

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 | iv |
| 32 Glossary | xvii |
| 33 Preface | xx |
| 34 Chapter 1: The Standard Model of Particle Physics | 1 |
| 35 1.1 Introduction: Particles and Fields | 1 |
| 36 1.2 Quantum Electrodynamics | 4 |
| 37 1.3 An Aside on Group Theory | 9 |
| 38 1.4 Quantum Chromodynamics | 11 |
| 39 1.5 The Weak Interaction | 14 |
| 40 1.6 The Higgs Potential and the SM | 20 |
| 41 1.7 The Standard Model: A Summary | 26 |
| 42 Chapter 2: Di-Higgs Phenomenology and Physics Beyond the Standard Model | 30 |
| 43 2.1 Intro to Di-Higgs | 30 |
| 44 2.2 Resonant HH Searches | 31 |
| 45 2.3 Non-resonant HH Searches | 33 |
| 46 Chapter 3: Experimental Apparatus | 37 |
| 47 3.1 The Large Hadron Collider | 37 |
| 48 3.2 The ATLAS Experiment | 40 |
| 49 Chapter 4: Simulation | 49 |
| 50 4.1 Event Generation | 49 |
| 51 4.2 Detector Simulation | 52 |
| 52 4.3 Correlated Fluctuations in FastCaloSim | 54 |

| | | |
|----|--|-----|
| 53 | 4.4 Outlook | 62 |
| 54 | Chapter 5: Reconstruction | 64 |
| 55 | 5.1 Jets | 64 |
| 56 | 5.2 Flavor Tagging | 66 |
| 57 | Chapter 6: Setting up the $HH \rightarrow b\bar{b}b\bar{b}$ Analysis | 76 |
| 58 | 6.1 Data and Monte Carlo Simulation | 78 |
| 59 | 6.2 Triggers and Object Definitions | 80 |
| 60 | Chapter 7: Analysis Selection | 83 |
| 61 | 7.1 Analysis Selection | 83 |
| 62 | 7.2 Background Reduction and Top Veto | 91 |
| 63 | 7.3 Kinematic Region Definition | 94 |
| 64 | Chapter 8: Background Estimation | 100 |
| 65 | 8.1 The Two Tag Region | 101 |
| 66 | 8.2 Kinematic Reweighting | 101 |
| 67 | Chapter 9: Uncertainties and Validation | 161 |
| 68 | 9.1 Statistical Uncertainties and Bootstrapping | 161 |
| 69 | 9.2 Background Shape Uncertainties | 166 |
| 70 | 9.3 Signal Uncertainties | 169 |
| 71 | 9.4 Background Validation | 171 |
| 72 | Chapter 10: Results | 174 |
| 73 | 10.1 m_{HH} Distributions | 174 |
| 74 | 10.2 Statistical Analysis | 185 |
| 75 | 10.3 Results | 186 |
| 76 | Chapter 11: Comparisons with other channels | 192 |
| 77 | Chapter 12: Conclusions | 195 |
| 78 | Appendix A: Overview of Other $b\bar{b}b\bar{b}$ Channels | 211 |

| | | |
|----|--|-----|
| 79 | Appendix B: Future Ideas for $HH \rightarrow b\bar{b}b\bar{b}$ | 214 |
| 80 | B.1 pairAGraph: A New Method for Jet Pairing | 214 |
| 81 | B.2 Background Estimation with Mass Plane Interpolation | 219 |

LIST OF FIGURES

| | Figure Number | Page |
|--|---------------|------|
| 83 1.1 Diagram of the elementary particles described by the Standard Model [6]. | 3 | |
| 84 2.1 Representative diagrams for the gluon-gluon fusion production of spin-0 (X) 85 and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs 86 bosons. The spin-0 resonance considered for this thesis is a generic narrow 87 width resonance which may be interpreted in the context of two Higgs doublet 88 models [23], whereas the spin-2 resonance is considered as a Kaluza-Klein 89 graviton within the bulk Randall-Sundrum (RS) model [24, 25]. | 31 | |
| 90 2.2 Dominant contributing diagrams for non-resonant gluon-gluon fusion produc- 91 tion of HH . κ_λ and κ_t represent ratios of the Higgs self-coupling and coupling 92 to top quarks respectively, relative to the values predicted by the Standard 93 Model. | 32 | |
| 94 2.3 Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common 95 decay mode, representing 34 % of all HH events produced at the LHC. | 33 | |
| 96 2.4 Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demon- 97 strating the impact of the interference between the two diagrams of Figure 2.2 98 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram con- 99 tribution, demonstrating the shape of the box diagram contribution, whereas 100 for $\kappa_\lambda = 10$, the triangle diagram dominates, with a strong low mass peak. 101 The interplay between the two is quite evident for other values, resulting in, 102 e.g., the double peaked structure present for $\kappa_\lambda = 2$ (near maximal destruc- 103 tive interference) and $\kappa_\lambda = -5$. At $\kappa_\lambda = 5$, the interference leads to a deficit at 104 high m_{HH} , resulting in a narrower distribution (and thus a more pronounced 105 low mass peak) than the $\kappa_\lambda = 10$ case. [26] | 35 | |
| 106 3.1 Diagram of the ATLAS detector [46] | 40 | |
| 107 3.2 Cross section of the ATLAS detector showing how particles interact with 108 various detector components [48] | 42 | |
| 109 3.3 2D projections of the ATLAS coordinate system | 43 | |

| | | | |
|-----|-----|--|----|
| 110 | 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. [62] | 51 |
| 117 | 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 [65]. | 53 |
| 120 | 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) [67]. | 55 |
| 123 | 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. | 57 |
| 128 | 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. | 58 |
| 134 | 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. | 59 |
| 142 | 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. | 60 |

| | | |
|-----|--|----|
| 145 | 4.8 Comparison of the Gaussian fluctuation model to the default FCSV2 version | 89 |
| 146 | and to G4 simulation, using pions of 65 GeV energy and $0.2 < \eta < 0.25$. | |
| 147 | Variables shown relate to calorimeter clusters, three-dimensional spatial group- | |
| 148 | ings of cells [72] which provide powerful insight on the structure of energy | |
| 149 | deposits in the calorimeter. Variables considered include number and energy | |
| 150 | of clusters, the 2nd moment in lambda, ($\langle \lambda^2 \rangle$), which describes the square | |
| 151 | of the longitudinal extension of a cluster, where λ is the distance of a cell from | |
| 152 | the shower center along the shower axis, and a cluster moment is defined as | |
| 153 | $\langle x^n \rangle = \frac{\sum E_i x_i}{\sum E_i}$, and the distance ΔR , between the cluster and the true pion. | |
| 154 | With the correlated fluctuations, variables demonstrate improved modeling | |
| 155 | relative to default FastCaloSimV2. | 61 |
| 156 | 5.1 Illustration of an interaction producing two light jets and one b -jet in the | 69 |
| 157 | transverse plane. While the light jets decay “promptly”, coinciding with | |
| 158 | the primary vertex of the proton-proton interaction, the longer lifetime of B | |
| 159 | hadrons leads to a secondary decay vertex, displaced from the primary vertex | |
| 160 | by length L_{xy} . This is typically a few mm, and therefore is not directly visible | |
| 161 | in the detector, but leads to a large transverse impact parameter, d_0 , for the | |
| 162 | resulting tracks. [80] | 69 |
| 163 | 5.2 Performance of the various low and high level flavor tagging algorithms in | 73 |
| 164 | $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light | |
| 165 | and c -jet rejection. The high level taggers demonstrate significantly better | |
| 166 | performance than any of the individual low level taggers, with DL1 offering | |
| 167 | slight improvements over MV2 due to the inclusion of additional input variables. [85] | 73 |
| 168 | 5.3 Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demon- | 74 |
| 169 | strating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c | |
| 170 | controls the importance of c -jet rejection in the discriminating variable, and | |
| 171 | values shown have been optimized separately for each DL1 configuration. DL1r | |
| 172 | demonstrates a significant improvement in both light and c jet rejection over | |
| 173 | MV2 and DL1. [85] | 74 |
| 174 | 7.1 Comparison of m_{H1} vs m_{H2} planes for the full Run 2 2 b dataset with different | 89 |
| 175 | pairings. As evidenced, this choice significantly impacts where events fall in | |
| 176 | this plane, and therefore which events fall into the various kinematic regions | |
| 177 | defined in this plane (see Section 7.3). The signal regions for the resonant/early | |
| 178 | Run 2 analysis are shown for reference for the BDT and D_{HH} pairings, while | |
| 179 | the the min ΔR signal region shifted is shifted slightly up and to the right to | |
| 180 | match the non-resonant selection. Note that the band structure around 80 GeV | |
| 181 | in both m_{H1} and m_{H2} is introduced by the top veto described in Section 7.2. | 89 |

