 26.6 s, of which FastCaloSim itself is only 0.015 s, with the majority
 1389 of the remaining simulation time due to GEANT4. Good physics modeling is achieved, and
 1390 the correlated fluctuations method shows good proof of concept improvement for the modeling
 1391 of hadronic showers.

1392 **4.4 Outlook**

1393 There has been significant effort in the community to develop a set of fast simulation tools,
 1394 with the use of machine learning methods at the forefront of such approaches (e.g. [69], [70]).
 1395 Most fast simulation approaches generally are based on parametrizations of fully simulated
 1396 events, but fall into two paradigms - a “by hand” simulation, which focuses on the modeling
 1397 of individual detector effects, or a fully parametrized simulation, in which a generative model
 1398 (e.g. a Generative Adversarial Network or Variational Autoencoder) is trained to directly
 1399 reproduce the input events. Both approaches can be extremely powerful, but each suffer from
 1400 certain drawbacks. The “by hand” approach offers the advantage of direct encoding of expert
 1401 knowledge – if an effect needs to be modeled, a new parametrization is introduced. However,
 1402 by the same token, it requires dedicated parametrizations for each effect. Fully parametrizing
 1403 the simulation with a generative model offloads this burden onto the network itself. However,

¹⁴⁰⁴ by doing so, the ability to use expert knowledge is diminished – the network is required to
¹⁴⁰⁵ learn all relevant effects.

¹⁴⁰⁶ The method presented here represents an effort to step towards a hybrid between these two
¹⁴⁰⁷ approaches, leveraging the power of machine learning techniques for individual parametriza-
¹⁴⁰⁸ tions within the by hand framework. Such hybrid solutions have the potential to be extremely
¹⁴⁰⁹ powerful, and further work on the development of these solutions is an interesting direction
¹⁴¹⁰ of future study.

1411

Chapter 5

1412

RECONSTRUCTION

1413 Chapter 3 discusses how a proton-proton collision may be captured by a physical detector
 1414 and turned into data that may be stored and analyzed. Chapter 4 discusses the simulation
 1415 of this same process. At this most basic level, however, the ATLAS detector is only a
 1416 machine for turning particles into a set of electrical signals, albeit in a very sophisticated,
 1417 physics motivated way. This chapter discusses the step of turning these electrical signals into
 1418 objects which may be identified with the underlying physics processes, and therefore used to
 1419 make statements about what occurred within a given collision event. This process is termed
 1420 *reconstruction*, and we will focus particularly on jets and flavor tagging, as the most relevant
 1421 pieces for this thesis work.

1422 **5.1 Jets**

1423 As discussed in Chapters 3 and 4, the production of particles with color charge from a
 1424 proton-proton interaction is complicated both by parton showering and by confinement: a
 1425 quark produced from a hard scatter is not seen as a quark, but rather, as a spray of particles
 1426 with a variety of hadrons in the final state, which subsequently shower upon interaction with
 1427 the calorimeter in a complicated way.

1428 For hard scatter electrons, photons, or muons on the other hand, the picture is much
 1429 clearer: there is no parton showering, and each has a distinct signature in the detector:
 1430 photons have no tracks and a very localized calorimeter shower, electrons are associated
 1431 with tracks and are similarly localized in the calorimeter, and muons are associated with
 1432 tracks, pass through the calorimeter due to their large mass, and leave signatures in the muon
 1433 spectrometer.

Jets are a tool to deal with the messiness of quarks and gluons. The basic concept is to group the multitude of particles produced by hadronization into a single object. Such an object then has associated properties, including a four-vector, which may be identified with the corresponding initial state particle. In practice a variety of information from the ATLAS detector is used for such a reconstruction. The analysis considered in this thesis uses particle flow jets [71], which combines information from both the tracker and the calorimeter, where the combined objects may be identified with underlying particles. However, jets built from clusters of calorimeter cells [72] as well as from charged particle tracks [73] have also been used very effectively.

A variety of algorithms are used to associate detector level objects to a given jet. The most commonly used in ATLAS is the anti- k_T algorithm [74], which is a successor to the k_T algorithm, among others [75], and develops as follows. Both algorithms are sequential recombination algorithms, which begin with the smallest distance, d_{ij} between considered objects (e.g. particles or intermediate groupings of particles). If d_{ij} is less than a parameter d_{iB} (B for “beam”) object i is combined with object j , the distance d_{ij} is recomputed, and the process repeats. This proceeds until $d_{ij} \geq d_{iB}$, at which point the jet is “complete” and removed from the list of considered objects.

The definitional difference between k_T and anti- k_T is these distance parameters. In general form, these are defined as

$$d_{ij} = \min(p_{Ti}^{2p}, p_{Tj}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \quad (5.1)$$

$$d_{iB} = p_{Ti}^{2p} \quad (5.2)$$

where p_{Ti} is the transverse momentum of object i , ΔR_{ij} is the angular distance between objects i and j , R is a radius parameter, and p controls the tradeoff between the p_T and angular distance terms. For the k_T algorithm $p = 1$; for the anti- k_T algorithm, $p = -1$. This is a simple change, but results in significantly different behavior.

The anti- k_T algorithm can be understood as follows: for a single high p_T particle (p_{T1}) surrounded by a bunch of low p_T particles, the low p_T particles will be clustered with the

high p_T one if

$$d_{1j} = \frac{1}{p_{T1}^2} \frac{\Delta R_{1j}^2}{R^2} < \frac{1}{p_{T1}^2} \quad (5.3)$$

$$\implies \Delta R_{1j} < R. \quad (5.4)$$

1455 Therefore, a single high p_T particle (p_{T1}) surrounded by a bunch of low p_T particles results in
 1456 a perfectly conical jet. This shape may change with the presence of other high momentum
 1457 particles, but the key feature of the dynamics is that the jet shape is determined by high p_T
 1458 objects due to the $\frac{1}{p_T}$ nature of this definition. In contrast, the k_T algorithm results in jets
 1459 influenced by low momentum particles, which results in a less regular shape. This property,
 1460 of regular jet shapes determined by high momentum objects, as well as demonstrated good
 1461 practical performance, makes the anti- k_T algorithm the favored jet algorithm in ATLAS.

1462 Because jets are composed of multiple objects, a useful property of jets is jet *substructure*,
 1463 that is, acknowledging that jets are composite objects, analyzing the structure of a given
 1464 jet to infer physics information. This leads to the use of *subjets*; that is, after running jet
 1465 clustering, often to create a “large-R”, $R = 1.0$ anti- k_T jet, a smaller radius jet clustering
 1466 algorithm is run within the jet. Subjets are often chosen using the k_T algorithm, which again
 1467 is sensitive to lower momentum particles, with $R = 0.2$ or 0.3 . For the boosted version of this
 1468 thesis analysis, such a strategy is used, in which the subjets are *variable radius* and depend
 1469 on the momentum of the (proto)jet in question. Beyond this thesis work, substructure is
 1470 used in a large variety of analyses, with a set of associated variables and tools developed for
 1471 exploiting this structure *TODO: Cite some?*.

1472 5.2 Flavor Tagging

1473 For this this thesis, the physics process being considered is $HH \rightarrow b\bar{b}b\bar{b}$. From the previous
 1474 section, we know that the standard practice is to identify these b quarks (or, rather, the
 1475 resulting B hadrons, due to confinement) with jets – in our case, these b -jets are $R=0.4$
 1476 anti- k_T particle flow jets (see Chapter 7). However, not all jets produced at the LHC are
 1477 from B hadrons; rather, there are a variety of different types of jets corresponding to different

1478 flavors of quarks. These are often classified as light jets (from u , d , or s quarks, or gluons)
1479 or as other *heavy flavor* jets, e.g. c -jets, involving c quarks. Distinguishing between these
1480 different categories is a very active area of work in ATLAS, termed *flavor tagging*, with much
1481 focus on *b-tagging* in particular, that is, the identification of jets from B hadron decays. We
1482 here briefly describe the techniques used for flavor tagging in ATLAS.

1483 What distinguishes a b -jet from any other jet? This question is fundamental to the
1484 design of the various b -tagging algorithms, and has two major answers: (1) B hadrons have
1485 long lifetimes, and (2) B hadrons have large masses. It is most illustrative to compare
1486 the B hadron properties to a common light meson, e.g. π^0 , the neutral pion, with quark
1487 content $\frac{1}{\sqrt{2}}(u\bar{u} - d\bar{d})$. B hadrons have lifetimes around 1.5 ps, corresponding to a decay
1488 length $c\tau \approx 0.45$ mm. In contrast, π^0 has a lifetime of 8.4×10^{-5} ps, which is around 20,000
1489 times shorter! Theoretically, this comes from CKM suppression of the b to c transition, which
1490 dominates the B decay modes. Experimentally, this difference pops up as shown in Figure
1491 5.1 – light flavor initiated jets decay almost immediately at the proton-proton interaction
1492 point, whereas b -jets are distinguished by a displaced secondary vertex, corresponding to
1493 the 5 mm decay length calculated above. This displaced vertex falls short of the detector
1494 itself, but may be inferred from larger transverse (perpendicular to beam) and longitudinal
1495 (parallel to beam) impact parameters of the resulting tracks, termed d_0 and z_0 respectively.

1496 Coming to the mass, B mesons have masses of around 5.2 GeV, whereas the π^0 mass
1497 is around 0.134 GeV, (around 40 times lighter). This higher mass gives access to a larger
1498 decay phase space, leading to a high multiplicity for b -jets (average of 5 charged particles per
1499 decay).

1500 One final distinguishing feature of B hadrons is their *fragmentation function*, a function
1501 describing the production of an observed final state. For B hadrons, this is particularly
1502 “hard”, with the B hadrons themselves contributing to an average of around 75 % of the b -jet
1503 energy. Thus, the identification of b -jets with B hadrons is, in some sense, descriptive.

1504 We have contrasted b -jets and light jets, demonstrating that there are several handles
1505 available for making this distinction. c -jets are slightly more similar to b -jets, but the same

1506 handles still apply – c -hadron lifetimes are between 0.5 and 1 ps, a factor of 2 smaller than B
1507 hadrons. Their mass is around 1.9 GeV, 2 to 3 times smaller than B hadrons, and c -hadrons
1508 contribute to an average of around 55 % of c -jet energy. Therefore, while the gap is slightly
1509 smaller, a distinction may still be made.

1510 The ATLAS flavor tagging framework [77] relies on developing a suite of “low-level”
1511 taggers, which use a variety of information about tracks and vertices as inputs. The output
1512 of these lower level taggers are then fed into a higher level tagger, which aggregates these
1513 results into a high level discriminant. Each of these taggers is described below.

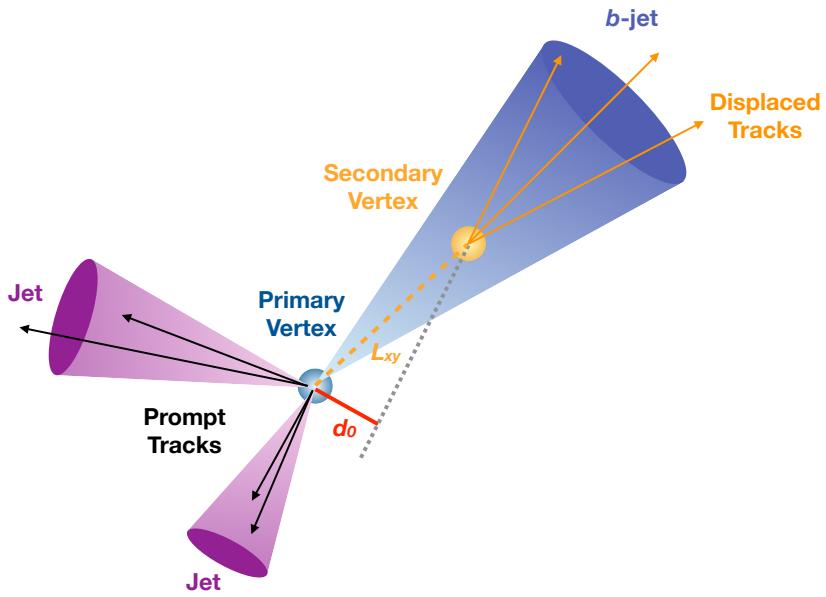


Figure 5.1: Illustration of an interaction producing two light jets and one b -jet in the transverse plane. While the light jets decay “promptly”, coinciding with the primary vertex of the proton-proton interaction, the longer lifetime of B hadrons leads to a secondary decay vertex, displaced from the primary vertex by length L_{xy} . This is typically a few mm, and therefore is not directly visible in the detector, but leads to a large transverse impact parameter, d_0 , for the resulting tracks. [76]

1514 5.2.1 IP2D/3D

1515 IP2D and IP3D are taggers based on the large track impact parameter (IP) nature of B
 1516 hadron decays. Both are based on histogram templates derived from Monte Carlo simulation,
 1517 which are used as probability density functions to construct log-likelihood discriminants.
 1518 IP2D incorporates just the transverse impact parameter information using 1D histogram
 1519 templates, whereas IP3D uses both transverse and longitudinal impact parameters in a 2D
 1520 template, which accounts for correlations. Importantly, these are *signed* impact parameters,
 1521 with sign based on the angle between the impact parameter and the considered jet – positive
 1522 impact parameters are consistent with a track extrapolation in front of the jet (angle between
 1523 impact parameter line and jet $< 90^\circ$), and therefore more consistent with tracks originating
 1524 from a displaced decay.

1525 Rather than using the impact parameters directly, an impact parameter *significance*
 1526 is used which incorporates an uncertainty on the impact parameter – tracks with a lower
 1527 uncertainty but the same impact parameter will contribute more in the calculation. This
 1528 signed significance is what is used to sample from the PDF templates, with the resulting
 1529 discriminants the sum of probability ratios between given jet hypotheses over tracks associated
 1530 to a given jet, namely $\sum_{i=1}^N \log \frac{p_b}{p_{light}}$ between b -jet and light jet hypotheses, where p_b and
 1531 p_{light} are the per-track probabilities. Similar discriminants are defined between b - and c -jets
 1532 and c and light jets. *TODO: show distributions?*

1533 5.2.2 SV1

1534 SV1 is an algorithm which aims to find a secondary vertex (SV) in a given jet. Operating
 1535 on all vertices associated with a considered jet, the algorithm discards tracks based on a
 1536 variety of cleaning requirements. It then proceeds to first construct all two-track vertices,
 1537 and then iterates over all the tracks involved in these two track vertices to try to fit a single
 1538 secondary vertex, which would then be identified with the secondary vertex from the b or c
 1539 hadron decay. This fit proceeds by evaluating a χ^2 for the association of a track and vertex,

removing the track with the largest χ^2 , and iterating until the χ^2 is acceptable and the vertex has an invariant mass of less than 6 GeV (for consistency with b or c hadron decay).

A variety of discriminating variables may then be constructed, including (1) invariant mass of the secondary vertex, (2) number of tracks associated with the secondary vertex, (3) number of two-track vertices, (4) energy fraction of the tracks associated to the secondary vertex (relative to all of the tracks associated to the jet), and various metrics associated with the secondary vertex position and decay length, including (5) transverse distance between the primary and secondary vertex, (6) distance between the primary and secondary vertex (7) distance between the primary and secondary vertex divided by its uncertainty, and (8) ΔR between the jet axis and the direction of the secondary vertex relative to the primary vertex.

While all eight of these variables are used as inputs to the higher level taggers, the number of two-track vertices, the vertex mass, and the vertex energy fraction are additionally used with 3D histogram templates to evaluate flavor tagging performance by constructing log-likelihood discriminants, similar to the procedure for IP2D/3D.

5.2.3 JetFitter

Rather than focusing on a particular aspect of the B hadron or D hadron decay topology (e.g impact parameter or secondary vertex), the third low level tagger, JETFITTER [78], tries to reconstruct the full decay chain, including all involved vertices. This is structured around a Kalman filter formalism [79], and has the strong underlying assumption that all tracks which stem from B and D hadron decay must intersect a common flight path. This assumption provides significant constraints, allowing for the reconstruction of vertices from even a single track, reducing the number of degrees of freedom in the fit, and allowing the use of “downstream” information, e.g., compatibility of tracks with a $B \rightarrow D$ -like decay. The constructed topology, including primary vertex location and B -hadron flight path, is iteratively updated over tracks associated to a given jet, and a variety of discriminating variables related to the resulting topology and reconstructed decay are used as inputs to the high level taggers.

1567 5.2.4 *RNNIP*

1568 The IP2D and IP3D algorithms rely on per-track probabilities, and the final discriminating
 1569 variables (and inputs to the higher level taggers) are sums (products) over these independently
 1570 considered quantities. In practice, however, the tracks are not independent – this is merely a
 1571 simplifying assumption to allow for the use of a binned likelihood, as treatment of all of the
 1572 interdependencies in such a framework quickly becomes intractable. To address this issue, a
 1573 recurrent neural network-based algorithm, RNNIP [80], is used, which takes as input a variety
 1574 of per-track variables, including the signed impact parameter significances (as in IP3D) as
 1575 well as track momentum fraction relative to the jet and ΔR between the track and the jet.
 1576 RNNs are sequence-based, and vectors of input variables corresponding to tracks for a given
 1577 jet are ordered by magnitude of transverse impact parameter significance and then passed
 1578 to the network, which outputs class probabilities corresponding to b-jet, c-jet, light-jet, and
 1579 τ -jet hypotheses. Such a procedure allows the network to learn interdependencies between
 1580 tracks, improving performance.

1581 5.2.5 *MV2 and DL1*

1582 Outputs from the above taggers are combined into high level taggers to aggregate all of the
 1583 information and improve discriminating power relative to the respective individual taggers as,
 1584 as shown in Figure 5.2. These high level taggers are primarily in two forms: MV2, which
 1585 uses a Boosted Decision Tree (BDT) for this aggregation, and DL1, which uses a deep neural
 1586 network. For the baseline versions of these taggers, only inputs from IP2D, IP3D, SV1, and
 1587 JetFitter are used. The tagger used for this thesis analysis, DL1r, additionally incorporates
 1588 RNNIP, demonstrating improved performance over the baseline DL1, as shown in Figure 5.3.
 1589 All high level taggers also include jet p_T and $|\eta|$.

DL1 offers a variety of improvements over MV2. Rather than a single discriminant output, as with MV2, DL1 has a multidimensional output, corresponding to probabilities for a jet to be a *b*-jet, *c*-jet, or light jet. This allows the trained network to be used for both *b*- and *c*-jet

tagging. The final discriminant for b -tagging with DL1 correspondingly takes the form

$$D_{\text{DL1}} = \ln \left(\frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \right) \quad (5.5)$$

where p_b , p_c , and p_{light} are the output b , c , and light jet probabilities, and f_c corresponds to an effective c -jet fraction, which may be tuned to optimize performance.

DL1 further includes an additional set of JETFITTER input variables relative to MV2 which correspond to c -tagging – notably properties of secondary and tertiary vertices, as would be seen in a $B \rightarrow D$ decay chain.

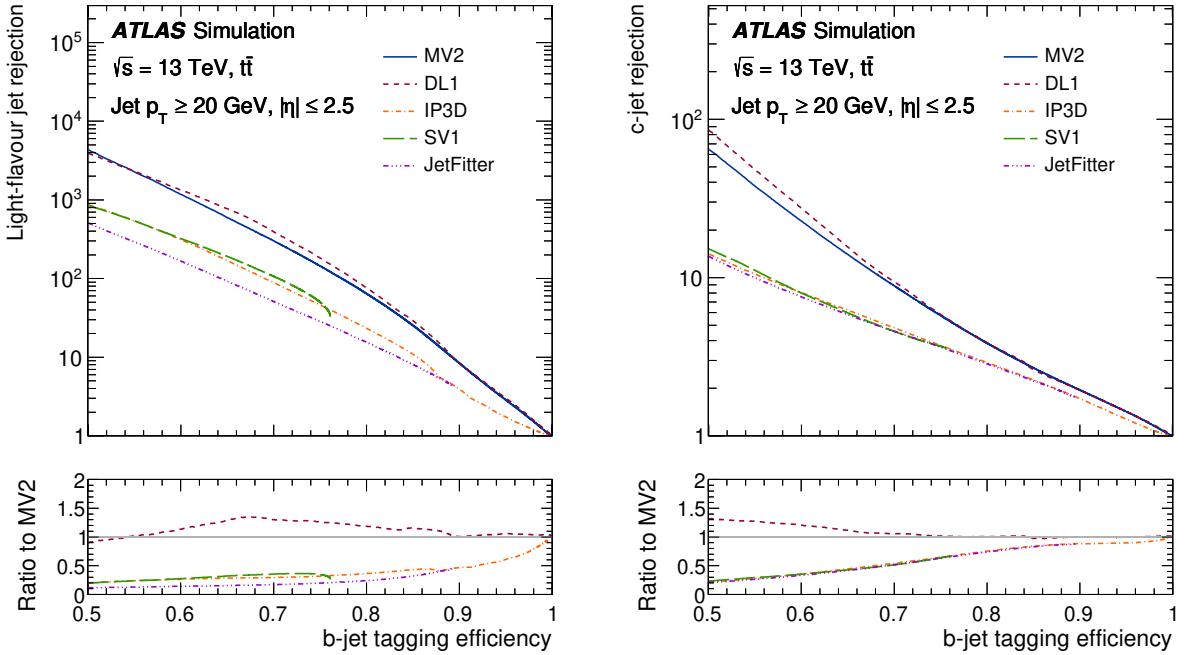


Figure 5.2: Performance of the various low and high level flavor tagging algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. The high level taggers demonstrate significantly better performance than any of the individual low level taggers, with DL1 offering slight improvements over MV2 due to the inclusion of additional input variables.

Figure 5.2 shows a comparison of the performance of the various taggers. The b -tagging performance of DL1 and MV2 is found to be similar, with some improvements in light jet and c -jet rejection from the additional variables used in DL1. The performance of these high level taggers additionally is seen to be significantly better than any of the individual low level ones, even in regimes where only a single low level tagger is relevant (such as high b -tagging efficiencies, where SV1 and JETFITTER are limited by selections on tracks entering the respective algorithms).

The inclusion of RNNIP offers a significant improvement on top of baseline DL1, as shown in Figure 5.3, strongly motivating the choice of DL1r for this thesis.

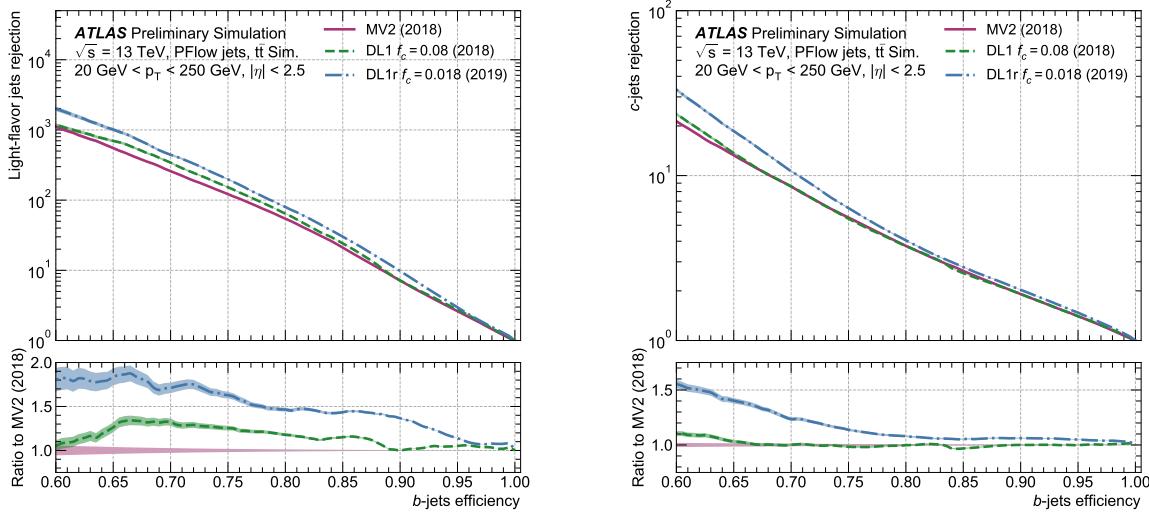


Figure 5.3: Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c controls the importance of c -jet rejection in the discriminating variable, and values shown have been optimized separately for each DL1 configuration. DL1r demonstrates a significant improvement in both light and c jet rejection over MV2 and DL1. [81]

1604 5.2.6 *Some Practical Notes*

1605 The b -tagging metrics presented in Figures 5.2 and 5.3 correspond to evaluating a tradeoff
1606 between b -jet efficiency and light jet and c -jet rejection. In this case, b -jet efficiency is defined
1607 such that, e.g. for a 77 % efficiency, 77 % of the real b -jets will be tagged as such. Somewhat
1608 counterintuitively, this means that a lower b -jet efficiency corresponds to a more aggressive
1609 (“tighter”) selection on the discriminating variable, while a higher b -jet efficiency corresponds
1610 to a less aggressive (“looser”) cut (100 % efficiency means no cut). Light and c jet efficiencies
1611 are defined similarly, with rejection defined as 1/ the corresponding efficiency.

1612 In ATLAS, the respective b -tagging efficiencies are used to define various b -tagging working
1613 points. The working point used for the nominal b -jet identification in this thesis is 77 % with
1614 DL1r. A loosened (less aggressive) selection at the 85 % working point is additionally used.
1615 See Chapter 7 for further details.

1616

Chapter 6

1617

THE ANATOMY OF AN LHC SEARCH

1618 In this thesis so far, we have set the theoretical foundation for the work carried out at the
 1619 LHC. We have described how one may translate between this theoretical foundation and what
 1620 we are actually able to observe with the ATLAS detector. We have further stepped through
 1621 the process of simulating production of specific physics processes and their appearance in
 1622 our detector, allowing us to describe how a hypothetical physics model would be seen in
 1623 our experiment. The question then becomes: all of these pieces are on the table, what do
 1624 we do with them? This chapter attempts to answer exactly that, setting up a roadmap for
 1625 assembling these pieces into a statement about the universe.

1626 ***6.1 Object Selection and Identification***

1627 As described in Chapter 5, there is a complicated set of steps for going from electrical signals
 1628 in a detector to physics objects.

1629 ***6.2 Defining a Signal Region***

1630 ***6.3 Background Estimation***

1631 ***6.4 Uncertainty Estimation***

1632 ***6.5 Hypothesis Testing***

1633

Chapter 7

1634

SEARCH FOR PAIR PRODUCTION OF HIGGS BOSONS IN THE $b\bar{b}b\bar{b}$ FINAL STATE

1635

This chapter presents two complementary searches for pair production of Higgs bosons in the final state. Such searches are separated based on the signal models being considered: resonant production, in which a new spin-0 or spin-2 particle is produced and decays to two Standard Model Higgs bosons, and non-resonant production, which is sensitive to the value of the Higgs self-coupling λ_{HHH} . Further information on the theory behind both channels can be found in Chapter 2.

While the searches face many similar challenges and proceed (in broad strokes) in a very similar manner, separate optimizations are performed to maximize the respective sensitivities for these two very different sets of signal hypotheses. More particularly, resonant signal hypotheses are (1) very peaked in values of the mass of the HH candidate system near the value of the resonance mass considered and (2) considered across a very broad range of signal mass hypotheses. The resonant searches are therefore split into resolved and boosted topologies based on Lorentz boost of the decay products, with the resolved channel as one of the primary focuses of this thesis. Further, several analysis design decisions are made to allow for sensitivity to a broad range of masses – in particular, though sensitivity is limited at lower values of m_{HH} relative to other channels *TODO: Combination, bbyy* due to the challenging background topology, retaining and properly reconstructing these low mass events allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity up until the kinematic threshold at 250 GeV.

In contrast, non-resonant signal hypotheses are quite broad in m_{HH} , and have a much more limited mass range, with Standard Model production peaking near 400 GeV, and the majority of the analysis sensitivity able to be captured with a resolved topology. Even for

1657 Beyond the Standard Model signal hypotheses, which may have more events at low m_{HH} ,
 1658 the non-resonant nature of the production allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity while
 1659 discarding much of the challenging low mass background. Such freedom allows for decisions
 1660 which focus on improved background modeling for the middle to upper HH mass regime,
 1661 resulting in improved modeling and smaller uncertainties than would be obtained with a
 1662 more generic approach.

1663 Both searches are presented in the following, with emphasis on particular motivations for,
 1664 and consequences of, the various design decisions involved for each respective set of signal
 1665 hypotheses.

1666 The analyses improve upon previous work [82] in several notable ways. The resonant
 1667 search leverages a Boosted Decision Tree (BDT) based pairing algorithm, offering improved
 1668 HH pairing efficiency over a broad mass spectrum. The non-resonant adopts a different
 1669 approach, with a simplified algorithm based on the minimum angular distance (ΔR) between
 1670 jets in a Higgs candidate. Such an approach very efficiently discards low mass background
 1671 events, resulting in an easier to estimate background with reduced systematic uncertainties.

1672 A particular contribution of this thesis is the background estimation, which uses a novel,
 1673 neural network based approach, offering improved modeling over previous methods, as well
 1674 as the ability to model correlations between observables. While all aspects of the analysis of
 1675 course contribute to the final result, the author of this thesis wishes to emphasize that the
 1676 background estimate, with the corresponding uncertainties and all other associated decisions,
 1677 really is the core of the $HH \rightarrow b\bar{b}b\bar{b}$ analysis – the development of this procedure, and all
 1678 associated decisions, is similarly the core of this thesis work.

1679 ATLAS has performed a variety of searches in complementary decay channels as well,
 1680 notably for early Run 2 in the $b\bar{b}W^+W^-$ [83], $b\bar{b}\tau^+\tau^-$ [84], $W^+W^-W^+W^-$ [85], $b\bar{b}\gamma\gamma$ [86],
 1681 and $W^+W^-\gamma\gamma$ [87] final states, which were combined along with $b\bar{b}b\bar{b}$ in [21]. ATLAS has
 1682 also released a variety of full Run 2 results, including boosted $b\bar{b}\tau^+\tau^-$ [88], VBF $b\bar{b}b\bar{b}$ [17],
 1683 $b\bar{b}\ell\nu\ell\nu$ [89], and $b\bar{b}\gamma\gamma$ [90].

1684 CMS has also performed searches for resonant production of Higgs boson pairs in the

1685 $b\bar{b}b\bar{b}$ final state (among others) at $\sqrt{s} = 8$ TeV [91] and $\sqrt{s} = 13$ TeV [92]. CMS have also
1686 performed a combination of their searches in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, and $b\bar{b}VV$ channels
1687 in [93].

1688 This analysis also benefits from improvements to ATLAS jet reconstruction and calibration,
1689 and flavour tagging [77]. In particular, this analysis benefits from the introduction of particle
1690 flow jets [71]. These make use of tracking information to supplement calorimeter energy
1691 deposits, improving the angular and transverse momentum resolution of jets by better
1692 measuring these quantities for charged particles in those jets.

1693 The analysis also benefits from the new DL1r ATLAS flavour tagging algorithm. Whereas
1694 the flavour tagging algorithm used in the previous analysis (MV2) used a boosted decision
1695 tree (BDT) to combine the output of various low level algorithms, DL1r (and the baseline
1696 DL1 algorithm) uses a deep neural network to do this combination. In addition to the low
1697 level algorithms used as inputs to MV2, DL1 includes a variety of additional variables used
1698 for c -tagging. DL1r further incorporates RNNIP, a recurrent neural network designed to
1699 identify b -jets using the impact parameters, kinematics, and quality information of the tracks
1700 in the jets, while also taking into account the correlations between the track features.

1701 The overall analysis sensitivity further benefits from a factor of ~ 4.6 increase in integrated
1702 luminosity.

1703 7.1 Data and Monte Carlo Simulation

1704 Both the resonant and non-resonant searches are performed on the full ATLAS Run 2 dataset,
1705 consisting of $\sqrt{s} = 13$ TeV proton-proton collision data taken from 2016 to 2018 inclusive.
1706 Data taken in 2015 is not used due to a lack of trigger jet matching information and b -jet
1707 trigger scale factors. The integrated luminosity collected and usable in this analysis¹ was:

- 1708 • 24.6 fb^{-1} in 2016

¹approximately 9 fb^{-1} of data was collected but could not be used in this analysis due to an inefficiency in the b -jet triggers at the start of 2016 [94]

₁₇₀₉ • 43.65 fb^{-1} in 2017

₁₇₁₀ • 57.7 fb^{-1} in 2018

₁₇₁₁ This gives a total integrated luminosity of 126 fb^{-1} . This is lower than the 139 fb^{-1} ATLAS
₁₇₁₂ collected during Run 2 [95] due to the inefficiency described in footnote 1 as well as the
₁₇₁₃ 3.2 fb^{-1} of 2015 data which is unused due to the trigger scale factor issue mentioned above.

₁₇₁₄ In this analysis, Monte Carlo samples are used purely for modelling signal processes. The
₁₇₁₅ background is strongly dominated by events produced by QCD multijet processes, which
₁₇₁₆ are difficult to correctly model in simulation. This necessitates the use of a data-driven
₁₇₁₇ background modelling technique, which is described in Section 7.6.

₁₇₁₈ The scalar resonance signal model is simulated at leading order in α_s using MADGRAPH
₁₇₁₉ [52]. Hadronization and parton showering are done using HERWIG 7 [53][54] with EVTGEN [56],
₁₇₂₀ and the nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a two Higgs
₁₇₂₁ doublet model where the new neutral scalar is produced through gluon fusion and required
₁₇₂₂ to decay to a pair of SM Higgs bosons. The heavy scalar is assigned a width much smaller
₁₇₂₃ than detector resolution, and the other 2HDM particles do not enter the calculation.

₁₇₂₄ Scalar samples are produced at resonance masses between 251 and 900 GeV and the
₁₇₂₅ detector simulation is done using AtlFast-II [61]. In addition the samples at 400 GeV and
₁₇₂₆ 900 GeV are also fully simulated to verify that the use of AtlFast-II is acceptable. For higher
₁₇₂₇ masses, as well as for the boosted analysis, samples are produced between 1000 and 5000 GeV,
₁₇₂₈ and the detector is fully simulated. As discussed in Chapter 4, an outstanding issue with
₁₇₂₉ AtlFast-II is the modeling of jet substructure. While such variables are not used for the
₁₇₃₀ resolved analysis, the boosted analysis begins at 900 GeV, motivating the different detector
₁₇₃₁ simulation in these two regimes.

₁₇₃₂ The spin-2 resonance signal model is also simulated at LO in α_s using MADGRAPH.
₁₇₃₃ Hadronization and parton showering are done using PYTHIA 8 [55] with EVTGEN, and the
₁₇₃₄ nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a Randall-Sundrum
₁₇₃₅ graviton with $c = 1.0$.

1736 Spin-2 resonance samples are produced at masses between 251 and 5000 GeV, and these
1737 samples are all produced with full detector simulation.

1738 For the non-resonant search, samples are produced at values of $\kappa_\lambda = 1.0$ and 10.0, and are
1739 simulated using POWHEG BOX v2 generator [49–51] at next-to-leading order (NLO), with full
1740 NLO corrections with finite top mass, using the PDF4LHC [96] parton distribution function
1741 (PDF) set. Parton showers and hadronization are simulated with PYTHIA 8.

1742 Alternative ggF samples are simulated at NLO using POWHEG BOX v2, but instead using
1743 HERWIG 7 [97] for parton showering and hadronization. The comparison between these two
1744 is used to assess an uncertainty on the parton showering.

1745 7.2 Triggers and Object Definitions

1746 To maximize analysis sensitivity, a combination of multi- b -jet triggers is used. Due to the use
1747 of events with two b -tagged jets in the background estimate, such triggers have a maximum
1748 requirement of two b -tagged jets. For the resonant analysis, a combination of triggers of
1749 various topologies is used, namely

- 1750 • 2b + HT, which requires two b -tagged jets and a minimum value of of H_T , defined to
1751 be the scalar sum of p_T across all jets in the event.
- 1752 • 2b + 2j, which requires two b -tagged jets and two other jets matching some kinematic
1753 requirements
- 1754 • 2b + 1j, which requires two b -tagged jets and one other jet matching some kinematic
1755 requirements
- 1756 • 1b, which requires one b -tagged jet

1757 Due to minimal contributions from some of these triggers for the Standard Model non-resonant
1758 signal, a simplified strategy relying entirely on 2b + 1j and 2b + 2j triggers is used for the
1759 non-resonant search.

1760 While the use of multiple triggers is beneficial for analysis sensitivity, it comes with some
 1761 complications. Namely, a set of scale factors must be assigned to simulated events account for
 1762 differences in trigger efficiency between real and simulated events. Because these scale factors
 1763 may differ between triggers, the use of multiple triggers becomes complicated: an event may
 1764 pass more than one trigger, while trigger scale factors are only provided for individual triggers.

1765 To simplify this calculation, a set of hierarchical offline selections is applied, closely
 1766 mimicking the trigger selection. Based on these selections, events are sorted into categories
 1767 (*trigger buckets*), after which the decision of a *single trigger* is checked.

1768 The resonant search applies such categorization in the following way, with selections
 1769 considered in order:

- 1770 1. If the leading jet is b -tagged with $p_T > 325 \text{ GeV}$, the event is in the $1b$ trigger category.
- 1771 2. Otherwise, if the leading jet is not b -tagged, but has $p_T > 168.75 \text{ GeV}$, the event is in
 1772 the $2b + 1j$ trigger category.
- 1773 3. If neither of the first two selections pass, if the scalar sum of jet p_T s, $H_T > 900 \text{ GeV}$,
 1774 the event falls into the $2b + HT$ trigger category.
- 1775 4. Events that do not pass any of the above offline selections are in the $2b + 2j$ trigger
 1776 category.

1777 Corresponding triggers are then checked in each category, and the final set of events consists
 1778 of those events that pass the trigger decision in their respective categories.

1779 For the resonant search, the $2b + 1j$ and $2b + 2j$ triggers are the dominant categories,
 1780 containing roughly 26 % and 49 % of spin-2 events, evaluated on MC16d samples with
 1781 resonance masses between 300 and 1200 GeV. Notably, the $1b$ trigger efficiency is largest at
 1782 high ($> 1 \text{ TeV}$) resonance masses.

1783 For the non-resonant search, it was noted that the $1b$ trigger has minimal contribution,
 1784 while the $2b + HT$ events are largely captured by the $2b + 2j$ trigger. Therefore, for, a

1785 simplified scheme is considered, with selections:

- 1786 1. If the 1st leading jet has $p_T > 170 \text{ GeV}$ and the 3rd leading jet has $p_T > 70 \text{ GeV}$, the
1787 event is in the $2b + 1j$ trigger category.
- 1788 2. Otherwise, the event is in the $2b + 2j$ trigger category.

1789 **7.3 Analysis Selection**

1790 After the trigger selections of Section 7.2, a variety of selections on the analysis objects are
1791 made, with the goal of (1) reconstructing a HH -like topology and (2) suppressing contributions
1792 from background processes.

1793 Both analyses begin with a common pre-selection, requiring at least four $R = 0.4$ anti- k_T
1794 jets with $|\eta| < 2.5$ and $p_T > 40 \text{ GeV}$. The $|\eta| < 2.5$ requirement is necessary for b -tagging
1795 due to the coverage of the ATLAS tracking detector (see Chapter 3), while the $p_T > 40 \text{ GeV}$
1796 requirement is motivated by the trigger thresholds. A low p_T category, which would include
1797 events failing the analysis selection due to this p_T cut, was considered for the non-resonant
1798 search, but was found to contribute minimal sensitivity. At least two of the jets passing this
1799 pre-selection are required to be b -tagged, and additional b -tagging requirements are made to
1800 define the following regions:

- 1801 • “2 b Region”: require exactly two b -tagged jets, used for background estimation
- 1802 • “4 b Region”: require at least (but possibly more) four b -tagged jets, used as a signal
1803 region for both resonant and non-resonant searches

1804 The non-resonant analysis additionally defines two 3 b regions:

- 1805 • “3 $b+1$ loose Region”: require exactly three b -tagged jets which pass the 77 % b-tagging
1806 working point (nominal) and one additional jet that fails the 77 % b-tagging working
1807 point but passes the *looser* 85 % b-tagging working point. Used as a signal region for
1808 the non-resonant search.

- 1809 • “3 b +1 fail Region”: complement of 3 b +1 loose. Require exactly three b -tagged jets
 1810 which pass the 77 % b-tagging working point, but require that none of the remaining jets
 1811 pass the 85 % b-tagging working point. Used as a validation region for the non-resonant
 1812 search.

1813 After these requirements, four jets are chosen, ranked first by b -tagging requirement and then
 1814 by p_T (e.g. for the 2 b region, the jets chosen are the two b -tagged jets and the two highest p_T
 1815 non-tagged jets; for the 4 b region, the jets are the four highest p_T b -tagged jets). To match
 1816 the topology of a $HH \rightarrow b\bar{b}b\bar{b}$ event, these four jets are then *paired* into *Higgs candidates*: the
 1817 four jets are split into two sets of two, and each of these pairs is used to define a reconstructed
 1818 object that is a proxy for a Higgs in a HH event.

1819 For four jets there are three possible pairings. For signal events, a correct pairing can be
 1820 identified (provided all necessary jets pass pre-selection) using the truth information of the
 1821 Monte Carlo simulation, and such information may be used to design/select an appropriate
 1822 pairing algorithm. This is only part of the story, however. The vast majority of the events in
 1823 data do *not* include a real HH decay (this is a search for a reason!), either because the event
 1824 originates from a background process (e.g. for 4 b events), or because the selection is not
 1825 designed to maximize the signal (e.g. 2 b events). As the pairing is part of the selection, it must
 1826 still be run on such events, such that various algorithms which achieve similar performance
 1827 in terms of pairing efficiency may have vastly different impacts in terms of the shape of the
 1828 background and the biases inherent in the background estimation procedure. The interplay
 1829 between these two facets of the pairing is an important part of the choices made for this
 1830 analysis.

1831 A comparison of different shapes due to three different paring strategies is shown in Figure
 1832 7.1.

1833 7.3.1 Resonant Pairing Strategy

1834 For the resonant analysis, a Boosted Decision Tree (BDT) is used for the pairing. The boosted
 1835 decision tree is given the total separation between the two jets in each of the two pairs (ΔR_1
 1836 and ΔR_2), the pseudo-rapidity separation between the two jets in each pair ($\Delta\eta_1$ and $\Delta\eta_2$),
 1837 and the angular separation between the two jets in each pair in the $x - y$ plane ($\Delta\phi_1$ and
 1838 $\Delta\phi_2$). The total separations (ΔR_s) are provided in addition to the components in order to
 1839 avoid requiring the boosted decision tree to reconstruct these variables in order to use them.
 1840 For these variables, pair 1 is the pair with the highest scalar sum of jet p_{T} s, and pair 2 the
 1841 other pair.

1842 The boosted decision tree is also parameterized on the di-Higgs mass (m_{HH}) by providing
 1843 this as an additional feature. Since the boosted decision tree is trained on correct and
 1844 incorrect pairings in signal events, there will be exactly one correct pairing and two incorrect
 1845 pairings in the training set for each m_{HH} value present in that set. As a result, this variable
 1846 cannot, in itself, distinguish a correct pairing from an incorrect pairing, and therefore the
 1847 inclusion of this variable simply serves to parameterize the BDT on m_{HH} ².

1848 The boosted decision tree was trained on one quarter of the total AFII simulated scalar
 1849 MC statistics, using the Gradient-based One Side Sampling (GOSS) algorithm which allows
 1850 rapid training with very large datasets. A preselection was applied requiring events to have
 1851 four jets with a p_{T} of at least 35 GeV. Note that this is a looser requirement than the 40 GeV
 1852 used in the analysis selection, and is meant to increase the available statistics for events with
 1853 low m_{HH} and to ensure a better performance as a function of that variable. Events were also
 1854 required to have four distinct jets that could be geometrically matched (to within $\Delta R \leq 0.4$)
 1855 to the b -quarks. The events used to train the BDT were not included when the analysis was
 1856 run on these signal simulations. The boosted decision tree was constructed with the following
 1857 hyperparameters:

1858 `min_data_in_leaf=50,`

²That is, the conditions placed on the other variables by the BDT vary with m_{HH} .

1859 num_leaves=180,
 1860 learning_rate=0.01

1861 These hyperparameters were optimized using a Bayesian optimization procedure [98].
 1862 Three fold cross-validation was used to perform this optimization without the need for an
 1863 additional sample, while avoiding over-training on signal events.

1864 *7.3.2 Non-resonant Pairing Strategy*

1865 For the non-resonant analysis, a simpler pairing algorithm is used, which proceeds as follows:
 1866 in a given event, Higgs candidates for each possible pairing are sorted by the p_T of the vector
 1867 sum of constituent jets. The angular separation, ΔR is computed between jets in the each of
 1868 the leading Higgs candidates, and the pairing with the smallest separation (ΔR_{jj}) is selected.
 1869 This method will be referred to as $\min \Delta R$ in the following.

1870 The primary motivation for the use of this pairing in the non-resonant search is a *smooth*
 1871 *mass plane*: by efficiently discarding low mass events, $\min \Delta R$ removes the background peak
 1872 present in the resonant search while maintaining good pairing efficiency for the Standard
 1873 Model non-resonant signal. This facilitates a background estimate with small kinematic bias
 1874 – the region in which the background estimate is derived is more similar to the signal region.

1875 Along with discarding low mass background, there is a corresponding loss of low mass
 1876 signal. This predominantly impacts points away from the Standard Model (see Figure 7.2),
 1877 but, because the $4b$ channel has the strongest contribution near the Standard Model and
 1878 because of the large low mass background present with other pairing methods, the impact on
 1879 analysis sensitivity is mitigated. The $\min \Delta R$ pairing is thus adopted for the non-resonant
 1880 search.

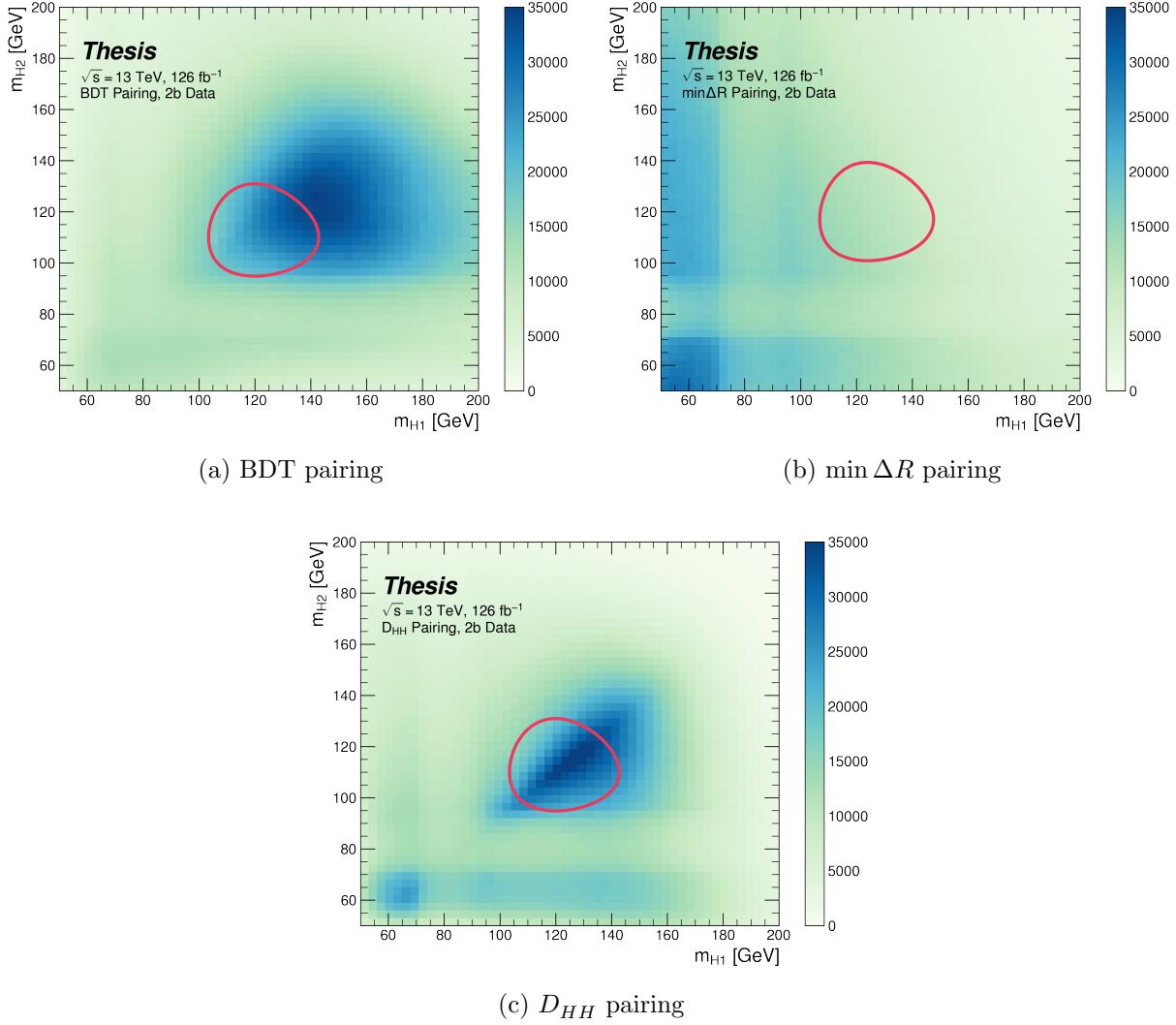


Figure 7.1: Comparison of m_{H1} vs m_{H2} planes for the full Run 2 2b dataset with different pairings. As evidenced, this choice significantly impacts where events fall in this plane, and therefore which events fall into the various kinematic regions defined in this plane (see Section 7.5). Respective signal regions are shown for reference, with the $\min \Delta R$ signal region shifted slightly up and to the right to match the non-resonant selection. Note that the band structure around 80 GeV in both m_{H1} and m_{H2} is introduced by the top veto described in Section 7.4.