| | | | |
|-----|-----|---|-----|
| 182 | 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. The distributions are quite similar at the Standard Model point, but for other variations, min ΔR does not pick up the low mass features. | 90 |
| 186 | 7.3 | Resonant search: Illustration of the impact of the top veto on $t\bar{t}$ Monte Carlo for the resonant analysis, with representative scalar signals shown for reference. The cut value used is 1.5, shown in the dashed black, and events below this value are discarded. This top veto clearly removes the bulk of $t\bar{t}$ events, and the value of the cut is chosen to retain analysis sensitivity, particularly for low mass. | 92 |
| 192 | 7.4 | Non-resonant search: Illustration of the impact of the top veto on $t\bar{t}$ Monte Carlo for the non-resonant analysis, with the Standard Model signal shown for reference. The cut value used is 1.5, shown in the dashed black, and events below this value are discarded. This top veto clearly removes the bulk of $t\bar{t}$ events. While this plot may seem to motivate a more aggressive cut on X_{Wt} , increasing the value of the cut was found to reduce analysis sensitivity. . . . | 93 |
| 198 | 7.5 | 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. | 95 |
| 203 | 7.6 | 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. | 97 |
| 208 | 7.7 | 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). | 98 |
| 212 | 8.1 | 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 (left) and after (right) CR derived reweighting for the 2018 Control Region. | 107 |
| 215 | 8.2 | 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 (left) and after (right) CR derived reweighting for the 2018 Control Region. | 108 |

| | | |
|-----|---|-----|
| 218 | 8.3 Resonant Search: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates. | 109 |
| 221 | 8.4 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region. | 110 |
| 224 | 8.5 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 Control Region. | 111 |
| 227 | 8.6 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region. | 112 |
| 230 | 8.7 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Control Region. | 113 |
| 233 | 8.8 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and sub-leading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region. | 114 |
| 236 | 8.9 Resonant Search: Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied | 115 |
| 239 | 8.10 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 (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 116 |
| 242 | 8.11 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 (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 117 |
| 245 | 8.12 Resonant Search: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates. | 118 |
| 248 | 8.13 Resonant Search: Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 119 |
| 251 | 8.14 Resonant Search: Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 120 |

| | | |
|-----|--|-----|
| 254 | 8.15 Resonant Search: Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 121 |
| 255 | | |
| 256 | | |
| 257 | 8.16 Resonant Search: Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 122 |
| 258 | | |
| 259 | | |
| 260 | 8.17 Resonant Search: Distributions of $\Delta\phi$ between jets in the leading and sub-leading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region. | 123 |
| 261 | | |
| 262 | | |
| 263 | 8.18 Resonant Search: Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied | 124 |
| 264 | | |
| 265 | | |
| 266 | 8.19 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 (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 125 |
| 267 | | |
| 268 | | |
| 269 | | |
| 270 | 8.20 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 (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 126 |
| 271 | | |
| 272 | | |
| 273 | 8.21 Non-resonant Search (4b): Distributions of the number of jets before (left) and after (right) 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. | 127 |
| 274 | | |
| 275 | | |
| 276 | 8.22 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 128 |
| 277 | | |
| 278 | | |
| 279 | 8.23 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 129 |
| 280 | | |
| 281 | | |
| 282 | 8.24 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 (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 130 |
| 283 | | |
| 284 | | |
| 285 | | |
| 286 | 8.25 Non-resonant Search (4b): Distributions of mass of the leading and sub-leading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 4b Control Region. | 131 |
| 287 | | |
| 288 | | |

| | | |
|-----|---|-----|
| 289 | 8.26 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading 290 and subleading Higgs candidates before (left) and after (right) CR derived 291 reweighting for the 2018 4b Control Region. | 132 |
| 292 | 8.27 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , 293 before (left) and after (right) CR derived reweighting for the 2018 4b Control 294 Region. Reweighting is done after the cut on this variable is applied. | 133 |
| 295 | 8.28 Non-resonant Search (4b): Distributions of ΔR between the closest Higgs 296 Candidate jets, ΔR between the other two, and average absolute value of HC 297 jet η before (left) and after (right) CR derived reweighting for the 2018 4b 298 Validation Region. | 134 |
| 299 | 8.29 Non-resonant Search (4b): Distributions of p_T of the 2nd and 4th leading 300 Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after 301 (right) CR derived reweighting for the 2018 4b Validation Region. | 135 |
| 302 | 8.30 Non-resonant Search (4b): Distributions of the number of jets before (left) 303 and after (right) CR derived reweighting for the 2018 4b Validation Region. A 304 minimum of 4 jets is required in each event in order to form Higgs candidates. | 136 |
| 305 | 8.31 Non-resonant Search (4b): Distributions of p_T of the 1st and 3rd leading 306 Higgs Candidate jets and ΔR between Higgs candidates before (left) and after 307 (right) CR derived reweighting for the 2018 4b Validation Region. | 137 |
| 308 | 8.32 Non-resonant Search (4b): Distributions of η of the 1st, 2nd, and 3rd lead- 309 ing Higgs Candidate jets before (left) and after (right) CR derived reweighting 310 for the 2018 4b Validation Region. | 138 |
| 311 | 8.33 Non-resonant Search (4b): Distributions of η of the 4th leading Higgs 312 Candidate jet and the p_T of the leading and subleading Higgs candidates before 313 (left) and after (right) CR derived reweighting for the 2018 4b Validation Region. | 139 |
| 314 | 8.34 Non-resonant Search (4b): Distributions of mass of the leading and sub- 315 leading Higgs candidates and of the di-Higgs system before (left) and after 316 (right) CR derived reweighting for the 2018 4b Validation Region. | 140 |
| 317 | 8.35 Non-resonant Search (4b): Distributions of $\Delta\phi$ between jets in the leading 318 and subleading Higgs candidates before (left) and after (right) CR derived 319 reweighting for the 2018 4b Validation Region. | 141 |
| 320 | 8.36 Non-resonant Search (4b): Distributions of the top veto variable, X_{Wt} , 321 before (left) and after (right) CR derived reweighting for the 2018 4b Validation 322 Region. Reweighting is done after the cut on this variable is applied. | 142 |

| | | |
|-----|--|-----|
| 323 | 8.37 Non-resonant Search (3b1l): Distributions of ΔR between the closest 324 Higgs Candidate jets, ΔR between the other two, and average absolute value 325 of HC jet η before (left) and after (right) CR derived reweighting for the 2018 326 3b1l Control Region. | 143 |
| 327 | 8.38 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading 328 Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after 329 (right) CR derived reweighting for the 2018 3b1l Control Region. | 144 |
| 330 | 8.39 Non-resonant Search (3b1l): Distributions of the number of jets before 331 (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region. 332 A minimum of 4 jets is required in each event in order to form Higgs candidates. | 145 |
| 333 | 8.40 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading 334 Higgs Candidate jets and ΔR between Higgs candidates before (left) and after 335 (right) CR derived reweighting for the 2018 3b1l Control Region. | 146 |
| 336 | 8.41 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd 337 leading Higgs Candidate jets before (left) and after (right) CR derived reweight- 338 ing for the 2018 3b1l Control Region. | 147 |
| 339 | 8.42 Non-resonant Search (3b1l): Distributions of η of the 4th leading Higgs 340 Candidate jet and the p_T of the leading and subleading Higgs candidates before 341 (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region. | 148 |
| 342 | 8.43 Non-resonant Search (3b1l): Distributions of mass of the leading and 343 subleading Higgs candidates and of the di-Higgs system before (left) and after 344 (right) CR derived reweighting for the 2018 3b1l Control Region. | 149 |
| 345 | 8.44 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the 346 leading and subleading Higgs candidates before (left) and after (right) CR 347 derived reweighting for the 2018 3b1l Control Region. | 150 |
| 348 | 8.45 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , 349 before (left) and after (right) CR derived reweighting for the 2018 3b1l Control 350 Region. Reweighting is done after the cut on this variable is applied. | 151 |
| 351 | 8.46 Non-resonant Search (3b1l): Distributions of ΔR between the closest 352 Higgs Candidate jets, ΔR between the other two, and average absolute value 353 of HC jet η before (left) and after (right) CR derived reweighting for the 2018 354 3b1l Validation Region. | 152 |
| 355 | 8.47 Non-resonant Search (3b1l): Distributions of p_T of the 2nd and 4th leading 356 Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after 357 (right) CR derived reweighting for the 2018 3b1l Validation Region. | 153 |

| | | |
|-----|---|-----|
| 358 | 8.48 Non-resonant Search (3b1l): Distributions of the number of jets before (left) and after (right) 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. | 154 |
| 362 | 8.49 Non-resonant Search (3b1l): Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. | 155 |
| 365 | 8.50 Non-resonant Search (3b1l): Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweight- ing for the 2018 3b1l Validation Region. | 156 |
| 368 | 8.51 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 (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. | 157 |
| 372 | 8.52 Non-resonant Search (3b1l): Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. | 158 |
| 375 | 8.53 Non-resonant Search (3b1l): Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. | 159 |
| 378 | 8.54 Non-resonant Search (3b1l): Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 3b1l Vali- dation Region. Reweighting is done after the cut on this variable is applied. . | 160 |
| 382 | 9.1 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 9.1(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 9.1(b) demonstrates the applied normalization correction, and Figure 9.1(c) shows the final band (normalized Figure 9.1(a) + Figure 9.1(b)). Comparing this with the IQR variation for the prediction from each bootstrap in each bin in Figure 9.1(d), the approximate envelope describes a very similar variation. | 165 |
| 393 | 9.2 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. . . | 167 |

| | | | |
|-----|------|---|-----|
| 395 | 9.3 | 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. | 168 |
| 399 | 9.4 | 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. | 172 |
| 406 | 9.5 | 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. | 173 |
| 411 | 10.1 | Resonant Search: Demonstration of the performance of the nominal reweighting in the control region on corrected m_{HH} , with Figure 10.1(a) showing $2b$ events normalized to the total $4b$ yield and Figure 10.1(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. | 175 |
| 417 | 10.2 | 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. | 176 |
| 422 | 10.3 | Resonant Search: Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The left plot overlays a variety of representative spin-0 signals, while the right does the same for spin-2. The background and data are identical between the two. 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} . Note that the spin-2 signals are significantly wider than the spin-0 signals. Near the kinematic threshold of 250 GeV, this leads to, e.g., the double peaked structure of the 280 GeV signal, which is understood to be an effect of the limited kinematic phase space in this region. | 177 |

| | | |
|-----|--|-----|
| 432 | 10.4 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 mis-modeling in the low $\Delta\eta_{HH}$ region near 600 GeV. | 179 |
| 436 | 10.5 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. | 180 |
| 439 | 10.6 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. | 181 |
| 443 | 10.7 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. | 182 |
| 446 | 10.8 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. | 183 |
| 452 | 10.9 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. | 184 |
| 458 | 10.10 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 $k/\bar{M}_{\text{Pl}} = 1$ [25] (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. Note that all results use the asymptotic approximation, though the validity of this approximation for the boosted results above 3 TeV is being evaluated. | 189 |

| | | |
|-----|--|-----|
| 504 | B.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. | 219 |
| 505 | | |
| 506 | | |
| 507 | | |
| 508 | | |
| 509 | B.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. | 220 |
| 510 | | |
| 511 | | |
| 512 | | |
| 513 | | |
| 514 | B.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. | 227 |
| 515 | | |
| 516 | | |
| 517 | | |
| 518 | B.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. | 229 |
| 519 | | |
| 520 | | |
| 521 | | |
| 522 | B.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. | 230 |
| 523 | | |
| 524 | | |
| 525 | B.6 Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible. | 231 |
| 526 | | |
| 527 | B.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. | 232 |
| 528 | | |
| 529 | | |
| 530 | | |
| 531 | | |
| 532 | | |
| 533 | | |
| 534 | | |
| 535 | | |
| 536 | B.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 %. | 233 |
| 537 | | |
| 538 | | |

GLOSSARY

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

ACKNOWLEDGMENTS

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

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

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

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

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

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

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

PREFACE

572 This thesis focuses primarily on searches for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$
573 final state. It begins with an overview of the relevant physics and experimental background
574 for such work, structured as follows: In Chapter 1, I provide an overview of the Standard
575 Model of particle physics, with discussion of the theoretical and experimental development of
576 such a model. Chapter 2 dives more into the details of Higgs boson pair production, as well as
577 the physics beyond the Standard Model relevant for this thesis. Chapter 3 then provides an
578 introduction to the experimental apparatus used for the presented searches, with an outline
579 of the Large Hadron Collider and the ATLAS detector. Chapter 4 details the procedure to
580 simulate the physics processes discussed in Chapters 1 and 2, including simulation of the
581 detector discussed Chapter 3. Finally, a review of the procedures to reconstruct objects used
582 for physics analysis is provided in Chapter 5, with a focus on jets and flavor-tagging.

583 The original contributions of this thesis are discussed in a variety of places. Chapter
584 4 includes my work on the development of methods to improve the modeling of hadronic
585 showers within a parametrized simulation of the ATLAS calorimeter. I entirely developed
586 both the method and the software for the Gaussian method discussed in Chapter 4, including
587 all of the validations presented there. The development of the Variational Autoencoder
588 method was done in collaboration with Dalila Salamani. This work has been published in a
589 set of proceedings [1] and implemented into ATLAS software. At the time of this writing, it
590 is a candidate for inclusion in the Run 3 simulation infrastructure.