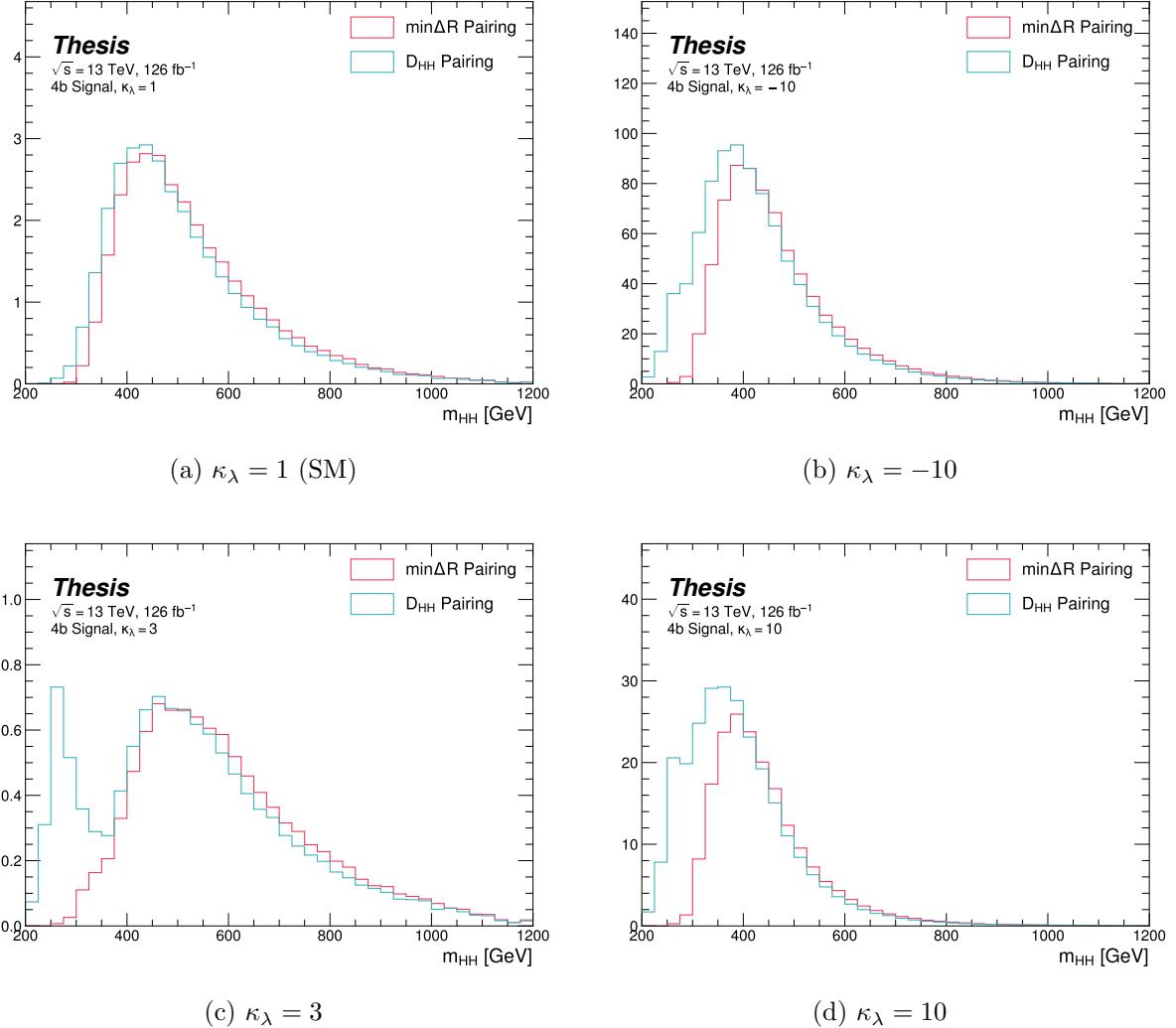


Figure 7.2: Comparison of signal distributions in the respective signal regions for the $\min \Delta R$ and D_{HH} pairing for various values of the Higgs trilinear coupling in the respective signal regions. The distributions are quite similar at the Standard Model point, but for other variations, $\min \Delta R$ does not pick up the low mass features.

1881 **7.4 Background Reduction and Top Veto**

1882 Choosing a pairing of the four b-tagged jets fully defines the di-Higgs candidate system used
1883 for each event in the remainder of the analysis chain. A requirement of $|\Delta\eta_{HH}| < 1.5$ on this
1884 di-Higgs candidate system mitigates QCD multijet background.

1885 In order to mitigate the hadronic $t\bar{t}$ background, a top veto is then applied, removing
1886 events consistent with a $t \rightarrow b(W \rightarrow q_1\bar{q}_2)$ decay.

1887 The jets in the event are separated into *HC jets* which are the four jets used to build the
1888 Higgs candidates, and *non-*HC jets**, the other jets (passing the p_T and $|\eta|$ requirements) in
1889 the event.

1890 W candidates are built by forming all possible pairs of all jets in each event. With n jets,
1891 there are $\binom{n}{2}$ such pairs. t candidates are then built by pairing each W candidate with each
1892 HC jet (for $4\binom{n}{2}$ combinations). Note that all jets in a t candidate must be distinct (i.e. a
1893 HC jet may not be used both on its own and in a W candidate).

With m_t denoting the invariant mass of the t candidate, and m_W the invariant mass of the W candidate, the quantity

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4 \text{ GeV}}{0.1 \cdot m_W}\right)^2 + \left(\frac{m_t - 172.5 \text{ GeV}}{0.1 \cdot m_t}\right)^2} \quad (7.1)$$

1894 is constructed for each combination.

1895 Events are then vetoed if the minimum X_{Wt} over all combinations is less than 1.5.

1896 The same definitions and procedures are used for both the resonant and non-resonant
1897 analyses. However, for the non-resonant search, the top candidates considered for X_{Wt} have
1898 the additional requirement that the jet used for the b is *b*-tagged. While this is identical to
1899 the resonant analysis by definition for 4*b* events, it does change the set of events considered in
1900 lower tag regions, in particular for the 2*b* events considered in the derivation of the background
1901 estimate. Such a change is found to reduce the impact of background systematics by increasing
1902 2*b* support in the high X_{Wt} kinematic region. *TODO: Insert plots of variables*

¹⁹⁰³ **7.5 Kinematic Region Definition**

As has been mentioned, an important piece of the analysis is the plane defined by the two Higgs candidate masses (the *Higgs candidate mass plane*). After the selection described above, a signal region is defined by requiring $X_{HH} < 1.6$, where:

$$X_{HH} = \sqrt{\left(\frac{m(H_1) - c_1}{0.1 \cdot m(H_1)}\right)^2 + \left(\frac{m(H_2) - c_2}{0.1 \cdot m(H_2)}\right)^2} \quad (7.2)$$

¹⁹⁰⁴ with $m(H_1)$, $m(H_2)$ the leading and subleading Higgs candidate masses, c_1 and c_2 correspond
¹⁹⁰⁵ to the center of the signal region, and the denominator provides a Higgs candidate mass
¹⁹⁰⁶ dependent resolution of 10 %. For consistency with the HH decay hypothesis, c_1 and c_2
¹⁹⁰⁷ are nominally (125 GeV, 125 GeV). However, these are allowed to vary due to energy loss,
¹⁹⁰⁸ with specific values chosen described below. The selection of these values is one of several
¹⁹⁰⁹ significant differences between the regions defined for the resonant and non-resonant search.
¹⁹¹⁰ We describe both below.

¹⁹¹¹ **7.5.1 Resonant Kinematic Regions**

¹⁹¹² For the resonant analysis, the signal region is centered at (120 GeV, 110 GeV) to account for
¹⁹¹³ energy loss leading to the Higgs masses being under-reconstructed. Note that leading and
¹⁹¹⁴ subleading Higgs candidates are defined according to the *scalar sum* of constituent jet p_T .

For the background estimation, two regions are defined which are roughly concentric around the signal region: a *validation region* which consists of those events not in the signal region, but which do pass

$$\sqrt{(m(H_1) - 1.03 \times 120 \text{ GeV})^2 + (m(H_2) - 1.03 \times 110 \text{ GeV})^2} < 30 \text{ GeV} \quad (7.3)$$

and a *control region* which consists of those events not in the signal or validation regions, but which do pass

$$\sqrt{(m(H_1) - 1.05 \times 120 \text{ GeV})^2 + (m(H_2) - 1.05 \times 110 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.4)$$

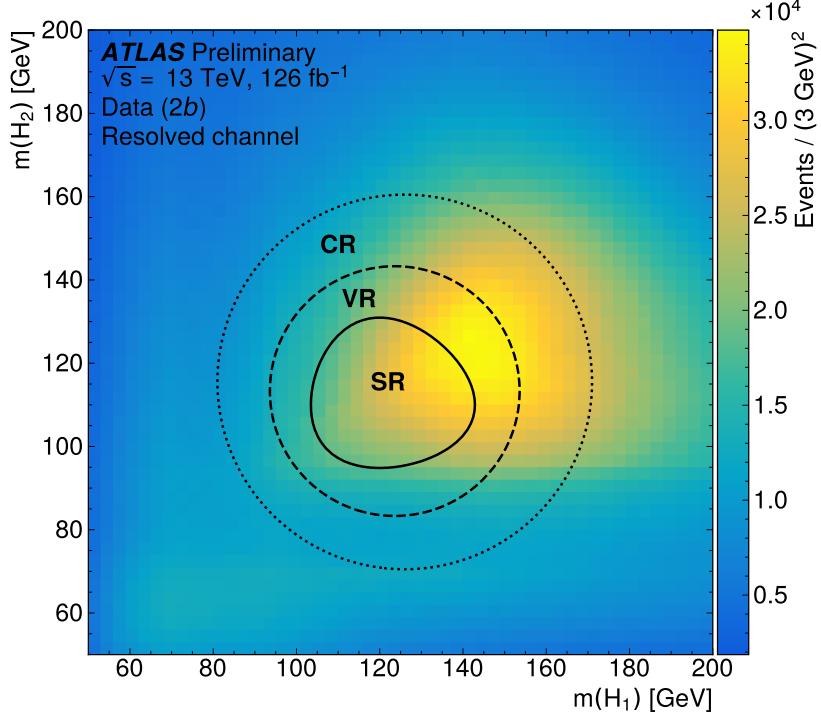


Figure 7.3: Regions used for the resonant search, shown on the $2b$ data mass plane. The outermost region (the “control region”) is used for derivation of the nominal background estimate. The innermost region is the signal region, where the signal extraction fit is performed. The region in between (the “validation region”) is used for the assessment of an uncertainty.

1915 For simplicity, the SR/VR/CR definitions from the early Run 2 paper [82] were chosen
 1916 for the resonant analysis, but were found to be close to optimal. These regions are shown in
 1917 Figure 7.3.

1918 7.5.2 Non-resonant Kinematic Regions

1919 For the non-resonant analysis the signal region is centered at $(124 \text{ GeV}, 117 \text{ GeV})$, corre-
 1920 sponding to the means of *correctly paired* Standard Model signal events. The shape of the
 1921 signal region (other than this change of center) was found to remain optimal.

1922 For the non-resonant search, leading and subleading Higgs candidates are defined according
 1923 to the *vector sum* of constituent jet p_T , more closely corresponding to the $1 \rightarrow 2$ decay
 1924 assumption behind the min ΔR pairing algorithm.

1925 Two areas for improvement were identified in the resonant analysis, which will be dis-
 1926 cussed in more detail below: *signal contamination* of the validation region (which impacts
 1927 the uncertainty assessed due to extrapolation) and *large nuisance parameter pulls* on this
 1928 uncertainty, corresponding to a rough assumption that the validation region is closer to the
 1929 signal region in the mass plane, and so offers a better estimate of the signal region.

To mitigate these two issues, a redesign of the control and validation regions was performed for the non-resonant analysis. The outer boundary defined by the shifted resonant control region:

$$\sqrt{(m(H_1) - 1.05 \times 124 \text{ GeV})^2 + (m(H_2) - 1.05 \times 117 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.5)$$

1930 is kept, roughly corresponding to combining the regions used for the resonant analysis. In
 1931 order to assess the variation of the background estimate, two sets of regions are desired, so
 1932 this combined region is split into *quadrants*, that is, divided into four pieces along axes that
 1933 intersect with the signal region center. To avoid kinematic bias, quadrants on opposite sides
 1934 of the signal region are paired, with these pairs corresponding to the non-resonant control
 1935 and validation regions.

1936 The particular orientation of the regions is chosen such that region centers align with the
 1937 leading and subleading Higgs candidate masses, corresponding to a set of axes rotated at
 1938 45° , with the “top” and “bottom” quadrants together comprising the control region, and the
 1939 other set (“left” and “right”) the validation region. These regions are shown in Figure 7.4

1940 This design of regions includes a set of events closer to the signal region in the mass plane,
 1941 leveraging the assumption that these events are more similar to signal region events, while
 1942 also including events further away from the signal region, mitigating signal contamination.
 1943 This region selection is found to have good performance in alternate validation regions (see
 1944 Section 7.8).

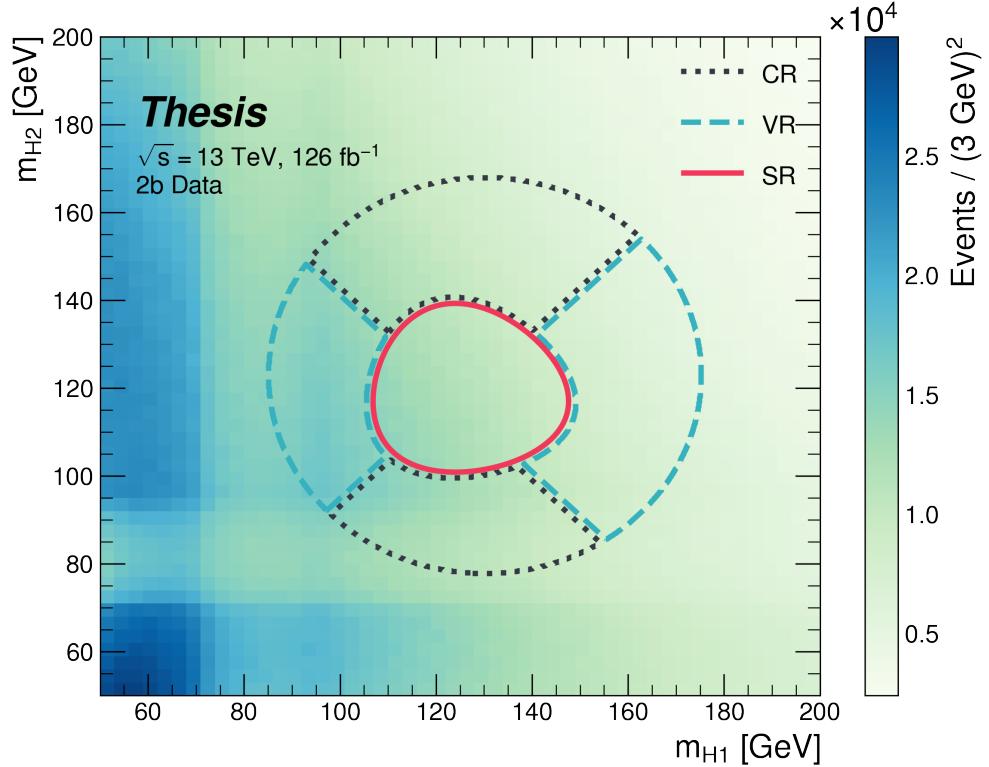


Figure 7.4: Regions used for the non-resonant search. The “top” and “bottom” quadrants together comprise the control region, in which the nominal background estimate is derived. The “left” and “right” quadrants together comprise the validation region, which is used to assess an uncertainty. The signal region, in the center, is where the signal extraction fit is performed.

1945 7.5.3 *Discriminating Variable*

1946 The discriminant used for the resonant analysis is *corrected* m_{HH} . This variable is calculated
 1947 by re-scaling the Higgs candidate four vectors such that each $m_H = 125 \text{ GeV}$. These re-scaled
 1948 four-vectors are then summed, and their invariant mass is the corrected m_{HH} . These re-scaled
 1949 four-vectors are not used for any other purpose. The effect of this correction, which sharpens
 the m_{HH} peak and correctly centres it, is shown in Figure 7.5.

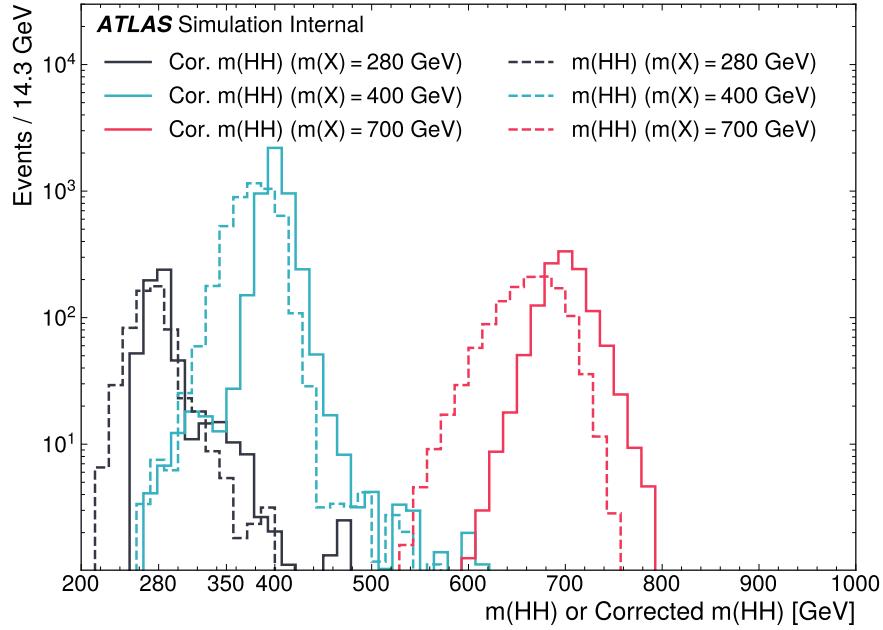


Figure 7.5: Impact of the m_{HH} correction on a range of spin-0 resonant signals. The corrected m_{HH} distributions (solid lines) are much sharper and more centered on the corresponding resonance masses than the uncorrected m_{HH} distributions (dashed).

1950

1951 For the non-resonant analysis, due to the broad nature of the signal in m_{HH} , such a
 1952 correction is not as motivated, and, indeed, is found to have very minimal impact. The
 1953 uncorrected m_{HH} (just referred to as m_{HH}) is therefore used as a discriminant. To maximize

1954 sensitivity, the non-resonant analysis additionally uses two variables for categorization: $\Delta\eta_{HH}$,
1955 an angular variable which, along with m_{HH} , fully characterizes the HH system [99], and X_{HH} ,
1956 the variable used for the signal region definition, which leverages the peaked structure of the
1957 signal in the $(m(H_1), m(H_2))$ plane to split the signal extraction fit into lower and higher
1958 purity regions (highest purity near $X_{HH} = 0$, the center of the signal region). Distributions
1959 of these variables are shown in *TODO: plots*. The categorization used for this thesis has been
1960 optimized to be 2×2 in these variables, with corresponding selections $0 \leq \Delta\eta_{HH} \leq 0.75$ and
1961 $0.75 \leq \Delta\eta_{HH} \leq 1.5$ for $\Delta\eta_{HH}$, and $0 \leq X_{HH} \leq 0.95$ and $0.95 \leq X_{HH} \leq 1.6$ for X_{HH} .

¹⁹⁶² **7.6 Background Estimation**

¹⁹⁶³ After the event selection described above there are two major backgrounds, QCD and $t\bar{t}$.
¹⁹⁶⁴ A very similar approach is used for both the resonant and the non-resonant analyses, with
¹⁹⁶⁵ some small modifications. This approach is notably fully data-driven, which is warranted due
¹⁹⁶⁶ to the flexibility of the estimation method, as well as the high relative proportion of QCD
¹⁹⁶⁷ background ($> 90\%$), and allows for the use of machine learning methods in the construction
¹⁹⁶⁸ of the background estimate. However, it sacrifices an explicit treatment of the $t\bar{t}$ component.
¹⁹⁶⁹ Performance of the background estimate on the $t\bar{t}$ component is checked explicitly, and
¹⁹⁷⁰ minimal impact due to $t\bar{t}$ mismodeling is seen.

¹⁹⁷¹ Contributions of single Higgs processes and ZZ are found to be negligible, and the
¹⁹⁷² Standard Model HH background is found to have no impact on the resonant search.

¹⁹⁷³ The foundation of the background estimate lies in the derivation of a reweighting function
¹⁹⁷⁴ which matches the kinematics of events with exactly two b -tagged jets to those of events in
¹⁹⁷⁵ the higher tagged regions (events with three or four b -tagged jets). The reweighting function
¹⁹⁷⁶ and overall normalization are derived in the control region. Systematic bias of this estimate
¹⁹⁷⁷ is assessed in the validation region.

¹⁹⁷⁸ For the resonant analysis, the systematic bias is a bias due to extrapolation: the validation
¹⁹⁷⁹ region lies between the control and signal regions. For the non-resonant analysis, the bias
¹⁹⁸⁰ instead comes from different possible interpolations of the signal region kinematics – given the
¹⁹⁸¹ choice of nominal estimate, the validation region is a conceptually equivalent, but maximally
¹⁹⁸² different, signal region estimate.

¹⁹⁸³ **7.6.1 The Two Tag Region**

¹⁹⁸⁴ Events in data with exactly two b -tagged jets are used for the data driven background estimate.
¹⁹⁸⁵ The hypothesis here is that, due to the presence of multiple b -tagged jets, the kinematics of
¹⁹⁸⁶ such events are similar to the kinematics of events in higher b -tagged regions (i.e. events
¹⁹⁸⁷ with three and four b -tagged jets, respectively), and any differences can be corrected by a

1988 reweighting procedure. The region with three b -tagged jets is split into two b -tagging regions,
 1989 as described in Section 7.3, with the $3b + 1$ loose region used as an additional signal region.
 1990 The lower tagged $3b$ component ($3b + 1$ fail) is reserved for validation of the background
 1991 modelling procedure. Events with fewer than two b -tagged jets are not used for this analysis,
 1992 as they are relatively more different from the higher tag regions.

1993 The nominal event selection requires at least four jets in order to form Higgs candidates.
 1994 For the four tag region, these are the four highest p_T b -tagged jets. For the three tag regions,
 1995 these jets are the three b -tagged jets, plus the highest p_T jet satisfying a loosened b -tagging
 1996 requirement. Similarly, and following the approach of the resonant analysis, the two tag region
 1997 uses the two b -tagged jets and the two highest p_T non-tagged jets to form Higgs candidates.
 1998 Combinatoric bias from selection of different numbers of b -tagged jets is corrected as a part
 1999 of the kinematic reweighting procedure through the reweighting of the total number of jets in
 2000 the event. In this way, the full event selection may be run on two tagged events.

2001 7.6.2 Kinematic Reweighting

2002 The set of two tagged data events is the fundamental piece of the data driven background
 2003 estimate. However, kinematic differences from the four tag region exist and must be corrected
 2004 in order for this estimate to be useful. Binned approaches based on ratios of histograms
 2005 have been previously considered [82], [17], but are limited in their handling of correlations
 2006 between variables and by the “curse of dimensionality”, i.e. the dataset becomes sparser and
 2007 sparser in “reweighting space” as the number of dimensions in which to reweight increases,
 2008 limiting the number of variables used for reweighting. This leads either to an unstable fit
 2009 result (overfitting with finely grained bins) or a lower quality fit result (underfitting with
 2010 coarse bins).

2011 Note that even machine learning methods such as Boosted Decision Trees (BDTs) [100],
 2012 may suffer from this curse of dimensionality, as the depth of each decision tree used is limited
 2013 by the available statistics after each set of corresponding selections (cf. binning in a more
 2014 sophisticated way), limiting the expressivity of the learned reweighting function.

2015 To solve these issues, a neural network based reweighting procedure is used here. This
 2016 is a truly multivariate approach, allowing for proper treatment of variable correlations. It
 2017 further overcomes the issues associated with binned approaches by learning the reweighting
 2018 function directly, allowing for greater sensitivity to local differences and helping to avoid the
 2019 curse of dimensionality.

2020 *Neural Network Reweighting*

Let $p_{4b}(x)$ and $p_{2b}(x)$ be the probability density functions for four and two tag data respectively across some input variables x . The problem of learning the reweighting function between two and four tag data is then the problem of learning a function $w(x)$ such that

$$p_{2b}(x) \cdot w(x) = p_{4b}(x) \quad (7.6)$$

from which it follows that

$$w(x) = \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (7.7)$$

This falls into the domain of density ratio estimation, for which there are a variety of approaches. The method considered here is modified from [101, 102], and depends on a loss function of the form

$$\mathcal{L}(R(x)) = \mathbb{E}_{x \sim p_{2b}}[\sqrt{R(x)}] + \mathbb{E}_{x \sim p_{4b}}\left[\frac{1}{\sqrt{R(x)}}\right]. \quad (7.8)$$

where $R(x)$ is some estimator dependent on x and $\mathbb{E}_{x \sim p_{2b}}$ and $\mathbb{E}_{x \sim p_{4b}}$ are the expectation values with respect to the 2b and 4b probability densities. A neural network trained with such a loss function has the objective of finding the estimator, $R(x)$, that minimizes this loss. It is straightforward to show that

$$\arg \min_R \mathcal{L}(R(x)) = \frac{p_{4b}(x)}{p_{2b}(x)} \quad (7.9)$$

2021 which is exactly the form of the desired reweighting function.

In practice, to avoid imposing explicit positivity constraints, the substitution $Q(x) \equiv \log R(x)$ is made. The loss function then takes the equivalent form

$$\mathcal{L}(Q(x)) = \mathbb{E}_{x \sim p_{2b}}[\sqrt{e^{Q(x)}}] + \mathbb{E}_{x \sim p_{4b}}\left[\frac{1}{\sqrt{e^{Q(x)}}}\right], \quad (7.10)$$

with solution

$$\arg \min_Q \mathcal{L}(Q(x)) = \log \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (7.11)$$

2022 Taking the exponent then results in the desired reweighting function.

2023 Note that similar methods for density ratio estimation are available [103], e.g. from a

2024 more standard binary cross-entropy loss. However, these were found to perform no better
2025 than the formulation presented here.

2026 *Variables and Results*

2027 The neural network is trained on a variety of variables sensitive to two vs. four tag differences.

2028 To help bring out these differences, the natural logarithm of some of the variables with a
2029 large, local change is taken. The set of training variables used for the resonant analysis is

2030 1. $\log(p_T)$ of the 4th leading Higgs candidate jet

2031 2. $\log(p_T)$ of the 2nd leading Higgs candidate jet

2032 3. $\log(\Delta R)$ between the closest two Higgs candidate jets

2033 4. $\log(\Delta R)$ between the other two Higgs candidate jets

2034 5. Average absolute value of Higgs candidate jet η

2035 6. $\log(p_T)$ of the di-Higgs system.

2036 7. ΔR between the two Higgs candidates

2037 8. $\Delta\phi$ between the jets in the leading Higgs candidate

2038 9. $\Delta\phi$ between the jets in the subleading Higgs candidate

2039 10. $\log(X_{Wt})$, where X_{Wt} is the variable used for the top veto

2040 11. Number of jets in the event.

2041 The non-resonant analysis uses an identical set of variables with two notable changes

2042 1. The definition of X_{Wt} differs from the resonant definition (as described in Section 7.4).

2043 2. An integer encoding of the two trigger categories is used as an input (variable which
 2044 takes on the value 0 or 1 corresponding to each of the two categories). This was found
 2045 to improve a mismodeling near the tradeoff in m_{HH} of the two buckets.

2046 The neural network used for both resonant and non-resonant reweighting has three densely
 2047 connected hidden layers of 50 nodes each with ReLU activation functions and a single node
 2048 linear output. This configuration demonstrates good performance in the modelling of a variety
 2049 of relevant variables, including m_{HH} , when compared to a range of networks of similar size.

2050 In practice, a given training of the reweighting neural network is subject to variation
 2051 due to training statistics and initial conditions. An uncertainty is assigned to account for
 2052 this (Section 7.7), which relies on training an ensemble of reweighting networks [104]. To
 2053 increase the stability of the background estimate, the median of the predicted weight for each
 2054 event is calculated across the ensemble. This median is then used as the nominal background
 2055 estimate. This approach is indeed seen to be much more stable and to demonstrate a better
 2056 overall performance than a single arbitrary training. Each ensemble used for this analysis
 2057 consists of 100 neural networks, trained as described in Section 7.7.

2058 The training of the ensemble used for the nominal estimate is done in the kinematic
 2059 Control Region. The prediction of these networks in the Signal Region is then used for the
 2060 nominal background estimate. In addition, a separate ensemble of networks is trained in the
 2061 Validation Region. The difference between the prediction of the nominal estimate and the

2062 estimate from the VR derived networks in the Signal Region is used to assign a systematic
 2063 uncertainty. Further details on this systematic uncertainty are shown in Section 7.7. Note
 2064 that although the same procedure is used for both Control and Validation Region trained
 2065 networks, only the median estimate from the VR derived reweighting is used for assessing a
 2066 systematic – no additional “uncertainty on the uncertainty” from VR ensemble variation is
 2067 applied.

2068 Each reweighted estimate is normalized such that the reweighted $2b$ yield matches the $4b$
 2069 yield in the corresponding training region. Note that this applies to each of the networks used
 2070 in each ensemble, where the normalization factor is also subject to the procedure described in
 2071 Section 7.7. As the median over these normalized weights is not guaranteed to preserve this
 2072 normalization, a further correction is applied such that the $2b$ yield, after the median weights
 2073 are applied, matches the $4b$ yield in the corresponding training region. As no preprocessing
 2074 is applied to correct for the class imbalance between $2b$ and $4b$ events entering the training,
 2075 this ratio of number of $4b$ events ($n(4b)$) over number of $2b$ events ($n(2b)$) is folded into the
 2076 learned weights. Correspondingly, the set of normalization factors described above is near 1
 2077 and the learned weights are centered around $n(4b)/n(2b)$ (roughly 0.01 over the full dataset).
 2078 This normalization procedure applies for all instances of the reweighting (e.g. those used for
 2079 validations in Section 7.8), with appropriate substitutions of reweighting origin (here $2b$) and
 2080 reweighting target (here $4b$).

2081 Note that, due to different trigger and pileup selections during each year, the reweighting
 2082 is trained on each year separately. An approach of training all of the years together with
 2083 a one-hot encoding was explored, but was found to have minimal benefit over the split
 2084 years approach, and in fact to increase the systematic bias of the corresponding background
 2085 estimate. Because of this, and because trigger selections for each year significantly impact
 2086 the kinematics of each year, such that categorizing by year is expected to reflect groupings
 2087 of kinematically similar events and to provide a meaningful degree of freedom in the signal
 2088 extraction fit, the split-year approach is kept.

2089 The control region closure for the 2018 dataset is shown for the resonant search in Figures

2090 7.6 through 7.14 and for the non-resonant search in Figures 7.24 through 7.32 for 4b and
 2091 Figures 7.42 through 7.50 for 3b1l. The impact of this control region derived reweighting
 2092 on the validation region is shown in Figures 7.15 through 7.23 for the resonant search and
 2093 Figures 7.33 through 7.41 for 4b and Figures 7.51 through 7.59 for 3b1l for the non-resonant
 2094 search. Generally good performance is seen, with some occasional mis-modeling. For the
 2095 resonant search, this is most notable in the case of individual jet p_T . Such mis-modeling
 2096 may be corrected by including the variables in the input set, but this was found to not
 2097 improve the modeling of m_{HH} , and so is not done here. This mis-modeling is notable for the
 2098 non-resonant search in the leading Higgs candidate jet p_T , and is a direct consequence of the
 2099 trigger category input, which improves modeling of m_{HH} . Results are similar for other years,
 2100 but are not included here for brevity.

2101 One other salient feature of the non-resonant plots is the distributions of m_{H1} and m_{H2} ,
 2102 which emphasize the quadrant region definitions – the control region has a peak around
 2103 125 GeV in m_{H1} , which may be thought of as “signal region-like”, motivating this alignment,
 2104 though consequently the distribution of m_{H2} is quite bimodal. The reverse is true for the
 2105 validation region.

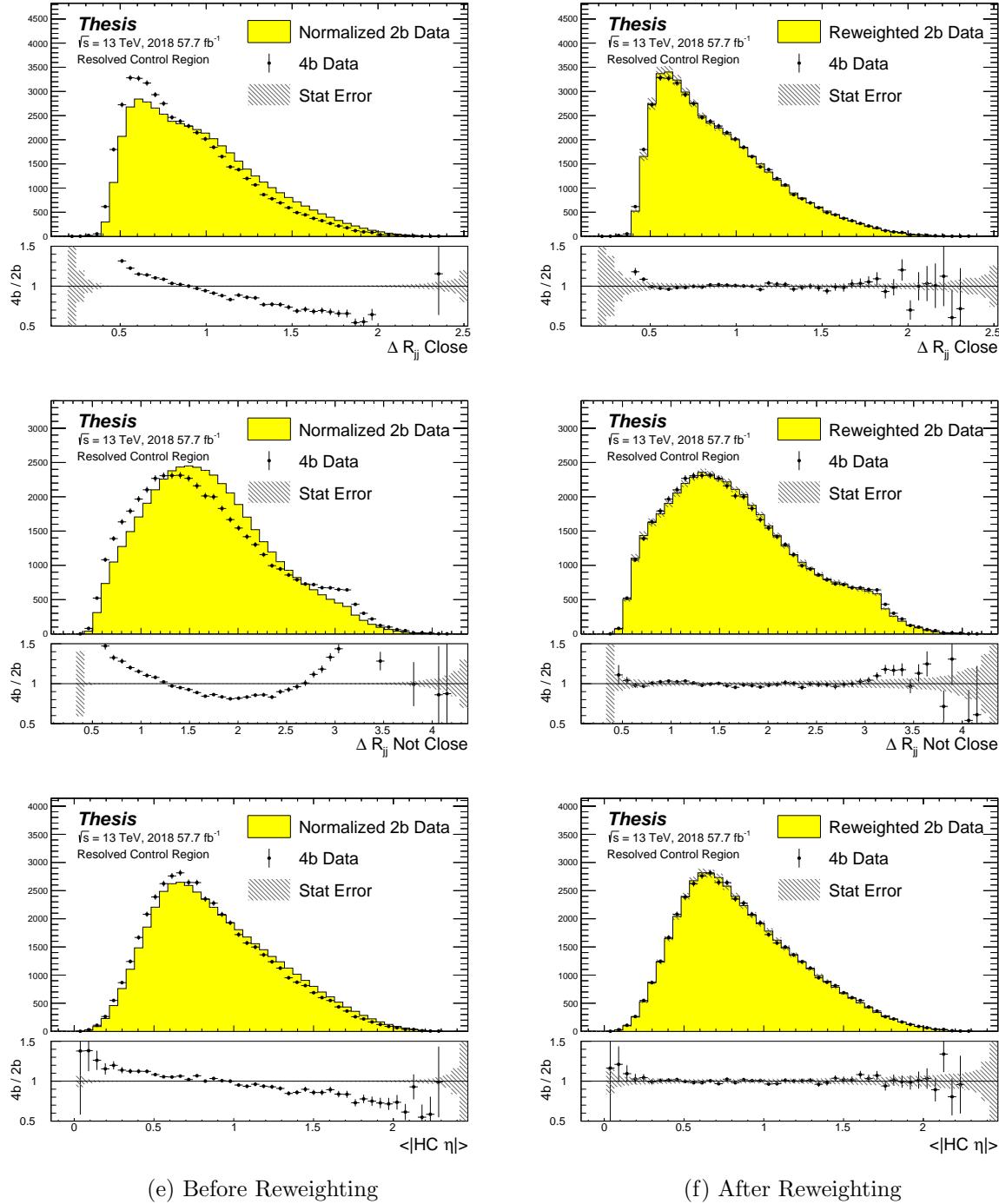


Figure 7.6: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Control Region.

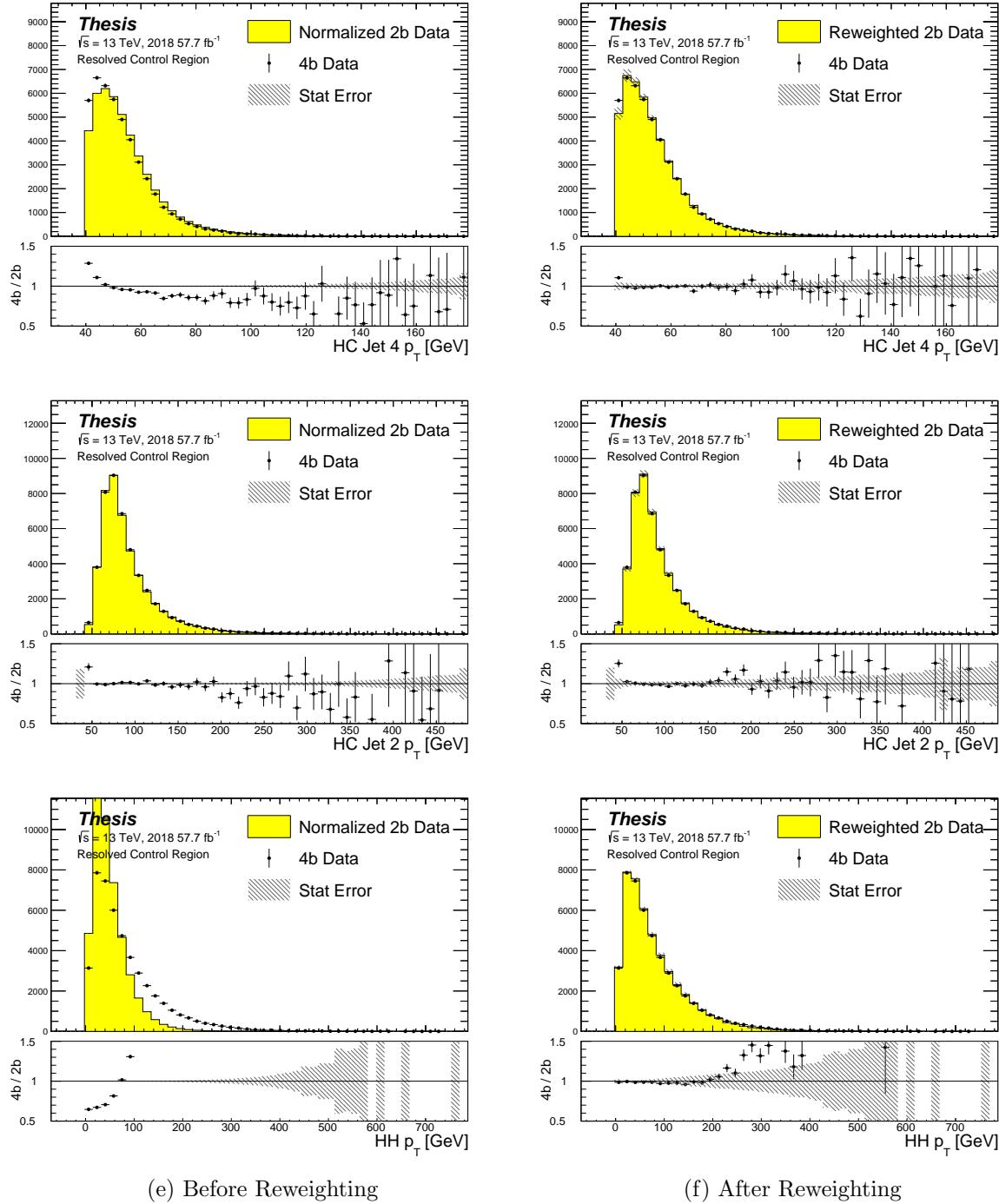


Figure 7.7: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.

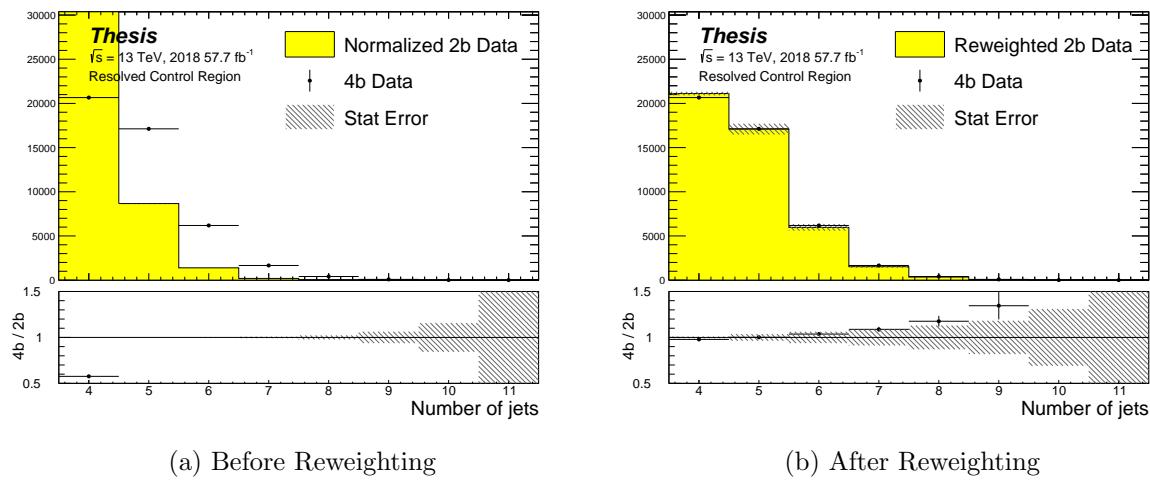


Figure 7.8: Resonant Search: Distributions of the number of jets before and after CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

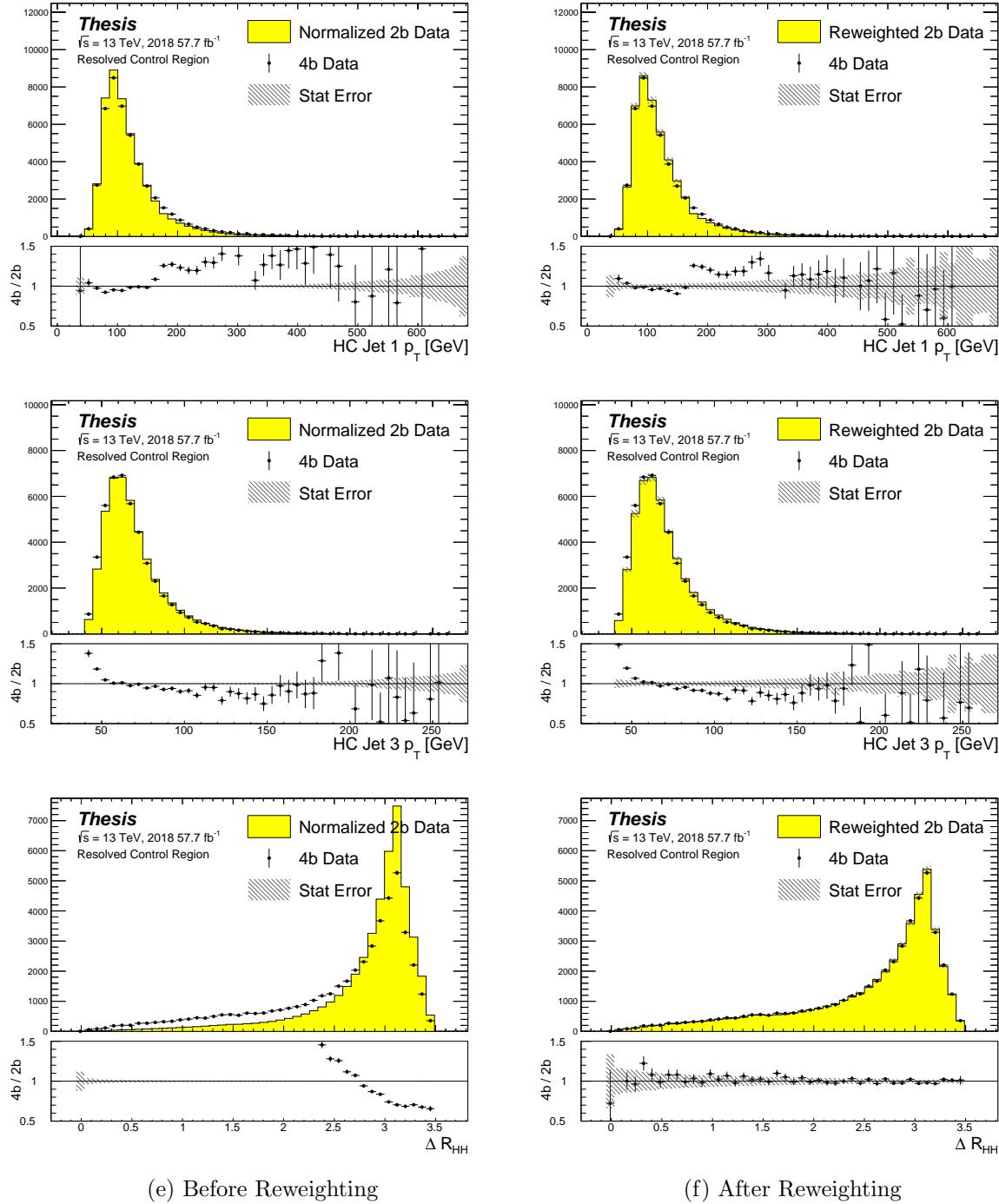


Figure 7.9: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

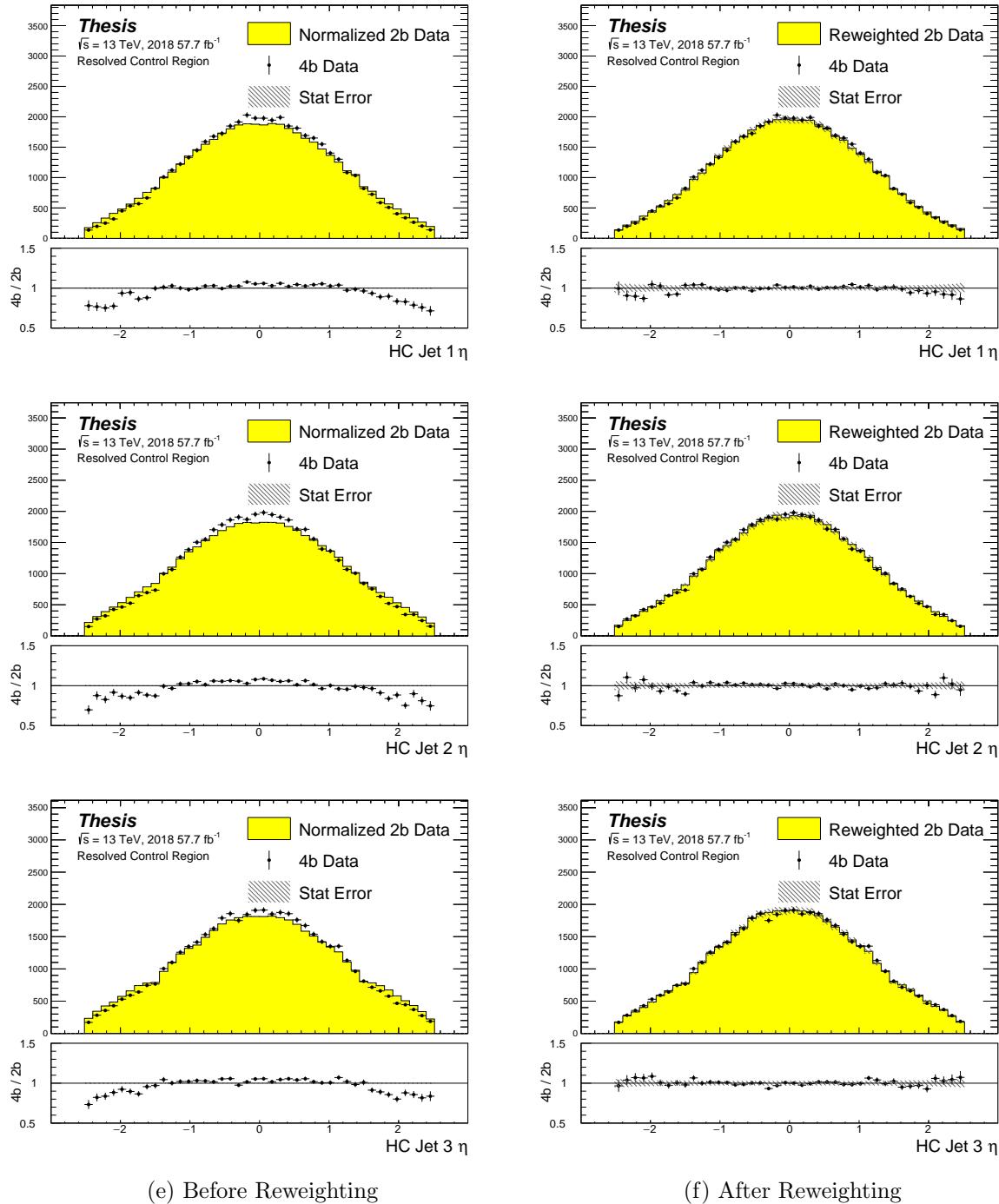


Figure 7.10: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Control Region.