591 Chapters 6 through 10 detail searches for resonant and non-resonant pair production of
592 Higgs bosons in the $b\bar{b}b\bar{b}$ final state. I was one of the main analyzers for both of these searches,
593 responsible for much of the development of the methods, infrastructure, and documentation.

594 My most major contribution was the development of the background estimation procedure
595 and the associated uncertainties, which I spearheaded both conceptually and practically. This
596 is quite a significant contribution for both the resonant and non-resonant, as it is the core of
597 much of the analysis design, with the most direct impact on the final results – to paraphrase
598 Georges Aad during the resonant review process, “This is the analysis.”

599 This was not my only contribution – for the resonant search, I contributed to the
600 development of the analysis selection and codebase, performed many of the necessary cross-
601 checks, and was the co-editor of the ATLAS internal documentation, along with Beojan
602 Stanislaus, who developed the BDT pairing and much of the analysis software. Credit goes as
603 well to Lucas Borgna, for much of the work behind the development of the trigger strategy.

604 The resonant search follows many of the procedures of the early Run 2 analysis [2], with
605 the pairing method and background estimation method constituting the two biggest analysis-
606 level differences from that work. The non-resonant analysis has several additional changes,
607 which include a variety of new kinematic variable and region definitions, as well as a different
608 pairing method than both the early Run 2 search and the resonant search. I was responsible
609 for a large majority of the studies behind each of these decisions. I am also responsible for
610 the development of much of the modern $4b$ software infrastructure, including, of course, the
611 background estimation framework, a new limit setting framework, and a new centralized
612 plotting framework, the latter of which greatly facilitates both studies and documentation for
613 the more complicated non-resonant analysis strategy.

614 At the time of this writing, the preliminary resonant results have been published [3], with
615 a paper soon to follow, pending some additional studies on the high mass ($> 3 \text{ TeV}$) results
616 in the boosted analysis channel ¹. The non-resonant results are more preliminary, but the
617 analysis strategy presented in this thesis is approximately final, and the analysis is beginning

¹This thesis focuses on the resolved analysis channel, so these additional studies do not impact the final results of this thesis work. The boosted channel is included in the limits presented in Figure 10.10, but in no other plots or results. See Appendix A for a description of the boosted analysis selection.

618 internal ATLAS review.

619 While these above results are the main results of this thesis, proof-of-concept studies for
620 two novel $4b$ analysis methods are included in Appendix B. This work, done in collaboration
621 primarily with Nicole Hartman, was not used for the $4b$ results presented here, but I encourage
622 the interested reader to consider these for further study in future iterations of the $4b$ analysis.
623 I note as well that, while these methods are promising in the context of the $4b$ analysis, they
624 are also methodologically interesting, and conceptually related results have been published
625 concurrently with the development of the work presented in this thesis in [4] and [5].

626

DEDICATION

627

To family, both given and found

628

Chapter 1

629

THE STANDARD MODEL OF PARTICLE PHYSICS

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

638 **1.1 Introduction: Particles and Fields**

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

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

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

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

659 Particles are observed to behave in different ways under different forces. These behaviors
660 respect certain *symmetries*, which are most naturally described in the language of group
661 theory. The respective fields, charges, and generators of these symmetry groups are the basic
662 pieces of the SM Lagrangian, which describes the full dynamics of the theory. In the following,
663 we will build up the basic components of this Lagrangian. The treatment presented here relies
664 heavily on Jackson's Classical Electrodynamics [7] for the build-up, and Thomson's Modern
665 Particle Physics [8] for the rest, with reference to Srednicki's Quantum Field Theory [9], and
666 some personal biases and interjections.

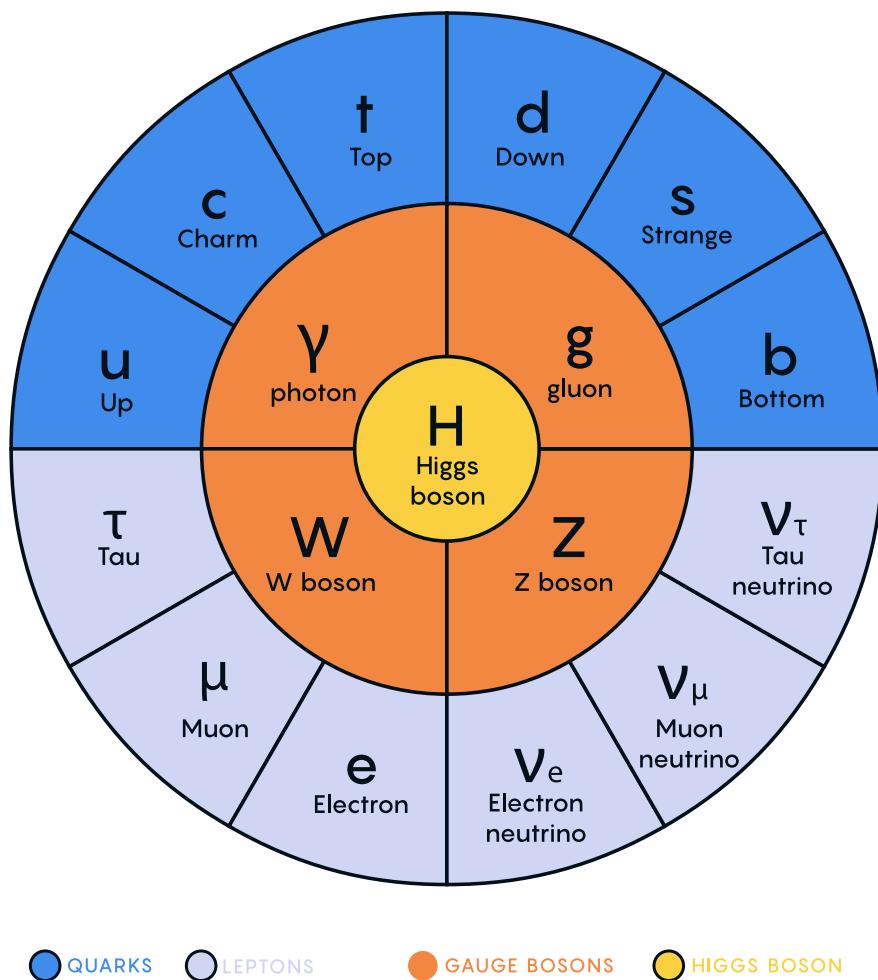


Figure 1.1: Diagram of the elementary particles described by the Standard Model [6].

⁶⁶⁷ **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:

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

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

⁶⁶⁸ 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:

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

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

674 We know that electrons experience electromagnetic interactions with other objects. Given
 675 this, and the fact that such interactions must be transmitted *somewhat* between e.g. two
 676 electrons, it seems natural that these interactions are facilitated by electromagnetic radiation.
 677 More specifically, we may think of photons as *mediators* of the electromagnetic force. It
 678 follows, then, that a description of electromagnetism on the level of particles must involve a
 679 description of both the "source" particles (e.g. electrons), the mediators (photons), and their
 680 interactions. Further, this description must be (1) relativistic and (2) consistent with the
 681 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

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

682 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

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

683 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:

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

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

685 $-q\bar{\psi}\gamma^\mu\psi A_\mu$.

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

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

686 which is exactly the quantum electrodynamics Lagrangian.