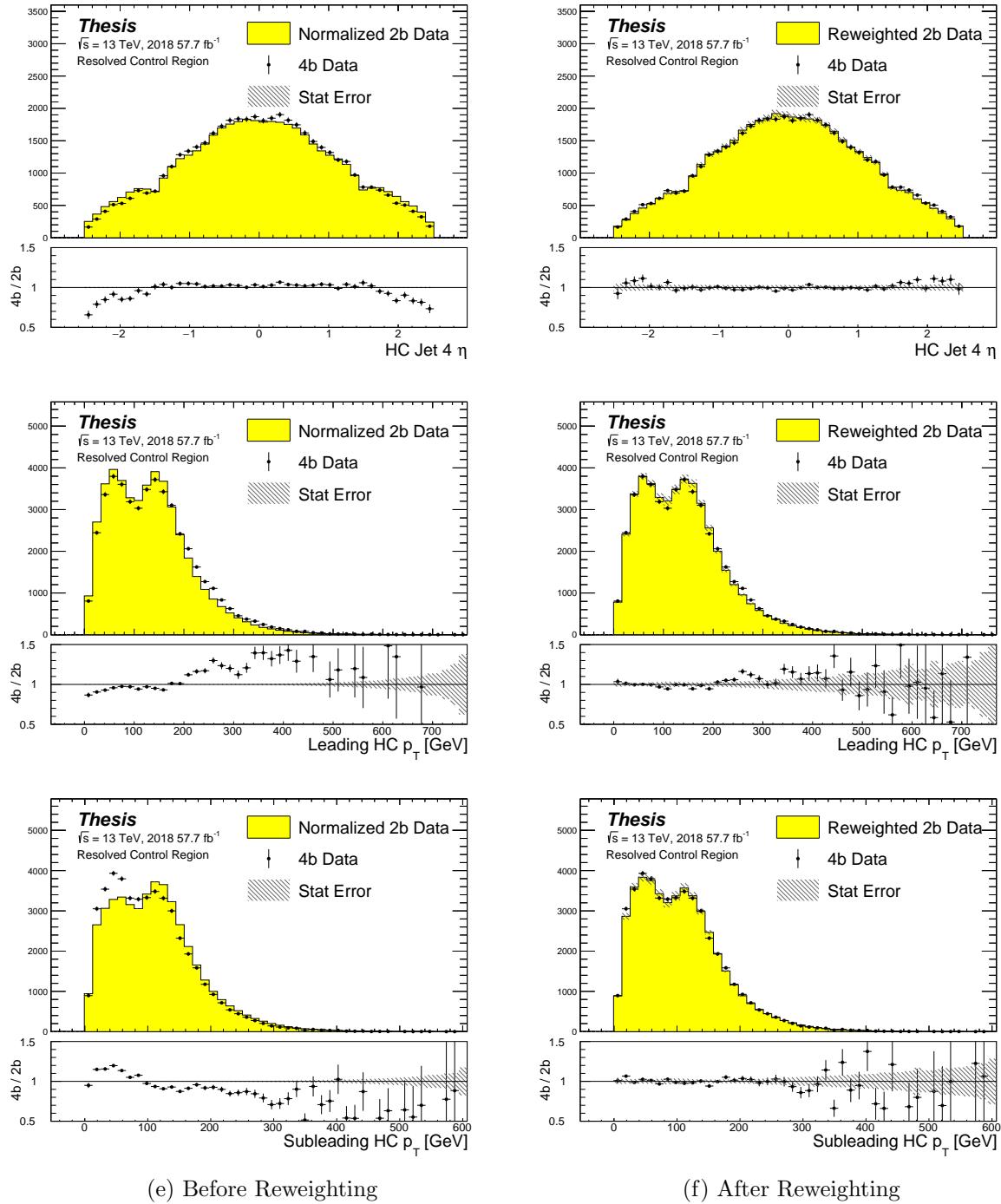


Figure 7.11: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

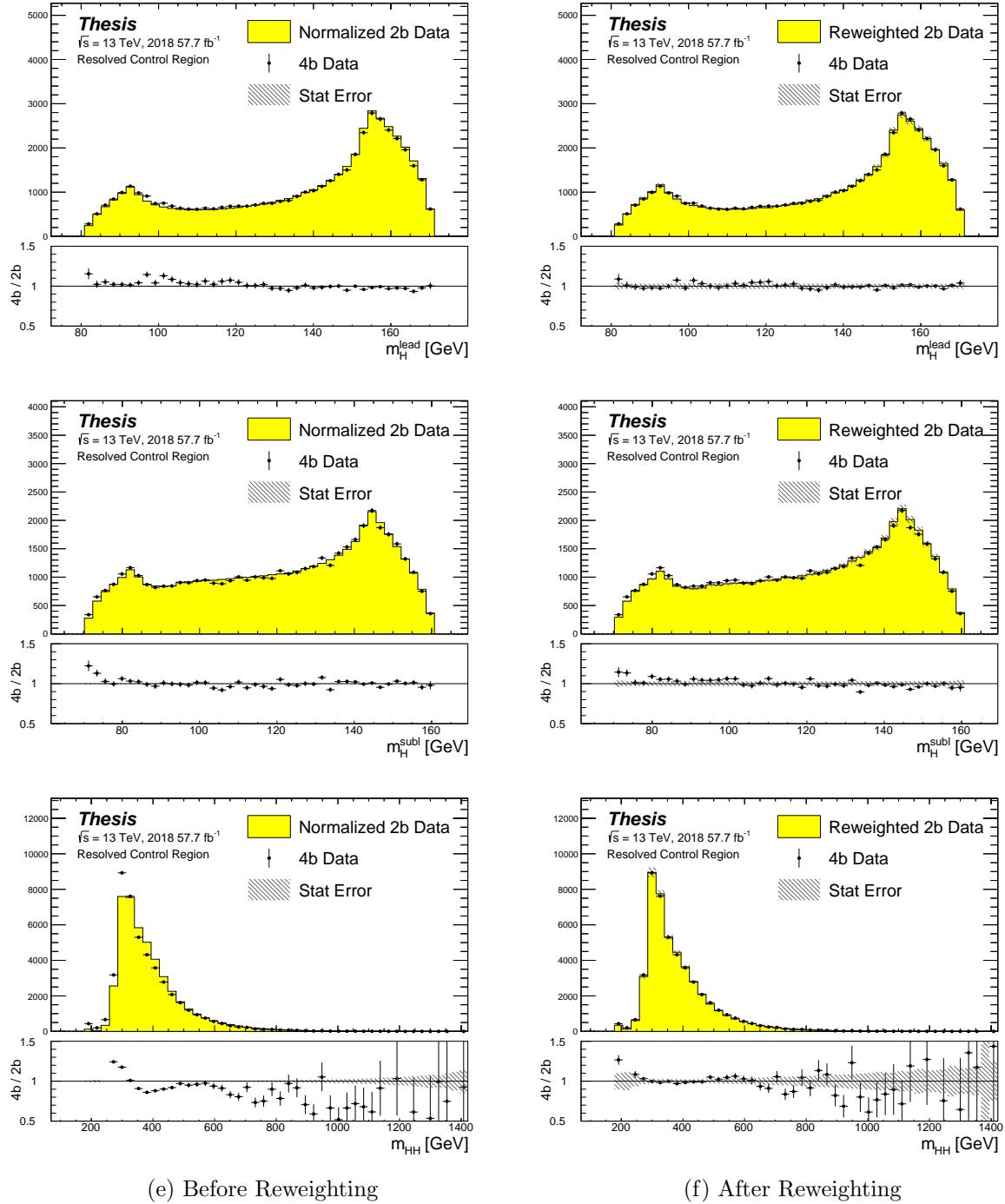


Figure 7.12: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Control Region.

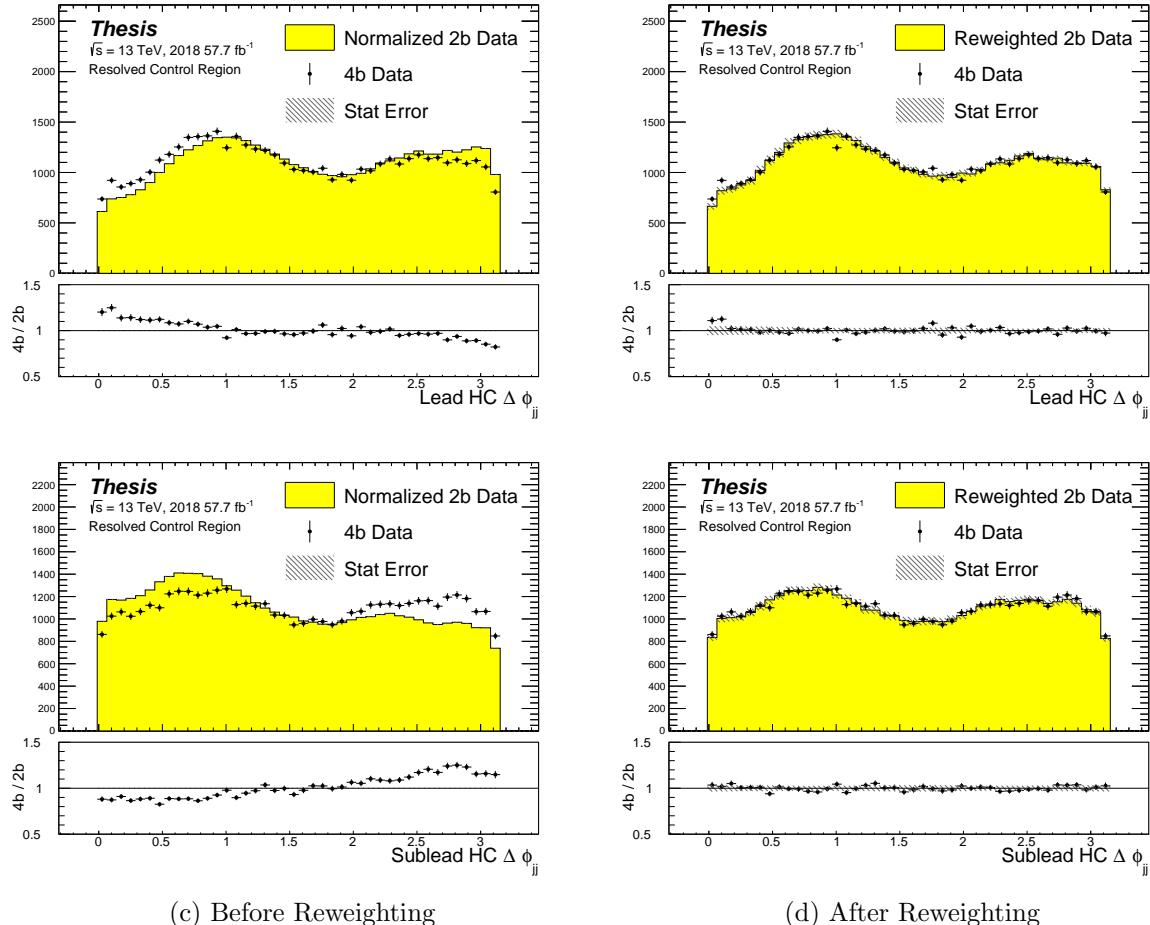


Figure 7.13: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Control Region.

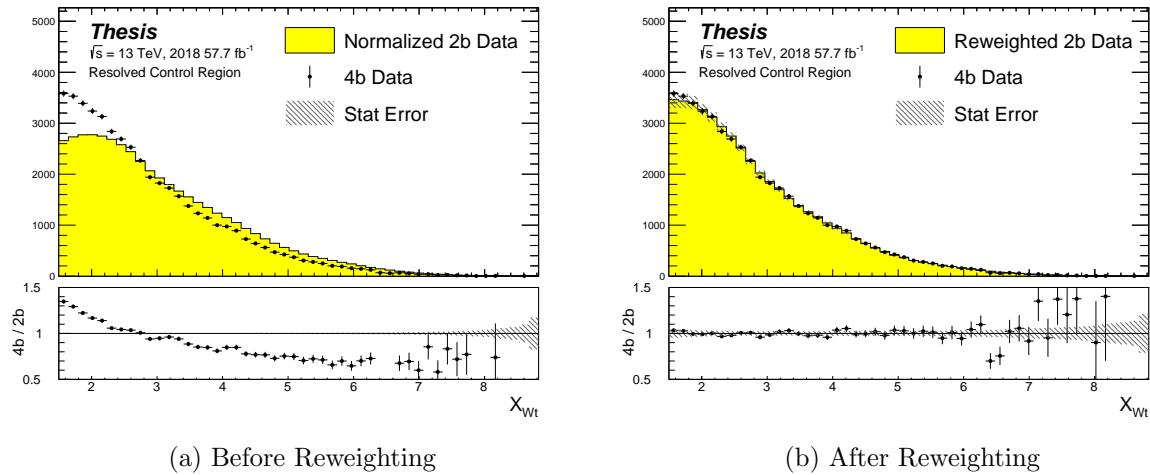


Figure 7.14: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied

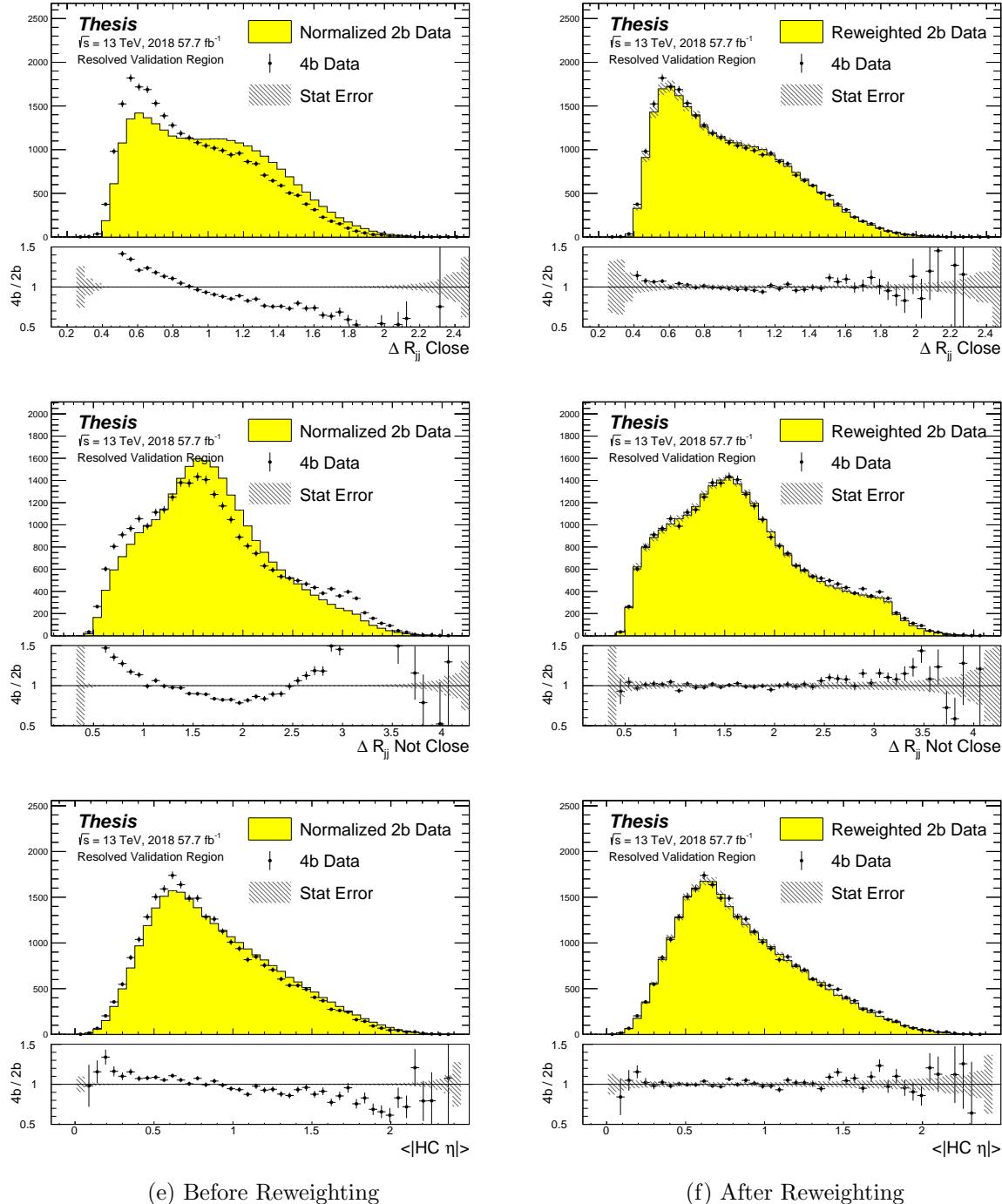


Figure 7.15: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 Validation Region.

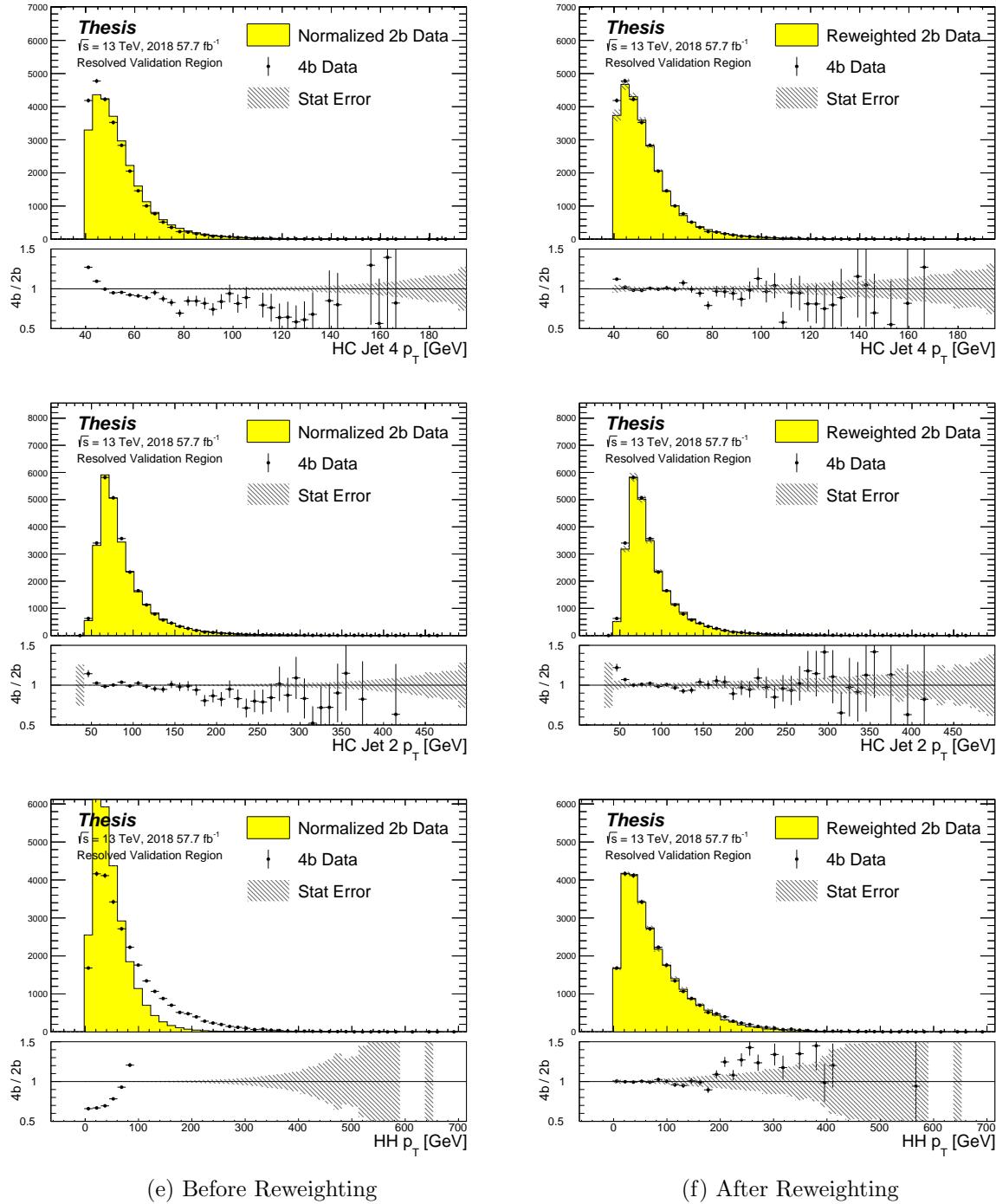


Figure 7.16: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

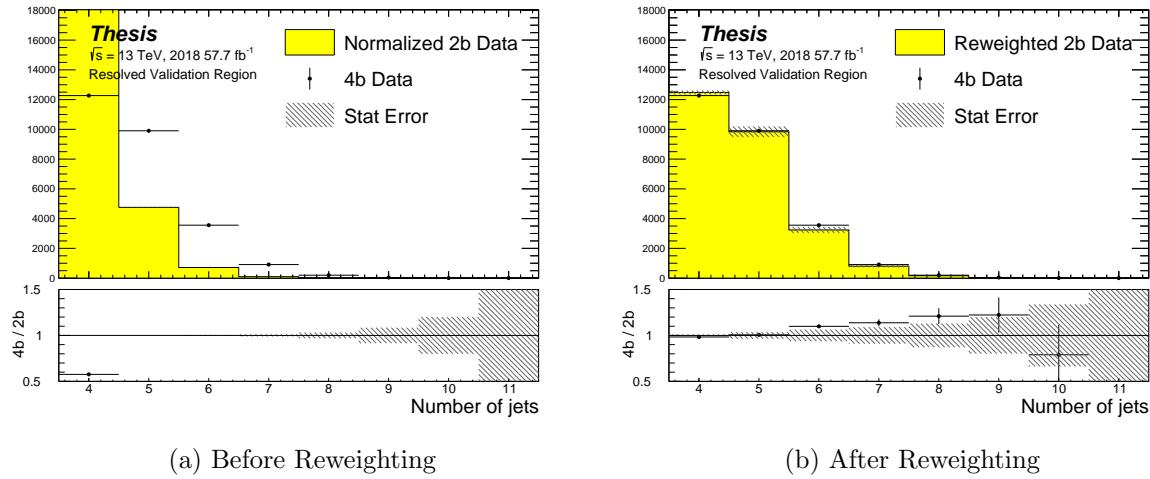


Figure 7.17: **Resonant Search:** Distributions of the number of jets before and after CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

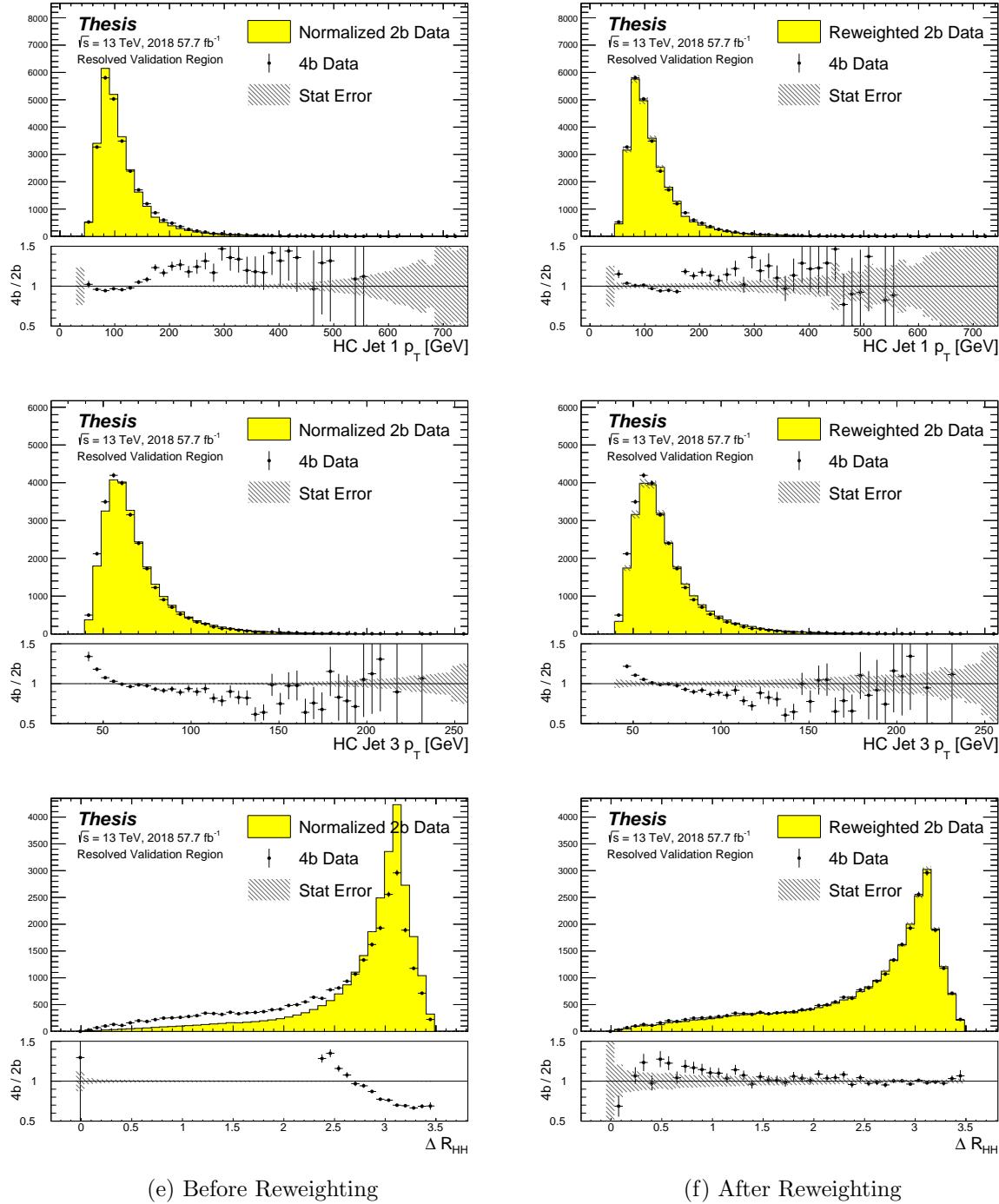


Figure 7.18: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

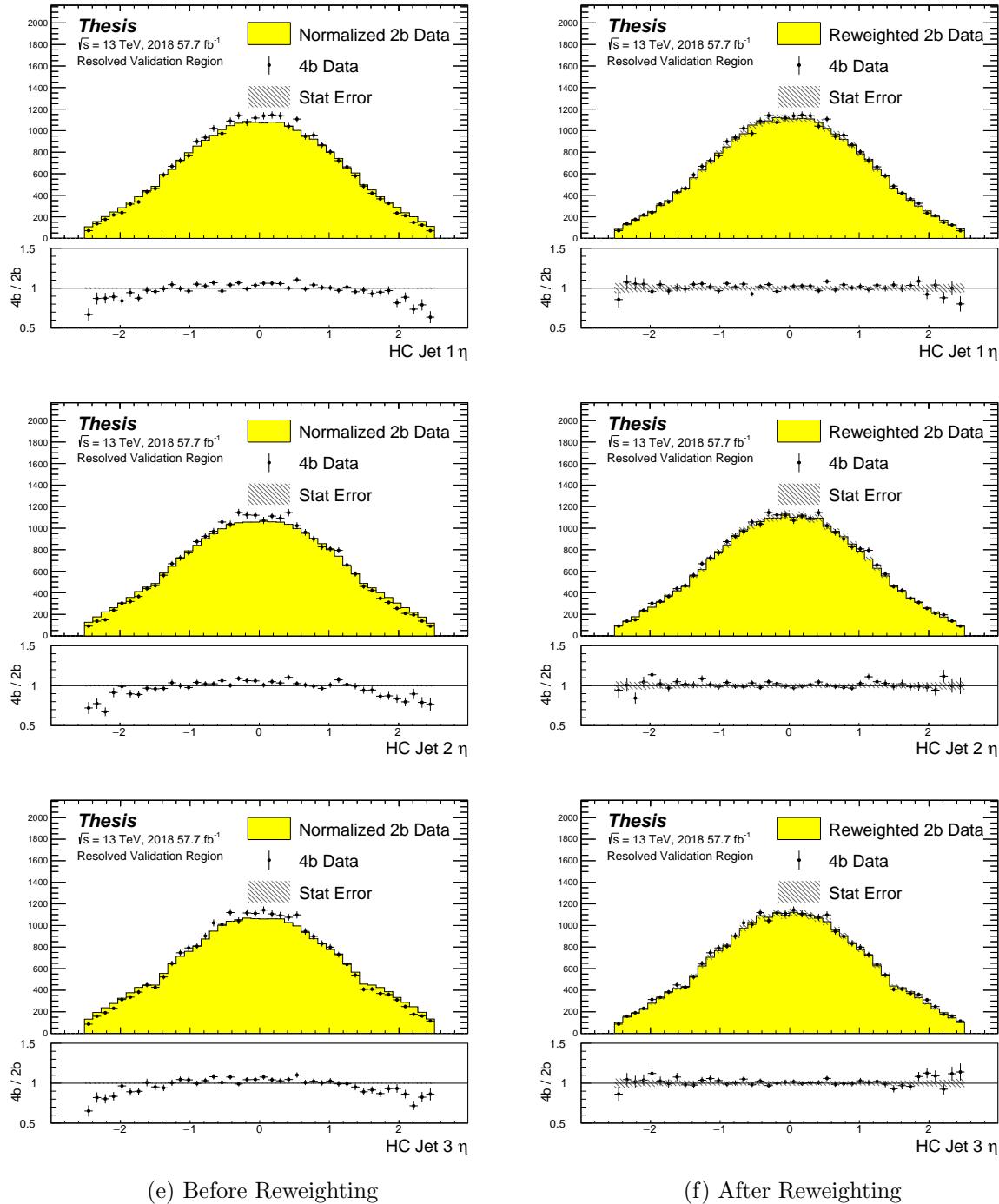


Figure 7.19: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 Validation Region.

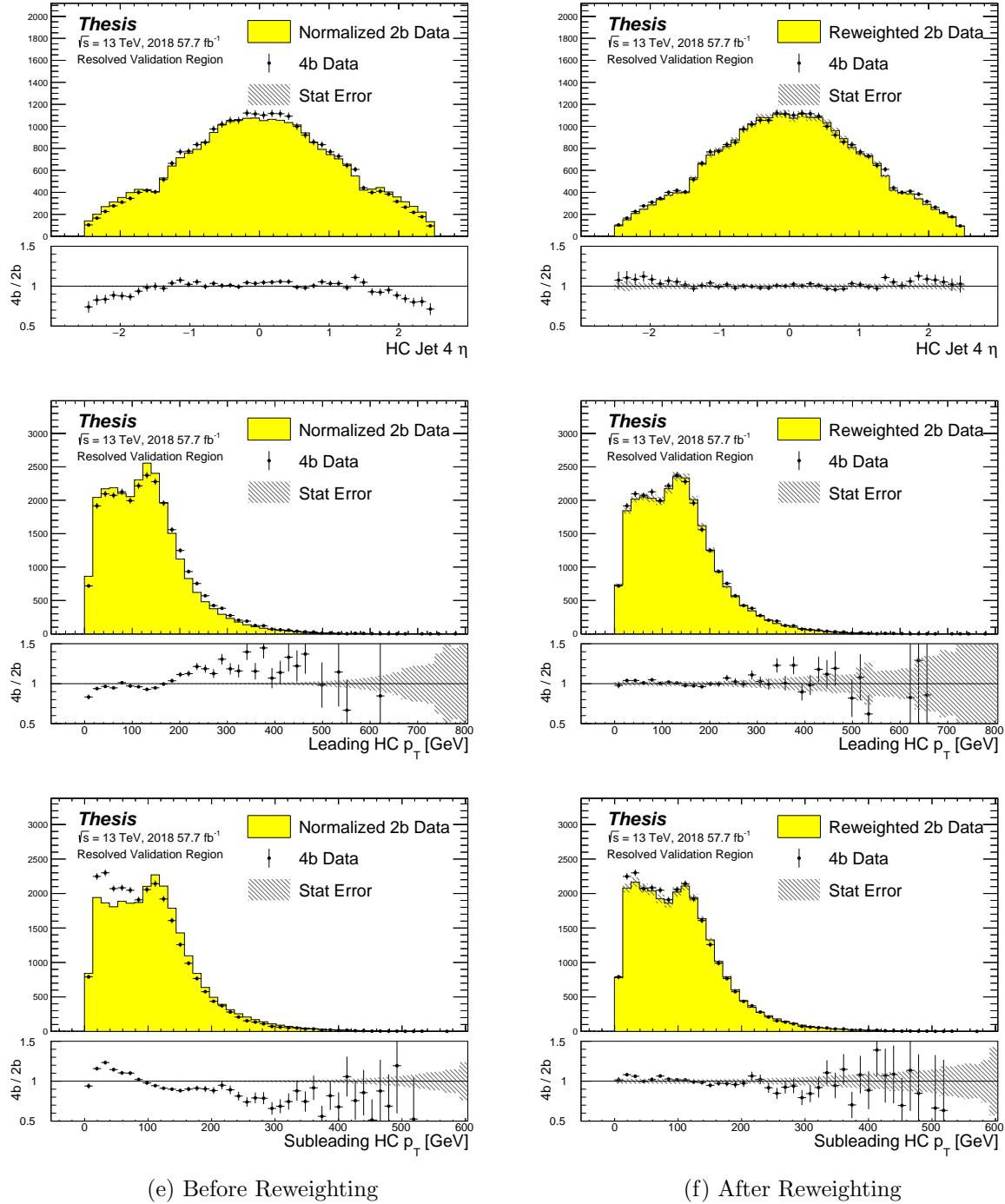


Figure 7.20: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

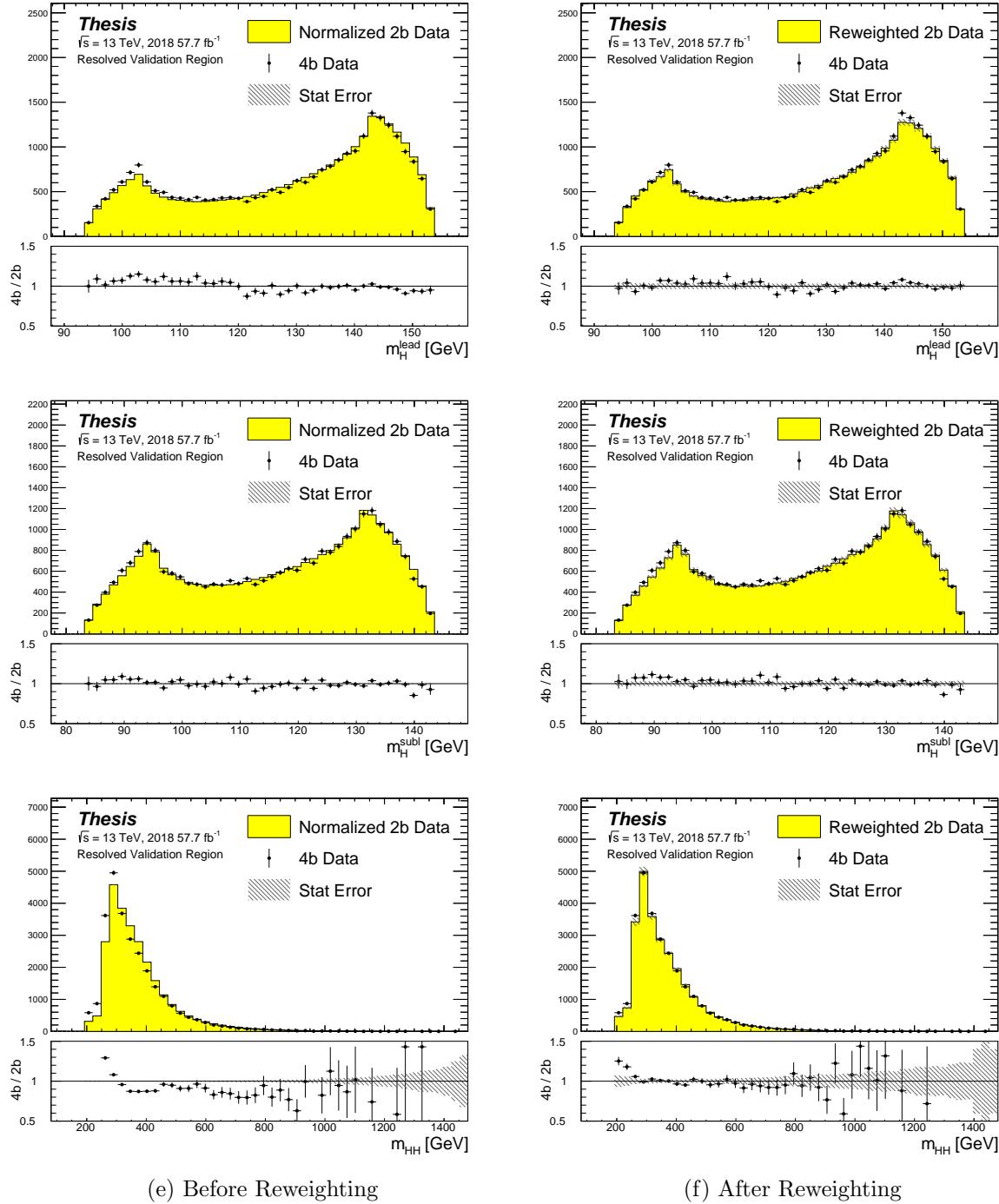


Figure 7.21: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 Validation Region.

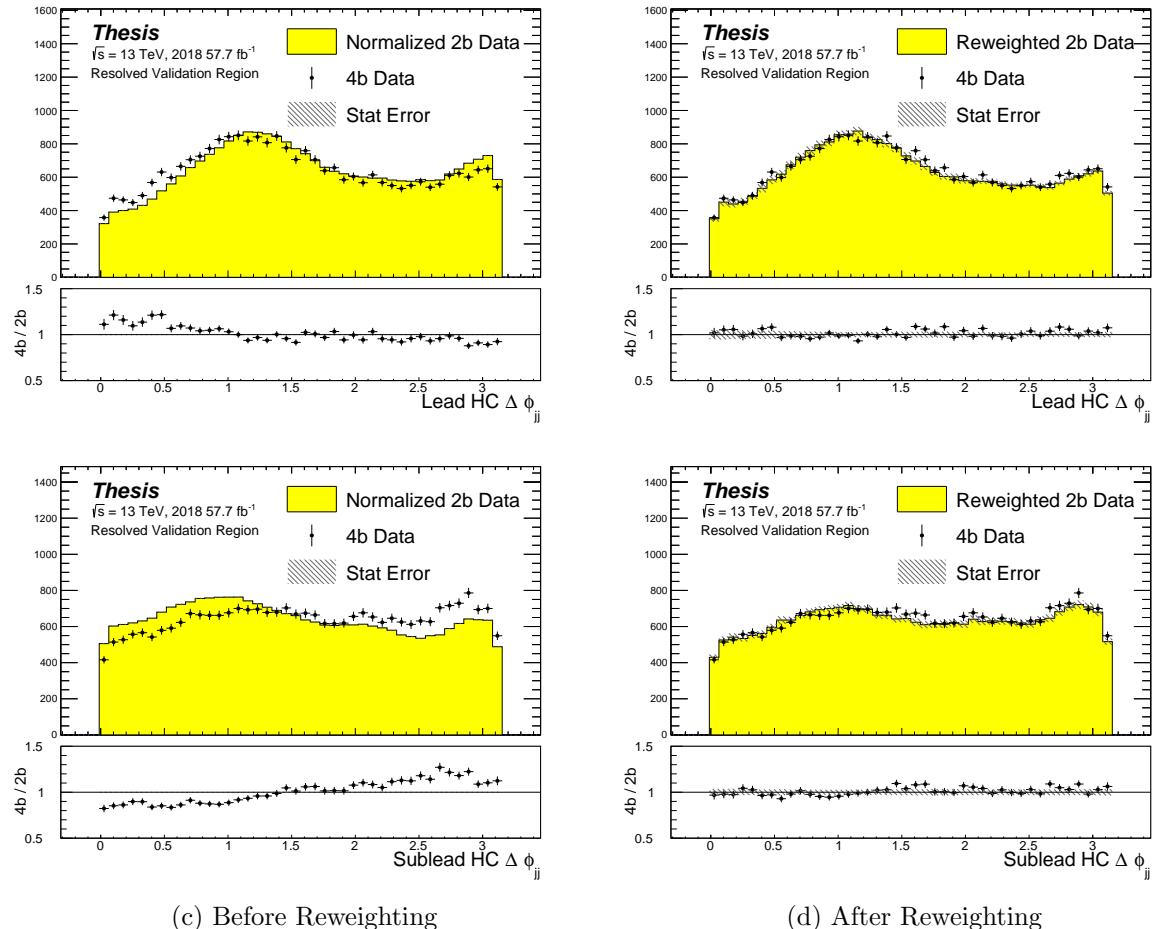


Figure 7.22: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 Validation Region.

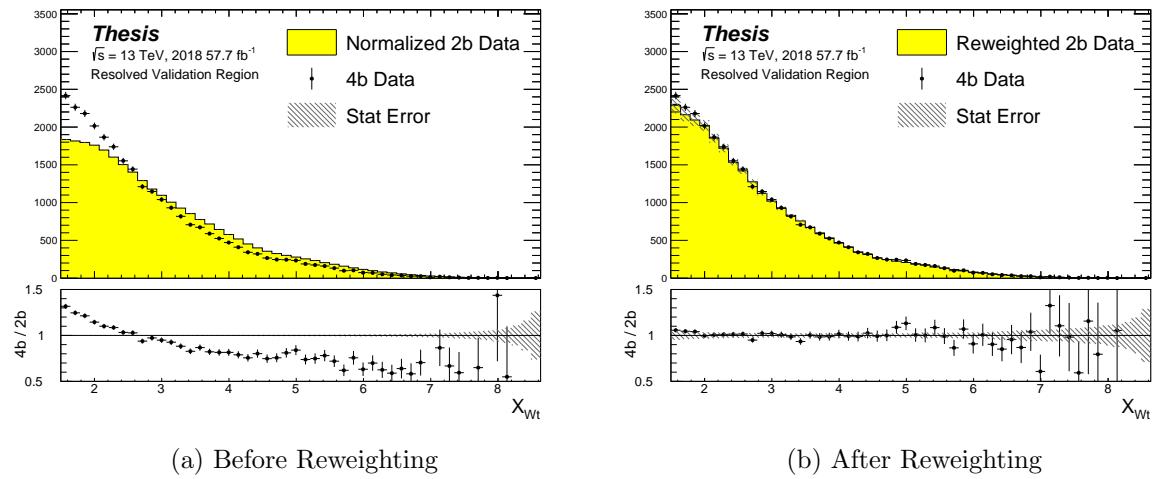


Figure 7.23: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied

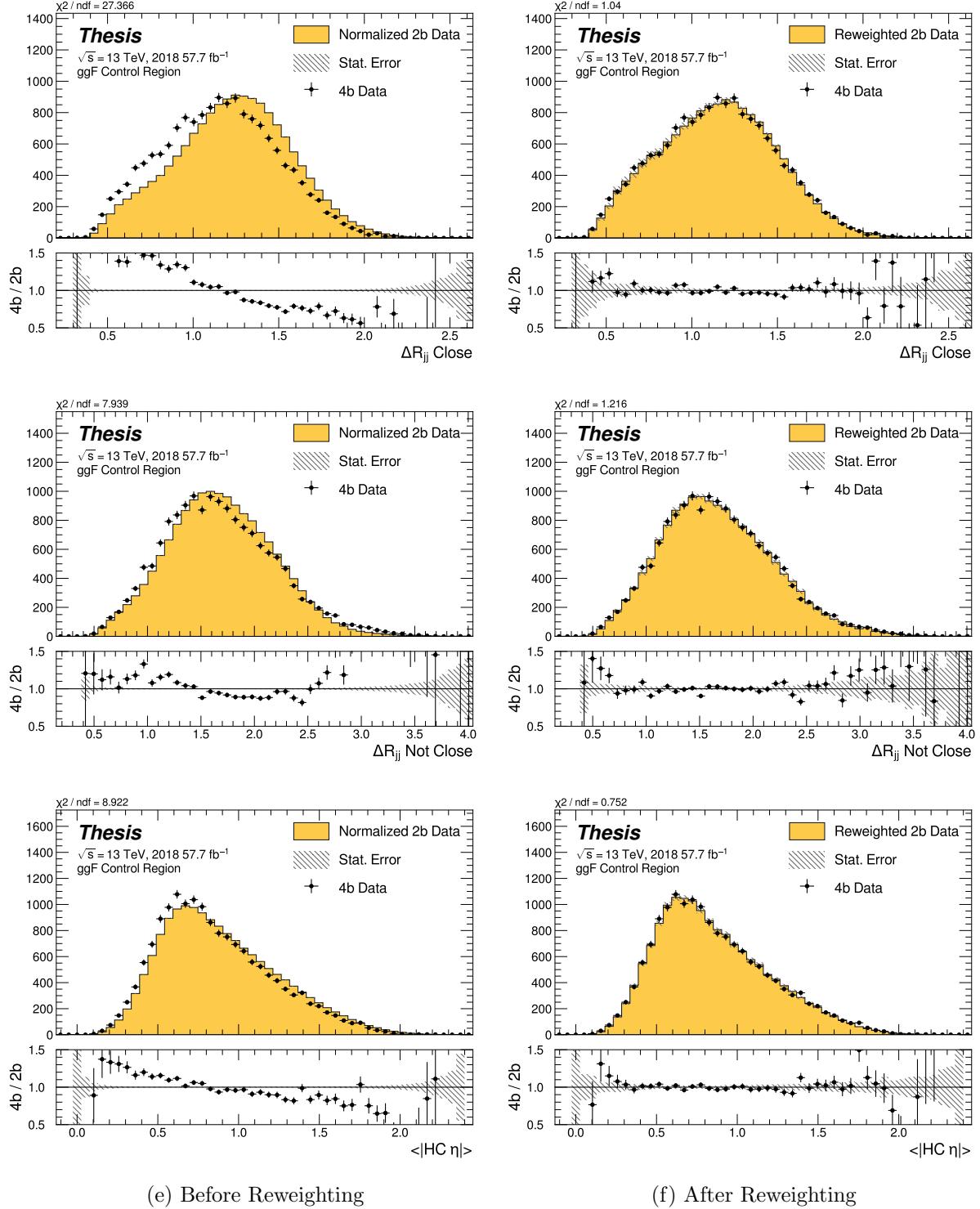


Figure 7.24: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Control Region.

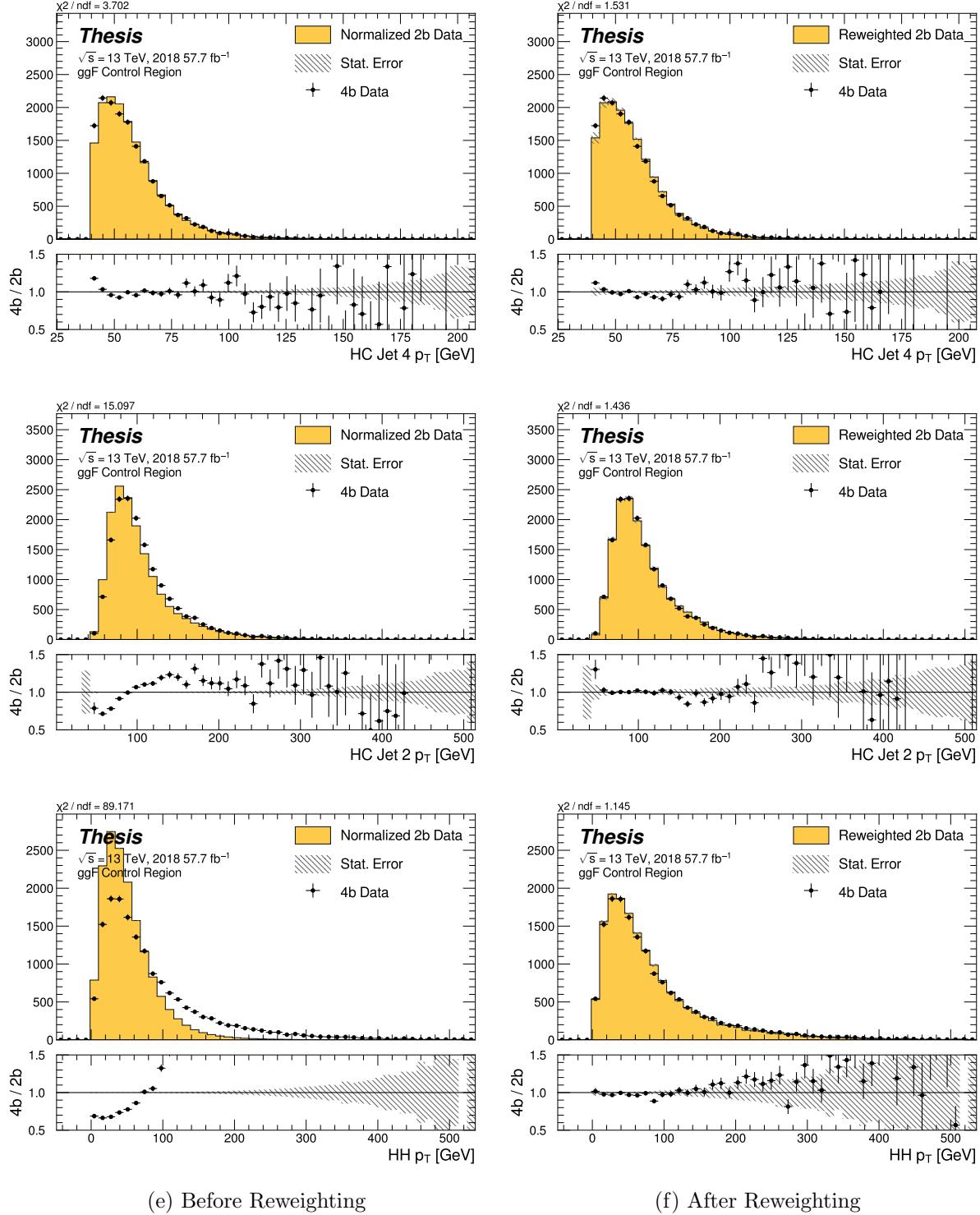


Figure 7.25: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

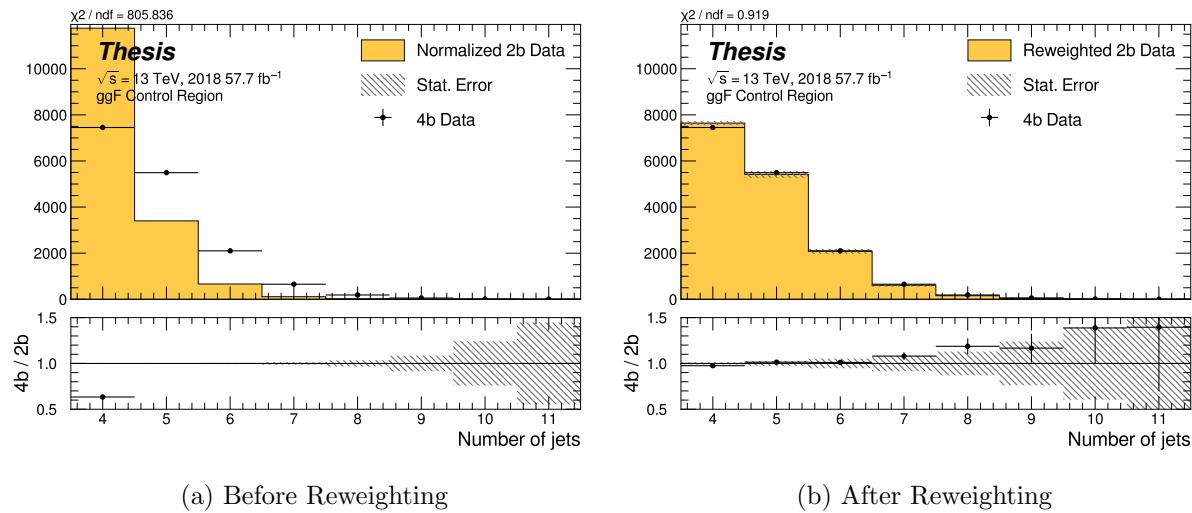


Figure 7.26: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

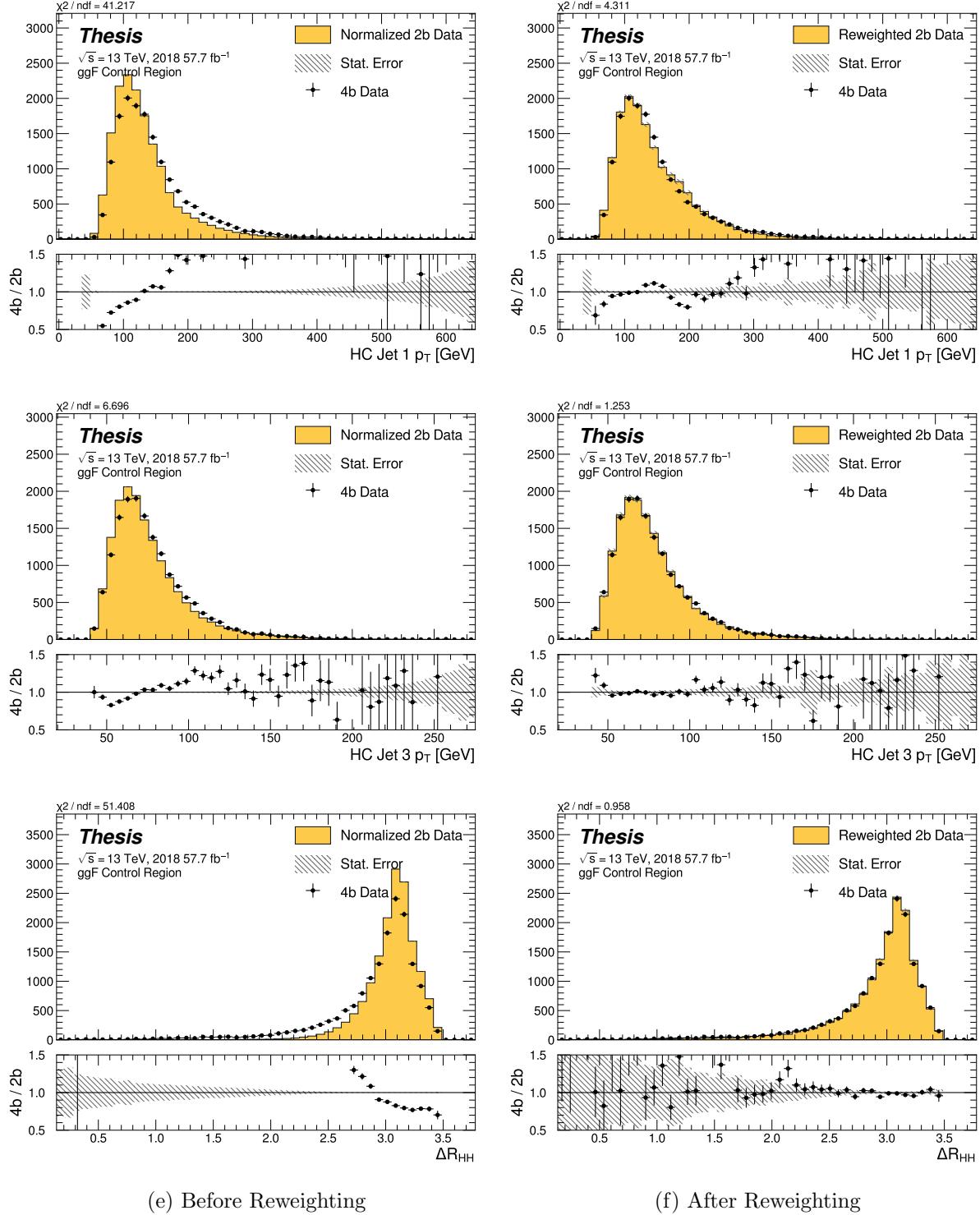


Figure 7.27: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

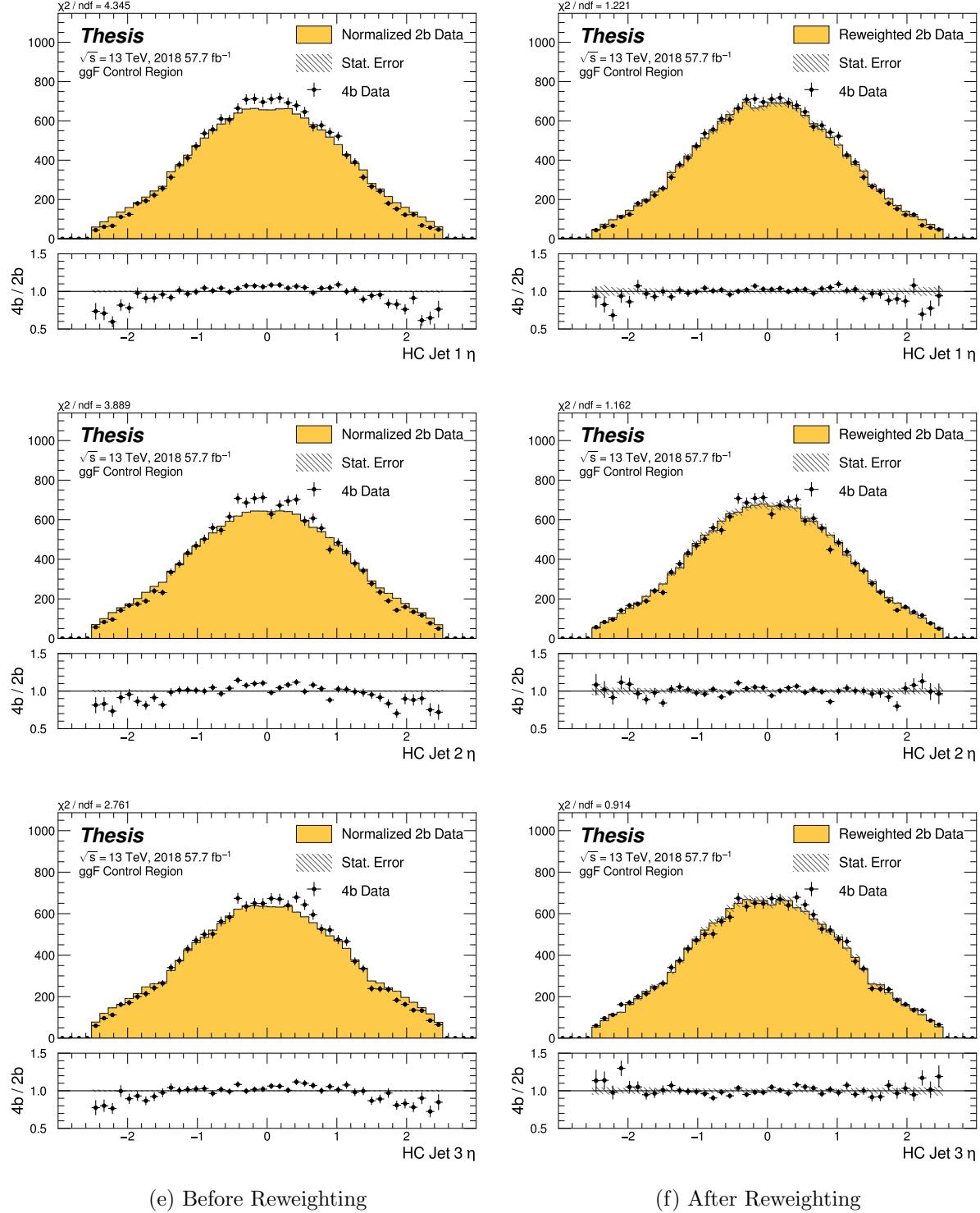


Figure 7.28: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Control Region.

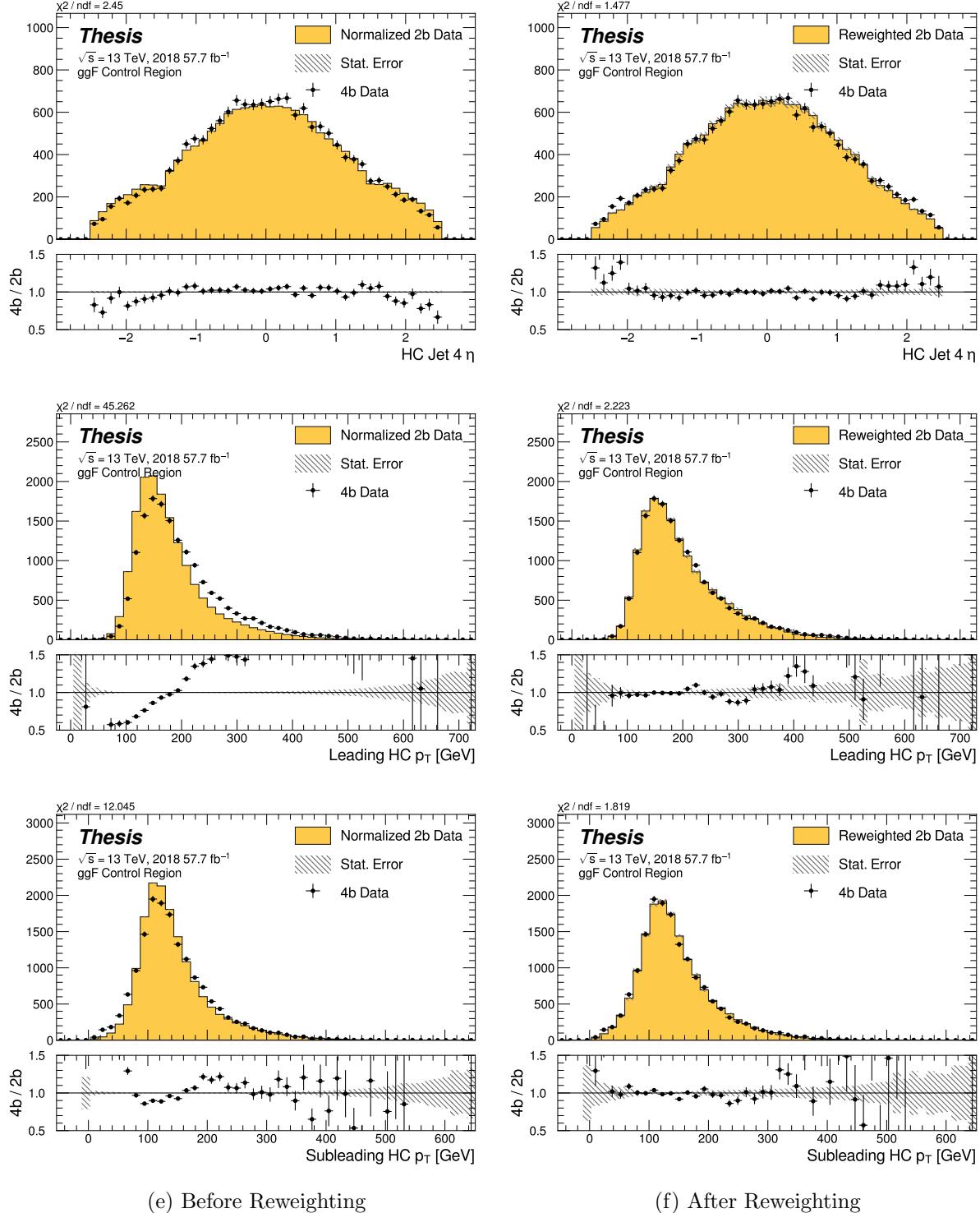


Figure 7.29: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

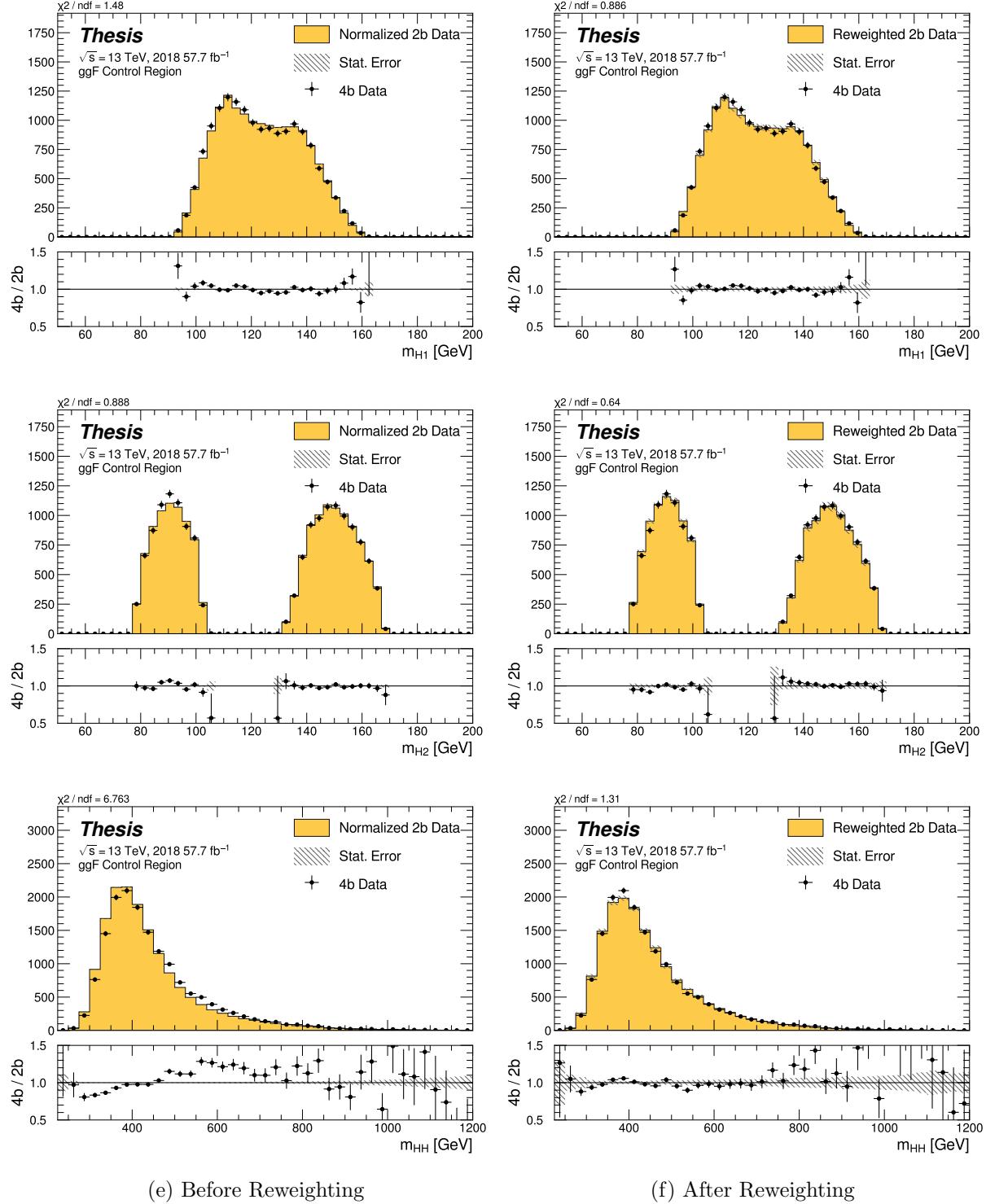


Figure 7.30: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Control Region.

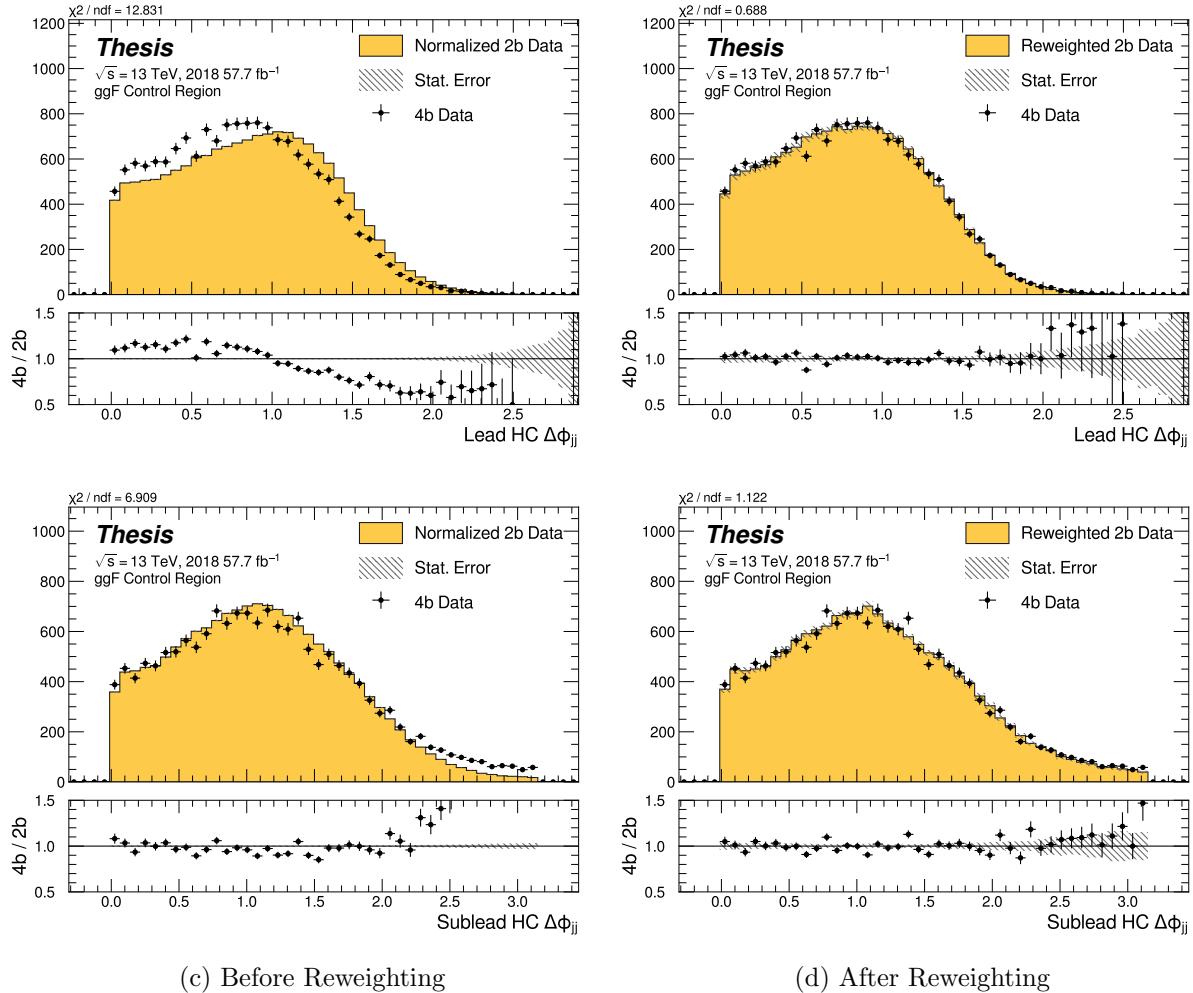


Figure 7.31: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Control Region.

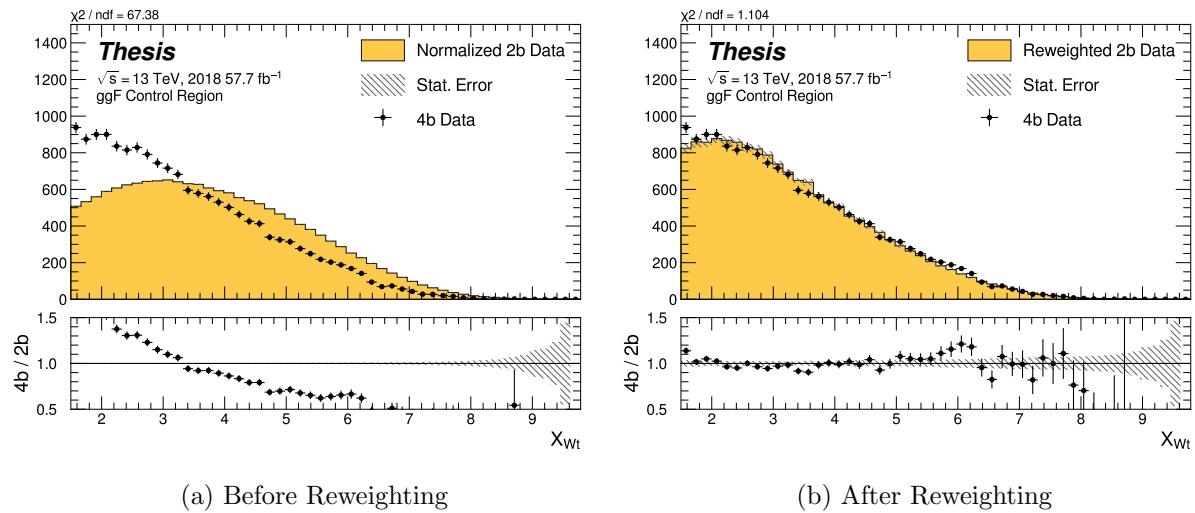


Figure 7.32: **Non-resonant Search (4b):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.

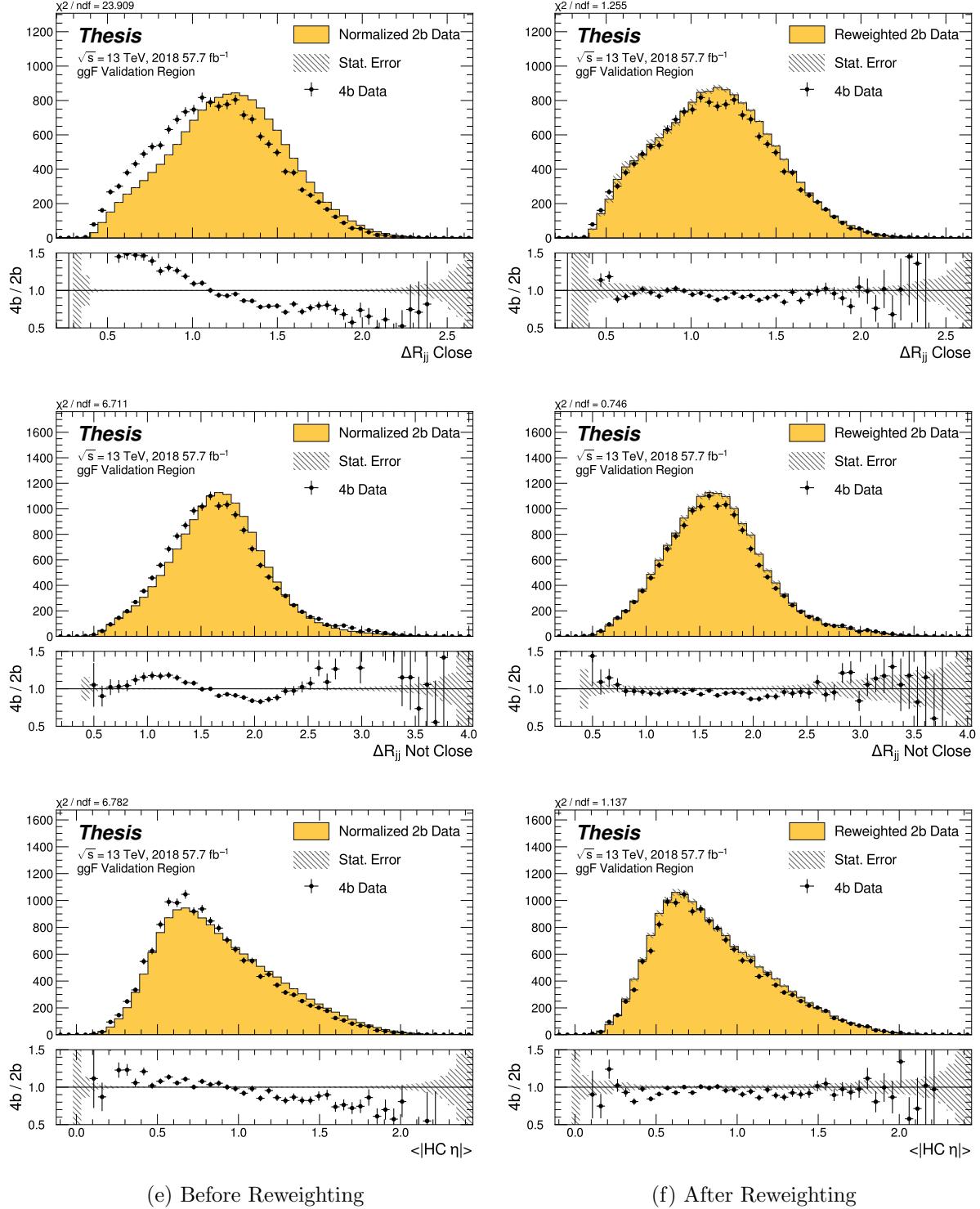


Figure 7.33: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 4b Validation Region.

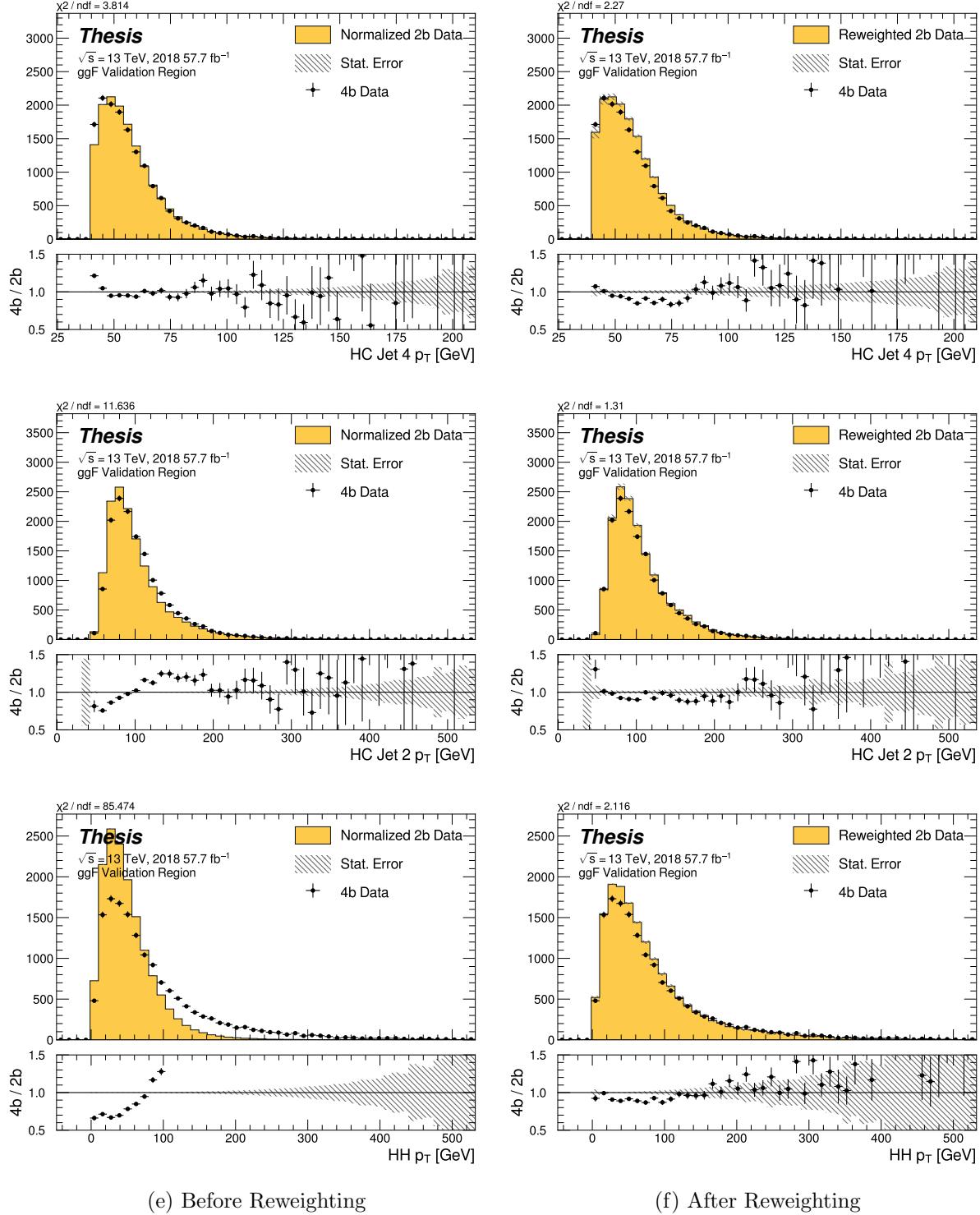


Figure 7.34: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

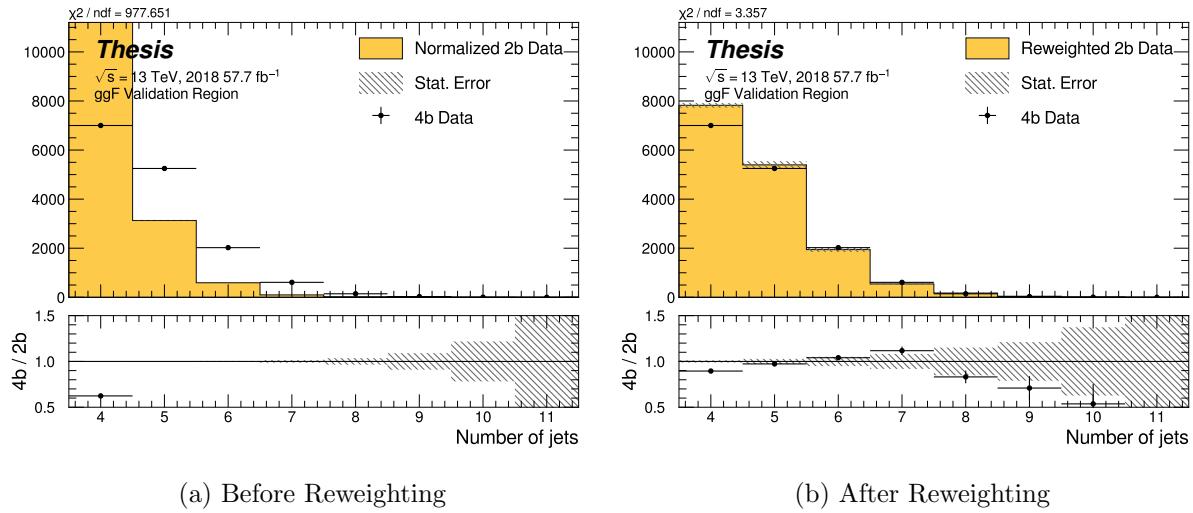


Figure 7.35: **Non-resonant Search (4b):** Distributions of the number of jets before and after CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

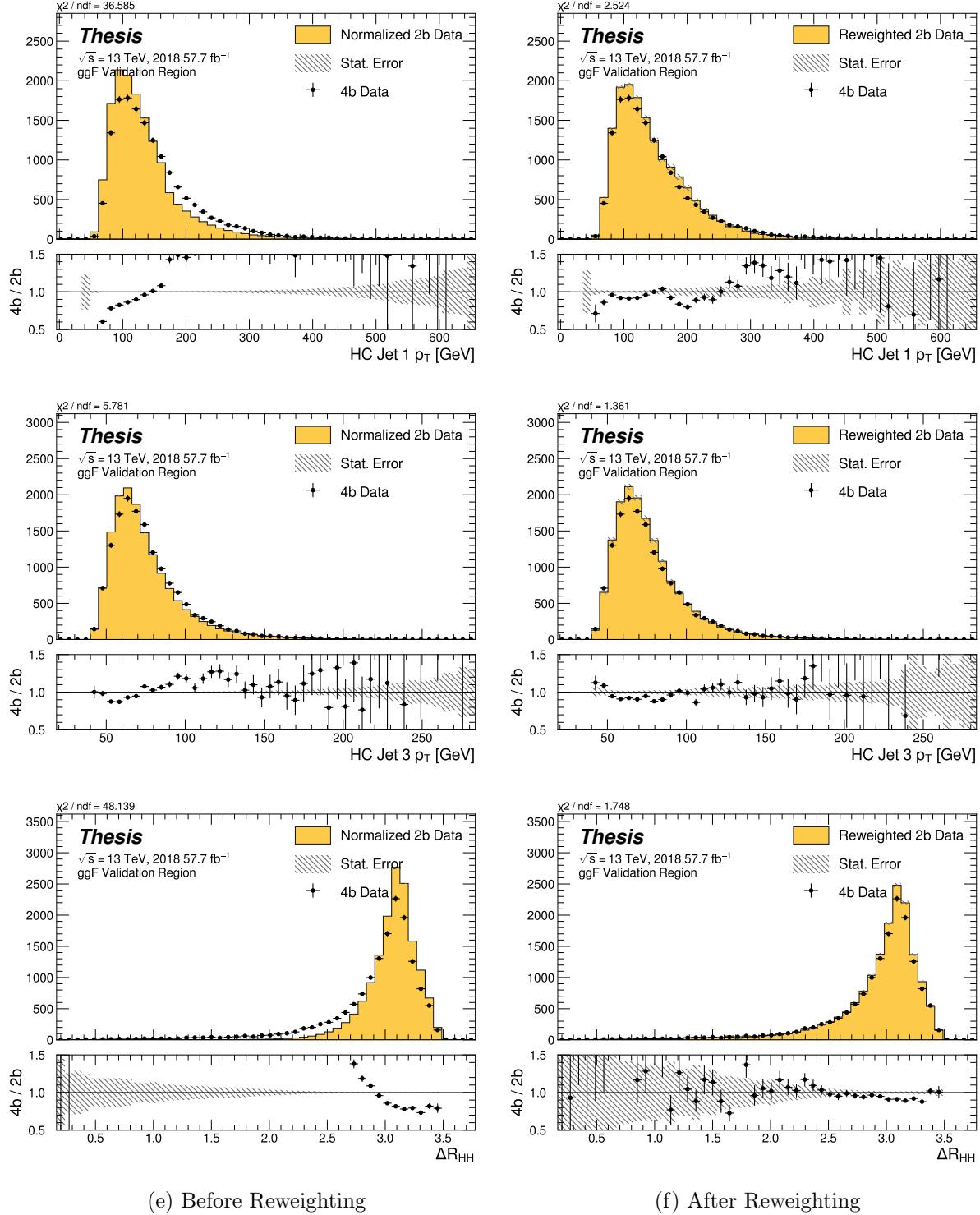


Figure 7.36: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

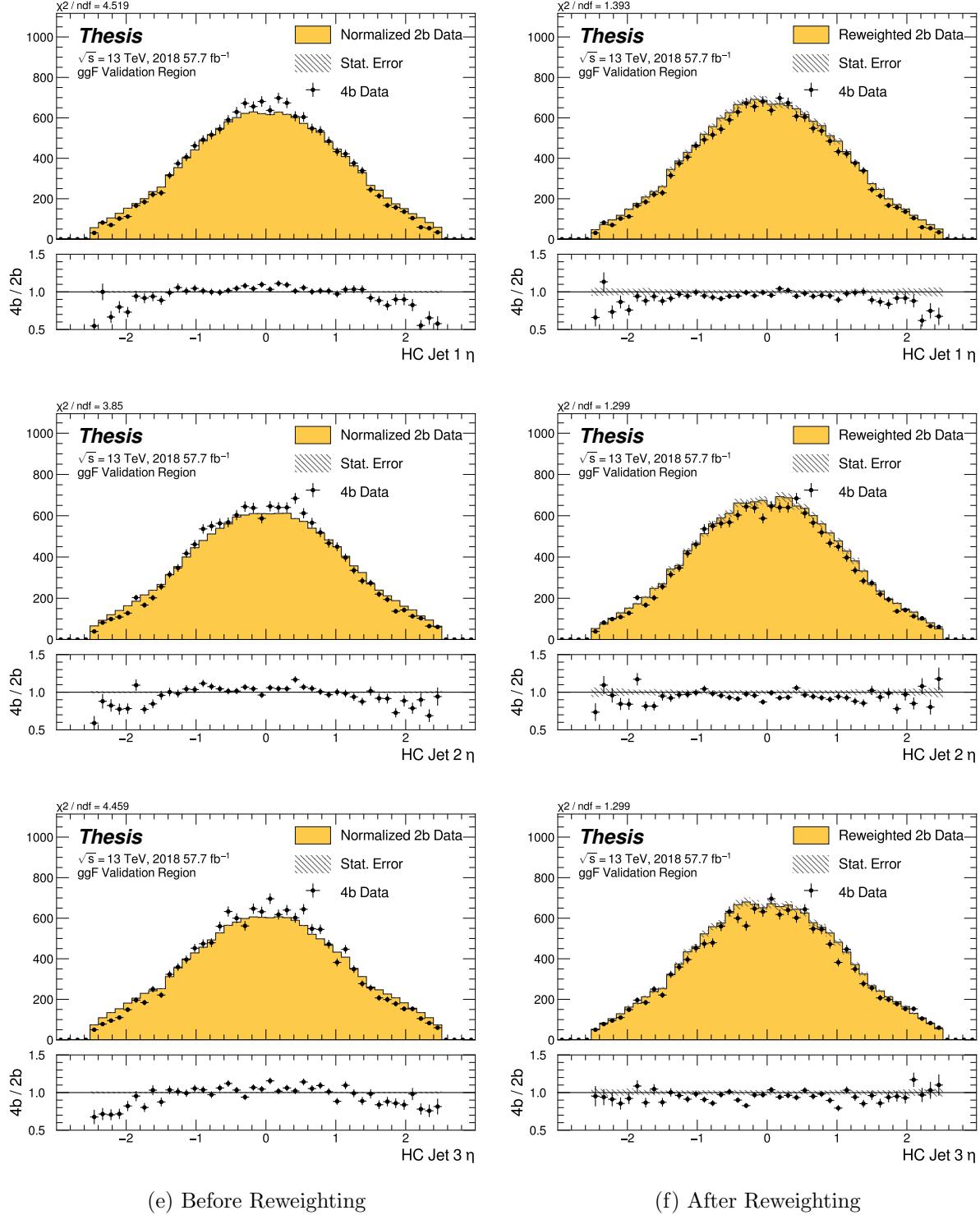


Figure 7.37: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 4b Validation Region.

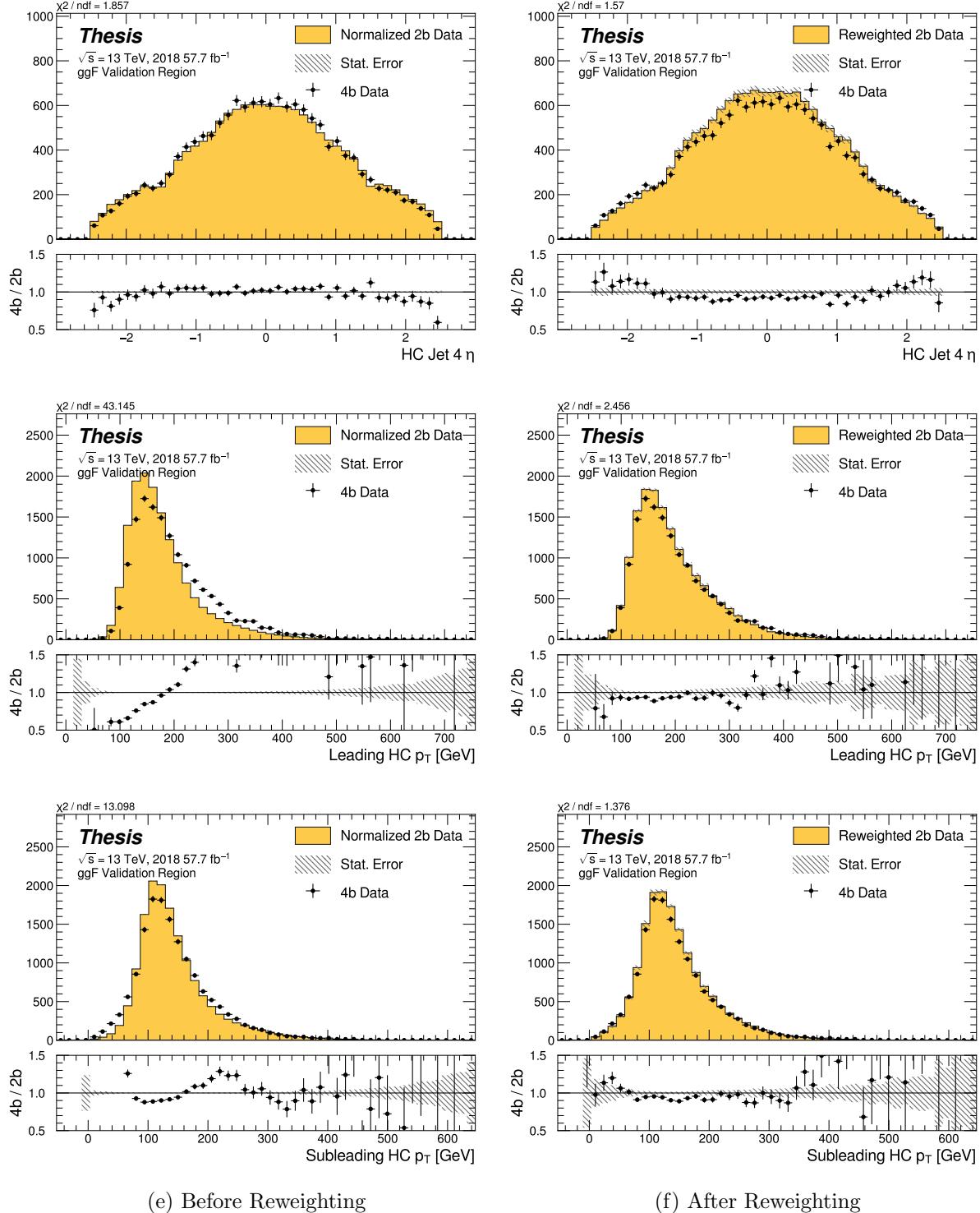


Figure 7.38: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

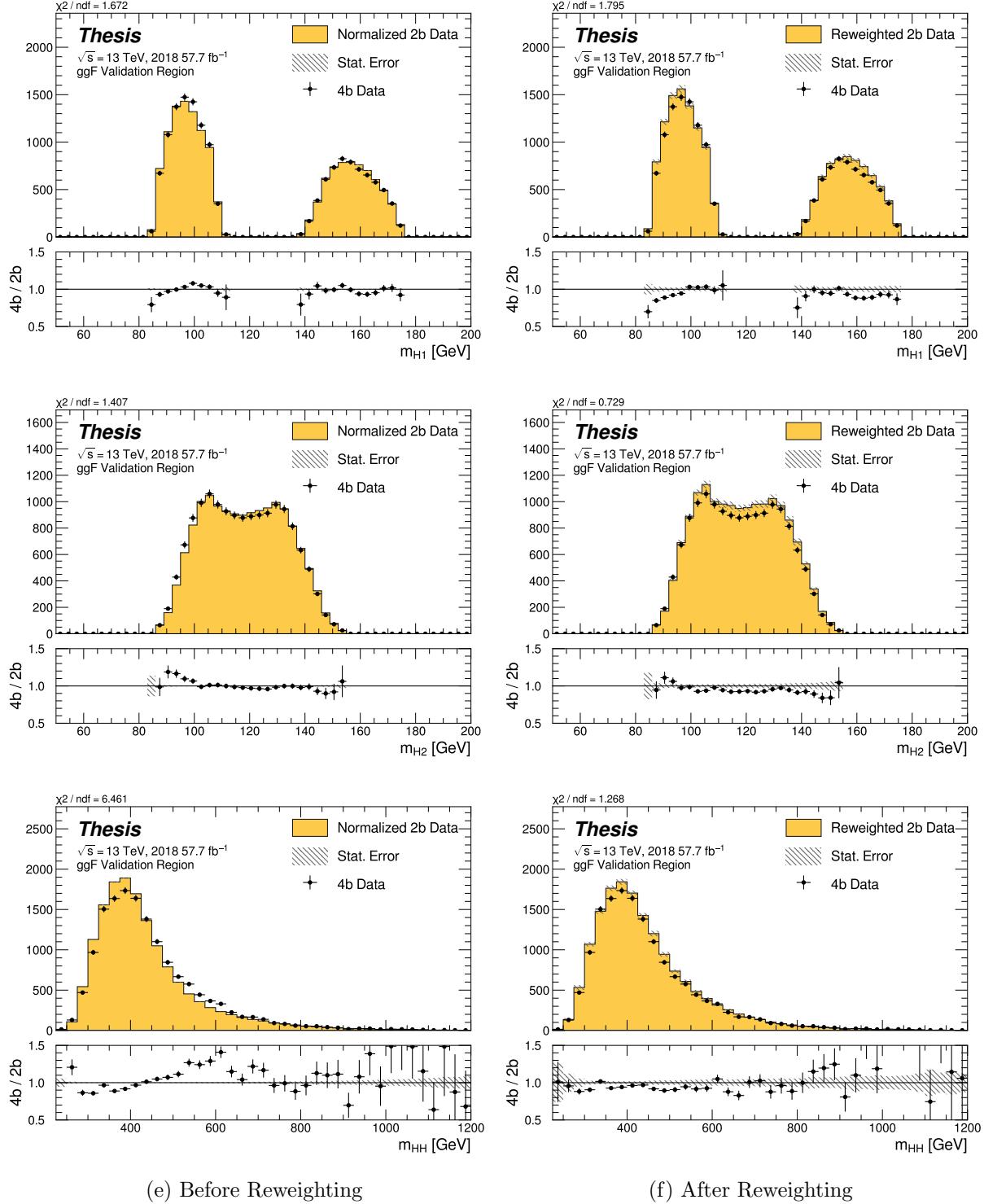


Figure 7.39: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 4b Validation Region.

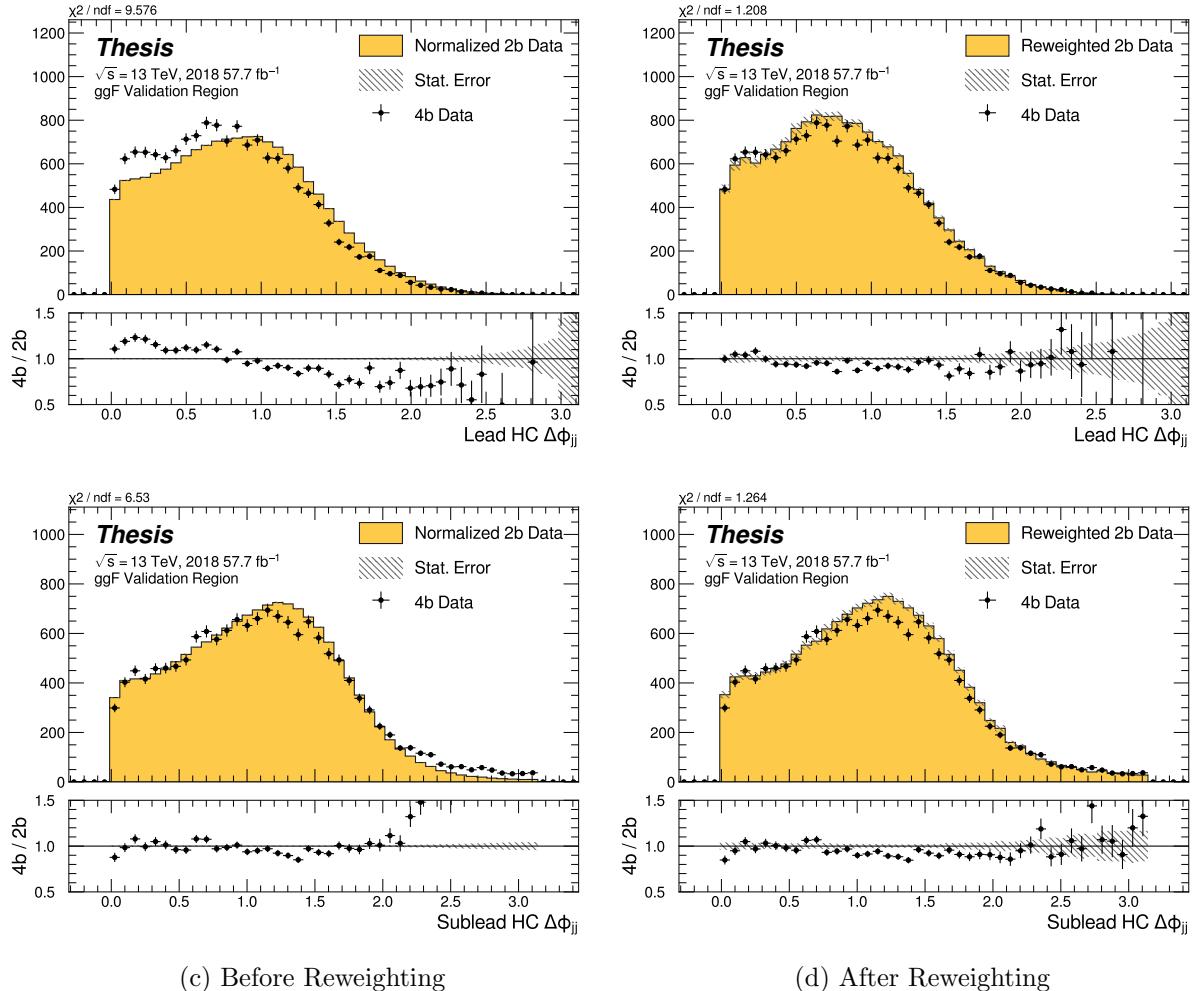


Figure 7.40: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 4b Validation Region.

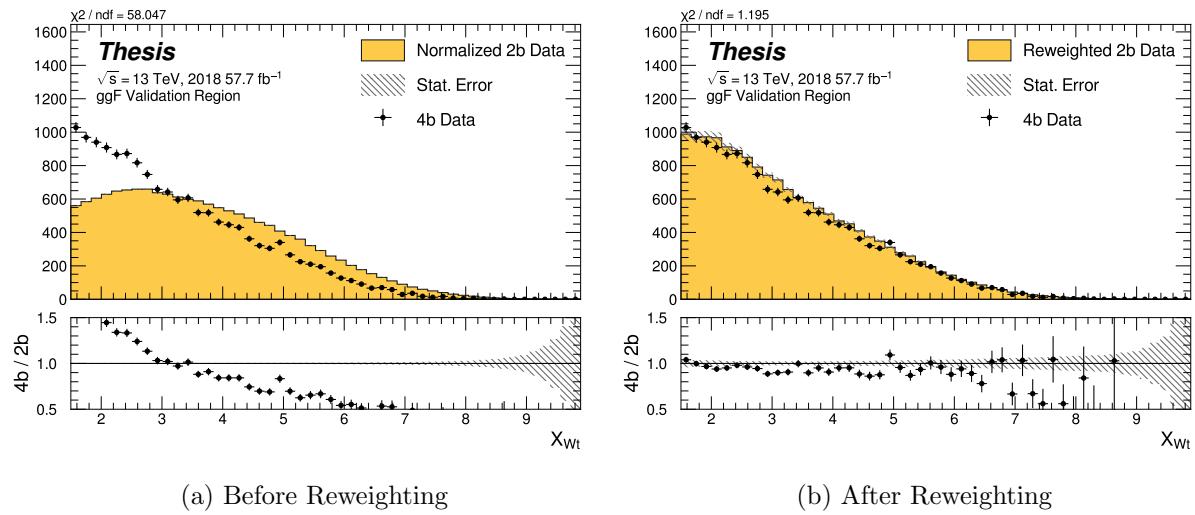


Figure 7.41: **Non-resonant Search (4b):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.