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

691 $\mathcal{L}(q, \dot{q}, t)$ and $\dot{q} = \frac{\partial q}{\partial t}$, any field q which has no time derivative in the Lagrangian has 0
692 conjugate momentum, and thus no dynamics.

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

697 What does this mean? A^μ is a four component object, but it would appear that only three
698 of the components have dynamics: we have too many degrees of freedom in the theory. This
699 is the principle behind *gauge symmetry* – an extra constraint on A^μ (a *gauge condition*) must
700 be defined such that a unique A^μ defines the theory and satisfies the condition. However,
701 we are free to choose this extra condition – the physics content of the theory should be
702 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

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

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

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

705 Under the same transformation of \vec{A} , the electric field \vec{E} becomes $-\nabla\phi - \frac{\partial\vec{A}}{\partial t} - \frac{\partial\nabla\lambda}{\partial t} =$
706 $-\nabla(\phi + \frac{\partial\lambda}{\partial t}) - \frac{\partial\vec{A}}{\partial t}$, such that, for the \vec{E} field to be unchanged, we must additionally apply
707 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

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

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

⁷¹⁰ 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:

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

⁷¹¹ since by the arguments above, the physics content of the Lagrangian is invariant under the
⁷¹² transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, we may directly make this transformation, and remove this
⁷¹³ 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
⁷¹⁴ this same transformation of A_μ , so we may say that the QED Lagrangian is invariant under
⁷¹⁵ local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$.

⁷¹⁶ These arguments illuminate some important concepts which will serve us well going forward.
⁷¹⁷ First, while we have remained grounded in the “familiar” physics of electromagnetism for the
⁷¹⁸ above, arguments of the “top down” variety would lead us to the exact same conclusions.
⁷¹⁹ That is, suppose we wanted to construct a theory of spin-1/2 particles that was invariant
⁷²⁰ under local transformations of the form $\psi \rightarrow e^{iq\lambda(x)}\psi$. More broadly, we could say that we
⁷²¹ desire this theory to be invariant under local $U(1)$ transformations, where $U(1)$ is exactly
⁷²² this group, under multiplication, of complex numbers with absolute value 1. By very similar

723 arguments as above, we would see that, to achieve invariance, this theory would necessitate
724 an additional degree of freedom, A_μ , with the exact properties that are familiar to us from
725 electrodynamics. These arguments based on symmetries are extremely powerful in building
726 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 ,

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

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

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

734 1.3 An Aside on Group Theory

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

⁷⁴⁰ objects, ψ , of the form $e^{iq\lambda(x)}$. Note that such transformations, though they are local (i.e. a
⁷⁴¹ function of spacetime) are purely *phase* transformations. Relatedly, $U(1)$ is an Abelian group,
⁷⁴² meaning that group elements commute.

⁷⁴³ To set up language to generalize beyond $U(1)$, note that we may equivalently write $U(1)$
⁷⁴⁴ 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,
⁷⁴⁵ with each $\alpha^a(x)$ an associated scalar function to generator t^a , and g is some scalar strength
⁷⁴⁶ parameter. Of course this is a bit silly for $U(1)$, which has a single generator, and thus
⁷⁴⁷ reduces to the transformation we discussed above. However, this becomes much more useful
⁷⁴⁸ for groups of higher degree, with more generators and degrees of freedom.

⁷⁴⁹ To discuss these groups in a bit more detail, note that $U(n)$ is the unitary group of degree
⁷⁵⁰ n , and corresponds to the group of $n \times n$ unitary matrices (that is, $U^\dagger U = UU^\dagger = 1$). Given
⁷⁵¹ that group elements are $n \times n$, this means that there are n^2 degrees of freedom: n^2 generators
⁷⁵² are needed to characterize the group.

⁷⁵³ For $U(1)$, this is all consistent with what we have said above – the group of 1×1 unitary
⁷⁵⁴ matrices have a single generator, and the phases we identify above clearly satisfy unitarity.
⁷⁵⁵ Note that these degrees of freedom for the gauge group also characterize the number of gauge
⁷⁵⁶ bosons we need to satisfy the local symmetry: for $U(1)$, we need one gauge boson, the photon.

⁷⁵⁷ Of relevance for the Standard Model are also the special unitary groups $SU(n)$. These
⁷⁵⁸ are defined similarly to the unitary groups, with the additional requirement that group
⁷⁵⁹ elements have determinant 1. This extra constraint removes 1 degree of freedom: groups are
⁷⁶⁰ characterized by $n^2 - 1$ generators.

⁷⁶¹ In particular, we will examine the groups $SU(2)$ in the context of the weak interaction,
⁷⁶² with an associated $2^2 - 1 = 3$ gauge bosons (cf. the W^\pm and Z bosons), and $SU(3)$, with an
⁷⁶³ associated $3^2 - 1 = 8$ gauge bosons (cf. gluons of different flavors). Note that these groups
⁷⁶⁴ are non-Abelian (2×2 or 3×3 matrices do not, in general, commute), leading to a variety of
⁷⁶⁵ complications. However, both of these theories feature interactions with spin-1/2 particles,
⁷⁶⁶ with transformations of a very similar form: $\psi \rightarrow e^{ig\vec{\alpha}(x)\cdot\vec{T}}\psi$, and the general framing of the
⁷⁶⁷ arguments for QED will serve us well in the following.

768 **1.4 Quantum Chromodynamics**

769 In some sense, the simplest extension the development of QED is quantum chromodynamics
770 (QCD). QCD is a theory in which, once the basic dynamics are framed (a non-trivial task!) the
771 group structure becomes apparent. The quark model, developed by Murray Gell-Mann [10]
772 and George Zweig [11], provided the fundamental particles involved in the theory, and had
773 great success in explaining the expanding zoo of experimentally observed hadronic states.

774 Some puzzles were still apparent – the Δ^{++} baryon, e.g., is composed of three up quarks,
775 u , with aligned spins. As quarks are fermions, such a state should not be allowed by the
776 Pauli exclusion principle. The existence of such a state in nature implies the existence
777 of another quantum number, and a triplet of values, called *color charge* was proposed by
778 Oscar Greenberg [12]. With these pieces in place, the structure becomes more apparent, as
779 elucidated by Han and Nambu [13].

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

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

792 Counting up the available types of gluons, then, we come up with nine. Six of mixed
793 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,
794 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

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

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

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

⁸¹² for $G_{\mu\nu}^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a - f^{abc}G_\mu^b G_\nu^c$.

⁸¹³ This gives the component of the field strength corresponding to a particular gauge field a ,
⁸¹⁴ where the first two terms have the familiar form of the QED field strength, while the last
⁸¹⁵ term is new, and explicitly related to the group structure via the f^{abc} constants. In terms
⁸¹⁶ of the physics content of the theory, this latter term gives rise to a gluon *self-interaction*, a
⁸¹⁷ distinguishing feature of QCD.

⁸¹⁸ Similarly as in QED, a Lorentz invariant combination of field strength tensors may be made
⁸¹⁹ as $G_{\mu\nu}G^{\mu\nu}$. However, this is not manifestly gauge invariant. Under a gauge transformation

820 U , the covariant derivative behaves as $D^\mu \rightarrow UD^\mu U^{-1}$, corresponding to $G^{\mu\nu} \rightarrow UG^{\mu\nu}U^{-1}$.
821 The cyclic property of the trace thus ensures the gauge invariance of $\text{tr}(G_{\mu\nu}G^{\mu\nu})$, which we
822 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

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

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

The Lagrangian then becomes

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

824 and we have constructed QCD.

825 1.5 The Weak Interaction

826 One of the first theories of the weak interaction was from Enrico Fermi [14], in an effort to
827 explain beta decay, a process in which an electron or positron is emitted from an atomic
828 nucleus, resulting in the conversion of a neutron to a proton or proton to a neutron respectively.
829 Fermi's hypothesis was of a direct interaction between four fermions. However, in the advent of
830 QED, it is natural to wonder if a theory based on mediator particles and gauge symmetries
831 applies to the weak force as well. The modern formulation of such a theory is due to Sheldon

832 Glashow, Steven Weinberg, and Abdus Salam [15], and is what we will describe in the
833 following.

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

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

846 This leads to three degrees of freedom for the gauge symmetry of the weak interactions, so
847 we thus need a theory which is locally invariant under transformations of a group with three
848 generators. The simplest such choice is $SU(2)$. We may follow a very similar prescription as
849 for QED and QCD: $SU(2)$ has three generators, which implies the existence of three gauge
850 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
851 in this case the generators are for $SU(2)$, which may be written in terms of the familiar Pauli
852 matrices: $\vec{T} = \frac{1}{2}\vec{\sigma}$. The structure constants for $SU(2)$ are the antisymmetric Levi-Civita
853 tensor, so the corresponding gauge covariant derivative is $D_\mu \equiv \partial_\mu + ig_W W_\mu^k t^k$, and the field
854 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

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

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

856 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:

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

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

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

for right and left-handed chiralities respectively – acting on a Dirac field with such operators 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:

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

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

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:

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

888 for left-handed doublets Q and L for quarks and electron fields respectively and right handed
889 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

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

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

This then leaves our Lagrangian as

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

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

891 where we have expanded the covariant derivative for clarity. You may note that we have
892 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

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

893 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

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

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

895 Note that in the $SU(2)_L \times U(1)_Y$ theory, the Lagrangian must be invariant under all of
896 the local gauge transformations. In particular, this means that the hypercharge must be the
897 same for fermion fields in each weak doublet to preserve $U(1)_Y$ invariance. This gives insight
898 into the relation between the charges of $SU(2)_L \times U(1)_Y$ and electric charge. In particular
899 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,

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

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

902 **1.6 The Higgs Potential and the SM**

903 In the above, we have neglected a discussion of masses. However there are several things to
904 sort out here. In the first place, we know experimentally that the weak interactions occur
905 over very short ranges at low energies (e.g., why Fermi's effective four fermion interaction was
906 such a good description). This is consistent with massive W^\pm and Z bosons (and indeed, this
907 is seen experimentally). However, requiring local gauge invariance forbids mass terms in the
908 Lagrangian. In the simple $U(1)$ QED example, such a term would have the form $\frac{1}{2}m_\gamma^2 A_\mu A^\mu$,
909 which is not invariant under the transformation $A_\mu \rightarrow A_\mu - \partial_\mu \lambda$, and similar arguments hold
910 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:

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

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

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

911 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
912 and right-handed particles transform differently under $SU(2)_L$, this is manifestly not gauge
913 invariant.

914 The question then becomes: how do we include particle masses while preserving the
915 gauge properties of our theory? The answer, due to Robert Brout and François Englert [16],
916 Peter Higgs [17], and Gerald Guralnik, Richard Hagen, and Tom Kibble [18] comes via the
917 Higgs mechanism, which we will describe in the following. Importantly for this thesis, this
918 mechanism predicts the existence of a physical particle, the Higgs boson, and a particle
919 consistent with the Higgs boson was seen by both ATLAS [19] and CMS [20] 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

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

920 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

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

921 we may notice that this has some interesting properties. Considering, as illustration, a similar
922 potential for a real scalar field, $\mu^2 \chi^2 + \lambda \chi^4$, taking the derivative and setting it equal to 0
923 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
924 unique minimum at $\chi = 0$, and for $\mu^2 < 0$ there are degenerate minima at $\chi = \pm\sqrt{-\frac{\mu^2}{2\lambda}}$.
925 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

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

926 where we have explicitly chosen a real, non-zero vacuum expectation value for the neutral
927 component of the Higgs doublet to maintain a massless photon, as we shall see. Physically,
928 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

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

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

⁹³⁶ in a similar way, completing the story of three massless degrees of freedom (Goldstone bosons)
⁹³⁷ 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

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

⁹³⁸ 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

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

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

⁹³⁹ As identified above, $Y = 2(Q - I_W^{(3)})$. The Higgs has 0 electric charge, and the lower doublet
⁹⁴⁰ 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

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

941 which is proportional to the W mass – an analysis with the form of the physical Z boson
942 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

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

943 is manifestly not gauge invariant because right and left handed particles transform differently
944 under $SU(2)_L$. However, because the Higgs is constructed via an $SU(2)_L$ doublet, ϕ , writing
945 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 [21] terms

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

946 where g_f is a corresponding Yukawa coupling, f_1 and f_2 have been used to denote components
947 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

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

948 Notably here, the fermion coupling to the Higgs boson scales with the mass of the fermion, a
949 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

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

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

952 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

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

with the sum running over 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 many verified predictions and many cross checks. While we have spent much time in this chapter on the theoretical components of the Standard Model, we have not discussed the corresponding experimental discoveries in detail, though this thesis itself participates in an experimental cross check of the Standard Model.

As listed in Figure [6], there are 17 particles in the Standard Model, and the history of interplay between theoretical prediction and experimental discoveries surrounding each of these is paramount to the development of the field of particle physics, and of the way we understand the universe.

Indicative of the importance and strength of electromagnetism in the everyday world, the electron and photon were foundational discoveries that began the theoretical flurry which resulted in the Standard Model. While electric charge was observed by even the ancient Greeks (and, in fact, the word electric is derived from the Greek word for amber, which picks up a charge when rubbed with fur), the connection of this charge to a subatomic particle came later, with J.J. Thompson the first (in 1897) to definitively show the existence of electrons, using cathode ray tubes to demonstrate a particle with a mass much smaller than hydrogen and with a charge to mass ratio independent the of material used in the cathode.

977 The discovery of the photon is much talked about in any introductory quantum mechanics
 978 course via the dual wave/particle nature of light. The assumption in 1900 of Max Planck that
 979 electromagnetic radiation could only be emitted or absorbed in discrete quantities (“quanta”)
 980 resolved the ultraviolet catastrophe, a classical prediction that energy emitted by a black
 981 body diverges for high frequencies. Soon after, in 1905, Einstein postulated that such quanta
 982 corresponded to physical particles, explaining, for instance, the photoelectric effect.

983 These two foundational particles led to the development of both atomic theory and
 984 quantum mechanics. In 1936, Carl D. Anderson and Seth Neddermeyer, while studying
 985 cosmic radiation, observed a particle that behaved similarly to an electron but had a shallower
 986 curvature in a magnetic field (though a sharper curvature than protons). With an assumption
 987 of the same electric charge, this difference is indicative of a particle with mass in between
 988 that of an electron and a proton, and this was the first observation of the muon.

989 In 1968, deep inelastic scattering experiments at SLAC, in which a beam of electrons is
 990 fired at atomic nuclei to probe internal structure of protons and neutrons, confirmed the
 991 existence of internal proton structure, the first observation of what would be identified as
 992 quarks. The proton contains two up quarks and a down quark – however the existence of up
 993 and down quarks, in conjunction with the observation of kaons and pions and the “eightfold
 994 way” of Gell-Mann and Zweig, indirectly confirmed the existence of the strange quark.