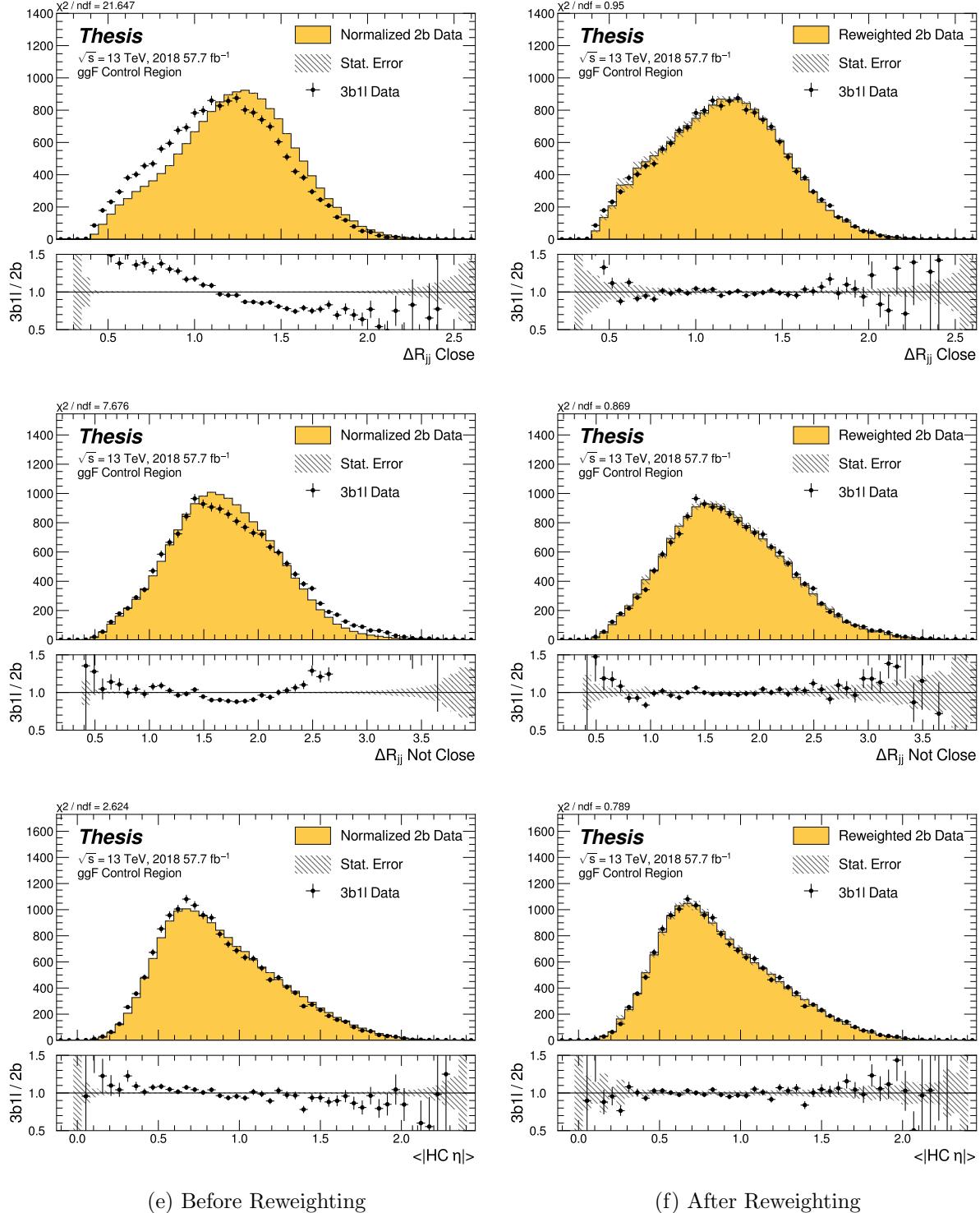


Figure 7.42: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Control Region.

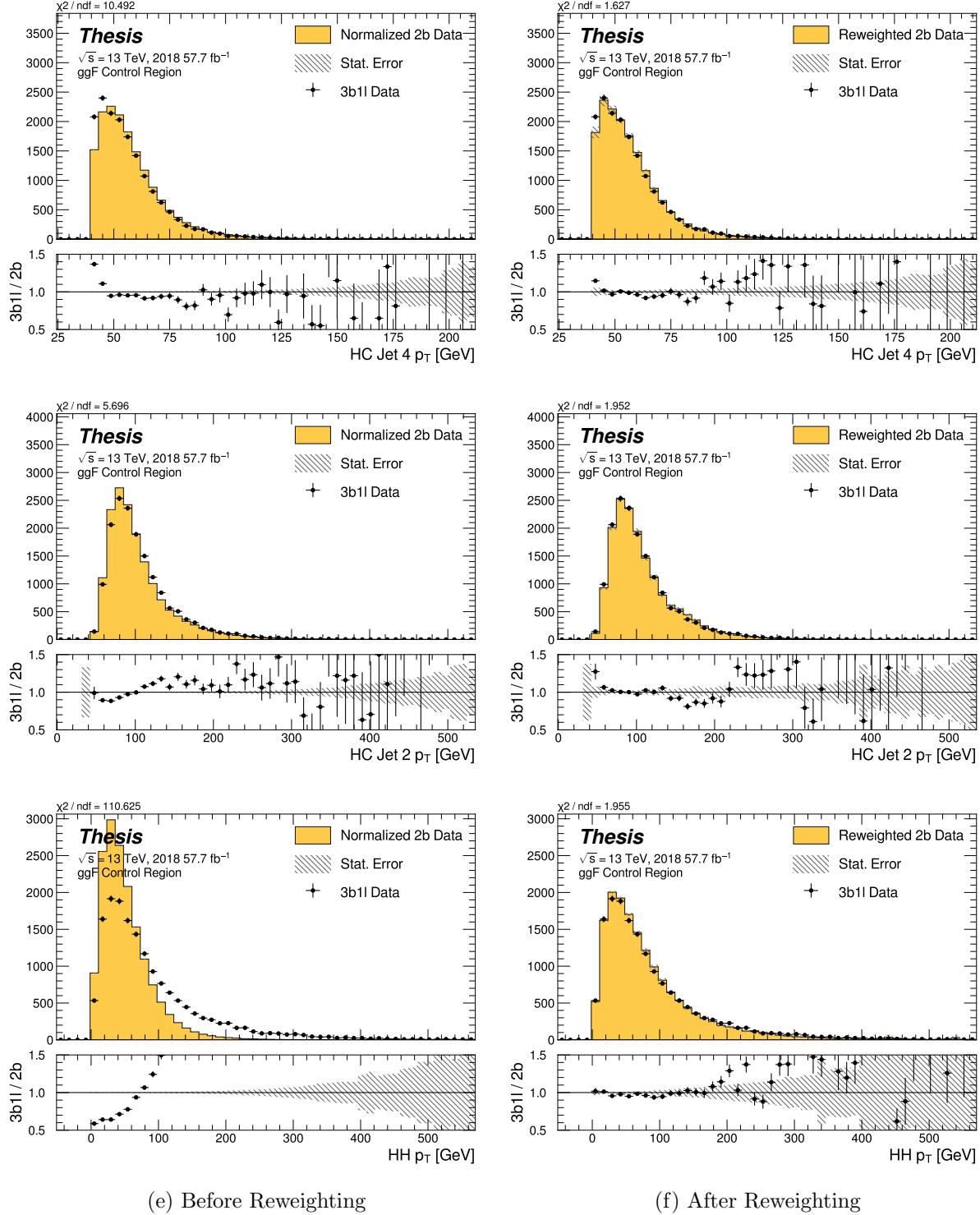


Figure 7.43: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

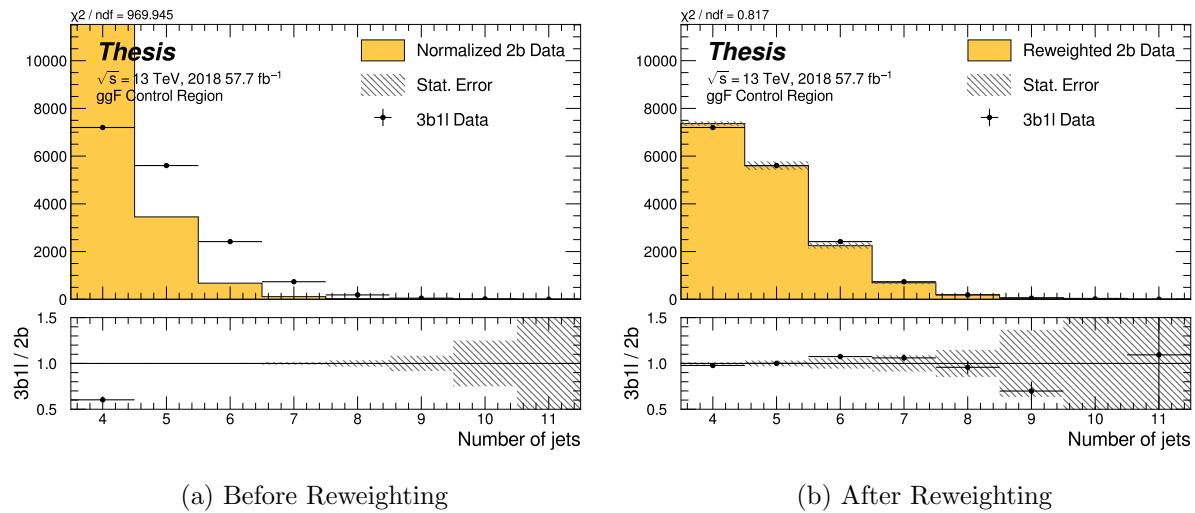


Figure 7.44: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

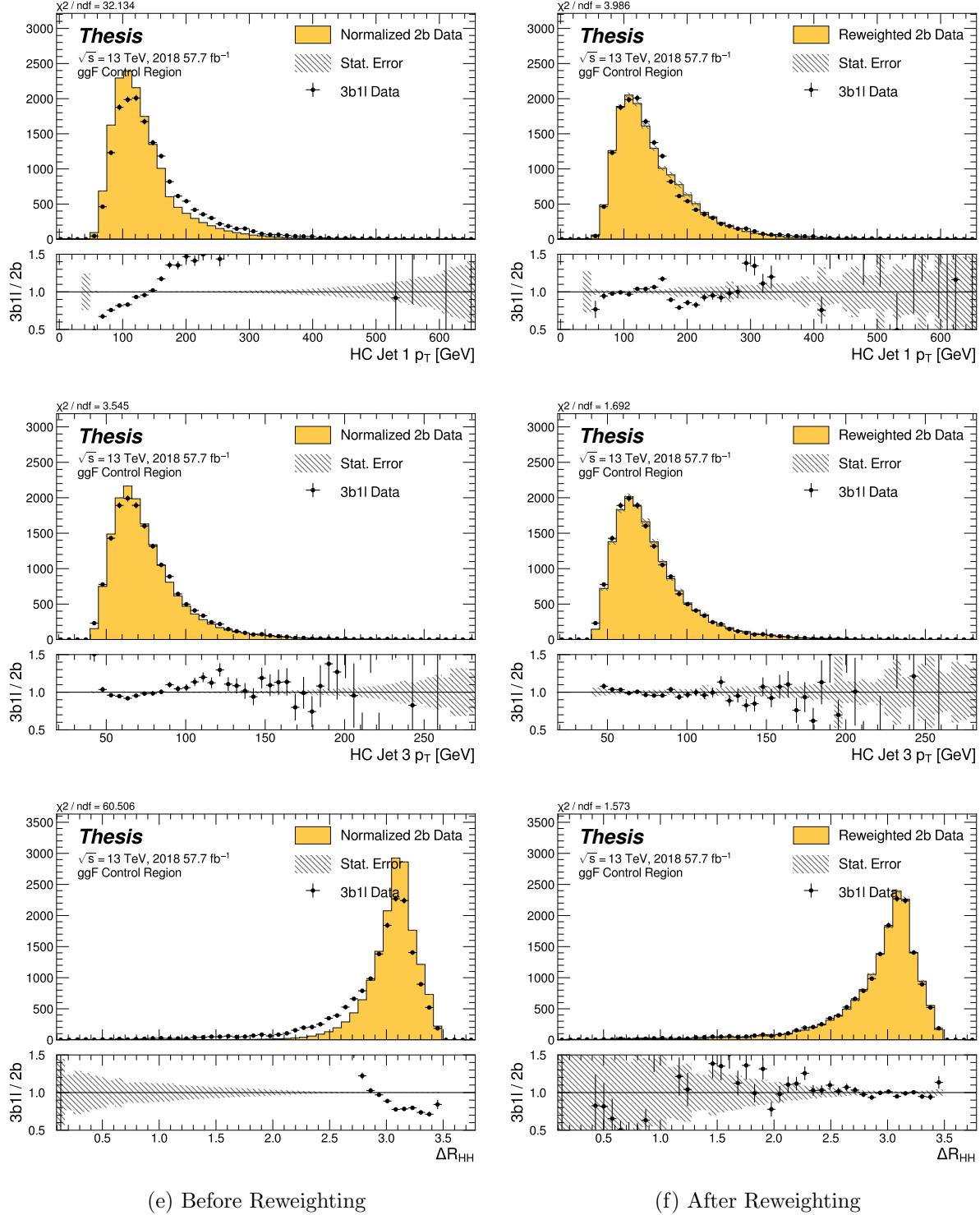


Figure 7.45: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

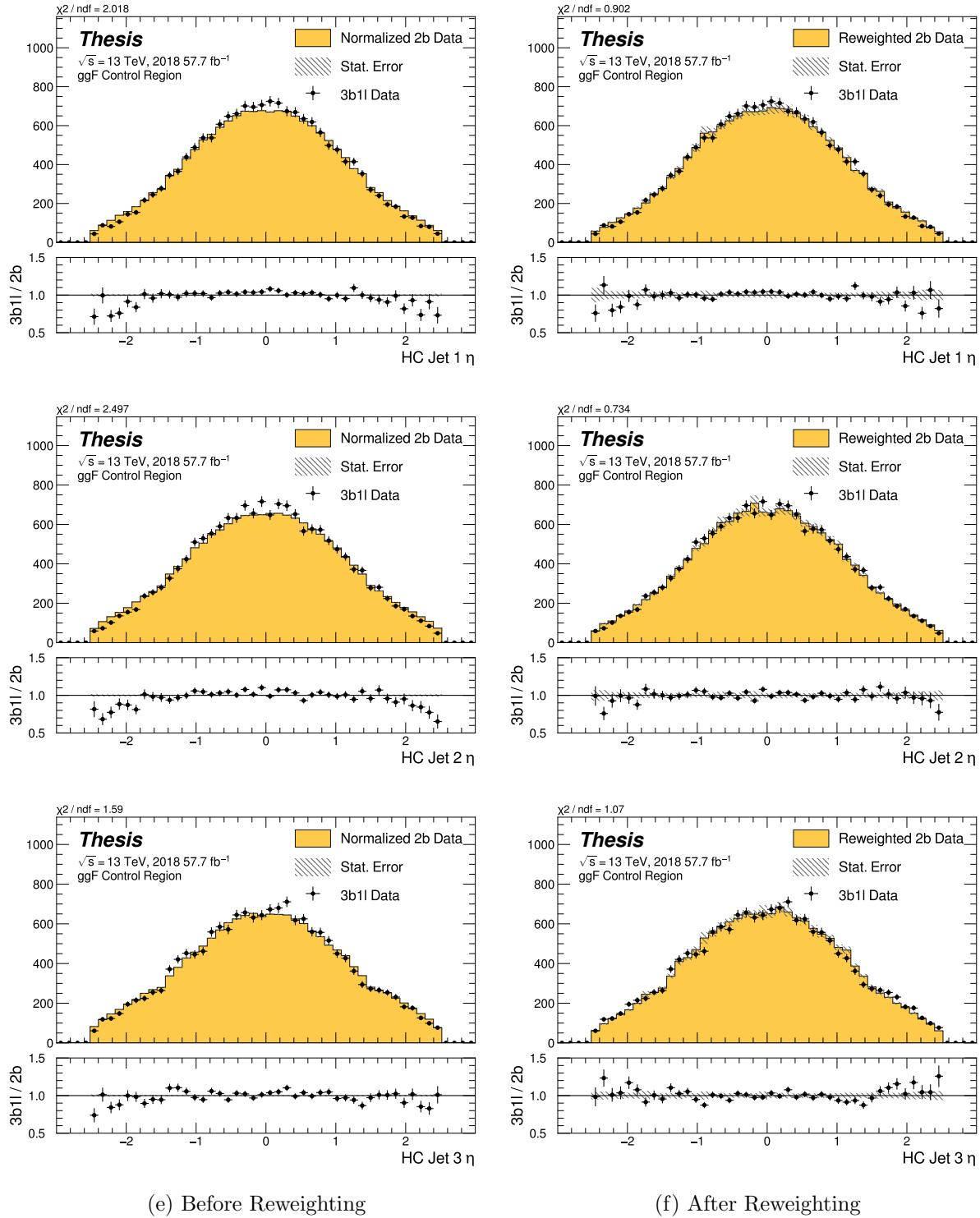


Figure 7.46: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Control Region.

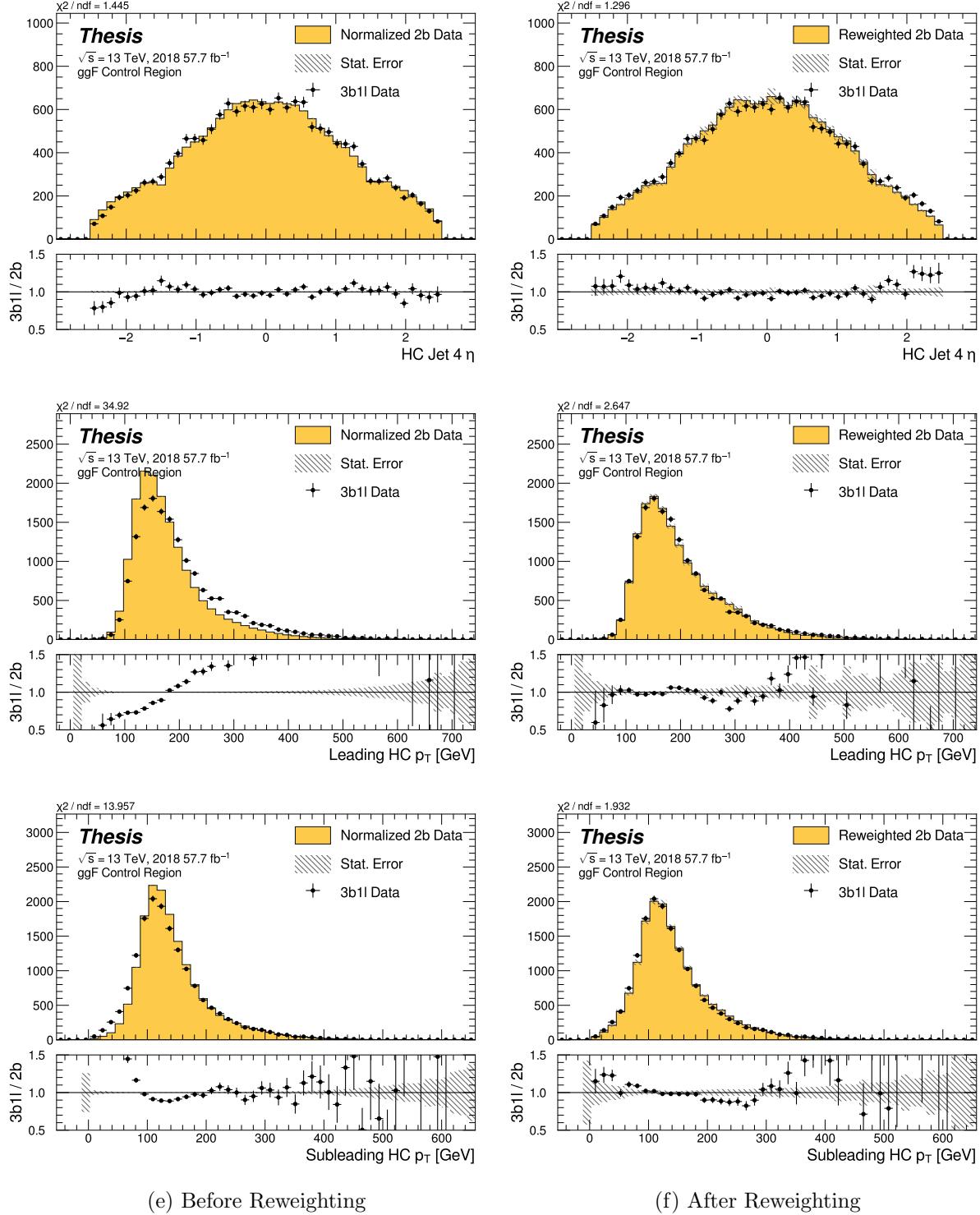


Figure 7.47: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

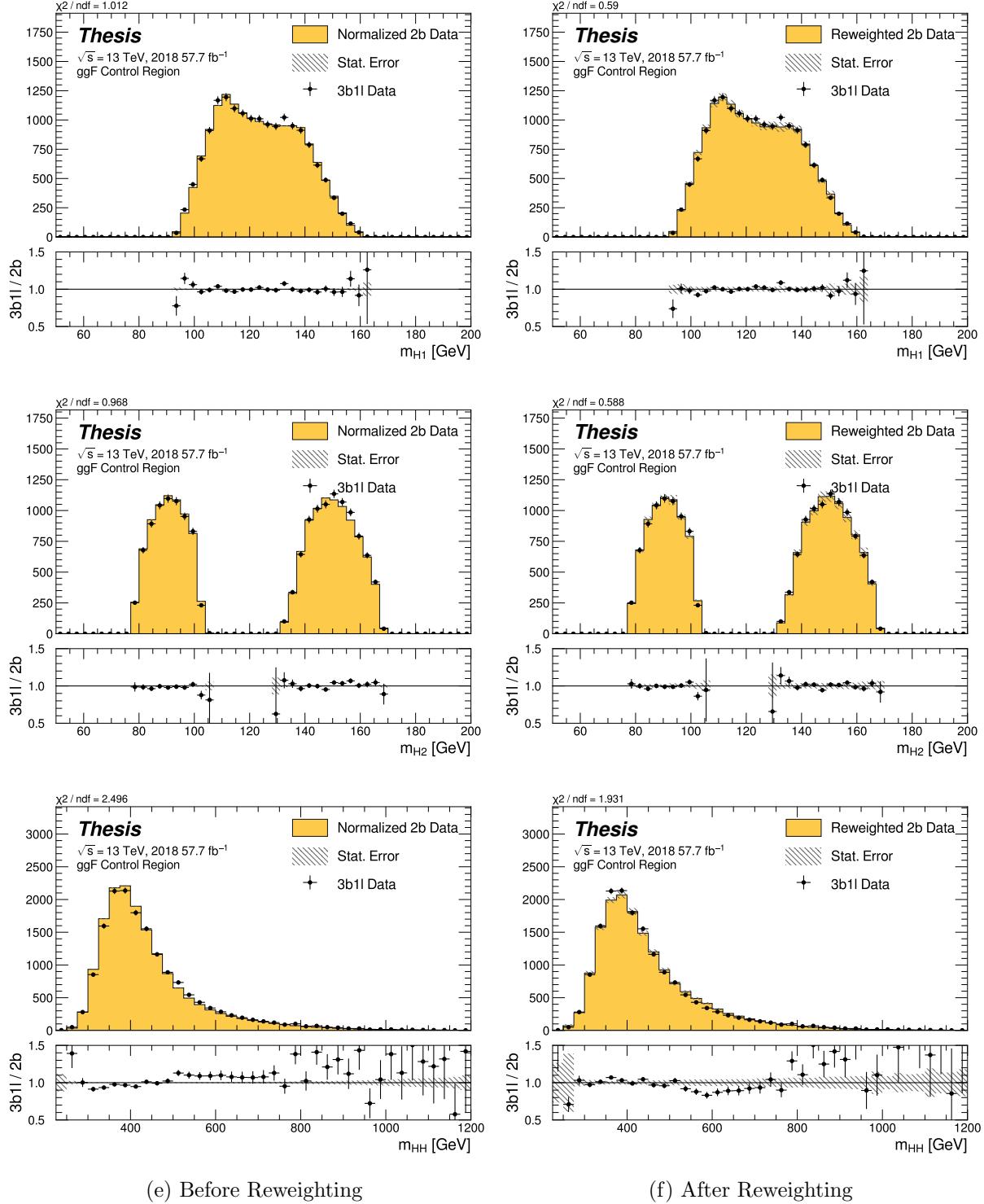


Figure 7.48: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Control Region.

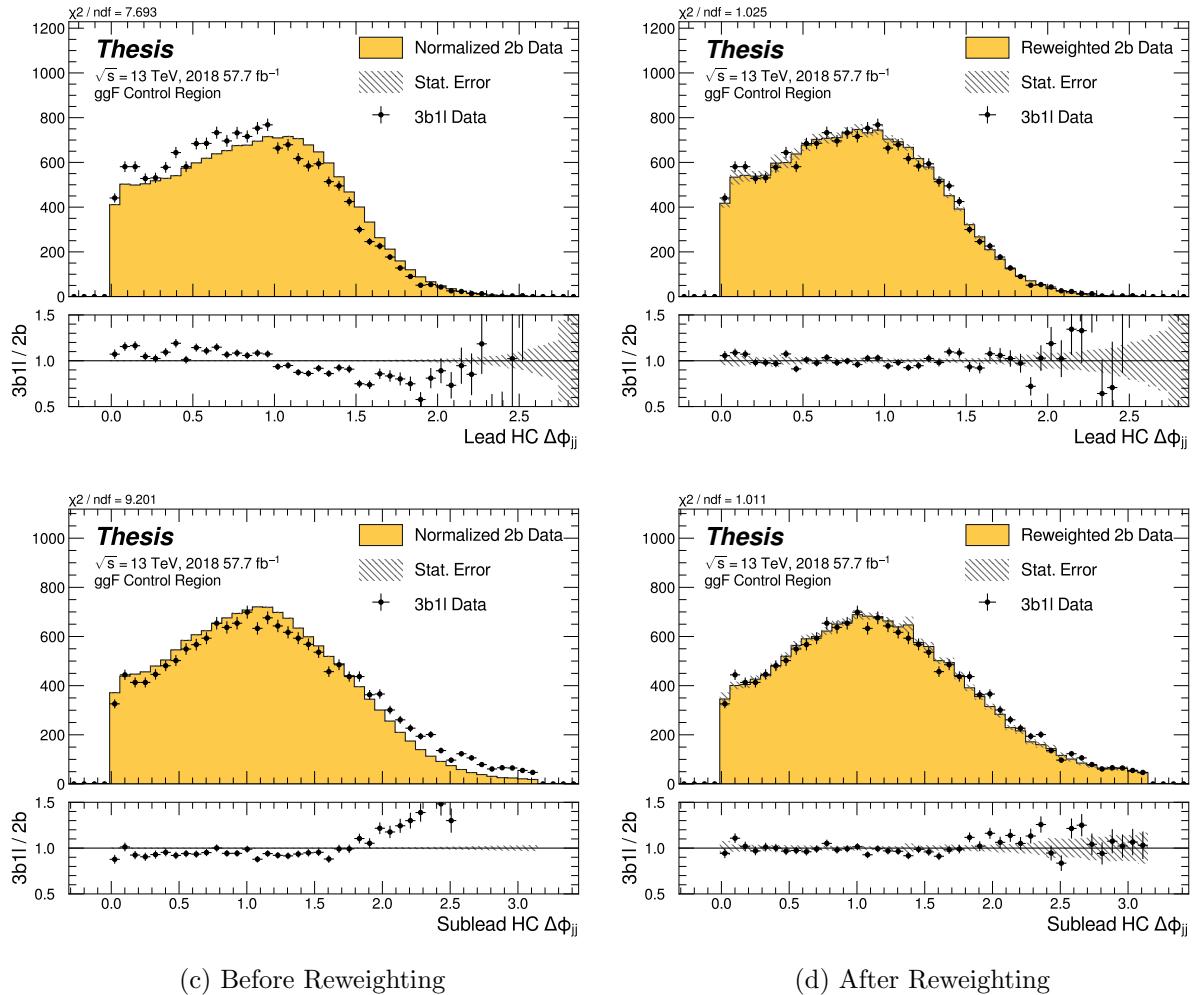


Figure 7.49: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Control Region.

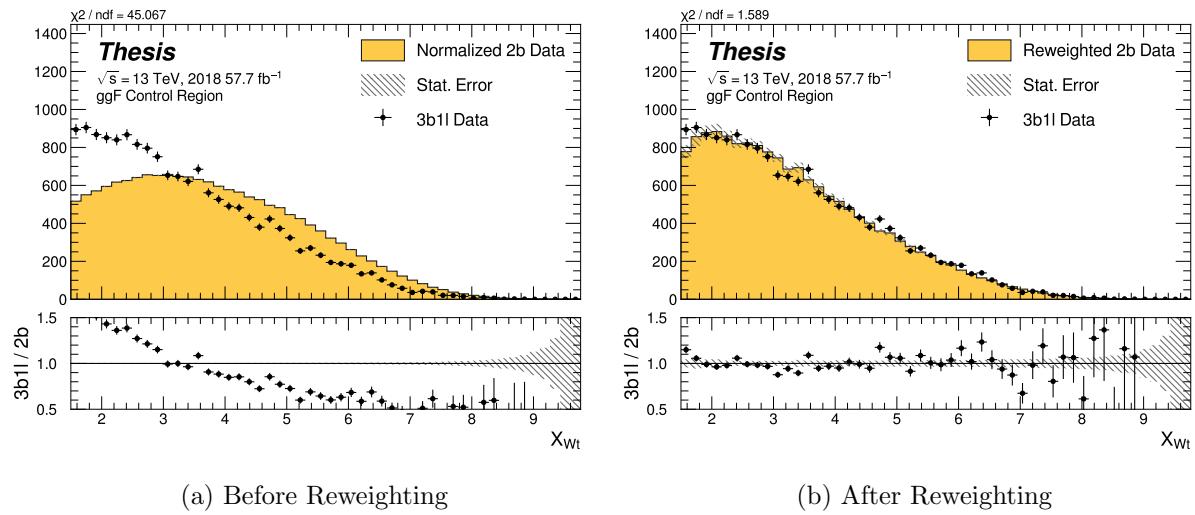


Figure 7.50: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Control Region. Reweighting is done after the cut on this variable is applied.

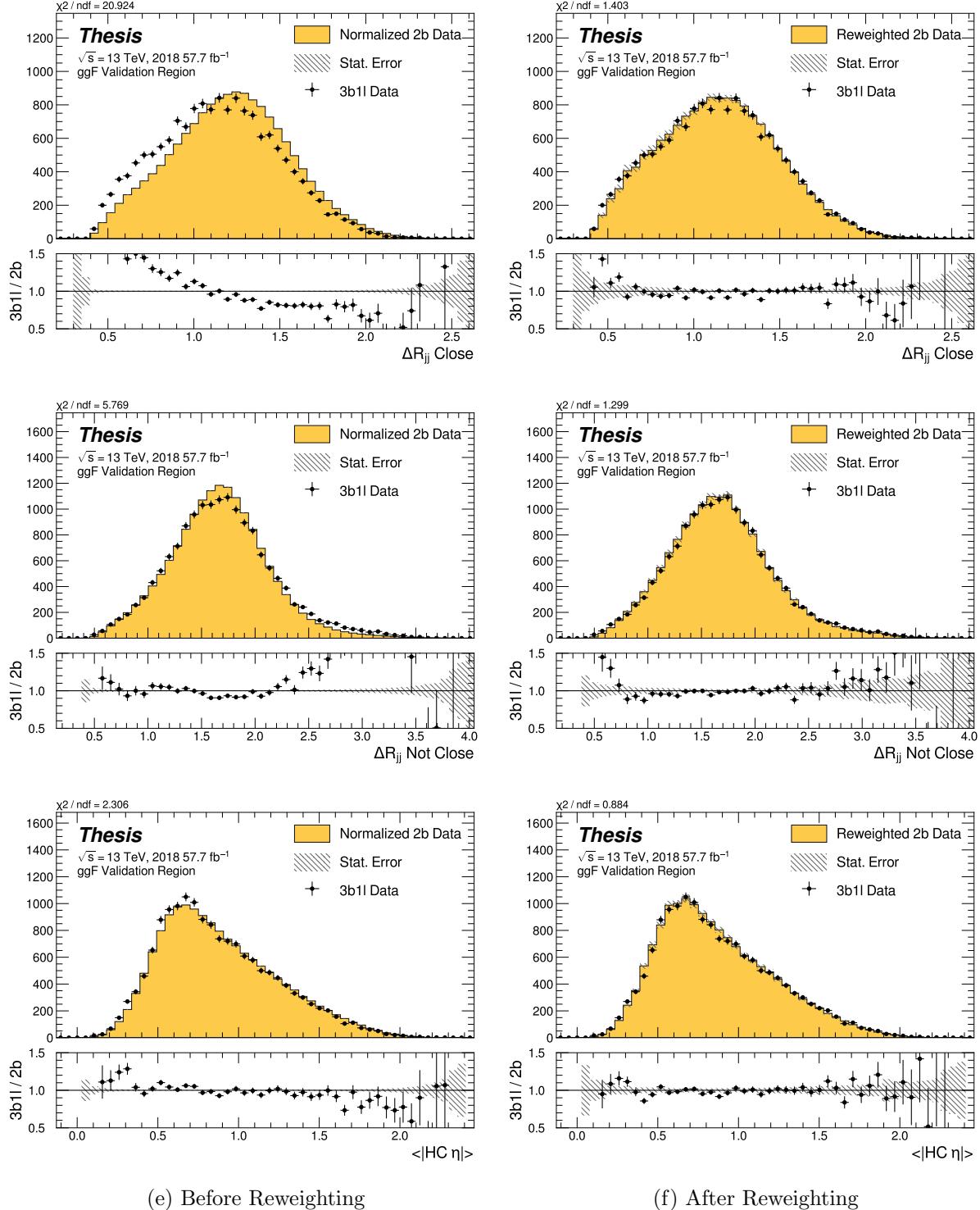


Figure 7.51: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before and after CR derived reweighting for the 2018 3b1l Validation Region.

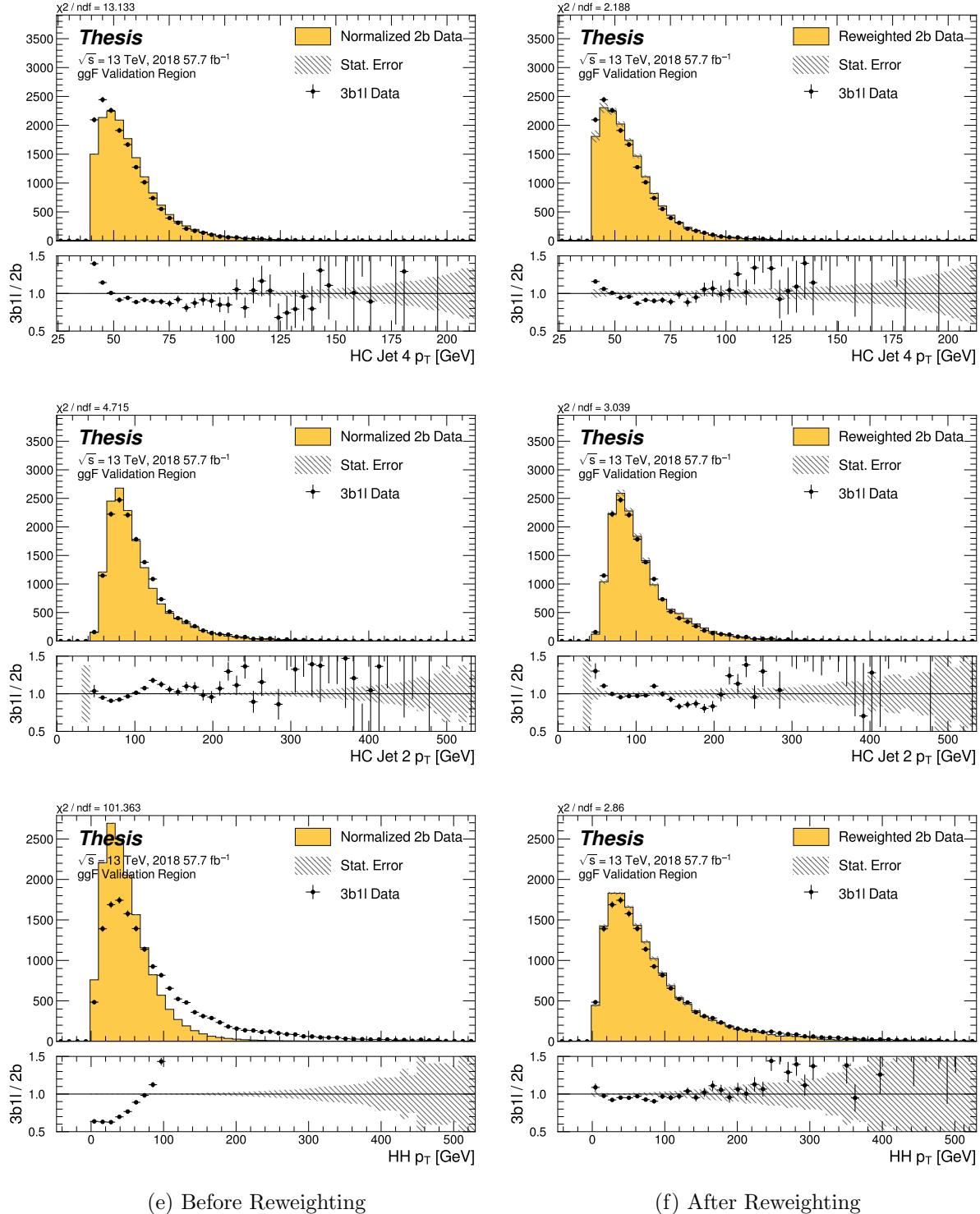


Figure 7.52: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

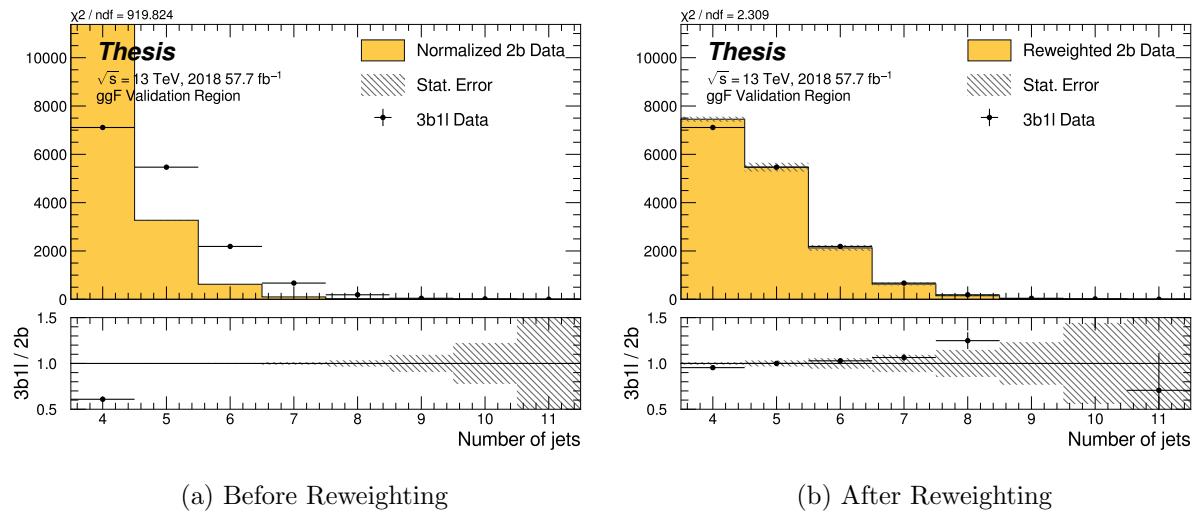


Figure 7.53: **Non-resonant Search (3b1l):** Distributions of the number of jets before and after CR derived reweighting for the 2018 3b1l Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

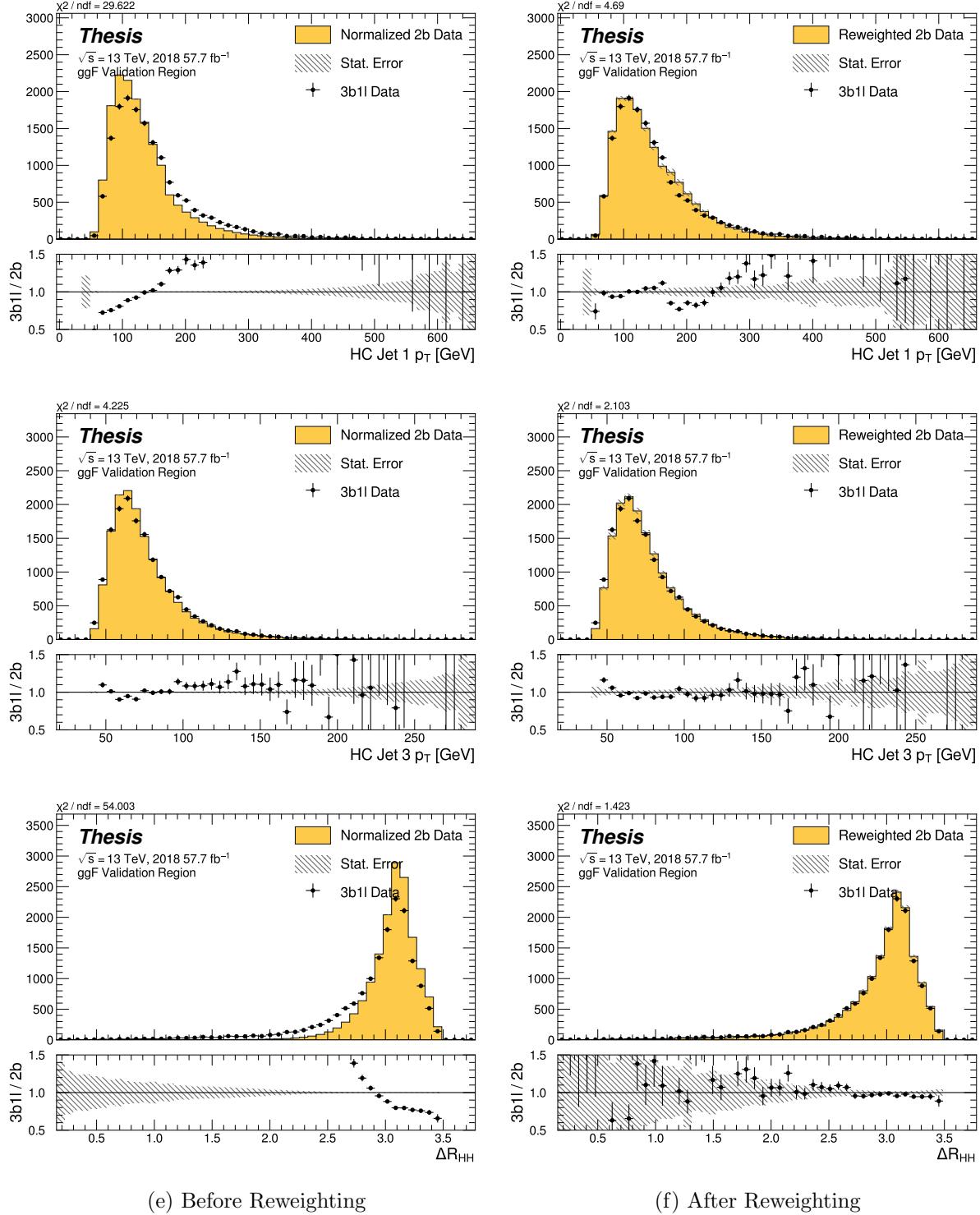


Figure 7.54: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

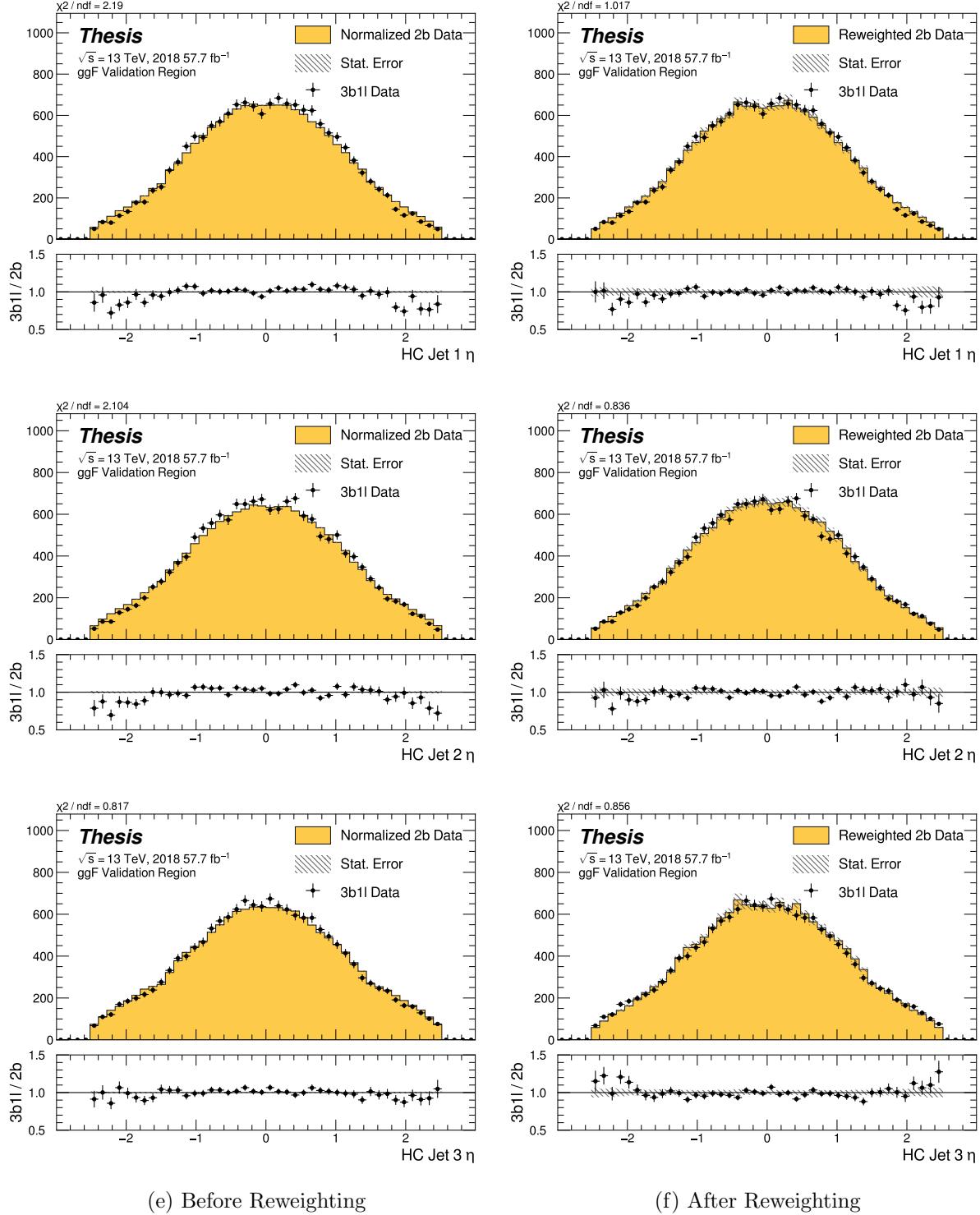


Figure 7.55: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before and after CR derived reweighting for the 2018 3b1l Validation Region.

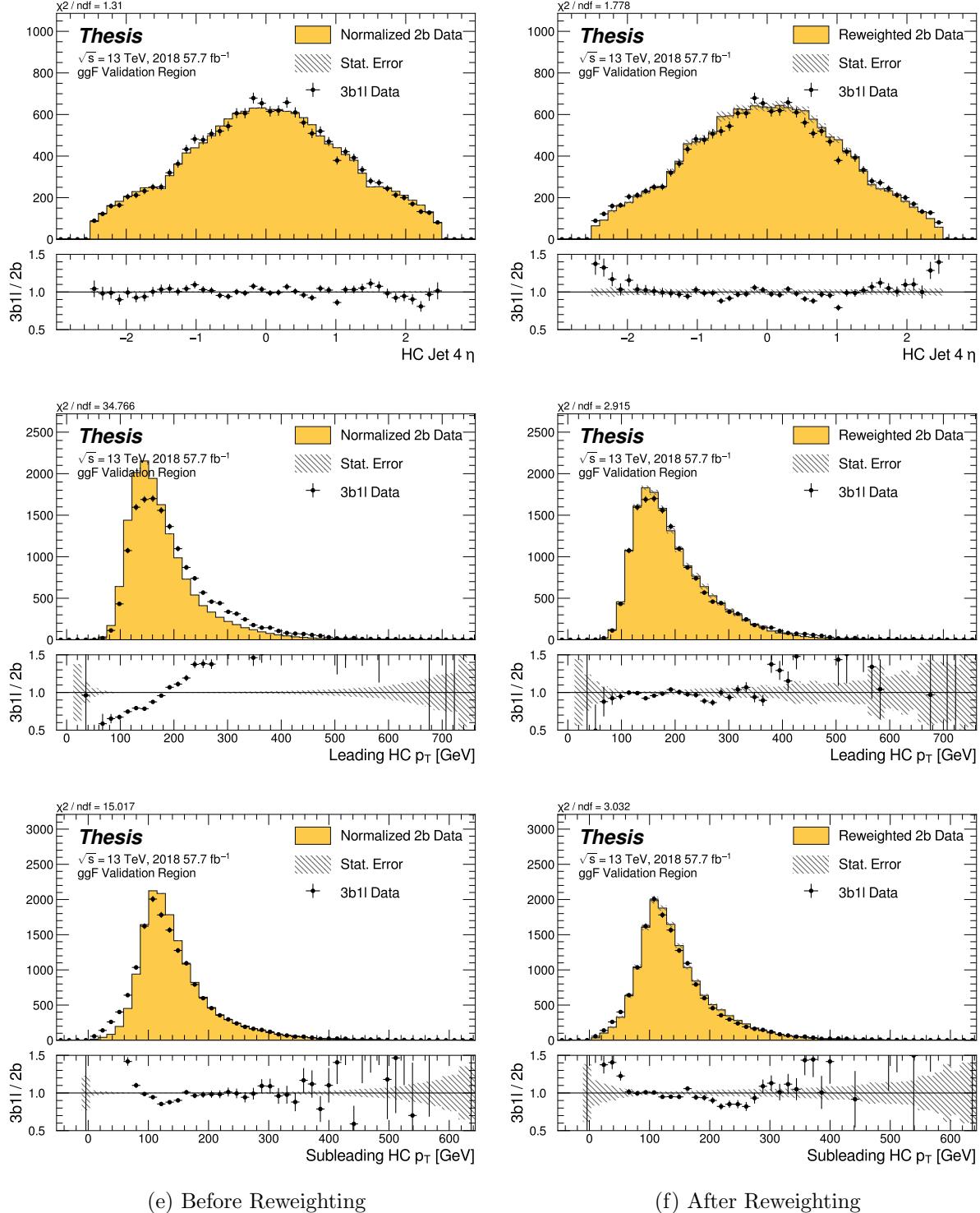


Figure 7.56: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

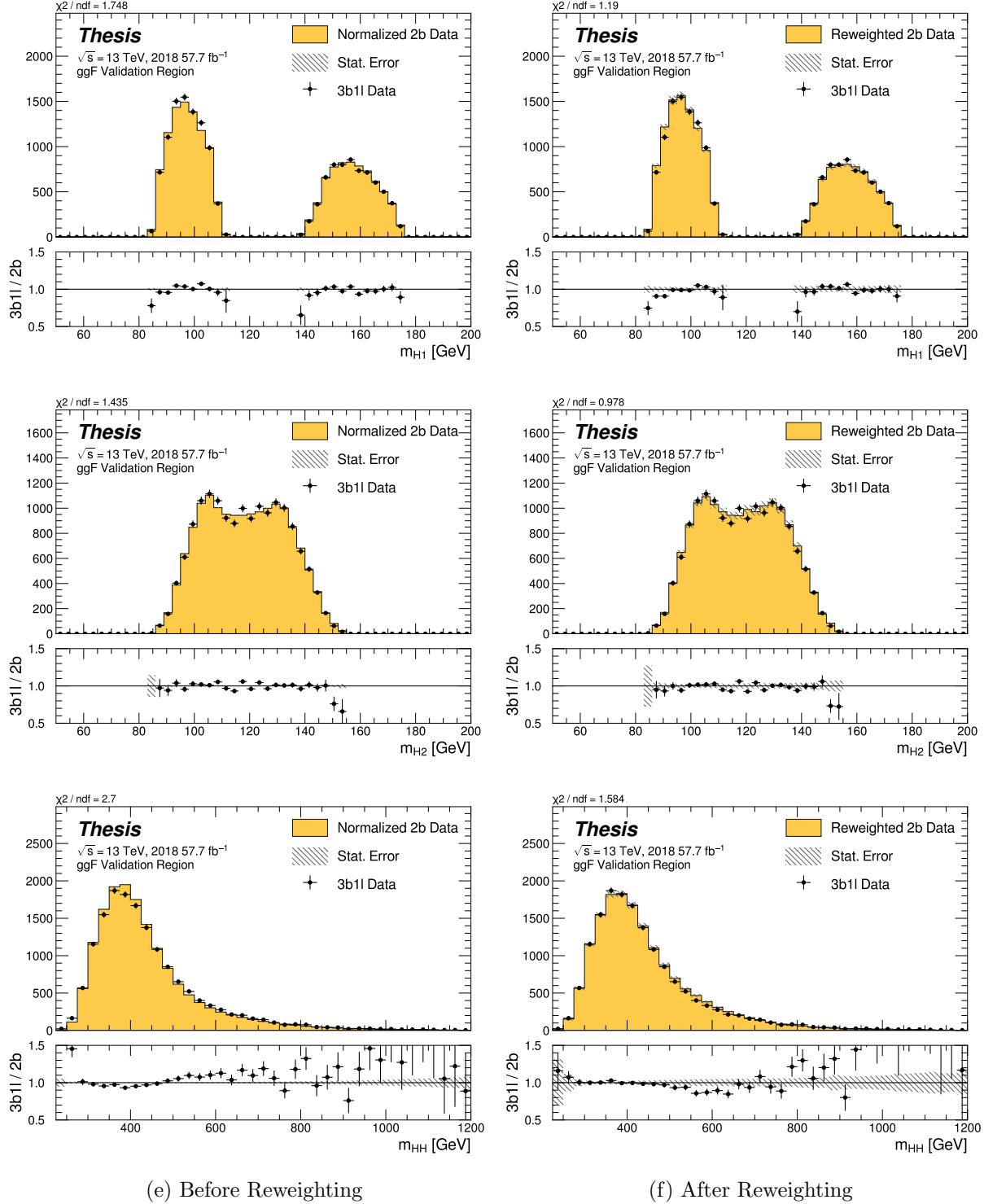


Figure 7.57: **Non-resonant Search (3b1l):** Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before and after CR derived reweighting for the 2018 3b1l Validation Region.