995 The charm quark was discovered via the observation of a charm anti-charm meson, called
 996 J/ψ , by Burton Richter and Samuel Ting in 1974, with the dual name a consequence of
 997 the shared, but independent, discovery. Richter’s group at SLAC made the discovery with
 998 SPEAR, an electron-positron collider, whereas Ting’s group utilized fixed target collisions of
 999 a proton beam. Both observed a new resonance near 3 GeV.

1000 SPEAR was additionally used for the discovery of the tau by Martin Lewis Perl in
 1001 experiments between 1974 and 1977, via the detection of anomalous events requiring the
 1002 production and decay of a new particle pair $\tau^+\tau^-$.

1003 In 1977, the bottom quark was discovered at Fermilab by Leon Lederman via the obser-
 1004 vation of a resonance near 9.5 GeV produced by fixed target proton beam collisions. This

1005 resonance, the Υ meson, consists of a bottom quark and an anti-bottom quark, and was
 1006 observed in the di-muon decay channel.

1007 The same resonance was important in the discovery of the gluon, this time in electron-
 1008 positron collisions, first by the PLUTO detector at DORIS (DESY) in 1978 and then by
 1009 the TASSO, MARK-J, JADE, and PLUTO experiments at PETRA (DESY) in 1979. The
 1010 1978 observation demonstrated excellent consistency with a three-gluon decay topology for
 1011 the $\Upsilon(9.46\text{ GeV})$ decay, but the mass of the $\Upsilon(9.46\text{ GeV})$ is not high enough to resolve three
 1012 distinct jets. Operating at $\sqrt{s} = 27.4\text{ GeV}$, the experiments in 1979 demonstrated a three jet
 1013 topology consistent (at these higher energies) with gluon bremsstrahlung, that is $e^+e^- \rightarrow q\bar{q}g$,
 1014 providing the first evidence for the existence of the gluon.

1015 At CERN in 1983, proton-antiproton collisions led to the discovery of the W and Z bosons
 1016 with the UA1 and UA2 experiments, for which Carlo Rubbia and Simon van der Meer received
 1017 the Nobel Prize in 1984.

1018 The top quark was discovered in 1995 at the Tevatron at Fermilab, a proton anti-proton
 1019 collider, by the CDF and DØ experiments, offering a center of mass energy of 1.8 TeV.

1020 The final piece of the puzzle was the Higgs boson, discovered by ATLAS and CMS at the
 1021 Large Hadron Collider in 2012. *TODO: add neutrinos and citations*

1022 The Standard Model, for all of its power, is notably not a complete theory of the universe
 1023 – there is no inclusion of gravity, for instance, though a consistent description may be provided
 1024 with the introduction of a spin-2 particle. Neutrino oscillations demonstrate that neutrinos
 1025 have mass, but right-handed neutrinos have not been observed, leading to questions about
 1026 whether there is a different mechanism to provide neutrinos with mass than that described
 1027 above. Cosmology tells us that dark matter exists, but there is no corresponding particle
 1028 within the Standard Model. This thesis therefore also participates in searches for physics
 1029 beyond the Standard Model. We will provide a sketch of the relevant theories in the following
 1030 chapter, though a detailed theoretical discussion is beyond the scope of this work.

1031

Chapter 2

1032

DI-HIGGS PHENOMENOLOGY AND PHYSICS BEYOND THE STANDARD MODEL

1033

1034

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

1041

2.1 Intro to Di-Higgs

1042

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

1046

The focus of this thesis is gluon initiated processes – in the case of di-Higgs this is termed gluon-gluon fusion (ggF). HH production may also occur via vector boson fusion [22]. However the cross section for such production is significantly smaller. Representative Feynman diagrams are shown for gluon-gluon fusion resonant production in Figure 2.1 and for non-resonant production in Figure 2.2.

1051

As shown in Chapter 1, the Higgs coupling to fermions scales with particle mass. As the 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 kinematically disfavored, $H \rightarrow b\bar{b}$ is the dominant fermionic Higgs decay mode, and, in fact,

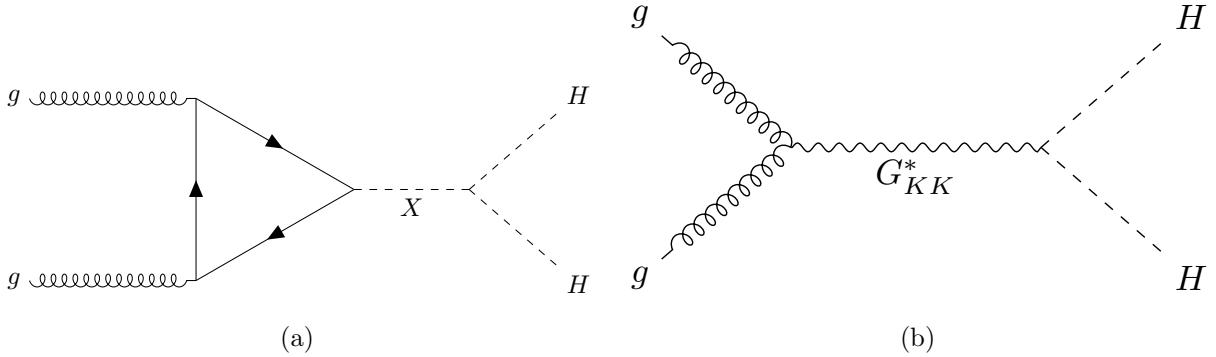


Figure 2.1: Representative diagrams for the gluon-gluon fusion production of spin-0 (X) and spin-2 (G_{KK}^*) resonances which decay to two Standard Model Higgs bosons. The spin-0 resonance considered for this thesis is a generic narrow width resonance which may be interpreted in the context of two Higgs doublet models [23], whereas the spin-2 resonance is considered as a Kaluza-Klein graviton within the bulk Randall-Sundrum (RS) model [24, 25].

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.

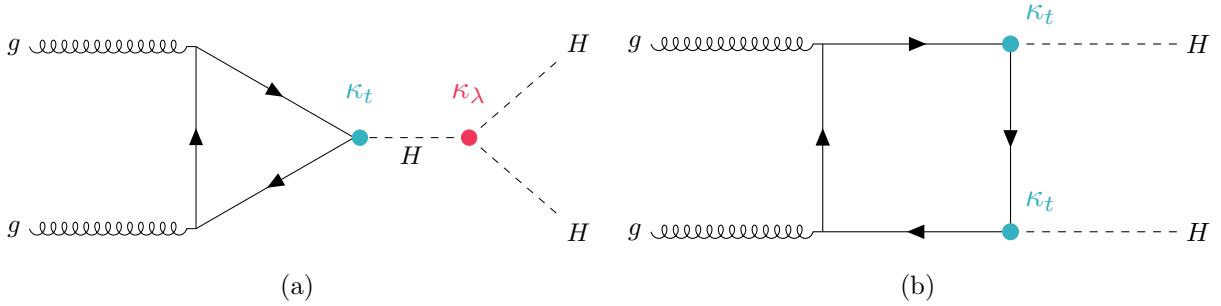


Figure 2.2: Dominant contributing diagrams for non-resonant gluon-gluon fusion production of HH . κ_λ and κ_t represent ratios of the Higgs self-coupling and coupling to top quarks respectively, relative to the values predicted by the Standard Model.