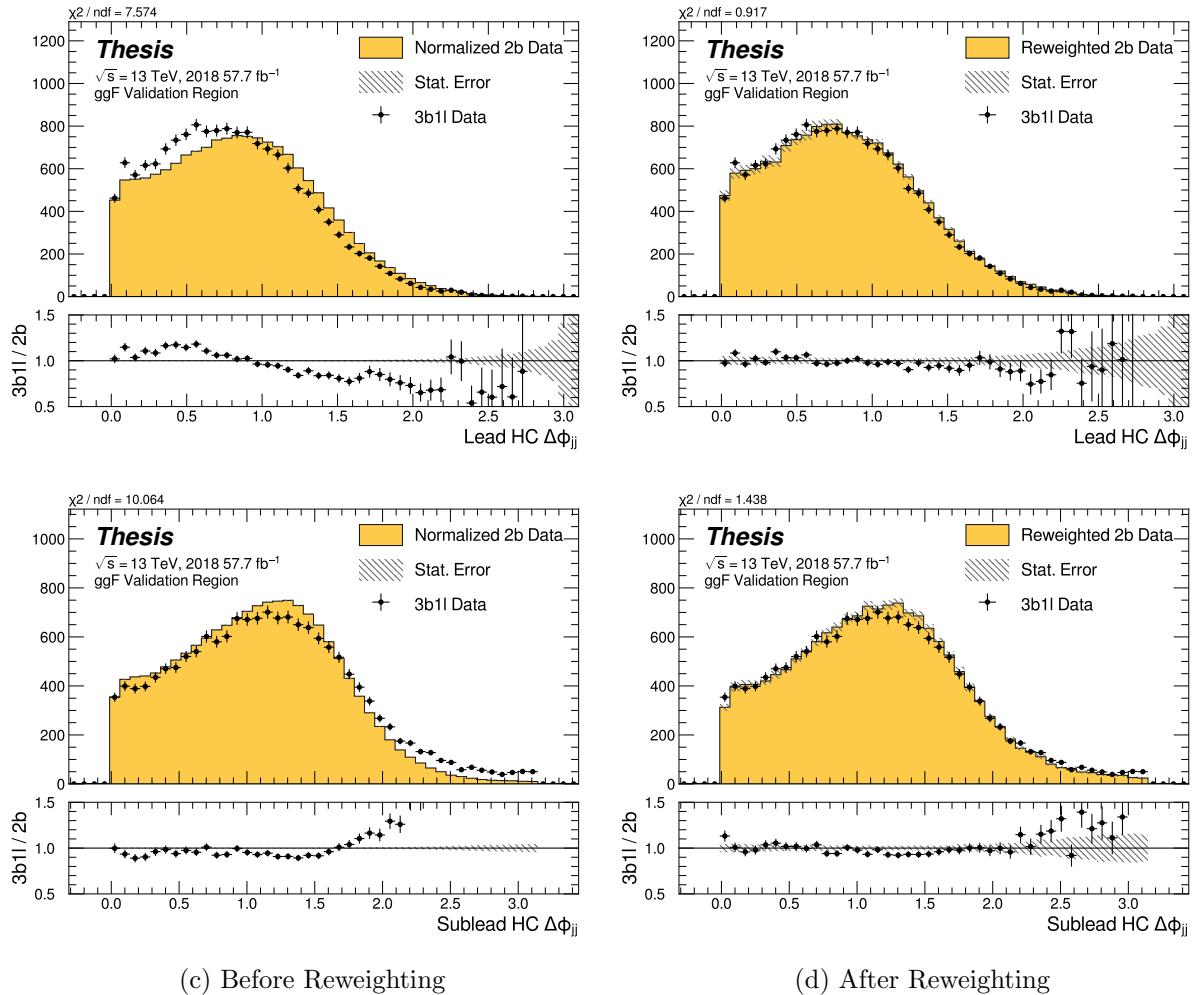


Figure 7.58: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before and after CR derived reweighting for the 2018 3b1l Validation Region.

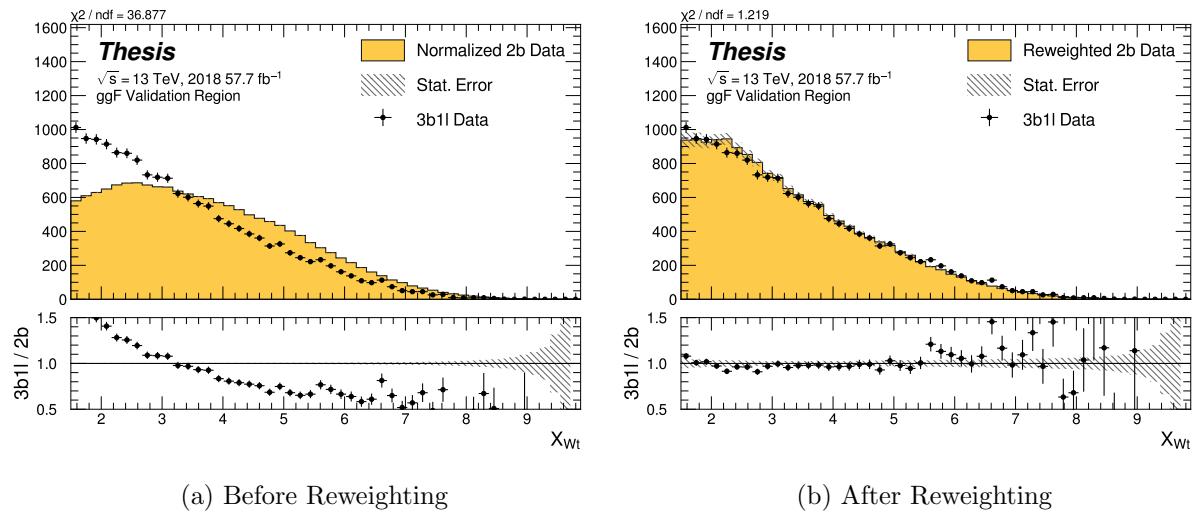


Figure 7.59: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before and after CR derived reweighting for the 2018 3b1l Validation Region. Reweighting is done after the cut on this variable is applied.

2106 **7.7 Uncertainties**

2107 A variety of uncertainties are assigned to account for known biases in the underlying methods,
2108 calibrations, and objects used for this analysis. The largest such uncertainty is associated
2109 with the kinematic bias inherent in deriving the background estimate outside of the signal
2110 region. However, a statistical biasing of this same estimate also has a significant impact.
2111 Additionally, due to the use of Monte Carlo for signal modelling and b -tagging calibration,
2112 uncertainties related to mismodellings in simulation must also be accounted for. Note that
2113 the results for the non-resonant analysis presented here are preliminary and only include
2114 background systematic, such that the discussion of the signal systematics *only* applies for
2115 the resonant search. However, these background systematics are expected to be by far the
2116 dominant uncertainties.

2117 *7.7.1 Statistical Uncertainties and Bootstrapping*

2118 There are two components to the statistical error for the neural network background estimate.
2119 The first is standard Poisson error, i.e., a given bin, i , in the background histogram has value
2120 $n_i = \sum_{j \in i} w_j$, where w_j is the weight for an event j which falls in bin i . Standard techniques
2121 then result in statistical error $\delta n_i = \sqrt{\sum_{j \in i} w_j^2}$, which reduces to the familiar \sqrt{N} Poisson error
2122 when all w_j are equal to 1.

2123 However, this procedure does not take into account the statistical uncertainty on the
2124 w_j due to the finite training dataset. Due to the large size difference between the two tag
2125 and four tag datasets, it is the statistical uncertainty due to the four tag training data that
2126 dominates that on the background. A standard method for estimating this uncertainty is the
2127 bootstrap resampling technique [105]. Conceptually, a set of statistically equivalent sets is
2128 constructed by sampling with replacement from the original training set. The reweighting
2129 network is then trained on each of these separately, resulting in a set of statistically equivalent
2130 background estimates. Each of these sets is below referred to as a replica.

2131 In practice, as the original training set is large, the resampling procedure is able to

2132 be simplified through the relation $\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$, which dictates that
 2133 sampling with replacement is approximately equivalent to applying a randomly distributed
 2134 integer weight to each event, drawn from a Poisson distribution with a mean of 1.

2135 Though the network configuration itself is the same for each bootstrap training, the
 2136 network initialization is allowed to vary. It should therefore be noted that the bootstrap
 2137 uncertainties implicitly capture the uncertainty due to this variation in addition to the
 2138 previously mentioned training set variation.

2139 The variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty
 2140 which is treated as a statistical uncertainty in the fit. Due to practical constraints, a
 2141 procedure for approximating the full bootstrap error band is developed which demonstrates
 2142 good agreement with the full bootstrap uncertainty. This procedure is described below.

2143 *Calculating the Bootstrap Error Band*

2144 The standard procedure to calculate the bootstrap uncertainty would proceed as follows: first,
 2145 each network trained on each bootstrap replica dataset would be used to produce a histogram
 2146 in the variable of interest. This would result in a set of replica histograms (e.g. for 100
 2147 bootstrap replicas, 100 histograms would be created). The nominal estimate would then be
 2148 the mean of bin values across these replica histograms, with errors set by the corresponding
 2149 standard deviation.

2150 In practice, such an approach is inflexible and demanding both in computation and in
 2151 storage, in so far as we would like to produce histograms in many variables, with a variety
 2152 of different cuts and binnings. This motivates a derivation based on event-level quantities.
 2153 However, due to non-trivial correlations between replica weights, simple linear propagation of
 2154 event weight variation is not correct.

2155 We therefore adopt an approach which has been empirically found to produce results
 2156 (for this analysis) in line with those produced by generating all of the histograms, as in the
 2157 standard procedure. This approach is described below. Note that, for robustness to outliers
 2158 and weight distribution asymmetry, the median and interquartile range (IQR) are used for

2159 the central value and width respectively (as opposed to the mean and standard deviation).

2160 The components involved in the calculation have been mentioned in Section 7.6 and are
2161 as follows:

2162 1. Replica weight (w_i): weight predicted for a given event by a network trained on replica
2163 dataset i .

2164 2. Replica norm (α_i): normalization factor for replica i . This normalizes the reweighting
2165 prediction of the network trained on replica dataset i to match the correponding target
2166 yield.

3. Median weight (w_{med}): median weight for a given event across replica datasets, used
for the nominal estimate. Defined (for 100 bootstrap replicas) as

$$w_{med} \equiv \text{median}(\alpha_1 w_1, \dots, \alpha_{100} w_{100}) \quad (7.12)$$

2167 4. Normalization correction (α_{med}): normalization factor to match the predicted yield of
2168 the median weights (w_{med}) to the target yield in the training region.

2169 As mentioned in Section 7.6, the *nominal estimate* is constructed from the set of median
2170 weights and the normalization correction, i.e. $\alpha_{med} \cdot w_{med}$.

2171 For the bootstrap error band, a “varied” histogram is then generated by applying, for
2172 each event, a weight equal to the median weight (with no normalization correction) plus half
2173 the interquartile range of the replica weights: $w_{varied} = w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})$.

2174 This varied histogram is scaled to match the yield of the nominal estimate. To account
2175 for variation of the nominal estimate yield, a normalization variation is calculated from the
2176 interquartile range of the replica norms: $\frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100})$. This variation, multiplied into
2177 the nominal estimate, is used to set a baseline for the varied histogram described above.

Denoting $H(\text{weights})$ as a histogram constructed from a given set of weights, $Y(\text{weights})$

as the predicted yield for a given set of weights, the final varied histogram is thus:

$$H(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})) \cdot \frac{Y(\alpha_{med} w_{med})}{Y(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100}))} + \frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100}) \cdot H(\alpha_{med} w_{med}) \quad (7.13)$$

where the first term roughly describes the behaviour of the bootstrap variation across the distribution of the variable of interest while the second term describes the normalization variation of the bootstrap replicas.

The difference between the varied histogram and the nominal histogram is then taken to be the bootstrap statistical uncertainty on the nominal histogram.

Figure 7.60 demonstrates how each of the components described above contribute to the uncertainty envelope for the non-resonant 2017 Control Region and compares this approximate band to the variation of histograms from individual bootstrap estimates. The error band constructed from the above procedure is seen to provide a good description of the bootstrap variation.

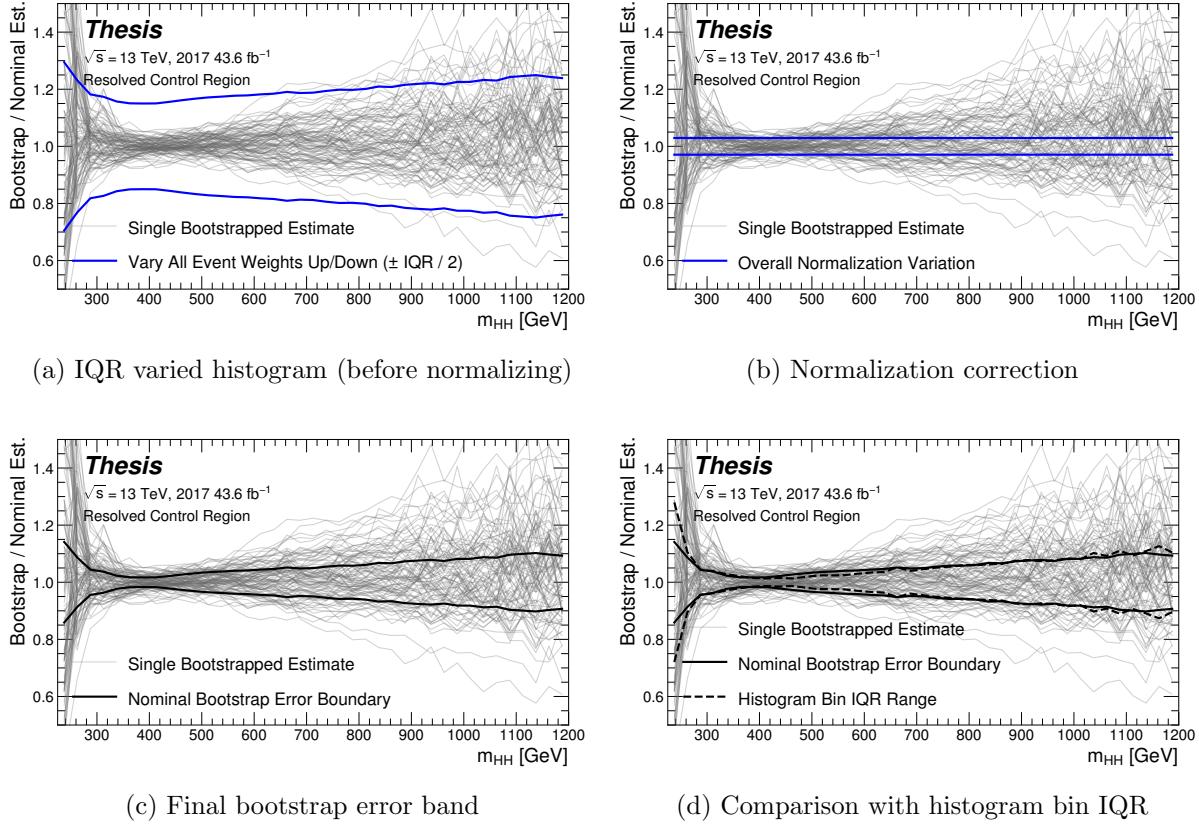


Figure 7.60: Illustration of the approximate bootstrap band procedure, shown as a ratio to the nominal estimate for the 2017 non-resonant background estimate. Each grey line is from the m_{HH} prediction for a single bootstrap training. Figure 7.60(a) shows the variation histograms constructed from median weight \pm the IQR of the replica weights. It can be seen that this captures the rough shape of the bootstrap envelope, but is not good estimate for the overall magnitude of the variation. Figure 7.60(b) demonstrates the applied normalization correction, and Figure 7.60(c) shows the final band (normalized Figure 7.60(a) + Figure 7.60(b)). Comparing this with the IQR variation for the prediction from each bootstrap in each bin in Figure 7.60(d), the approximate envelope describes a very similar variation.

2188 7.7.2 *Background Shape Uncertainties*

2189 To account for the systematic bias associated with deriving the reweighting function in the
2190 control region and extrapolating to the signal region, an alternative background model is
2191 derived in the validation region. Because of the fully data-driven nature of the background
2192 model, this is an uncertainty assessed on the full background. The alternative model and
2193 the baseline are consistent with the observed data in their training regions, and differences
2194 between the alternative and baseline models are used to define a shape uncertainty on the
2195 m_{HH} spectrum, with a two-sided uncertainty defined by symmetrizing the difference about
2196 the baseline.

2197 For the resonant analysis, this uncertainty is split into two components to allow for two
2198 independent variations of the m_{HH} spectrum: : a low- H_T and a high- H_T component, where
2199 H_T is the scalar sum of the p_T of the four jets constituting the Higgs boson candidates, and
2200 serves as a proxy for m_{HH} , while avoiding introducing a sharp discontinuity. The boundary
2201 value is 300 GeV. The low- H_T shape uncertainty primarily affects the m_{HH} spectrum below
2202 400 GeV (close to the kinematic threshold) by up to around 5%, and the high- H_T uncertainty
2203 mainly m_{HH} above this by up to around 20% relative to nominal. These separate m_{HH}
2204 regimes are by design – the H_T split is introduced to prevent low mass bins from constraining
2205 the high mass uncertainty and vice-versa.

2206 This was the *status quo* shape uncertainty decomposition from the Early Run 2 analysis.
2207 A decomposition in terms of orthogonal polynomials, which would provide increased flexibility,
2208 was also evaluated. This study revealed that both decompositions are able to account for the
2209 systematic deviations between four tag data and the background estimate (evaluated in the
2210 kinematic validation region), and produce almost identical limits. The simpler *status quo*
2211 decomposition is therefore kept.

2212 For the non-resonant analysis, the quadrant nature of the background estimation leads to
2213 a natural breakdown of the nuisance parameters: quadrants are defined in the signal region
2214 along the same axes as those used for the control and validation region definitions. Variations

are then assessed in each of these signal region quadrants, corresponding to regions that are “closer to” and “further away from” the nominal and alternate estimate regions, fully leveraging the power of the two equivalent but systematically different estimates.

Figure 7.61 shows an example of the variation in each H_T region for the 2018 resonant analysis. Figure 7.62 shows the example quadrant variation for the 2018 4 b non-resonant analysis.

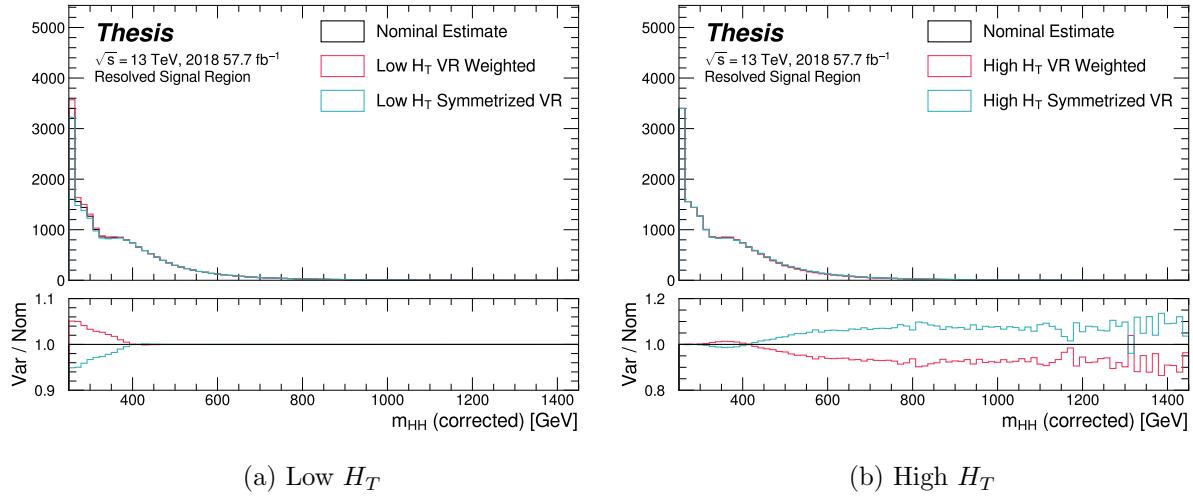
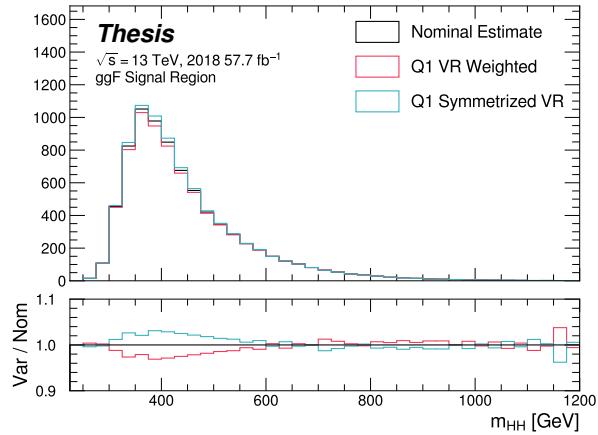
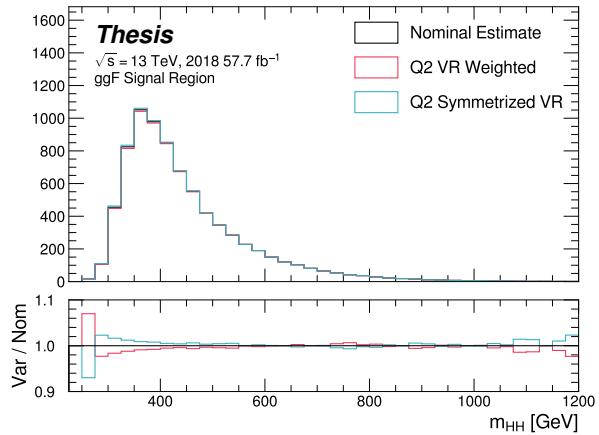


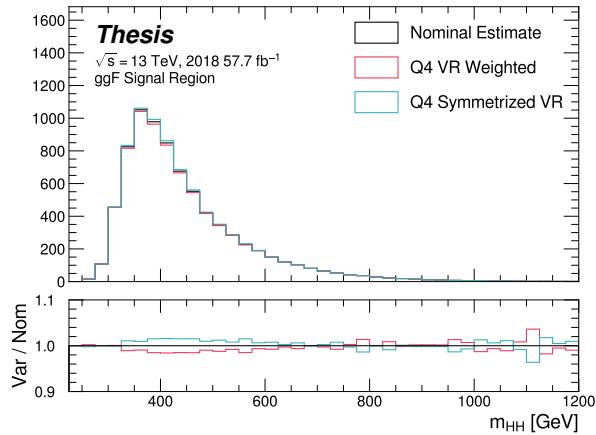
Figure 7.61: **Resonant Search:** Example of CR vs VR variation in each H_T region for 2018. The variation nicely factorizes into low and high mass components.



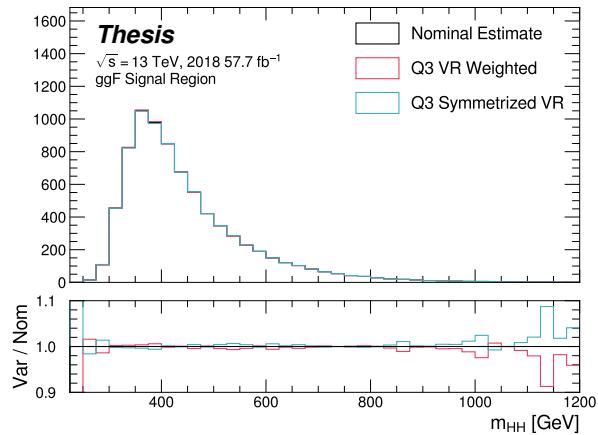
(a) Q1 (top)



(b) Q2 (left)



(c) Q4 (right)



(d) Q3 (bottom)

Figure 7.62: **Non-resonant Search (4b):** Example of CR vs VR variation in each signal region quadrant for 2018. Significantly different behavior is seen between quadrants, with the largest variation in quadrant 1 and the smallest in quadrant 4.

2221 7.7.3 *Signal Uncertainties*

2222 A variety of uncertainties are assessed on the the signal Monte Carlo simulation. As the
 2223 background estimate is fully data driven, such uncertainties are not needed for the background
 2224 estimate. Note again that the results presented for the non-resonant search only include the
 2225 background systematics described above.

2226 Detector modeling and reconstruction uncertainties account for differences between Monte
 2227 Carlo simulation and real data due to mismodelling of the detector as well as due to the
 2228 different performance of algorithms on simulation compared to data. In this analysis they
 2229 consist of uncertainties related to jet properties and uncertainties stemming from the flavor
 2230 tagging procedure. The jet uncertainties are treated according to the prescription in [106] and
 2231 are implemented as variations of the jet properties. These cover uncertainty in jet energy scale
 2232 and resolution. Uncertainties in b -tagging efficiency are treated according to the prescription
 2233 in Ref. [77] and implemented as scale factors applied to the Monte Carlo event weights. A
 2234 systematic related to the PtReco b -jet energy correction has been studied in the $HH \rightarrow \gamma\gamma b\bar{b}$
 2235 analysis [107] and found to be negligible compared to the other jet uncertainties. Following
 2236 this example, such a systematic is therefore neglected here.

2237 Trigger uncertainties stem from imperfect knowledge of the ratio between the efficiency of
 2238 a given trigger in data to its efficiency in Monte Carlo simulation. This ratio is applied as a
 2239 scale factor to all simulated events, with the systematic variations produced by varying the
 2240 scale factor up or down by one sigma. Such variations are evaluated based on measurements
 2241 of per-jet online efficiencies for both jet reconstruction and b -tagging, and these are used to
 2242 compute event-level uncertainties. These are then applied as overall weight variations on the
 2243 simulated events.

2244 An uncertainty on the total integrated luminosity used in this analysis is also applied, ans
 2245 is measured to be 1.7% [95], obtained using the LUCID-2 detector for the primary luminosity
 2246 measurements [108].

2247 A variety of theoretical uncertainties are also assessed on the signal. Such uncertainties

2248 are assessed by generating samples following the configuration of the baseline samples, but
 2249 with modifications to probe various aspects of the simulation. These include uncertainties in
 2250 the parton density functions (PDFs); uncertainties due to missing higher order terms in the
 2251 matrix elements; and uncertainties in the modelling of the underlying event, which includes
 2252 multi-parton interactions, of hadronic showers and of initial and final state radiation.

2253 Uncertainties due to modelling of the parton shower and the underlying event are eval-
 2254 uated by comparing results from using two different generators, namely HERWIG 7.1.3 and
 2255 PYTHIA 8.235. No significant dependence on the variable of interest, m_{HH} , is observed.
 2256 Therefore, a 5% flat systematic uncertainty is assigned to all signal samples, extracted from
 2257 the acceptance comparison for the full 4-tag selection.

2258 Uncertainties in the matrix element calculation are evaluated by varying the factorization
 2259 and renormalization scales used in the generator up and down by a factor of two, both
 2260 independently and simultaneously. This results in an effect smaller than 1% for all variations
 2261 and all masses; the impact of such uncertainties is therefore neglected.

2262 PDF uncertainties are evaluated using the PDF4LHC_NLO_MC set [96] by calculating
 2263 the signal acceptance for each PDF replica and taking the standard deviation. In all cases,
 2264 these uncertainties result in an effect smaller than 1% on the signal acceptance; therefore
 2265 these are also neglected.

2266 Theoretical uncertainties on the $H \rightarrow b\bar{b}$ branching ratio [109] are also included.

2267 **7.8 Background Validation**

2268 In addition to checking the performance of the background estimate in the control and
2269 validation regions, a variety of alternative selections are defined to allow for a full “dress
2270 rehearsal” of the background estimation procedure.

2271 Both the resonant and non-resonant analyses make use of a *reversed* $\Delta\eta$ region, in which
2272 the kinematic cut on $\Delta\eta_{HH}$ is reversed, so that events are required to have $\Delta\eta_{HH} > 1.5$.
2273 This is orthogonal to the nominal signal region and has minimal sensitivity, allowing for the
2274 comparison of the background estimate $4b$ data in the corresponding “signal region”. For
2275 this validation, a new reweighting is trained following nominal procedures, but entirely in the
2276 $\Delta\eta_{HH} > 1.5$ region.

2277 The non-resonant analysis additionally makes use of the $3b + 1$ fail region mentioned
2278 above, which again is orthogonal to the nominal signal regions and has minimal sensitivity.
2279 The reweighting in this case is between $2b$ and $3b + 1$ fail events rather than between $2b$
2280 and $3b + 1$ loose or $2b$ and $4b$. However, the kinematic selections of signal region events are
2281 otherwise identical, allowing for a complementary test of the background estimate.

2282 *TODO: Add shifted regions if they’re ready*

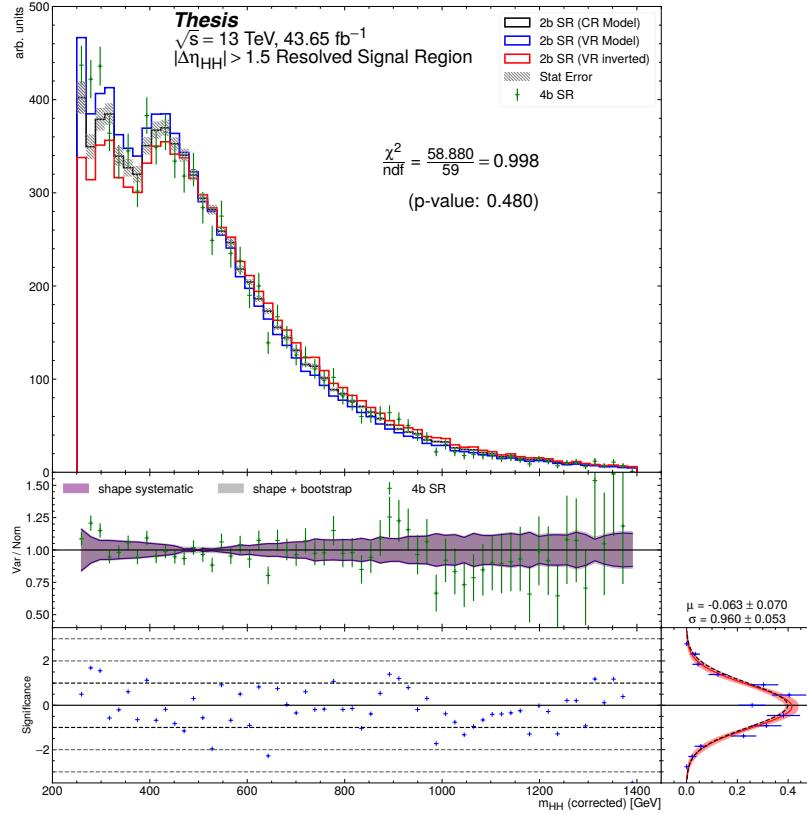


Figure 7.63: **Resonant Search:** Performance of the background estimation method in the resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background estimate is trained following nominal procedures entirely within the reversed $\Delta\eta_{HH}$ region, and the resulting model, including uncertainties, is compared with $4b$ data in the corresponding signal region. Good agreement is shown. The quoted p -value uses the χ^2 test statistic, and demonstrates no evidence that the data differs from the assessed background.

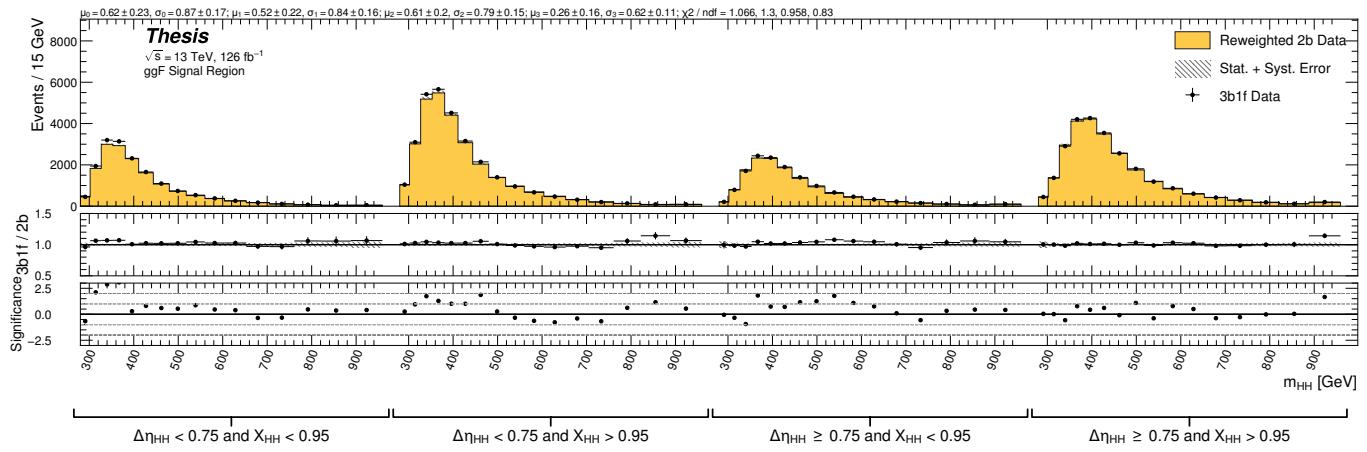


Figure 7.64: **Non-resonant Search:** Performance of the background estimation method in the $3b + 1$ fail validation region. A new background estimate is trained following nominal procedures but with a reweighting from $2b$ to $3b + 1$ fail events. Generally good agreement is seen, though there is some deviation at very low masses in the low $\Delta\eta_{HH}$ low X_{HH} category.

2283 **7.9 Overview of Other $b\bar{b}b\bar{b}$ Channels**

2284 The results discussed above have been developed in conjunction with (1) a boosted channel
2285 for the resonant search and (2) a vector boson fusion (VBF) channel for the non-resonant
2286 search. Detailed discussions of these two channels are beyond the scope of this thesis, though
2287 a combined set of resolved and boosted results are presented below. The VBF results are not
2288 included in this thesis, but much of this thesis work has been useful in the development of
2289 that result. For completeness, we therefore briefly summarize both analyses here.

2290 **7.9.1 Resonant: Boosted Channel**

2291 The boosted analysis selection targets resonance masses from 900 GeV to 5 TeV. In such
2292 events, H decays have a high Lorentz boost, such that the $b\bar{b}$ decays are very collimated. The
2293 resolved analysis fails to reconstruct such HH events, as the $R = 0.4$ jets start to overlap.

2294 The boosted analysis instead reconstructs H decays as large radius, $R = 1.0$ jets, with
2295 corresponding b -quarks identified with variable radius subjets, that is jets with a radius that
2296 scales as ρ/p_T , the p_T is that of the jet in question, and ρ is a fixed parameter, here chosen
2297 to be 30 GeV, which is optimized to maintain truth-level double b -labelling efficiency across
2298 the full range of Higgs jet p_T [73].

2299 Due to limited boosted b -tagging efficiency and to maintain sensitivity even when b -jets
2300 are highly collimated, the boosted analysis is divided into three categories based on the
2301 number of b -tagged jets associated to each large radius jet:

- 2302 • 4 b category: two b -tagged jets in each
- 2303 • 2 $b - 1$ category: two b -tagged jets in one, one in the other
- 2304 • 1 $b - 1$ category: one b -tagged jet in each

2305 The analysis then proceeds in each of these categories.

2306 The resolved and boosted channels are combined for resonance masses from 900 GeV to
2307 1.5 TeV inclusive. To keep the channels statistically independent, the boosted channel vetos
2308 events passing the resolved analysis selection.

2309 *7.9.2 Non-resonant: VBF Channel*

2310 The vector boson fusion channel is only considered for the non-resonant search. While the
2311 sensitivity is in general much more limited than the gluon-gluon fusion analysis due to the
2312 much smaller production cross section, VBF is sensitive to a variety of Beyond the Standard
2313 Model physics, both complementary and orthogonal to the theoretical scope of gluon-gluon
2314 fusion.

2315 The VBF channel proceeds very similarly to the ggF, with the primary differences being
2316 the kinematic selections and the categorization, which are impacted by the presence of two
2317 *VBF jets*, resulting from the two initial state quarks. The ggF channel result presented here
2318 includes a veto on VBF events, such that if events pass the full VBF selection, they are not
2319 included in the set of events considered for the ggF result.

2320 Beginning with the assumption of four *HH* jets and two VBF jets, the VBF channel first
2321 requires an event to have a minimum six jets. The VBF jets are reconstructed as the two jets
2322 with the highest di-jet invariant mass, m_{jj} , out of the set of all non-tagged jets in the event.
2323 If no such pair exists (i.e., there are less than two non-tagged jets), the event is placed in the
2324 ggF channel. To reduce the number of background events, three cuts are then applied, VBF
2325 jets are required to have $\Delta\eta > 3$ and a combined invariant mass of $m_{jji} < 1000$ GeV. *HH* jets
2326 are identified as in the ggF channel, and the vector sum of the p_T of the *HH* and VBF jets is
2327 required to be less than 65 GeV. The remainder of the analysis proceeds similarly to the ggF
2328 channel, and events failing any stage of this selection are considered for ggF.

2329 Note that the background estimation for the VBF channel is inherited from the resonant
2330 and ggF analyses, an ancillary, but significant, contribution of this thesis work.

2331 **7.10 m_{HH} Distributions**

2332 *7.10.1 Resonant Search*

2333 The final discriminant used for the resonant search is corrected m_{HH} . Histogram binning
2334 was optimized for the resonant search to be 84 equal width bins from 250 GeV to 1450 GeV,
2335 corresponding to a bin width of 14.3 GeV, and overflow events (events above 1450 GeV) are
2336 included in the last bin. A demonstration of the performance of the reweighting on this
2337 distribution is shown in Figure 7.65 for the control region and Figure 7.66 for the validation region. A background-only profile likelihood fit is run for the distribution in the
2338 signal region, and results with spin-0 signals overlaid are shown in Figure 7.67. Note that the
2339 plots show the sum across all years, but the signal extraction fit and background estimate
2340 are run with the years separately. Agreement is generally good throughout.
2341

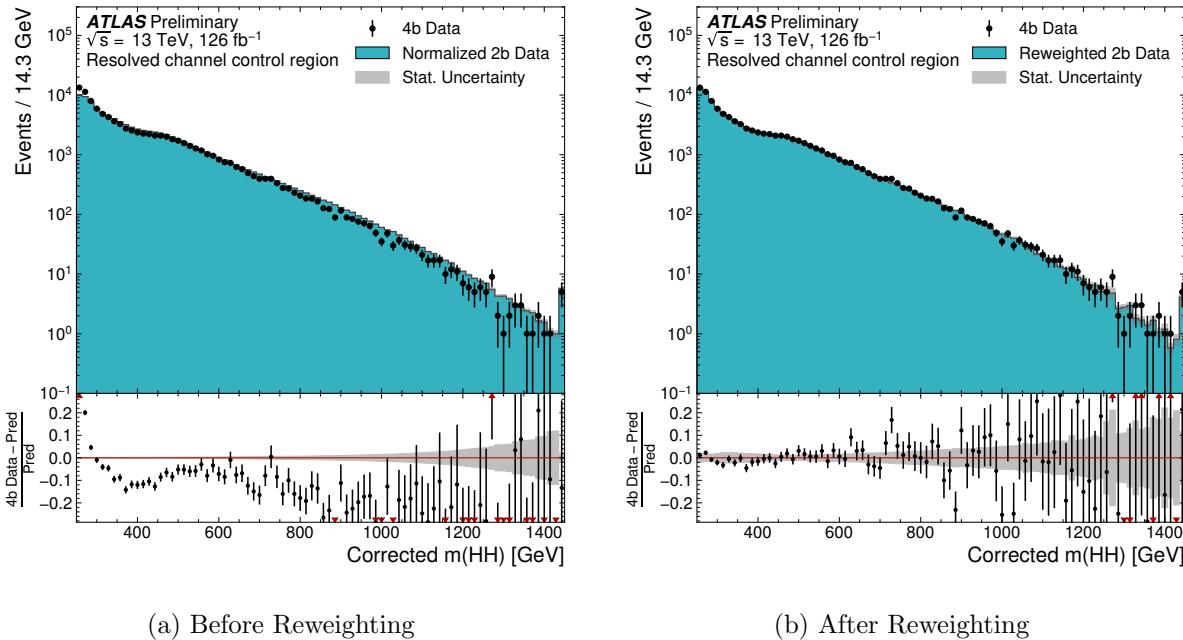


Figure 7.65: **Resonant Search:** Demonstration of the performance of the nominal reweighting in the control region on corrected m_{HH} , with Figure 7.65(a) showing $2b$ events normalized to the total $4b$ yield and Figure 7.65(b) applying the reweighting procedure. Agreement is much improved with the reweighting. Note that overall reweighted $2b$ yield agrees with $4b$ yield in the control region by construction.

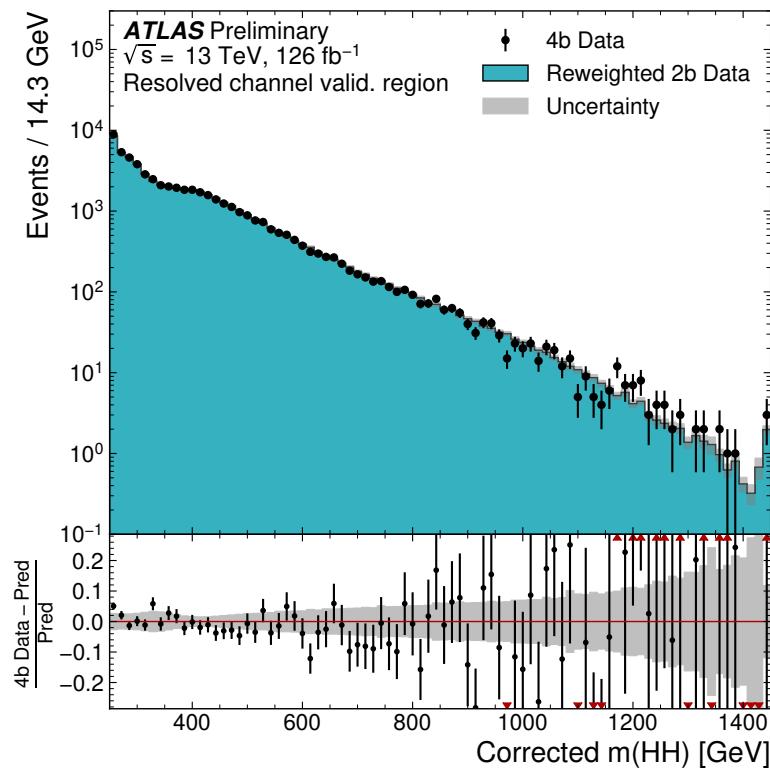


Figure 7.66: **Resonant Search:** Demonstration of the performance of the control region derived reweighting in the validation region on corrected m_{HH} . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined by a reweighting trained in the validation region.

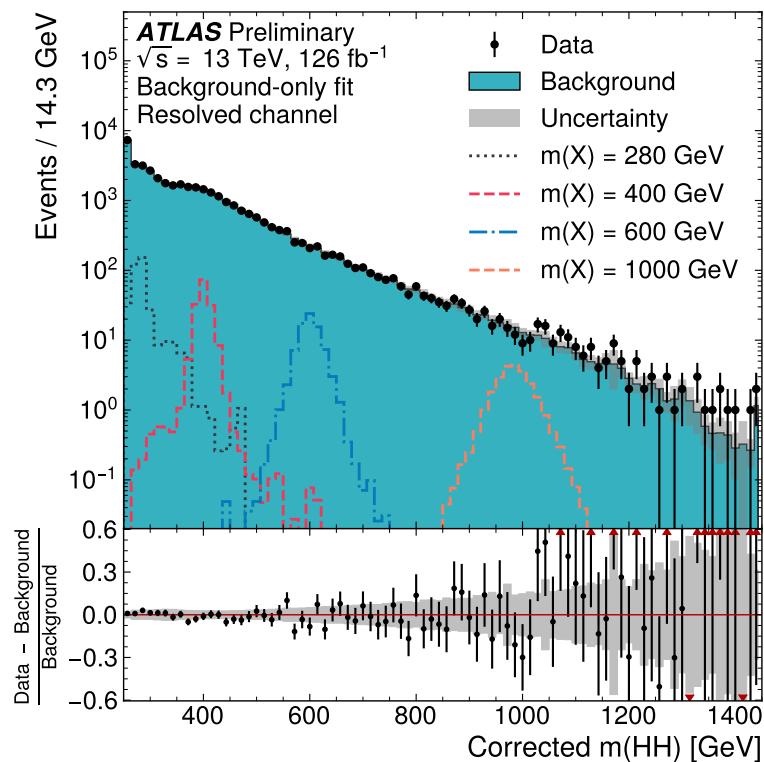


Figure 7.67: **Resonant Search:** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The closure is generally quite good, though there is an evident deficit in the background estimate relative to the data for higher values of corrected m_{HH} .

2342 7.10.2 Non-resonant Search

As discussed above, the non-resonant search splits the signal extraction into two categories of $\Delta\eta_{HH}$ ($0 \leq \Delta\eta_{HH} < 0.75$ and $0.75 \leq \Delta\eta_{HH} < 1.5$), and two categories of X_{HH} ($0 \leq X_{HH} < 0.95$ and $0.95 \leq X_{HH} < 1.6$). To maintain reasonable statistics in each bin entering the signal extraction fit, a variable width binning is considered defined by a resolution parameter, r , and a set range in m_{HH} , where bin edges are determined iteratively as

$$b_{low}^{i+1} = b_{low}^i + r \cdot b_{low}^i, \quad (7.14)$$

2343 where b_{low}^i is the low edge of bin i . The parameters used here are $r = 0.08$ over a range
2344 from 280 GeV to 975 GeV, and underflow and overflow are included in the intial and final
2345 bin contents respectively. m_{HH} with no correction is used as the final discriminant in each
2346 category.

2347 A demonstration of the performance of the reweighting on distributions unrolled across
2348 categories is shown in Figures 7.68 and 7.69 for the the control region and Figures 7.70
2349 and 7.71 for the validation region. A background-only profile likelihood fit is run for the
2350 distribution in the signal region, and results with the Standard Model HH signal and $\kappa_\lambda = 6$
2351 signal overlaid are shown for 4b in Figure 7.72 and 3b1l in Figure 7.73. Note that the plots
2352 show the sum across all years, but the signal extraction fit and background estimate are run
2353 with the years separately. All bins are normalized to represent a density of Events / 15 GeV.

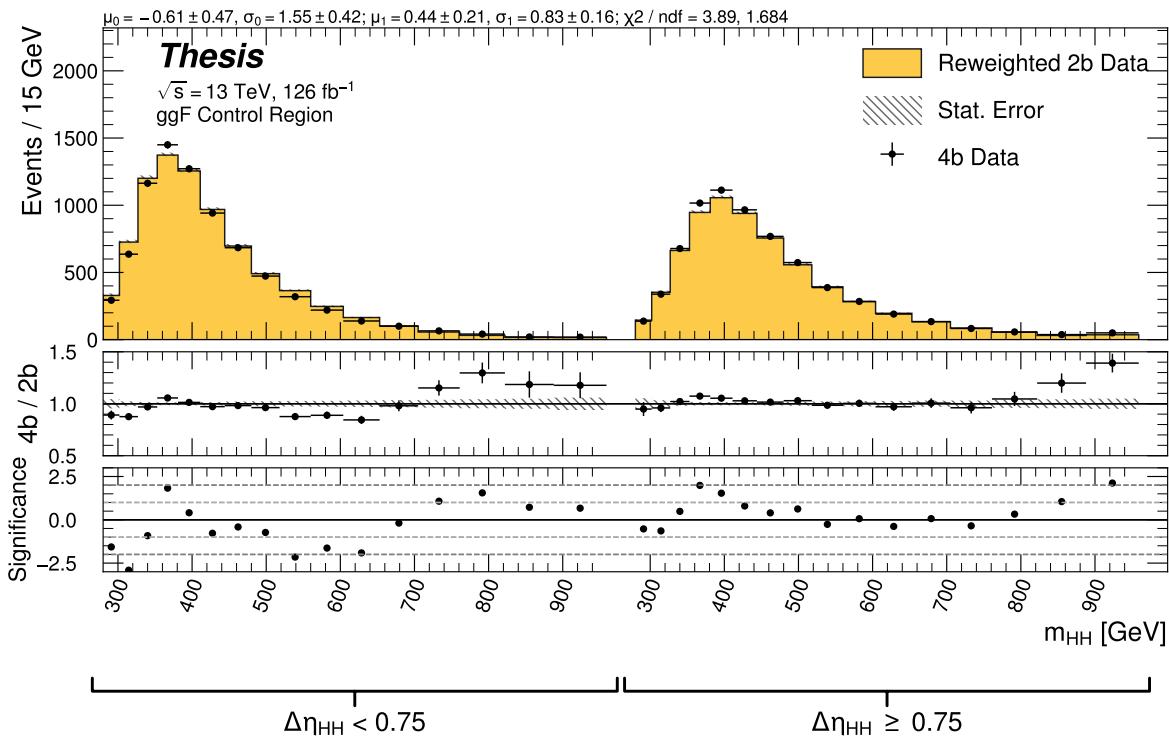


Figure 7.68: **Non-resonant Search (4b)**: Demonstration of the performance of the nominal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with some residual mismodeling in the low $\Delta\eta_{HH}$ region near 600 GeV.

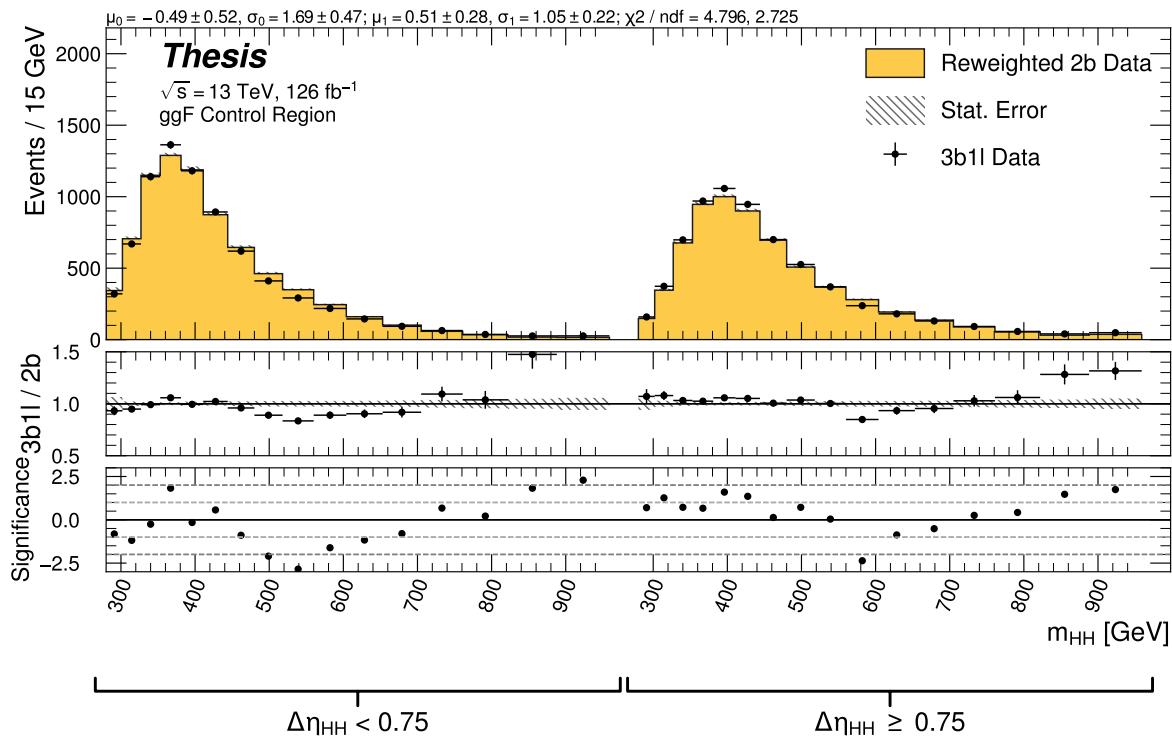


Figure 7.69: **Non-resonant Search (3b1l):** Demonstration of the performance of the nominal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with similar conclusions as for the $4b$ region.

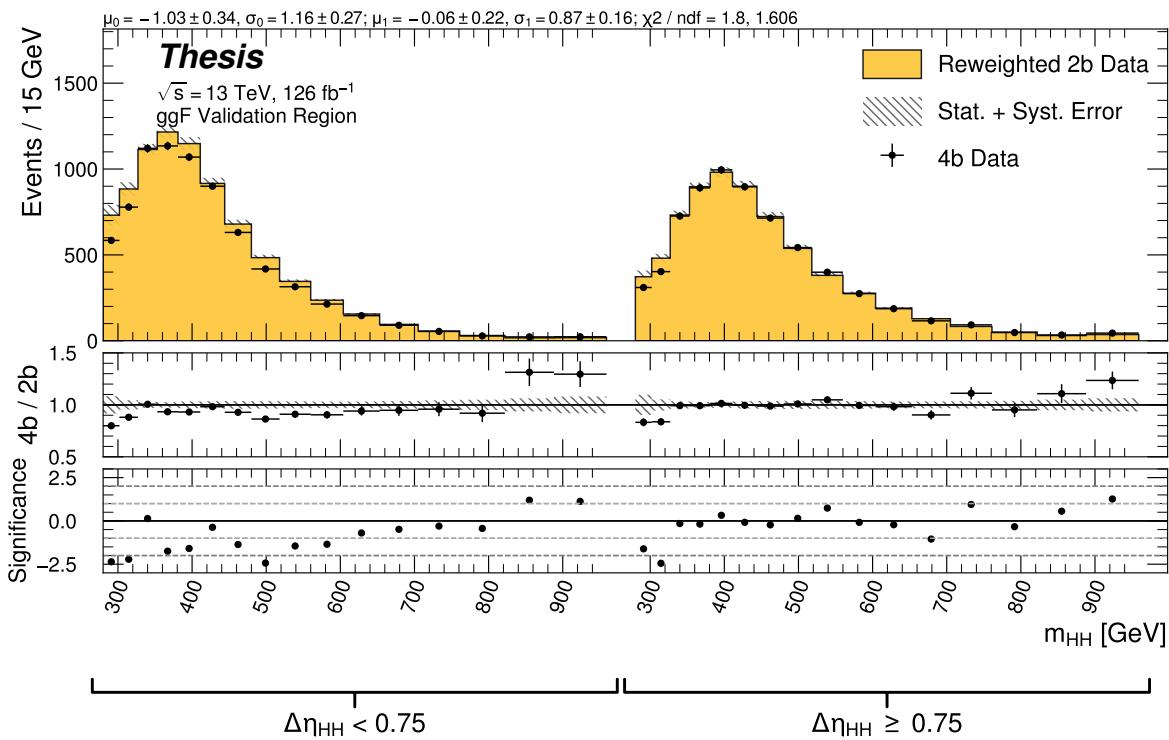


Figure 7.70: **Non-resonant Search (4b)**: Demonstration of the performance of the nominal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. The low $\Delta\eta_{HH}$ region is consistently overestimated, but, systematic uncertainties are defined via the difference between VR and CR estimates.

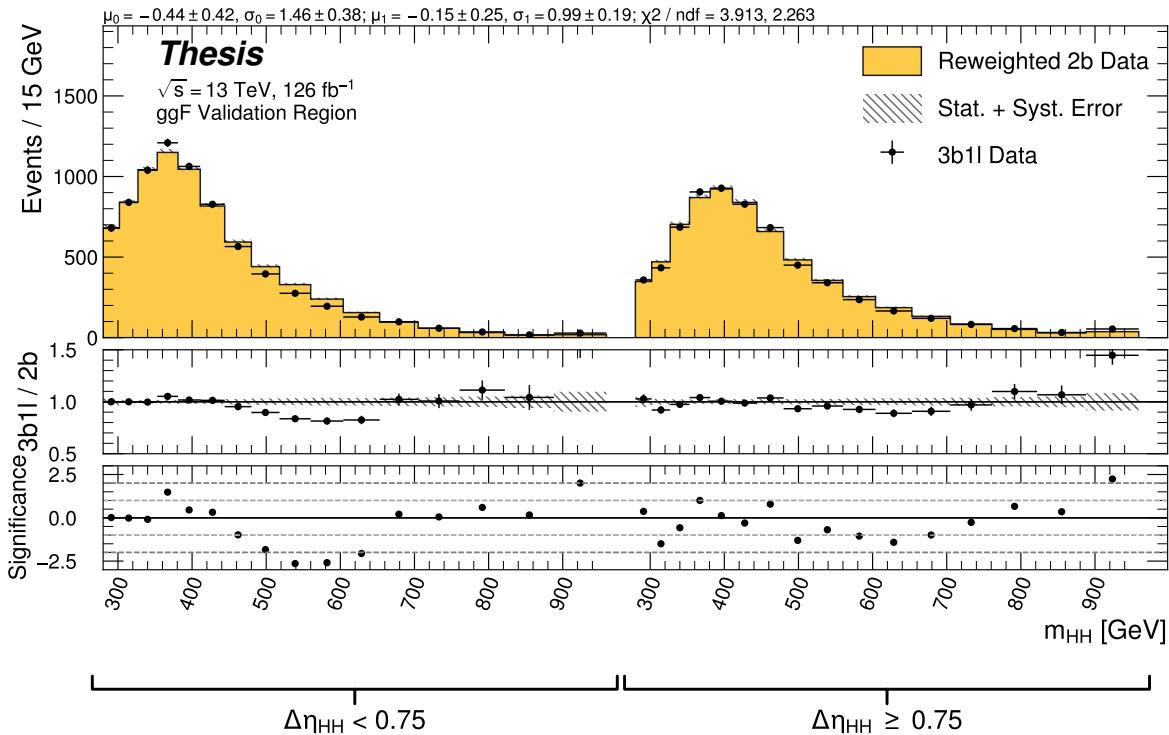


Figure 7.71: **Non-resonant Search (3b1l):** Demonstration of the performance of the nominal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. A deficit is present near 600 GeV, but agreement is fairly good otherwise.

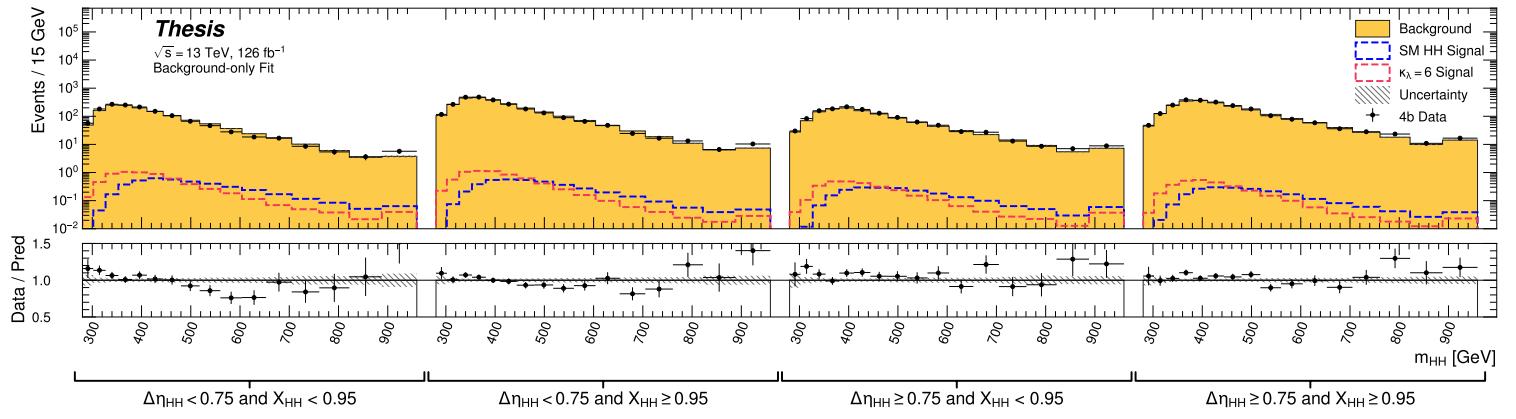


Figure 7.72: **Non-resonant Search (4b):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the $4b$ channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Modeling is generally quite good near the Standard Model peak, but disagreements are seen at very low and high masses. A deficit is present in low $\Delta\eta_{HH}$ bins near 600 GeV.

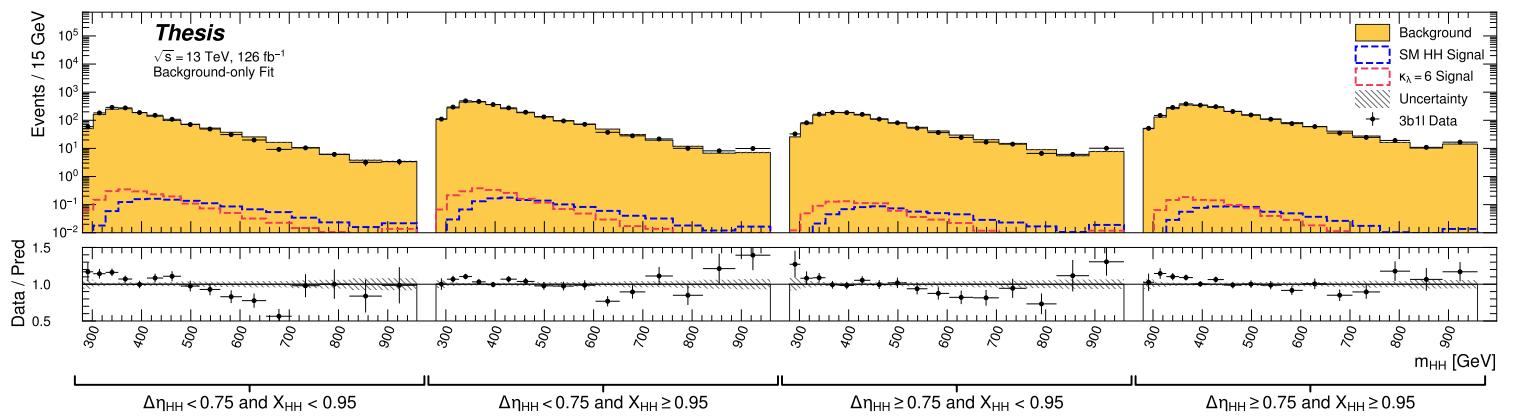


Figure 7.73: **Non-resonant Search (3b1l):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the $3b1l$ channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Conclusions are very similar to the $4b$ channels, with generally good modeling near the Standard Model peak, but disagreements at very low and high masses. A deficit is present near 600 GeV.

2354 **7.11 Statistical Analysis**

2355 The resonant analysis is used to set a 95% confidence level upper limit on the $pp \rightarrow X \rightarrow$
2356 $HH \rightarrow b\bar{b}b\bar{b}$ and $pp \rightarrow G_{KK}^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ cross-sections, while the non-resonant analysis
2357 is used to set a 95% confidence level upper limit on the $pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ cross sections for
2358 a variety of values of the trilinear Higgs coupling.

2359 The upper limit is extracted using the CL_s method [110]. The test statistic used is q_μ
2360 [111], where μ is the signal strength, and θ represents the nuisance parameters. A single
2361 hat represents the maximum likelihood estimate of a parameter, while $\hat{\theta}(x)$ represents the
2362 conditional maximum likelihood estimate of the nuisance parameters if the signal cross-section
2363 is fixed at x .

$$q_\mu = \begin{cases} -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad (7.15)$$

2364 CL_s for some test value of μ is then defined by

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p(q_\mu \geq q_{\mu, \text{obs}} | s+b)}{p(q_\mu \geq q_{\mu, \text{obs}} | b)}, \quad (7.16)$$

2365 where the p -values are calculated in the asymptotic approximation [111], which is valid in
2366 the large sample limit.

2367 The signal cross-section μ fb is excluded at the 95% confidence level if $CL_s < 0.05$.

Observed	-2σ	-1σ	Expected	$+1\sigma$	$+2\sigma$
4.4	3.1	4.2	5.9	8.2	11.0

Table 7.1: Limits on Standard Model $HH \rightarrow b\bar{b}b\bar{b}$ production, presented in units of the predicted Standard Model cross section. Results include background systematics only.

2368 7.12 Results

2369 Figure 7.74 shows the expected limit for the spin-0 and spin-2 resonant search. The resolved
 2370 channel covers the range between 251 and 1500 GeV and is combined with the boosted channel
 2371 between 900 and 1500 GeV. The boosted channel then extends to 3 TeV. The most significant
 2372 excess is seen for a signal mass of 1100 GeV, with local significance of 2.6σ for the spin-0
 2373 signal and 2.7σ for the spin-2 signal. This is reduced to 1.0σ and 1.2σ globally.

2374 The spin-2 bulk Randall-Sundrum model with $k/\overline{M}_{\text{Pl}} = 1$ is excluded for graviton masses
 2375 between 298 and 1440 GeV.

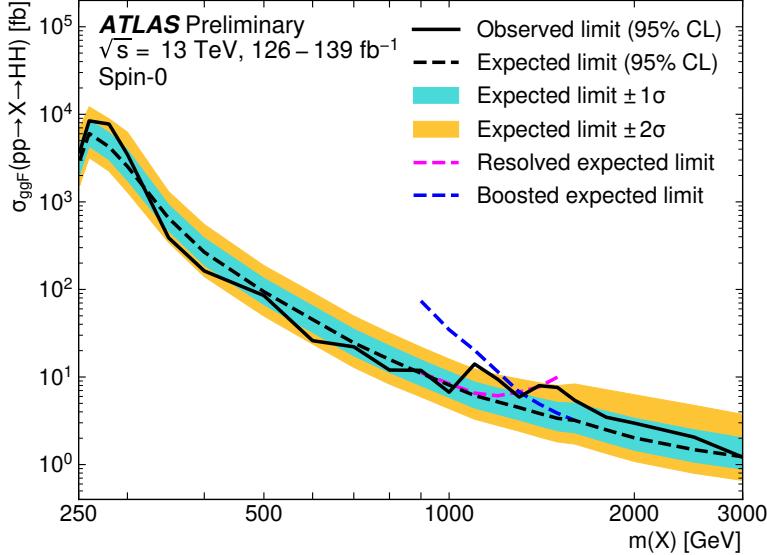
2376 Preliminary results are presented here for the gluon-gluon fusion non-resonant search,
 2377 combining results from the $4b$ and $3b + 1l$ signal regions in the 2×2 category scheme in
 2378 $\Delta\eta_{HH}$ and X_{HH} . These results will be further combined with a VBF channel as discussed,
 2379 but this is left for future work. Results shown here include background systematics only.
 2380 Limits are set for κ_λ values from -20 to 20 . The cross section limit for HH production is set
 2381 at 140 fb (180 fb) observed (expected), corresponding to an observed (expected) limit of 4.4
 2382 (5.9) times the Standard Model prediction (see Table 7.1). κ_λ is constrained to be within the
 2383 range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$ expected). These results are shown in
 2384 Figure 7.75.

2385 We note that this is a significant improvement over the early Run 2 result, which achieved
 2386 an observed (expected) limit of 12.9 (20.7) times the Standard Model prediction. The dataset
 2387 is 4.6 times larger, and a naive scaling of the early Run 2 result (Poisson statistics \implies a factor
 2388 of $1/\sqrt{4.6}$) would predict an observed (expected) limit of 6.0 (9.7) times the Standard Model.

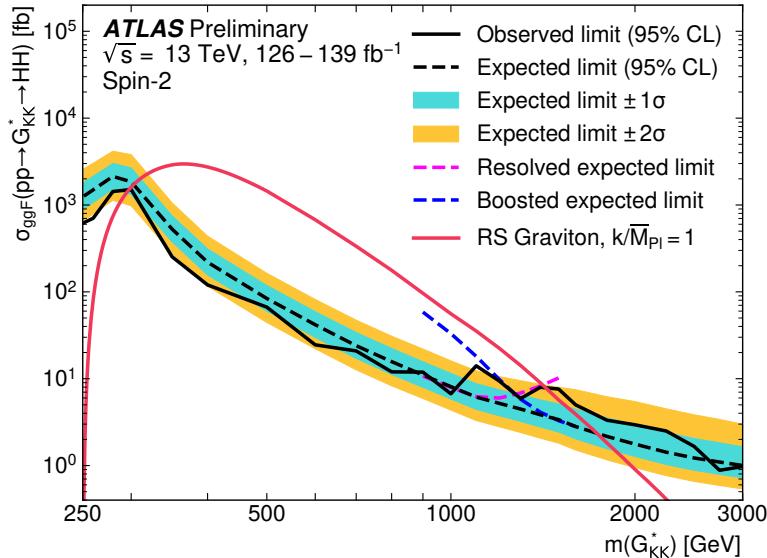
2389 The result of 4.4 (5.9) observed (expected) presented here is therefore both an improvement
 2390 by a factor of 3 (3.5) over the previous result and also beats the statistical scaling by around
 2391 30 (40) %, demonstrating the impact of the various analysis improvements presented here.
 2392 We note again that these results do not include the complete set of uncertainties – however
 2393 we expect the addition of the remaining uncertainties to have no more than a few percent
 2394 impact.

2395 The observed limits presented in Figure 7.75 are consistently above the 2σ band for values
 2396 of $\kappa_\lambda \geq 5$, peaking at a local significance of 3.8σ for $\kappa_\lambda = 6$. As this analysis is optimized for
 2397 points near the Standard Model, and as there is no excess present in more sensitive channels
 2398 in this same region (e.g. $HH \rightarrow bb\gamma\gamma$ *TODO: include comparison*), we do not believe this is a
 2399 real effect, but is rather due to a mis-modeling of the background at low mass, where the
 2400 $\min \Delta R$ pairing has poor signal efficiency and the assumption of well behaved background in
 2401 the mass plane breaks down. This is consistent with the location of the $\kappa_\lambda = 6$ signal in m_{HH} ,
 2402 as shown in Figures 7.72 and 7.73. It was considered, but not implemented, for this analysis
 2403 to impose a cut on m_{HH} near 350 or 400 GeV to avoid such a low mass modeling issue.

2404 To check the impact of if we would have imposed such a cut, and to verify that the excess
 2405 is due to the low mass regime, we therefore run the same set of limits without the low mass
 2406 bins. In this case, we choose to simply drop the first few bins in m_{HH} such that everything
 2407 else, including the higher mass bin edges, is kept the same. Due to the variable width binning,
 2408 this corresponds to an m_{HH} cut of 381 GeV. The results of this check are shown in Figure
 2409 7.76, overlaid with the limits of Figure 7.75 for reference. With the m_{HH} cut imposed, there
 2410 is a slight degradation in the expected limits for larger positive and negative values of κ_λ ,
 2411 but the points near the Standard Model are nearly identical. Further, the observed excess is
 2412 significantly reduced, with observed limits for $\kappa_\lambda \geq 5$ now falling entirely within the expected
 2413 1σ band. Due to the preliminary nature of these results, further study is left for future
 2414 work. However, we believe, in conjunction with the $HH \rightarrow bb\gamma\gamma$ results and our expectations
 2415 about the difficulty of the background estimation at low mass, that this is demonstrative of a
 2416 mismodeling rather than a real excess.



(a)



(b)

Figure 7.74: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of resonant production for spin-0 ($X \rightarrow HH$) and spin-2 $G_{KK}^* \rightarrow HH$. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the expected limits are shown in the colored bands. The resolved channel expected limit is shown in dashed pink and covers the range from 251 and 1500 GeV. It is combined with the boosted channel (dashed blue) between 900 and 1500 GeV. The theoretical prediction for the bulk RS model with $k/\bar{M}_{\text{Pl}} = 1$ [20] (solid red line) is shown, with the decrease below 350 GeV due to a sharp reduction in the $G_{KK}^* \rightarrow HH$ branching ratio. The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582.

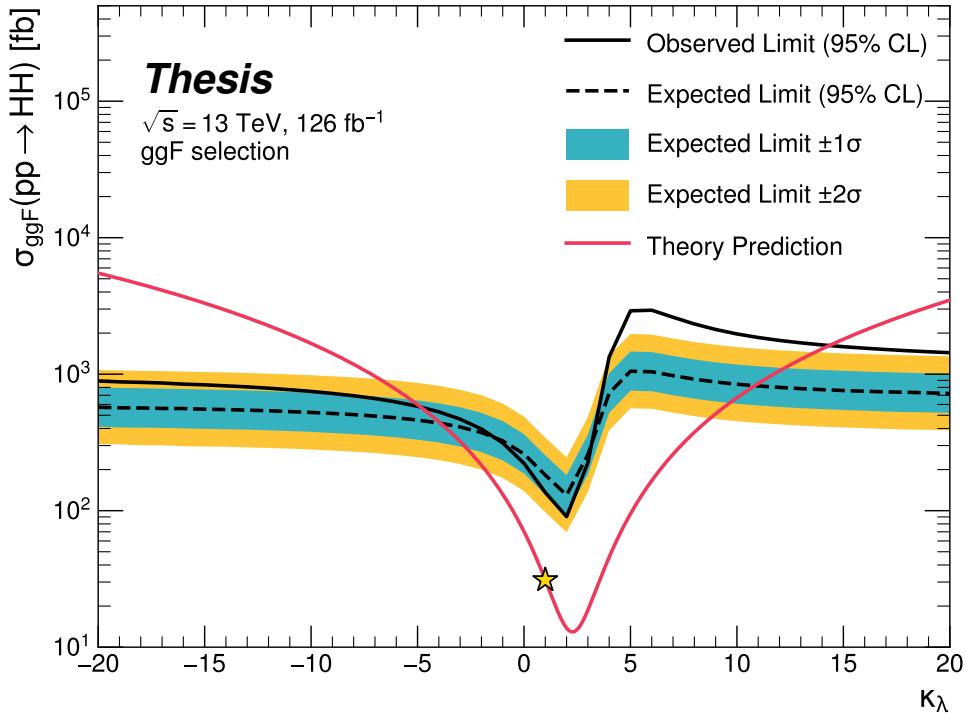


Figure 7.75: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of non-resonant production for a range of values of the Higgs self-coupling, with the Standard Model value ($\kappa_\lambda = 1$) illustrated with a star. The $\pm 1\sigma$ and $\pm 2\sigma$ ranges for the expected limits are shown in the colored bands. The cross section limit for HH production is set at 140 fb (180 fb) observed (expected), corresponding to an observed (expected) limit of 4.4 (5.9) times the Standard Model prediction. κ_λ is constrained to be within the range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$ expected). The nominal $H \rightarrow b\bar{b}$ branching ratio is taken as 0.582. We note that the excess present for $\kappa_\lambda \geq 5$ is thought to be due to a low mass background mis-modeling, present due to the optimization of this analysis for the Standard Model point, and is not present in more sensitive channels in this same region (e.g. $HH \rightarrow bb\gamma\gamma$).

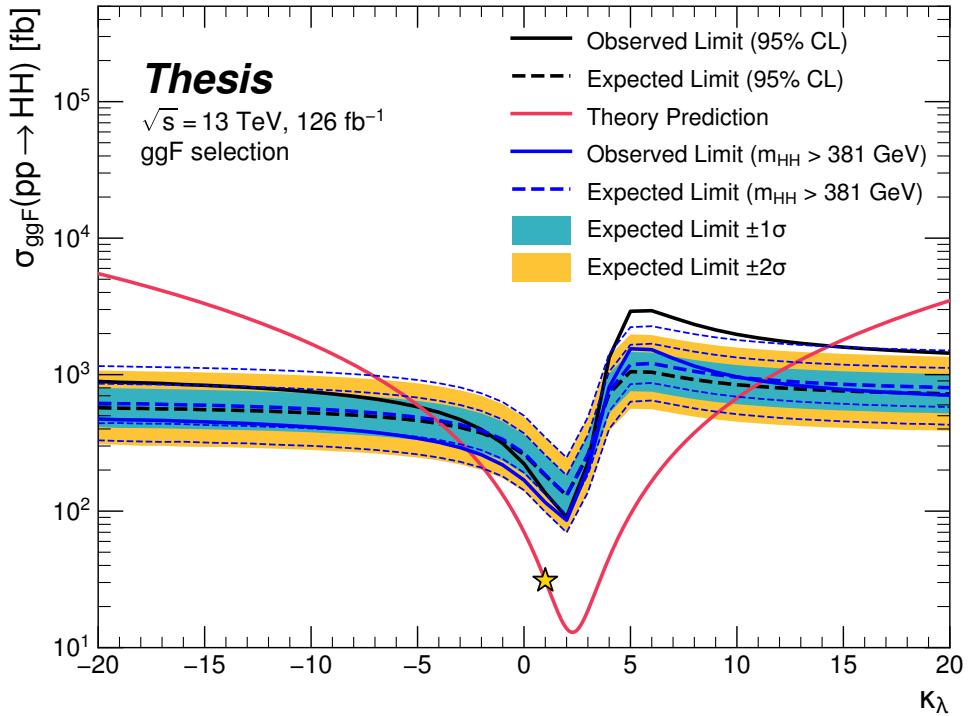


Figure 7.76: Comparison of the limits in Figure 7.75 with an equivalent set of limits that drop the m_{HH} bins below 381 GeV, with the value of 381 GeV determined by the optimized variable width binning. The expected limit band with this mass cut is shown in dashed blue, and the observed is shown in solid blue. The excess at and above $\kappa_\lambda = 5$ is significantly reduced, demonstrating that this is driven by low mass. Notably, there is minimal impact on the expected sensitivity with this m_{HH} cut.

Chapter 8

FUTURE IDEAS FOR $HH \rightarrow b\bar{b}b\bar{b}$

The searches presented in this thesis make use of a large suite of sophisticated techniques, selected through careful study and validation. During this process, a variety of interesting directions for the $HH \rightarrow b\bar{b}b\bar{b}$ analysis were explored by this thesis author, in collaboration with a few others¹, but were not used due to a variety of constraints. We present two such interesting directions here, with the hope of encouraging further exploration of these techniques in future work.

2425 8.1 pairAGraph: A New Method for Jet Pairing

As discussed in Chapter 7, one of the main problems to solve is the pairing of b -jets into Higgs candidates. Figure 7.1 demonstrates that the choice of the pairing method, while important for achieving good reconstruction of signal events, also significantly impacts the structure of non- HH events, leading to various biases in the background estimate. Evaluation of the pairing method therefore must take both of these factors into account. While we have presented some advantages in respective contexts for the pairing methods considered here, we of course would like to explore further improvements to this important component of the analysis.

To that end, we note that all of the pairing methods considered here share a common feature: four jets are selected, and the pairing is some discrimination between the available three pairings of these four jets. For the methods used in this analysis, the jet selection proceeds via a simple p_T ordering, with b -tagged jets receiving a higher priority than non-

¹Notably Nicole Hartman (SLAC), who spearheaded much of the development and proof of concept work, in collaboration with Michael Kagan and Rafael Teixeira De Lima.

2438 tagged jets.

2439 With the advent of a variety of machine learning methods for dealing with a variable number
2440 of inputs (e.g. recurrent neural networks [112], deep sets [113], graph neural networks [114],
2441 and transformers [115]), a natural place to improve on the pairing is to consider more than
2442 just four jets. The pairing and jet selection is then performed simultaneously, allowing for
2443 the incorporation of more event information in the pairing decision and the incorporation of
2444 jet correlation structure in the jet selection.

2445 In practice, the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets which pass the
2446 kinematic preselection, and any gain from this additional freedom would come from events
2447 with greater than or equal to five jets. However, this five jet topology is particularly exciting
2448 for scenarios such as events with initial state radiation (ISR), in which the $HH - > 4b$ jets
2449 are offset by a single jet with p_T similar in magnitude to that of the $HH - > 4b$ system.
2450 Such events have explicit event level information which is not encoded with the inclusion
2451 of only the $HH - > 4b$ jets, and are pathological if the ISR jet happens to pass b -tagging
2452 requirements.

2453 Additionally, with the use of lower tagged regions for background estimation and alternate
2454 signal regions, this extra flexibility in jet selection may provide a very useful bias – since the
2455 algorithm is trained on signal, the selected jets for the pairing will be the most “4b-like” jets
2456 available in the considered set.

2457 For the studies considered here, a transformer [115] based architecture is used. This is
2458 best visualized by considering the event as a graph with jets corresponding to nodes and edges
2459 corresponding to potential connections – for this reason, we term this algorithm “pairAGraph”.
2460 The approach is as follows: each jet, i , is represented by some vector of input variables, \vec{x}_i ,
2461 in our case the four-vector information, (p_T, η, ϕ, E) of each jet, plus information on the
2462 b -tagging decision. A multi-layer perceptron (MLP) is used to create a latent embedding,
2463 $\mathbf{h}(\vec{x}_i)$, of this input vector.

To describe the relationship between various jets in the event, we then define a vector \vec{z}_i

for each jet as

$$\vec{z}_i = \sum_j w_{ij} \mathbf{h}(\vec{x}_j) \quad (8.1)$$

where j runs over all jets in the event (including $i = j$), the w_{ij} can be thought of as edge weights, and $\mathbf{h}(\vec{x}_j)$ is the latent embedding for jet j mentioned above.

Within this formula, both \mathbf{h} and the w_{ij} are learnable. To learn an appropriate latent mapping and set of edge weights, we define a similarity metric corresponding to each possible jet pairing:

$$\vec{z}_{1a} \cdot \vec{z}_{1b} + \vec{z}_{2a} \cdot \vec{z}_{2b} \quad (8.2)$$

where subscripts $1a$ and $1b$ correspond to the two jets in pair 1, $2a$ and $2b$ to the jets in pair 2 for a given pairing of four distinct jets.

This similarity metric is calculated for all possible pairings, which are then passed through a softmax [116] activation function, which compresses these scores to between 0 and 1 with sum of 1, lending an interpretation as probability of each pairing.

In training, the ground truth pairing is set by *truth matching* jets to the b -jets in the HH signal simulation – a jet is considered to match if it is < 0.3 in ΔR away from a b -jet in the simulation record. Given this ground truth, a cross-entropy loss *TODO: cite* is used on the softmax outputs, and w_{ij} and \mathbf{h} are updated correspondingly. Training in such a way corresponds to updating w_{ij} and \mathbf{h} to maximize the similarity metric for the correct pairing.

In evaluation, the pairings with a higher score (and therefore higher softmax output) given the trained h and w_{ij} therefore correspond to the pairings that are most “ HH -like”. The maximum over these scores is therefore the pairing used as the predicted result from the algorithm.

Because the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets, it was found to be sufficient to only consider a maximum of 5 jets. Consideration of more is in principle possible, but the quickly expanding combinatorics leads to a rapidly more difficult problem. The jets considered are the five leading jets in p_T . Notably, this set of jets may include jets which are not b -tagged, even for the nominal 4 b region – therefore events with 4 b -tagged jets

2485 are not required to use all of them in the construction of Higgs candidates, in contrast to the
 2486 other algorithms used in this thesis.

2487 A comparison of the pairAGraph jet selection with the baseline selection used in Chapter 7
 2488 is considered in Table 8.1 for the MC16a Standard Model non-resonant signal. As a reminder,
 2489 the baseline selection orders jets by p_T , selecting first the highest p_T b -tagged jets (according
 2490 to the b -tag region definition) and then the highest p_T non-tagged jets. The first four jets in
 2491 this ordering are used.

2492 For the comparison presented in Table 8.1, only the leading five jets are considered in
 2493 applying both algorithms in order to compare results on more equal footing. The numbers
 2494 shown are the percent of the time that the correct jets are selected for the Higgs candidates
 2495 by each algorithm, given that the correct jets fall within these leading five jets, where “correct”
 2496 here means truth matched to the corresponding b -quarks. pairAGraph demonstrates a slight
 2497 improvement over the baseline for $4b$, which widens when considering lower b -tag categories.
 2498 Given that four b -quarks are present in all of these categories, this suggests that pairAGraph
 2499 is able to recover information in the case of, e.g., mis-tagged jets.

2500 Table 8.2 compares the HH pairing accuracy of a few different pairing algorithms for
 2501 the Standard Model signal. Notably, pairAGraph demonstrates a higher pairing accuracy
 2502 immediately after paring, but all methods are quite comparable after the full analysis selection.
 2503

2504 As mentioned in Chapter 7, though the pairing is quite important for signal events, it also
 2505 must be applied to events in data, where the overwhelming majority of events do not contain
 2506 HH . Though in general, pairing methods select for an HH -like topology, the additional
 2507 flexibility of pairAGraph to choose which jets enter the candidate HH system provides an
 2508 additional handle to shape the kinematics of events in data. Examples of this impact are seen
 2509 in Figures 8.1 and 8.2, which compare the $2b$ and $4b$ distributions of p_T of the HH candidate
 2510 system between BDT pairing and pairAGraph pairing before and after reweighting. $HH p_T$
 2511 was chosen as it is a variable which demonstrates both a large difference between $2b$ and $4b$
 2512 and a residual mis-modeling after reweighting. As can be seen in Figure 8.1, the $2b$ and $4b$

4b correct jets	96.7%	96.0%
3b+1 loose correct jets	96.3%	95.2%
3b correct jets	91.6%	83.2%

Table 8.1: Percent of the time that the correct jets are selected for the Higgs candidates by each algorithm, given that the correct jets fall within the set of considered jets, where “correct” here means truth matched to the corresponding b -quarks. Only the leading five jets are considered in the assessment of both algorithms. Definitions of the $4b$ and $3b + 1$ loose categories are as described in Section 7.3, where $3b$ requires three b -tagged jets and the fourth jet is untagged. pairAGraph demonstrates a slight improvement over the baseline for $4b$, which widens when considering lower b -tag categories. Given that four b -quarks are present in all of these categories, this suggests that pairAGraph is able to recover information in the case of, e.g., mis-tagged jets.

	After Pairing	After Full Selection
D_{HH}	71.8%	93.6%
$\min \Delta R$	69.7%	94.7%
pairAGraph	78.4%	94.2%

Table 8.2: Pairing accuracy evaluated for the Standard Model signal (MC16a), comparing D_{HH} and $\min \Delta R$ (discussed in Chapter 7) with pairAGraph trained on the Standard Model signal. Numbers are shown both immediately after pairing and after the full analysis selection. pairAGraph demonstrates a 7-8% higher accuracy than the other algorithms immediately after pairing, but all methods are quite comparable after the full analysis selection.

2513 distributions are more similar before reweighting with pairAGraph. Figure 8.2 further shows
 2514 that the residual mis-modeling after reweighting is reduced, along with the corresponding
 2515 uncertainty. While this is not fully conclusive, it provides a hint that the jets chosen for the
 2516 2b event HH candidate system may be more “4b-like” than the jets chosen with the baseline
 2517 selection.

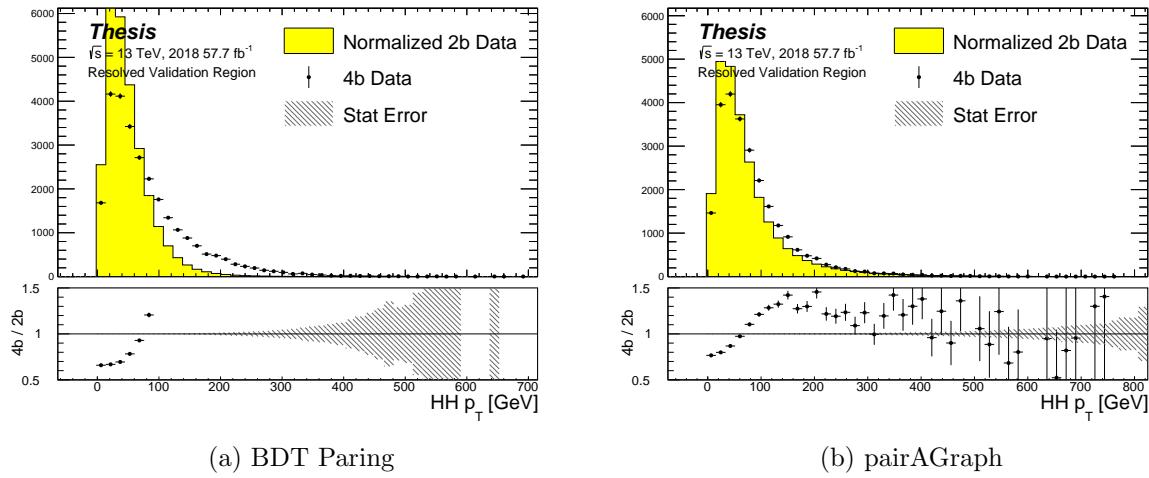


Figure 8.1: Comparison of distributions of $HH p_T$ in the 2018 resonant validation region before reweighting for BDT pairing (left) and pairAGraph (right). $HH p_T$ is a variable with a large difference between 2b and 4b, but the relative shapes seem to be more similar for pairAGraph than for BDT paring, corresponding to the hypothesis that pairAGraph chooses more “4b-like” jets.

2518 8.2 Background Estimation with Mass Plane Interpolation

2519 The choice of a pairing algorithm that results in a smooth mass plane (such as $\min \Delta R$)
 2520 opens up a variety of options for the background estimation. While the method based on
 2521 reweighting of 2b events used for this thesis performs well and has been extensively studied
 2522 and validated, it also relies on several assumptions. In particular, the reweighting is derived

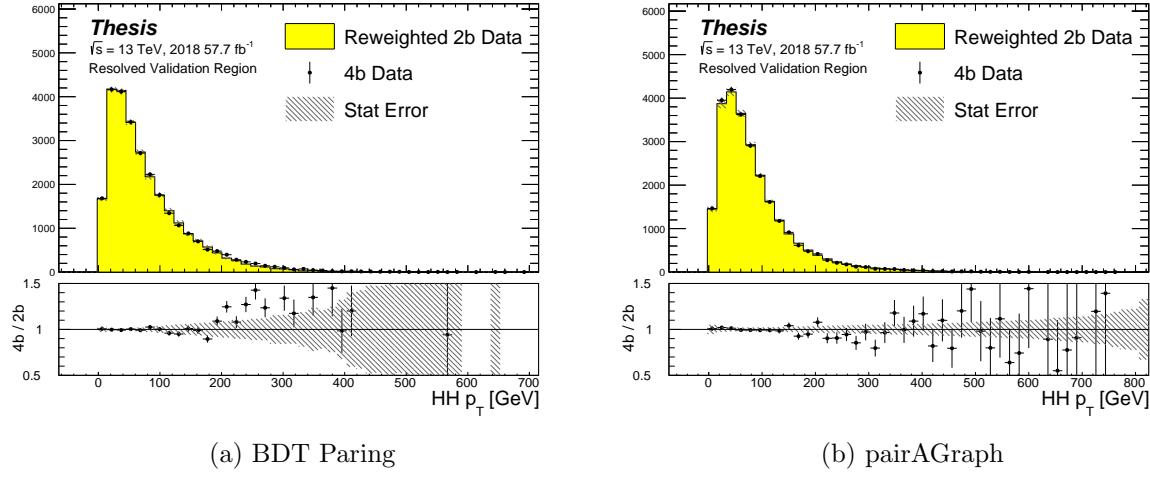


Figure 8.2: Comparison of distributions of $HH p_T$ in the 2018 resonant validation region after reweighting for BDT pairing (left) and pairAGraph (right). $HH p_T$ is a variable with a large difference between $2b$ and $4b$, and the reweighted agreement in the high p_T tail is significantly improved with pairAGraph, with a corresponding reduction in the assigned bootstrap uncertainty in that region.

2523 between e.g., $2b$ and $4b$ events *outside* of the signal region and then applied to $2b$ events *inside*
 2524 the signal region, with the assumption that the $2b$ to $4b$ transfer function will be sufficiently
 2525 similar in both regions of the mass plane. An uncertainty is assigned to account for the bias
 2526 due to this assumption, but the extrapolation in the mass plane is never explicitly treated in
 2527 the nominal estimate. While the approach of reweighting $2b$ events within the signal region
 2528 does have the advantage of incorporating explicit signal region information (that is, the $2b$
 2529 signal region events), the importance of the extrapolation bias motivates consideration of
 2530 a method that operates within the $4b$ mass plane. This additionally removes the reliance
 2531 on lower b -tagging regions, allowing for the use of, e.g. $3b$ triggers, and future-proofing the
 2532 analysis against trigger bandwidth constraints in the low tag regions.

2533 The pairAGraph pairing method discussed in the previous section was developed concur-

2534 rently with these studies and demonstrates good properties for an interpolated estimate (as
2535 shown below), and is therefore used in the following.

The method considered here relies on the following: for a given vector of input variables (event kinematics, etc), \vec{x} , the joint probability in the HH mass plane may be written as:

$$p(\vec{x}, m_{H1}, m_{H2}) = p(\vec{x}|m_{H1}, m_{H2})p(m_{H1}, m_{H2}) \quad (8.3)$$

2536 by the chain rule of probability. This means that the full dynamics of events in the HH mass
2537 plane may be described by (1) the conditional probability of considered variables \vec{x} , given
2538 values of m_{H1} and m_{H2} , and (2) the density of the mass plane itself.

2539 We present here an approach which uses normalizing flows [117] to model the conditional
2540 probabilities of events in the mass plane and Gaussian processes to model the mass plane
2541 density. These models are trained in a region around, but not including, the signal region,
2542 and the trained models are then used to construct an *interpolated* estimate of the signal
2543 region kinematics. This approach therefore explicitly treats event behavior within the mass
2544 plane, avoiding the concerns associated with a reweighted estimate. Validation of such a
2545 method, as well as assessing of closure and biases of the method, may be done in alternate
2546 b -tagging or kinematic regions, notably the now unused $2b$ region, results of which are shown
2547 below.

2548 8.2.1 Normalizing Flows

Normalizing flows model observed data $x \in X$, with $x \sim p_X$, as the output of an invertible, differentiable function $f : X \rightarrow Z$, with $z \in Z$ a latent variable with a simple prior probability distribution (often standard normal), $z \sim p_Z$. From a change of variables, given such a function, we may write

$$p_X(x) = p_Z(f(x)) \left| \det \left(\frac{d(f(x))}{dx} \right) \right| \quad (8.4)$$

2549 where $\left(\frac{d(f(x))}{dx} \right)$ is the Jacobian of f at x .

2550 The problem of normalizing flows then reduces to (1) choosing sets of f which are both
2551 tractable and sufficiently expressive to describe observed data, and (2) optimizing associated

sets of functional parameters on observed data via maximum likelihood estimation using the above formula. Sampling from the learned density is done by drawing from the latent distribution $z \sim p_Z$ (cf. inverse transform sampling) – the corresponding sample is then $x \sim p_X$ with $x = f^{-1}(z)$.

A standard approach to the definition of these f is as a composition of affine transformations (e.g. RealNVP [118]), that is, transformations of the form $\alpha z + \beta$, with α and β learnable parameter vectors. This can roughly be thought of as shifting and squeezing the input prior density in order to match the data density. However, this has somewhat limited expressivity, for instance in the case of a multi-modal density.

This work thus instead relies on neural spline flows [119] in which the functions considered are monotonic rational-quadratic splines, which have an analytic inverse. A rational quadratic function has the form of a quotient of two quadratic polynomials, namely,

$$f_j(x_i) = \frac{a_{ij}x_i^2 + b_{ij}x_{ij} + c_{ij}}{d_{ij}x_i^2 + e_{ij}x_i + f_{ij}} \quad (8.5)$$

with six associated parameters (a_{ij} through f_{ij}) per each piecewise bin j of the spline and each input dimension i . This is explicitly more flexible and expressive than a simple affine transformation, allowing, e.g., the treatment of multi-modality via the piecewise nature of the spline.

The rational quadratic spline is defined on a set interval. The transformation outside of this interval is set to the identity, with these linear tails allowing for unconstrained inputs. The boundaries between bins of the spline are set by coordinates called *knots*, with $K + 1$ knots for K bins – the two endpoints for the spline interval plus the $K - 1$ internal boundaries. The derivatives at these points are constrained to be positive for the internal knots, and boundary derivatives are set to 1 to match the linear tails.

The bin widths and heights are learnable ($2 \cdot K$ parameters) as are the internal knot derivatives ($K - 1$ parameters), and these $3K - 1$ outputs of the neural network are sufficient to define a monotonic rational-quadratic spline which passes through each knot and has the given derivative value at each knot.

2575 In the context of the $HH \rightarrow 4b$ analysis, a neural spline flow is used to model the four
 2576 vector information of each Higgs candidate, conditional on their respective masses. The
 2577 resulting flow is therefore five dimensional, with inputs $x = (p_{T,H1}, p_{T,H2}, \eta_{H1}, \eta_{H2}, \Delta\phi_{HH})$,
 2578 where the ATLAS ϕ symmetry has been encoded by assuming $\phi_{H1} = 0$. Conditional variables
 2579 m_{H1} and m_{H2} are not modeled by the flow, but “come along for the ride”. A standard normal
 2580 distribution in 5 dimensions is used for the underlying prior. Modeling of the four vectors
 2581 was chosen in order to reduce bias from modeling m_{HH} directly.

2582 The trained flow model then gives a model for $p(x|m_{H1}, m_{H2})$ which may be sampled
 2583 from to reconstruct distributions of HH kinematics given values of m_{H1} and m_{H2} .

2584 8.2.2 Gaussian Processes

2585 The second piece of this background estimate is the modeling of the mass plane density,
 2586 $p(m_{H1}, m_{H2})$. This is done using Gaussian process regression – note that a similar procedure
 2587 is used to define a systematic in the boosted $4b$ analysis. Generally, Gaussian processes
 2588 are a collection of random variables in which every finite collection of said variables is
 2589 distributed according to a multivariate normal distribution. For the context of Gaussian
 2590 process regression, what we consider is a Gaussian process over function space, that is, for a
 2591 collection of points, x_1, \dots, x_N , the space of corresponding function values, $(f(x_1), \dots, f(x_N))$
 2592 is Gaussian process distributed, that is, described by an N dimensional normal distribution
 2593 with mean μ , covariance matrix Σ .

2594 For a single point, this would correspond to a function space described entirely by a
 2595 normal distribution, with various samples from that distribution yielding various candidate
 2596 functions. For multiple points, a covariance matrix describes the relationship between each
 2597 pair of points – correspondingly, it is represented via a *kernel function*, $K(x, x')$. As, in
 2598 practice, μ may always be set to 0 via a centering of the data, the kernel function fully defines
 2599 the considered family of functions.

The considered family of functions describes a Bayesian *prior* for the data. This prior
 may be conditioned on a set of training data points (X_1, \vec{y}_1) . This conditional *posterior* may

then be used to make predictions $\vec{y}_2 = f(X_2)$ at a set of new points X_2 . Because of the Gaussian process prior assumption, \vec{y}_1 and \vec{y}_2 are assumed to be jointly Gaussian. We may therefore write

$$\begin{pmatrix} \vec{y}_1 \\ \vec{y}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(X_1, X_1) & K(X_1, X_2) \\ K(X_1, X_2) & K(X_2, X_2) \end{pmatrix} \right) \quad (8.6)$$

where we have used that the kernel function is symmetric and assumed prior mean 0.

By standard conditioning properties of Gaussian distributions,

$$\vec{y}_2 | \vec{y}_1 \sim \mathcal{N}(K(X_2, X_1)K(X_1, X_1)^{-1}\vec{y}_1, K(X_2, X_2) - K(X_2, X_1)K(X_1, X_1)^{-1}K(X_1, X_2)) \quad (8.7)$$

which is the sampling distribution for a Gaussian process given kernel K . In practice, the mean of this sampling distribution is used as the function estimate, with an uncertainty from the predicted variance at a given point.

The choice of kernel function has a very strong impact on the fitted curve, and must therefore be chosen to express the expected dynamics of the data. A common such choice is a radial basis function (RBF) kernel, which takes the form

$$K(x, x') = \exp \left(-\frac{d(x, x')^2}{2l^2} \right) \quad (8.8)$$

where $d(\cdot, \cdot)$ is the Euclidean distance and $l > 0$ is a length scale parameter. Conceptually, as distances $d(x, x')$ increase relative to the chosen length scale, the kernel smoothly dies off – further away points influence each other less.

Coming back to our case of the mass plane, the procedure runs as follows:

1. A binned 2d histogram of the blinded mass plane is created in a window around the “standard” analysis regions. Bins which have any overlap with the signal region are excluded.
2. A Gaussian process is trained using the bin centers, values as training points. The scikit-learn implementation [120] is used, with RBF kernel with anisotropic length scale (l is dimension 2). The length scale is initialized to $(50, 50)$ to cover the signal region,

2614 and optimized by minimizing the negative log-marginal likelihood on the training data,
 2615 $-\log p(\vec{y}|\theta)$. Training data is centered and scaled to mean 0, variance 1, and a statistical
 2616 error is included in the fit.

2617 3. The Gaussian process is then used to predict the density $p(m_{H1}, m_{H2})$ in the signal
 2618 region. This may then be sampled from via an inverse transform sampling to generate
 2619 values (m_{H1}, m_{H2}) according to the density (specifically, according to the mean of the
 2620 Gaussian process posterior). Though in principle the Gaussian process sampling is not
 2621 limited to bin centers, this is kept for simplicity, with a uniform smearing applied within
 2622 each sampled bin to approximate the continuous estimate, namely, if a bin is sampled
 2623 from, the returned value is drawn uniformly at random within the sampled bin.

2624 4. The sampling in the previous step can be arbitrary – to set the overall normalization,
 2625 a Monte Carlo sampling of the Gaussian process is done to approximate the relative
 2626 fraction of events predicted both inside (f_{in}) and outside (f_{out}) of the signal region,
 within the training box. The number of events outside of the signal region (n_{out}) is
 known, therefore, the number of events inside of the signal region, n_{in} , may be estimated
 as

$$n_{in} = \frac{n_{out}}{f_{out}} \cdot f_{in}. \quad (8.9)$$

2627 Note that the Monte Carlo sampling procedure is simply a set of samples of the Gaussian
 2628 process from uniformly random values of m_{H1}, m_{H2} , and is the most convenient approach
 given the irregular shape of the signal region.

2629 This procedure results in a generated set of predicted m_{H1}, m_{H2} values for signal region
 2630 background events, along with an overall yield prediction.

2629 8.2.3 The Full Prediction

2630 Given the normalizing flow parametrization of $p(x|m_{H1}, m_{H2})$ and the Gaussian process
 2631 generation of $(m_{H1}, m_{H2}) \sim p(m_{H1}, m_{H2})$ and prediction of the signal region yield, all of the

2632 pieces are in place to construct an interpolation background estimate. Namely

- 2633 1. Gaussian process sampled (m_{H1}, m_{H2}) values are provided to the normalizing flow to predict the other variables for the Higgs candidate four-vectors. These are used to construct the HH system (notably m_{HH}).
- 2636 2. These final distributions are normalized according to the predicted background yield.

2637 *8.2.4 Results*

2638 All of the following results use the pairAGraph pairing algorithm, and reweighted results use 2639 the region definitions from the resonant analysis.

2640 The Gaussian process sampling procedure is trained on a small fraction (0.01) of $2b$ data 2641 to mimic the available $4b$ statistics. This fraction of $2b$ data is blinded, and the prediction of 2642 the estimate trained on this blinded region may then be compared to real $2b$ data in the signal 2643 region. The predictions for signal region m_{H1} and m_{H2} individually are shown in Figure 8.3, 2644 and the resulting mass planes are compared in Figure 8.4. Good agreement is seen.

2645 The $4b$ region is kept blinded for this work, meaning that a direct comparison of the 2646 Gaussian process estimate in the $4b$ signal region is not done. However, a Gaussian process is 2647 trained on the blinded $4b$ region and compared to the corresponding reweighted $2b$ estimate, 2648 trained per the nominal procedures from the analyses above. The predictions for signal 2649 region m_{H1} and m_{H2} individually are shown in Figure 8.5, compared to both the control and 2650 validation region derived reweighting estimates, and the resulting signal region mass planes 2651 are compared in Figure 8.6. The estimates are seen to be compatible.

2652 The Gaussian process estimate may then be used as an input to the normalizing flow 2653 estimate to form a complete background estimate. Figure 8.7 shows such an estimate for the 2654 subsampled $2b$ signal region. Results for the prediction of the normalizing flow with inputs of 2655 real $2b$ signal region m_{H1} and m_{H2} are compared to the results of using Gaussian process 2656 predicted m_{H1} and m_{H2} , and are seen to be consistent, demonstrating the above closure of

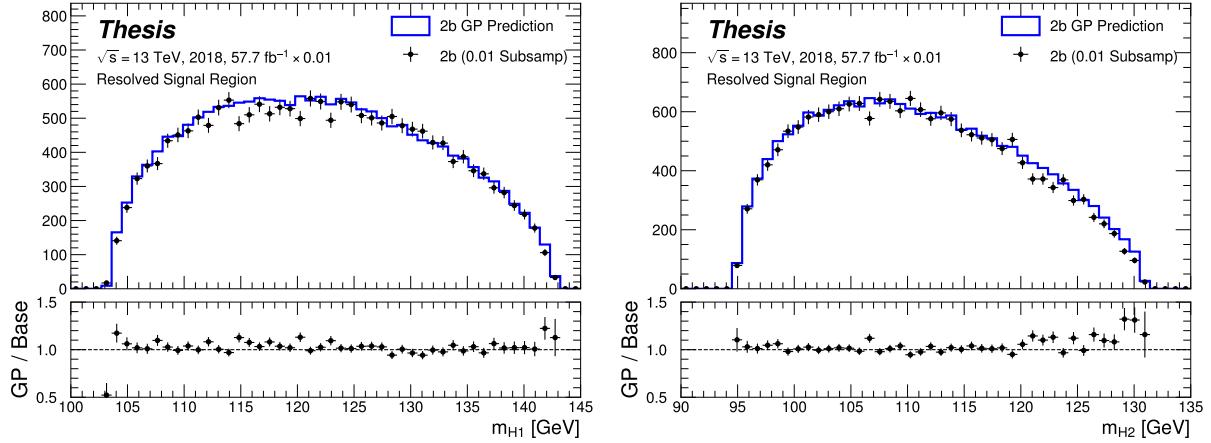


Figure 8.3: Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for $2b$ signal region events compared to real $2b$ signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics.

2657 the Gaussian process prediction. Reasonable agreement with real $2b$ signal region data is
2658 seen.

2659 Figure 8.8 demonstrates the application of this process to the $4b$ region, closely following
2660 how such an estimate would be used in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. As the $4b$ signal region
2661 is kept blinded for these studies, no direct evaluation is made, but results are compared to
2662 a resonant control region derived reweighting. Both signal region predictions are seen to
2663 be comparable, though there are some systematic differences. However, only the nominal
2664 estimates are compared here, with assessment of uncertainties on the interpolated estimate
2665 left for future work.

2666 8.2.5 Outstanding Points

2667 While good performance is demonstrated from the nominal interpolated background estimate,
2668 various uncertainties must be assigned according to the various stages of the estimate. These

2669 notably include

2670 • Assessing a statistical uncertainty on the normalizing flow training (cf. bootstrap
2671 uncertainty).

2672 • Propagation of the Gaussian process uncertainty through the sampling procedure.

2673 • Validation of the resulting estimate and assessment of necessary systematic uncertainties
2674 (e.g. from validation region non-closure).

2675 These are all quite tractable, but some, especially the choice of an appropriate systematic
2676 uncertainty, are certainly not obvious and require detailed study. In this respect, the
2677 reweighting validation work of the non-resonant analysis is certainly quite useful as a starting
2678 place in terms of the available regions and their correspondence to the nominal $4b$ signal
2679 region.

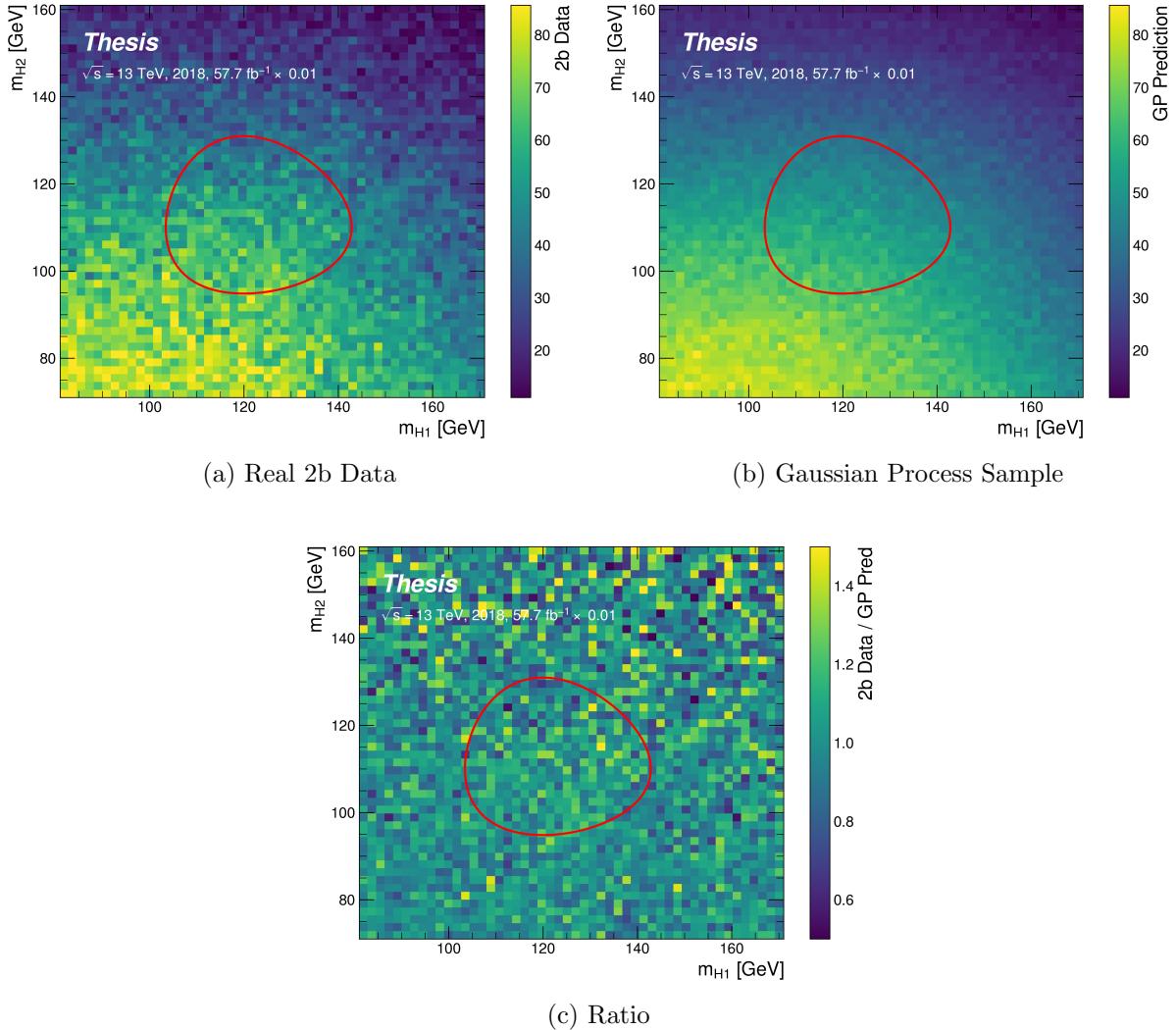


Figure 8.4: Gaussian process sampling prediction for the mass plane compared to the real 2b dataset for 2018. Only a small fraction (0.01) of the 2b dataset is used for both training and this final comparison to mimic 4b statistics. Good agreement is seen.

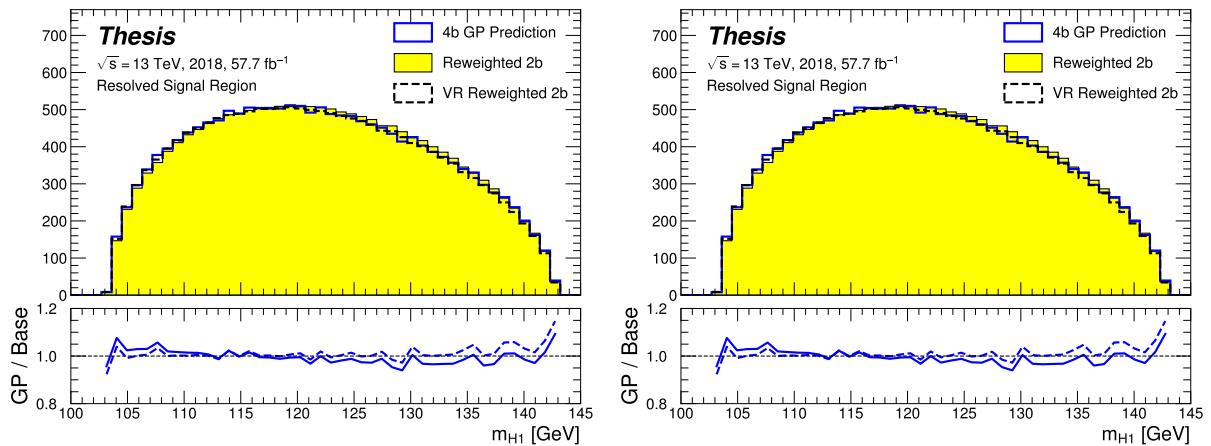


Figure 8.5: Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for 4b signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.

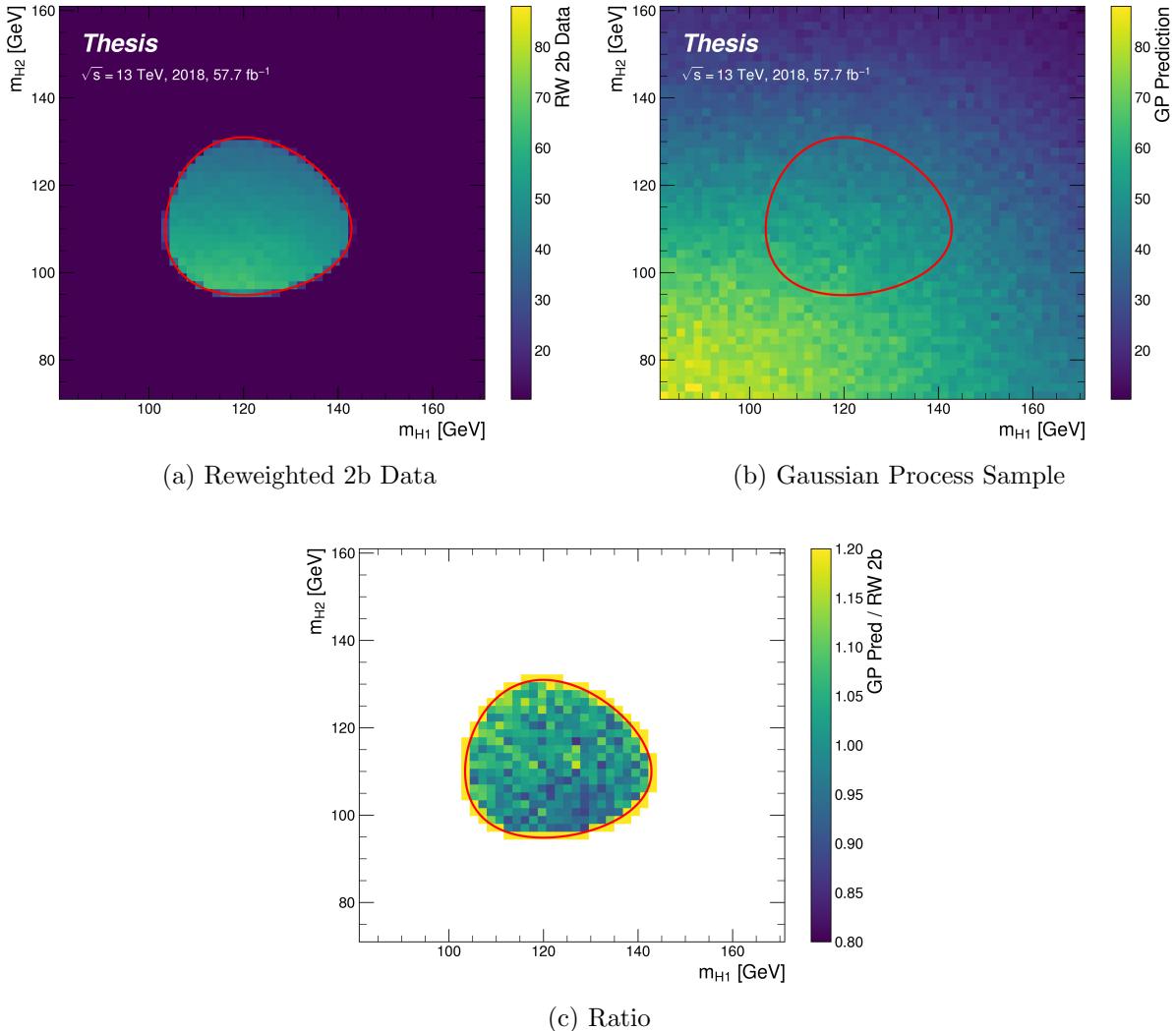


Figure 8.6: Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible.

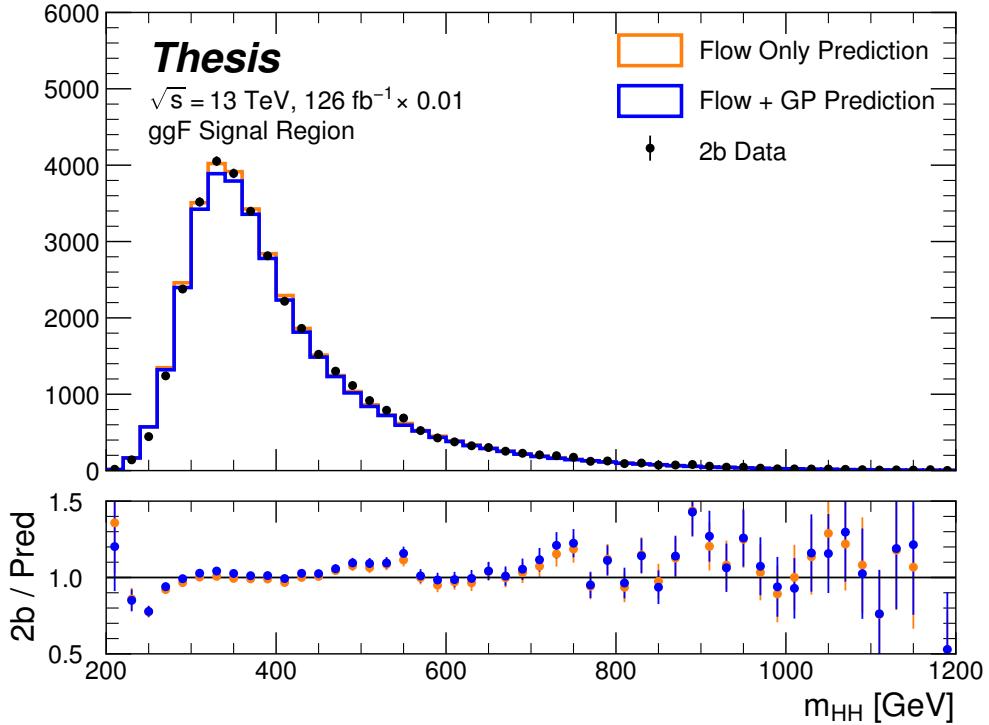


Figure 8.7: Comparison of the interpolation background estimate with real 2b data in the signal region. Only 1 % of 2b data is used in order to mimic 4b statistics, and results are presented here summed across years. The “Flow Only” prediction uses samples of actual 2b signal region data for the input values of m_{H_1} and m_{H_2} , whereas the “Flow + GP” prediction uses samples following the Gaussian process procedure above, more closely mimicking a the full background estimation procedure. The two predictions are quite comparable, demonstrating the closure of the Gaussian process estimate, and the predicted m_{HH} shape agrees well with 2b data. Only 2b statistical uncertainty is shown.

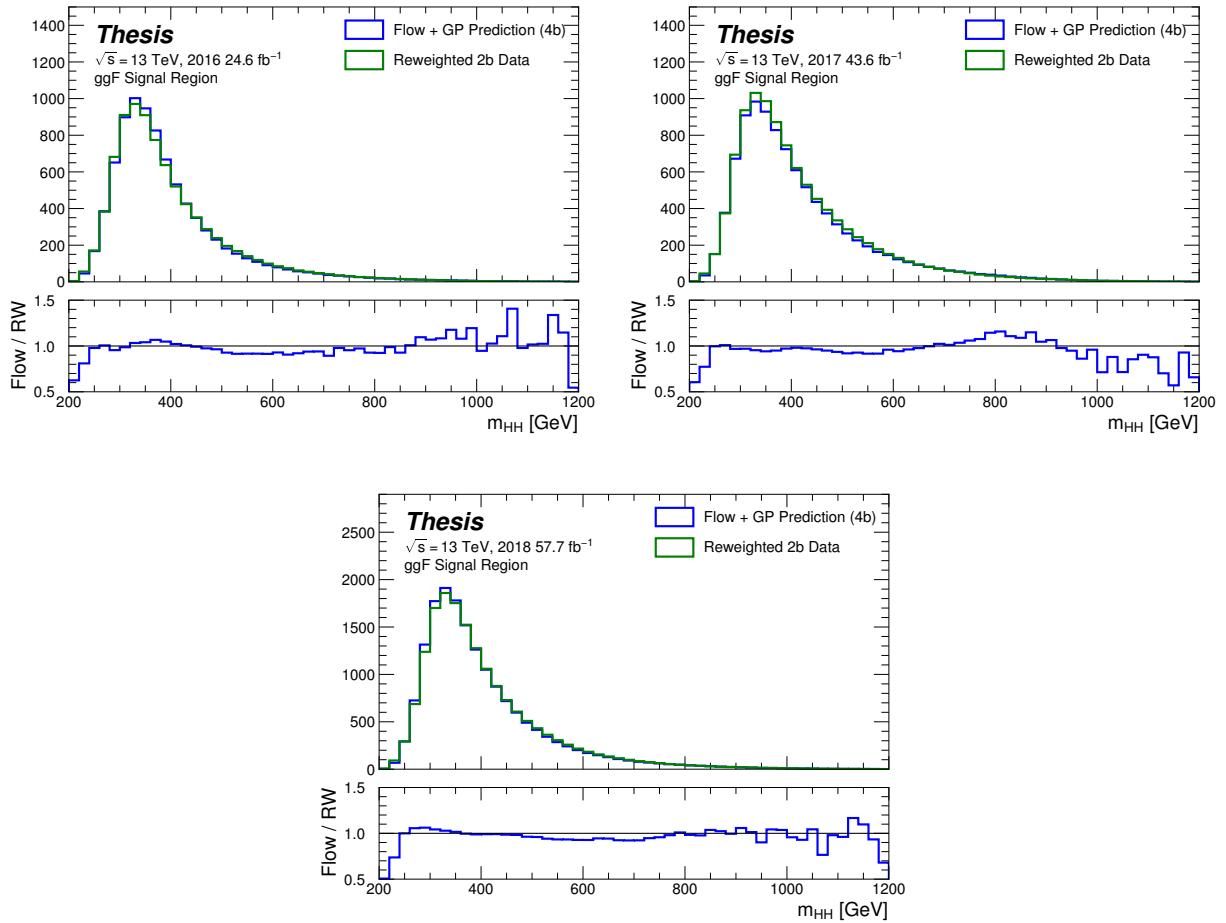


Figure 8.8: Comparison of the interpolation background estimate in the $4b$ signal region with the control region derived reweighted 2b estimate, shown for each year individually. Results are generally similar, within around 10 %.

2680

Chapter 9

2681

CONCLUSIONS

2682 This thesis has provided an overview of the Standard Model, with an emphasis on pair
2683 production of Higgs bosons and how this process may be used to both verify the Standard
2684 Model and to search for new physics. An overview of the Large Hadron Collider and the
2685 ATLAS detector has been provided, and the design and use of simulation infrastructure
2686 has been explained, including work to improve hadronic shower modeling in fast detector
2687 simulation. The translation of detector level information to analysis level information has
2688 been explained, with an emphasis on jets and the identification of B hadron decay. Finally,
2689 two searches for Higgs boson pair production have been presented, with a complete set of
2690 results for resonant production included, focusing on searches beyond the Standard Model,
2691 and a preliminary set of results for non-resonant production, targeting Standard Model
2692 production, with variations of the Higgs self-coupling. Two advanced techniques for the
2693 future of these analyses are further presented, along with proof-of-concept results.

2694

BIBLIOGRAPHY

- 2695 [1] *A New Map of All the Particles and Forces*, <https://www.quantamagazine.org/a-new-map-of-the-standard-model-of-particle-physics-20201022/>, Accessed:
2696
2697 2021-07-23 (cit. on p. 3).
- 2698 [2] J. D. Jackson, *Classical electrodynamics; 2nd ed.* Wiley, 1975 (cit. on p. 2).
- 2699 [3] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-
2700 107-03426-6 (cit. on p. 2).
- 2701 [4] M. Srednicki, *Quantum Field Theory*, Cambridge Univ. Press, 2007 (cit. on p. 2).
- 2702 [5] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964)
2703 214, ISSN: 0031-9163, URL: <https://www.sciencedirect.com/science/article/pii/S0031916364920013> (cit. on p. 11).
- 2704 [6] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*,
2705 tech. rep., CERN, 1964, URL: <https://cds.cern.ch/record/352337> (cit. on p. 11).
- 2706 [7] O. W. Greenberg, *Spin and Unitary-Spin Independence in a Paraquark Model of
2707 Baryons and Mesons*, Phys. Rev. Lett. **13** (20 1964) 598, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.598> (cit. on p. 11).
- 2708 [8] M. Y. Han and Y. Nambu, *Three-Triplet Model with Double $SU(3)$ Symmetry*, Phys.
2709 Rev. **139** (4B 1965) B1006, URL: <https://link.aps.org/doi/10.1103/PhysRev.139.B1006> (cit. on p. 11).
- 2710 [9] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36** (1968)
2711 1150, eprint: <https://doi.org/10.1119/1.1974382>, URL: <https://doi.org/10.1119/1.1974382> (cit. on p. 14).
- 2712
2713
2714
2715

- 2716 [10] *The Nobel Prize in Physics 1979*, <https://www.nobelprize.org/prizes/physics/1979/summary/>, Accessed: 2021-07-23 (cit. on p. 15).
- 2717
- 2718 [11] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2719 Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 20).
- 2720
- 2721 [12] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett.
2722 **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>
2723 (cit. on p. 20).
- 2724
- 2725 [13] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and
2726 Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 20).
- 2727
- 2728 [14] ATLAS Collaboration, *Observation of a new particle in the search for the Standard
2729 Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on p. 20).
- 2730
- 2731 [15] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS
2732 experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235)
(cit. on p. 20).
- 2733
- 2734 [16] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (21 1967) 1264, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264> (cit. on p. 25).
- 2735
- 2736 [17] ATLAS Collaboration, *Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion
2737 production using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
JHEP **07** (2020) 108, arXiv: [2001.05178 \[hep-ex\]](https://arxiv.org/abs/2001.05178) (cit. on pp. 28, 76, 95), Erratum:
2738 JHEP **01** (2021) 145.
- 2739
- 2740 [18] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics
Reports **516** (2012) 1, Theory and phenomenology of two-Higgs-doublet models,

- 2741 ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/S0370157312000695> (cit. on pp. 29, 30).
- 2742
- 2743 [19] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped gravitons at the CERN*
 2744 *LHC and beyond*, Phys. Rev. D **76** (3 2007) 036006, URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.036006> (cit. on pp. 29, 30).
- 2745
- 2746 [20] A. Carvalho, *Gravity particles from Warped Extra Dimensions, predictions for LHC*,
 2747 2018, arXiv: [1404.0102 \[hep-ph\]](https://arxiv.org/abs/1404.0102) (cit. on pp. 29, 30, 184).
- 2748 [21] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions*
 2749 *at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, Phys. Lett. B **800** (2020) 135103, arXiv:
 2750 [1906.02025 \[hep-ex\]](https://arxiv.org/abs/1906.02025) (cit. on pp. 30, 33, 76).
- 2751 [22] P. D. Group et al., *Review of Particle Physics*, Progress of Theoretical and Experimental
 2752 Physics **2020** (2020), 083C01, ISSN: 2050-3911, eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/34673722/ptaa104.pdf>, URL: <https://doi.org/10.1093/ptep/ptaa104> (cit. on p. 32).
- 2753
- 2754 [23] T. van Ritbergen and R. G. Stuart, *On the precise determination of the Fermi coupling*
 2755 *constant from the muon lifetime*, Nuclear Physics B **564** (2000) 343, ISSN: 0550-3213,
 2756 URL: [http://dx.doi.org/10.1016/S0550-3213\(99\)00572-6](https://dx.doi.org/10.1016/S0550-3213(99)00572-6) (cit. on p. 32).
- 2757
- 2758 [24] S. Dawson, A. Ismail, and I. Low, *What's in the loop? The anatomy of double Higgs*
 2759 *production*, Physical Review D **91** (2015), ISSN: 1550-2368, URL: [http://dx.doi.org/10.1103/PhysRevD.91.115008](https://dx.doi.org/10.1103/PhysRevD.91.115008) (cit. on p. 32).
- 2760
- 2761 [25] A. M. Sirunyan et al., *Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic*
 2762 *distributions in the dilepton final state in proton-proton collisions at $\sqrt{s}=13 \text{ TeV}$* ,
 2763 Physical Review D **102** (2020), ISSN: 2470-0029, URL: [http://dx.doi.org/10.1103/PhysRevD.102.092013](https://dx.doi.org/10.1103/PhysRevD.102.092013) (cit. on p. 32).
- 2764

- 2765 [26] S. Dawson, S. Dittmaier, and M. Spira, *Neutral Higgs-boson pair production at hadron*
2766 *colliders: QCD corrections*, Physical Review D **58** (1998), ISSN: 1089-4918, URL:
2767 <http://dx.doi.org/10.1103/PhysRevD.58.115012> (cit. on p. 33).
- 2768 [27] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading*
2769 *Order with Full Top-Quark Mass Dependence*, Physical Review Letters **117** (2016),
2770 ISSN: 1079-7114, URL: <http://dx.doi.org/10.1103/PhysRevLett.117.012001>
2771 (cit. on p. 33).
- 2772 [28] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass*
2773 *scheme*, The European Physical Journal C **79** (2019), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6973-3> (cit. on p. 33).
- 2775 [29] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-*
2776 *Leading Order in QCD*, Physical Review Letters **111** (2013), ISSN: 1079-7114, URL:
2777 <http://dx.doi.org/10.1103/PhysRevLett.111.201801> (cit. on p. 33).
- 2778 [30] D. Y. Shao, C. S. Li, H. T. Li, and J. Wang, *Threshold resummation effects in Higgs*
2779 *boson pair production at the LHC*, Journal of High Energy Physics **2013** (2013), ISSN:
2780 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)169](http://dx.doi.org/10.1007/JHEP07(2013)169) (cit. on p. 33).
- 2781 [31] D. de Florian and J. Mazzitelli, *Higgs pair production at next-to-next-to-leading*
2782 *logarithmic accuracy at the LHC*, 2015, arXiv: [1505.07122 \[hep-ph\]](https://arxiv.org/abs/1505.07122) (cit. on p. 33).
- 2783 [32] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*,
2784 Journal of High Energy Physics **2018** (2018), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2018\)059](http://dx.doi.org/10.1007/JHEP05(2018)059) (cit. on p. 33).
- 2786 [33] J. Baglio et al., *gg → HH: Combined uncertainties*, Physical Review D **103** (2021),
2787 ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.056002> (cit. on
2788 p. 33).

- 2789 [34] F. Dulat, A. Lazopoulos, and B. Mistlberger, *iHixs 2 — Inclusive Higgs cross sections*,
 2790 Computer Physics Communications **233** (2018) 243, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2018.06.025> (cit. on p. 34).
- 2791
 2792 [35] ATLAS Collaboration, *Measurement prospects of the pair production and self-coupling*
 2793 *of the Higgs boson with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-
 2794 2018-053, 2018, URL: <https://cds.cern.ch/record/2652727> (cit. on p. 34).
- 2795 [36] D. de Florian, I. Fabre, and J. Mazzitelli, *Triple Higgs production at hadron colliders*
 2796 *at NNLO in QCD*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL:
 2797 [http://dx.doi.org/10.1007/JHEP03\(2020\)155](http://dx.doi.org/10.1007/JHEP03(2020)155) (cit. on p. 34).
- 2798 [37] A. Papaefstathiou, T. Robens, and G. Tetlamatzi-Xolocotzi, *Triple Higgs boson*
 2799 *production at the Large Hadron Collider with Two Real Singlet scalars*, Journal of
 2800 High Energy Physics **2021** (2021), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP05\(2021\)193](http://dx.doi.org/10.1007/JHEP05(2021)193) (cit. on p. 34).
- 2801
 2802 [38] T. Robens, T. Stefaniak, and J. Wittbrodt, *Two-real-scalar-singlet extension of the*
 2803 *SM: LHC phenomenology and benchmark scenarios*, The European Physical Journal C
 2804 **80** (2020), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-020-7655-x> (cit. on p. 34).
- 2805
 2806 [39] D. Egana-Ugrinovic, S. Homiller, and P. Meade, *Multi-Higgs boson production probes*
 2807 *Higgs sector flavor*, Physical Review D **103** (2021), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.115005> (cit. on p. 34).
- 2808
 2809 [40] CERN, *CERN's accelerator complex*, <https://home.cern/science/accelerators/accelerator-complex>, Accessed: 27 July 2021 (cit. on p. 36).
- 2810
 2811 [41] J. Pequenao, “Computer generated image of the whole ATLAS detector”, 2008, URL:
 2812 <https://cds.cern.ch/record/1095924> (cit. on p. 38).
- 2813
 2814 [42] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
 JINST **3** (2008) S08003 (cit. on p. 39).

- 2815 [43] J. Pequenao and P. Schaffner, “How ATLAS detects particles: diagram of particle
 2816 paths in the detector”, 2013, URL: <https://cds.cern.ch/record/1505342> (cit. on
 2817 p. 40).
- 2818 [44] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-
 2819 TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633> (cit. on p. 41), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009,
 2820 2012, URL: <https://cds.cern.ch/record/1451888>.
- 2821 [45] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST
 2822 **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](https://arxiv.org/abs/1803.00844) (cit. on p. 41).
- 2823 [46] A. Outreach, “ATLAS Fact Sheet : To raise awareness of the ATLAS detector and
 2824 collaboration on the LHC”, 2010, URL: <https://cds.cern.ch/record/1457044>
 2825 (cit. on p. 43).
- 2826 [47] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, Eur. Phys.
 2827 J. C **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](https://arxiv.org/abs/1611.09661) (cit. on p. 44).
- 2828 [48] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-
 2829 PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 44).
- 2830 [49] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,
 2831 JHEP **11** (2004) 040, arXiv: [hep-ph/0409146 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0409146) (cit. on pp. 48, 79).
- 2832 [50] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton
 2833 shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: [0709.2092
 2834 \[hep-ph\]](https://arxiv.org/abs/0709.2092) (cit. on pp. 48, 79).
- 2835 [51] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO
 2836 calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010)
 2837 043, arXiv: [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581) (cit. on pp. 48, 79).
- 2838

- 2839 [52] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*
 2840 *differential cross sections, and their matching to parton shower simulations*, Journal of
 2841 High Energy Physics **2014** (2014), URL: [https://doi.org/10.1007/jhep07\(2014\)](https://doi.org/10.1007/jhep07(2014)079)
 2842 (cit. on pp. 48, 78).
- 2843 [53] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, The European Physical Journal
 2844 C **76** (2016), URL: <https://doi.org/10.1140/epjc/s10052-016-4018-8> (cit. on
 2845 pp. 48, 78).
- 2846 [54] M. Bähr et al., *Herwig++ physics and manual*, The European Physical Journal C
 2847 **58** (2008) 639, URL: <https://doi.org/10.1140/epjc/s10052-008-0798-9> (cit. on
 2848 pp. 48, 78).
- 2849 [55] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications
 2850 **191** (2015) 159, ISSN: 0010-4655, URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442> (cit. on pp. 48, 78).
- 2852 [56] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A
 2853 **462** (2001) 152 (cit. on pp. 48, 78).
- 2854 [57] S. Höche, *Introduction to parton-shower event generators*, 2015, arXiv: [1411.4085](https://arxiv.org/abs/1411.4085)
 2855 [[hep-ph](#)] (cit. on p. 49).
- 2856 [58] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods
 2857 in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated
 2858 Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 50).
- 2860 [59] P. Calafiura, J. Catmore, D. Costanzo, and A. Di Girolamo, *ATLAS HL-LHC Computing Conceptual Design Report*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2729668> (cit. on p. 50).
- 2862

- 2863 [60] A. Collaboration, *Computing and Software - Public Results*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults>, Accessed:
2864
2865 27 July 2021 (cit. on p. 51).
- 2866 [61] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, *Eur. Phys. J. C* **70**
2867 (2010) 823, arXiv: [1005.4568 \[physics.ins-det\]](https://arxiv.org/abs/1005.4568) (cit. on pp. 50, 78).
- 2868 [62] ATLAS Collaboration, *The new Fast Calorimeter Simulation in ATLAS*, ATL-SOFT-
2869 PUB-2018-002, 2018, URL: <https://cds.cern.ch/record/2630434> (cit. on pp. 52,
2870 53).
- 2871 [63] ROOT, *TPrincipal Class Reference*, <https://root.cern.ch/doc/master/classTPrincipal.html>, Accessed: 27 July 2021 (cit. on p. 52).
- 2872
2873 [64] M. Aharrouche et al., *Energy linearity and resolution of the ATLAS electromagnetic*
2874 *barrel calorimeter in an electron test-beam*, *Nuclear Instruments and Methods in*
2875 *Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated*
2876 *Equipment* **568** (2006) 601, ISSN: 0168-9002, URL: <http://dx.doi.org/10.1016/j.nima.2006.07.053> (cit. on p. 53).
- 2877
2878 [65] ATLAS Collaboration, *Evidence for prompt photon production in pp collisions at*
2879 $\sqrt{s} = 7 \text{ TeV}$ *with the ATLAS detector*, ATLAS-CONF-2010-077, 2010, URL: <https://cds.cern.ch/record/1281368> (cit. on p. 53).
- 2880
2881 [66] S. Gasiorowski and H. Gray, *Fast Calorimeter Simulation in ATLAS*, EPJ Web Conf.
2882 **245** (2020) 02002. 7 p, URL: <https://cds.cern.ch/record/2756298> (cit. on p. 54).
- 2883
2884 [67] D. P. Kingma and M. Welling, *Auto-Encoding Variational Bayes*, 2014, arXiv: [1312.6114 \[stat.ML\]](https://arxiv.org/abs/1312.6114) (cit. on p. 54).
- 2885
2886 [68] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance*
2887 *in LHC Run 1*, *The European Physical Journal C* **77** (2017), ISSN: 1434-6052, URL:
<http://dx.doi.org/10.1140/epjc/s10052-017-5004-5> (cit. on pp. 58, 59).

- 2888 [69] A. Ghosh and A. Collaboration, *Deep generative models for fast shower simulation*
 2889 *in ATLAS*, Journal of Physics: Conference Series **1525** (2020) 012077, URL: <https://doi.org/10.1088/1742-6596/1525/1/012077> (cit. on p. 60).
- 2890
 2891 [70] R. Di Sipio, M. F. Giannelli, S. K. Haghight, and S. Palazzo, *DijetGAN: a Generative-
 2892 Adversarial Network approach for the simulation of QCD dijet events at the LHC*,
 2893 Journal of High Energy Physics **2019** (2019), ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP08\(2019\)110](http://dx.doi.org/10.1007/JHEP08(2019)110) (cit. on p. 60).
- 2894
 2895 [71] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
 2896 *the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: [1703.10485](https://arxiv.org/abs/1703.10485) [hep-ex]
 2897 (cit. on pp. 63, 77).
- 2898
 2899 [72] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and*
 2900 *its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: [1603.02934](https://arxiv.org/abs/1603.02934)
 [hep-ex] (cit. on p. 63).
- 2901
 2902 [73] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
 2903 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017,
 URL: <https://cds.cern.ch/record/2268678> (cit. on pp. 63, 168).
- 2904
 2905 [74] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_T jet clustering algorithm*, Journal
 2906 of High Energy Physics **2008** (2008) 063, ISSN: 1029-8479, URL: <http://dx.doi.org/10.1088/1126-6708/2008/04/063> (cit. on p. 63).
- 2907
 2908 [75] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637,
 ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6>
 (cit. on p. 63).
- 2909
 2910 [76] A. Collaboration, *Configuration and performance of the ATLAS b-jet triggers in Run*
 2911 *2*, 2021, arXiv: [2106.03584](https://arxiv.org/abs/2106.03584) [hep-ex] (cit. on p. 67).

- 2912 [77] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency mea-*
 2913 *surement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019)
 2914 970, arXiv: 1907.05120 [hep-ex] (cit. on pp. 66, 77, 163).
- 2915 [78] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification*
 2916 *of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-
 2917 PHYS-PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (cit. on
 2918 p. 69).
- 2919 [79] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear
 2920 Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,
 2921 Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on
 2922 p. 69).
- 2923 [80] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
 2924 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL:
 2925 <https://cds.cern.ch/record/2255226> (cit. on p. 70).
- 2926 [81] *Expected performance of the 2019 ATLAS b-taggers*, <http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2019-005/>, Accessed: 2021-07-14 (cit. on
 2927 p. 72).
- 2928 [82] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final*
 2929 *state using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP
 01 (2019) 030, arXiv: 1804.06174 [hep-ex] (cit. on pp. 76, 89, 95).
- 2930 [83] ATLAS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay*
 2931 *mode at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **04** (2019) 092, arXiv: 1811.
 2932 04671 [hep-ex] (cit. on p. 76).
- 2933 [84] ATLAS Collaboration, *A search for resonant and non-resonant Higgs boson pair*
 2934 *production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the*

- 2938 *ATLAS detector*, Phys. Rev. Lett. **121** (2018) 191801, arXiv: 1808.00336 [hep-ex]
 2939 (cit. on p. 76), Erratum: Phys. Rev. Lett. **122** (2019) 089901.
- 2940 [85] ATLAS Collaboration, *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$*
 2941 *decay channel using ATLAS data recorded at $\sqrt{s} = 13 \text{ TeV}$* , JHEP **05** (2019) 124,
 2942 arXiv: 1811.11028 [hep-ex] (cit. on p. 76).
- 2943 [86] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state*
 2944 *with 13 TeV pp collision data collected by the ATLAS experiment*, JHEP **11** (2018)
 2945 040, arXiv: 1807.04873 [hep-ex] (cit. on p. 76).
- 2946 [87] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel*
 2947 *using pp collision data recorded at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, Eur. Phys.
 2948 J. C **78** (2018) 1007, arXiv: 1807.08567 [hep-ex] (cit. on p. 76).
- 2949 [88] ATLAS Collaboration, *Reconstruction and identification of boosted di- τ systems in*
 2950 *a search for Higgs boson pairs using 13 TeV proton–proton collision data in ATLAS*,
 2951 JHEP **11** (2020) 163, arXiv: 2007.14811 [hep-ex] (cit. on p. 76).
- 2952 [89] ATLAS Collaboration, *Search for non-resonant Higgs boson pair production in the*
 2953 *$b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13 \text{ TeV}$* , Phys.
 2954 Lett. B **801** (2020) 135145, arXiv: 1908.06765 [hep-ex] (cit. on p. 76).
- 2955 [90] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks*
 2956 *plus two final state with in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*,
 2957 ATLAS-CONF-2021-016, 2021, URL: <https://cds.cern.ch/record/2759683> (cit. on
 2958 p. 76).
- 2959 [91] CMS Collaboration, *Search for heavy resonances decaying to two Higgs bosons in final*
 2960 *states containing four b quarks*, Eur. Phys. J. C **76** (2016) 371, arXiv: 1602.08762
 2961 [hep-ex] (cit. on p. 77).

- 2962 [92] CMS Collaboration, *Search for production of Higgs boson pairs in the four b quark*
 2963 *final state using large-area jets in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , JHEP **01**
 2964 (2019) 040, arXiv: [1808.01473 \[hep-ex\]](https://arxiv.org/abs/1808.01473) (cit. on p. 77).
- 2965 [93] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in*
 2966 *Proton–Proton Collisions at $\sqrt{s} = 13 \text{ TeV}$* , Phys. Rev. Lett. **122** (2019) 121803, arXiv:
 2967 [1811.09689 \[hep-ex\]](https://arxiv.org/abs/1811.09689) (cit. on p. 77).
- 2968 [94] G. Bartolini et al., *Performance of the ATLAS b-jet trigger in $p\ p$ collisions at $\sqrt{s} =$*
 2969 *13 TeV* , tech. rep. ATL-COM-DAQ-2019-150, CERN, 2019, URL: <https://cds.cern.ch/record/2688819> (cit. on p. 77).
- 2971 [95] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054> (cit. on pp. 78, 163).
- 2974 [96] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43**
 2975 (2016) 023001, arXiv: [1510.03865 \[hep-ph\]](https://arxiv.org/abs/1510.03865) (cit. on pp. 79, 164).
- 2976 [97] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. **C58** (2008) 639, arXiv:
 2977 [0803.0883 \[hep-ph\]](https://arxiv.org/abs/0803.0883) (cit. on p. 79).
- 2978 [98] T. Head et al., *scikit-optimize/scikit-optimize: v0.5.2*, version v0.5.2, 2018, URL: <https://doi.org/10.5281/zenodo.1207017> (cit. on p. 84).
- 2980 [99] A. Carvalho et al., *Higgs pair production: choosing benchmarks with cluster analysis*,
 2981 Journal of High Energy Physics **2016** (2016) 1, ISSN: 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP04\(2016\)126](http://dx.doi.org/10.1007/JHEP04(2016)126) (cit. on p. 93).
- 2983 [100] A. Rogozhnikov, *Reweighting with Boosted Decision Trees*, Journal of Physics: Conference Series **762** (2016) 012036, ISSN: 1742-6596, URL: <http://dx.doi.org/10.1088/1742-6596/762/1/012036> (cit. on p. 95).
- 2986 [101] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood/Density*
 2987 *Ratio Estimation*, 2019, arXiv: [1911.00405 \[eess.SP\]](https://arxiv.org/abs/1911.00405) (cit. on p. 96).

- 2988 [102] T. Kanamori, S. Hido, and M. Sugiyama, *A Least-Squares Approach to Direct Im-*
 2989 *portance Estimation*, J. Mach. Learn. Res. **10** (2009) 1391, ISSN: 1532-4435 (cit. on
 2990 p. 96).
- 2991 [103] A. Andreassen and B. Nachman, *Neural networks for full phase-space reweighting*
 2992 *and parameter tuning*, Physical Review D **101** (2020), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.101.091901> (cit. on p. 97).
- 2994 [104] S. Fort, H. Hu, and B. Lakshminarayanan, *Deep Ensembles: A Loss Landscape Per-*
 2995 *spective*, 2020, arXiv: [1912.02757 \[stat.ML\]](https://arxiv.org/abs/1912.02757) (cit. on p. 98).
- 2996 [105] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1,
 2997 URL: <https://doi.org/10.1214/aos/1176344552> (cit. on p. 155).
- 2998 [106] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton*
 2999 *collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, (2020), arXiv: [2007 . 02645](https://arxiv.org/abs/2007.02645)
 3000 [[hep-ex](#)] (cit. on p. 163).
- 3001 [107] J. Adelman et al., *Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state with*
 3002 *the full Run 2 13 TeV pp collision data collected by the ATLAS Experiment Supporting*
 3003 *note.*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2711865> (cit. on
 3004 p. 163).
- 3005 [108] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring*
 3006 *in ATLAS*, JINST **13** (2018) P07017. 33 p, URL: <https://cds.cern.ch/record/2633501> (cit. on p. 163).
- 3008 [109] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature*
 3009 *of the Higgs Sector*, CERN Yellow Reports: Monographs, 869 pages, 295 figures, 248 ta-
 3010 *bles and 1645 citations*. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/>
 3011 CERN, 2016, URL: <https://cds.cern.ch/record/2227475> (cit. on p. 164).
- 3012 [110] A. L. Read, *Presentation of search results: the CLs technique*, Journal of Physics G:
 3013 Nuclear and Particle Physics **28** (2002) 2693, ISSN: 0954-3899 (cit. on p. 181).

- 3014 [111] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-*
3015 *based tests of new physics*, *The European Physical Journal C* **71** (2011), ISSN: 1434-6052
3016 (cit. on p. 181).
- 3017 [112] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-*
3018 *Term Memory (LSTM) network*, *Physica D: Nonlinear Phenomena* **404** (2020) 132306,
3019 ISSN: 0167-2789, URL: <http://dx.doi.org/10.1016/j.physd.2019.132306> (cit. on
3020 p. 188).
- 3021 [113] M. Zaheer et al., *Deep Sets*, 2018, arXiv: [1703.06114 \[cs.LG\]](https://arxiv.org/abs/1703.06114) (cit. on p. 188).
- 3022 [114] J. Zhou et al., *Graph Neural Networks: A Review of Methods and Applications*, 2021,
3023 arXiv: [1812.08434 \[cs.LG\]](https://arxiv.org/abs/1812.08434) (cit. on p. 188).
- 3024 [115] A. Vaswani et al., *Attention Is All You Need*, 2017, arXiv: [1706.03762 \[cs.CL\]](https://arxiv.org/abs/1706.03762)
3025 (cit. on p. 188).
- 3026 [116] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>,
3027 MIT Press, 2016 (cit. on p. 189).
- 3028 [117] I. Kobyzev, S. Prince, and M. Brubaker, *Normalizing Flows: An Introduction and*
3029 *Review of Current Methods*, *IEEE Transactions on Pattern Analysis and Machine*
3030 *Intelligence* (2020) **1**, ISSN: 1939-3539, URL: <http://dx.doi.org/10.1109/TPAMI.2020.2992934> (cit. on p. 194).
- 3032 [118] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using Real NVP*, 2017,
3033 arXiv: [1605.08803 \[cs.LG\]](https://arxiv.org/abs/1605.08803) (cit. on p. 195).
- 3034 [119] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, *Neural Spline Flows*, 2019,
3035 arXiv: [1906.04032 \[stat.ML\]](https://arxiv.org/abs/1906.04032) (cit. on p. 195).
- 3036 [120] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, *Journal of Machine*
3037 *Learning Research* **12** (2011) 2825 (cit. on p. 197).