The spin-2 signal considered is implemented within the bulk Randall-Sundrum (RS) model [24, 25], 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 [26], 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 [23], 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.

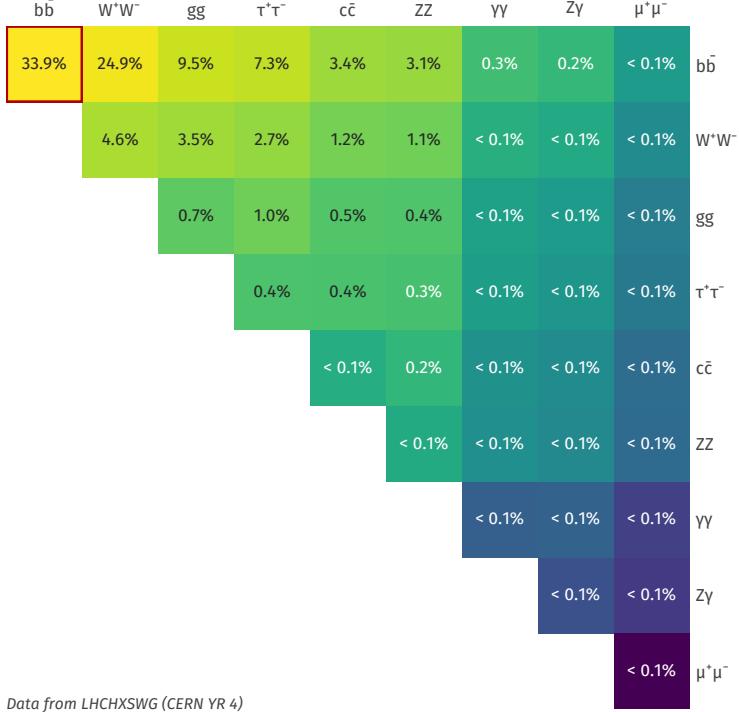


Figure 2.3: Illustration of dominant HH branching ratios. $HH \rightarrow b\bar{b}b\bar{b}$ is the most common decay mode, representing 34 % of all HH events produced at the LHC.

1082 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

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

1083 The mass of the SM Higgs boson has been experimentally measured to be 125 GeV [27],
 1084 and the vacuum expectation value $v = 246$ GeV has a precise determination from the muon
 1085 lifetime [28]. This coupling is therefore precisely predicted in the Standard Model, such that
 1086 an observed deviation from this prediction would be a clear sign of new physics.

1087 The relevant diagrams for non-resonant HH production are shown in Figure 2.2. Notably,
 1088 the diagrams *interfere* with each other, which can be easily seen by counting the fermion
 1089 lines. A detailed theoretical discussion is provided by, e.g. [29].

1090 For the searches presented here, the quark couplings to the Higgs are considered to be
 1091 consistent with the Standard Model value, with measurements of the dominant top Yukawa
 1092 coupling left to more sensitive direct measurements, e.g. from $t\bar{t}$ final states [30]. Variations of
 1093 the trilinear coupling away from the Standard Model are considered, however. Such variations
 1094 are parametrized via

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

1095 where λ_{HHH} is a varied coupling and λ_{HHH}^{SM} is the Standard Model prediction. As this
 1096 variation comes as a prefactor only with the *triangle* diagram, significant and interesting
 1097 effects are observed due to the interference. Examples of the impact of this tradeoff on the
 1098 di-Higgs invariant mass are shown in Figure 2.4. Generally speaking, the triangle diagram
 1099 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

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

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

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

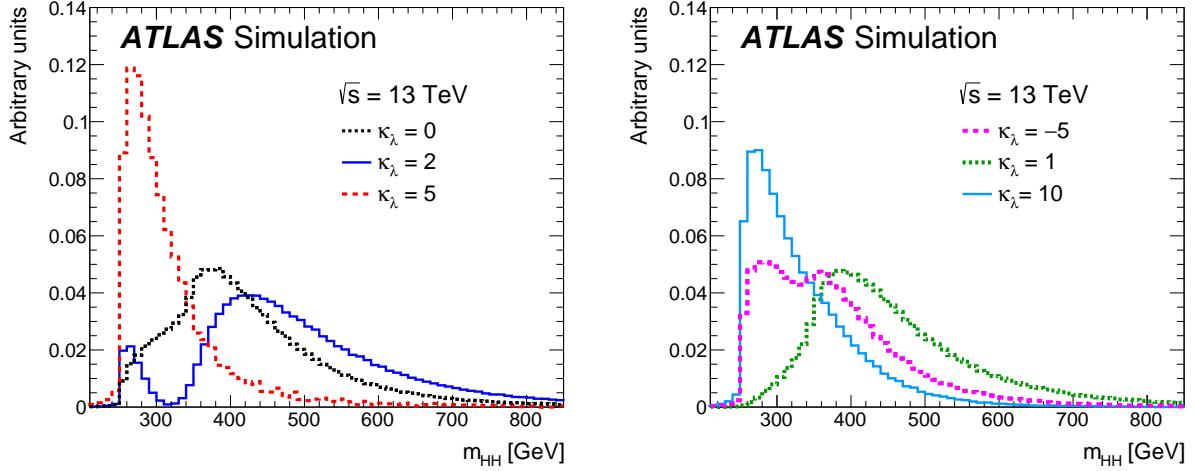


Figure 2.4: Monte Carlo generator level m_{HH} distributions for various values of κ_λ , demonstrating the impact of the interference between the two diagrams of Figure 2.2 on the resulting m_{HH} distribution. For $\kappa_\lambda = 0$ there is no triangle diagram contribution, demonstrating the shape 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 $\kappa_\lambda = -5$. At $\kappa_\lambda = 5$, the interference leads to a deficit at high m_{HH} , resulting in a narrower distribution (and thus a more pronounced low mass peak) than the $\kappa_\lambda = 10$ case. [26]

1100 and thus non-resonant HH production cross section may be parametrized as a second order
1101 polynomial in κ_λ .

1102 For positive values of κ_λ , due to the relative minus sign between the triangle and box
1103 diagrams, the interference between the two diagrams is *destructive*, with a maximum in-
1104 terference near $\kappa_\lambda = 2.3$, corresponding to the minimum cross section prediction. One
1105 may note that the Standard Model value of $\kappa_\lambda = 1$ is not far away from this minimum –
1106 correspondingly the Standard Model cross section for HH production is quite small, namely

1107 31.05 fb at $\sqrt{s} = 13 \text{ TeV}$ for production via gluon-gluon fusion [31–38] compared to, e.g.
 1108 single Higgs production, with a gluon-gluon fusion production cross section of 46.86 pb at
 1109 $\sqrt{s} = 13 \text{ TeV}$ [39] roughly 1500 times larger! For negative values of κ_λ , the interference is
 1110 constructive.

1111 ATLAS projections [40] of $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, and $b\bar{b}\tau^+\tau^-$ predict an expected signal strength
 1112 for Standard Model HH of 3.5σ with no systematic uncertainties and 3.0σ with systematic
 1113 uncertainties using the 3000 fb^{-1} of data from the HL-LHC (around $20\times$ the full Run 2
 1114 dataset considered in this thesis), constituting an *observation* of HH . As the cross section
 1115 for Standard Model HHH production, corresponding to the quartic Higgs interaction, is
 1116 much smaller (around 0.1 fb at $\sqrt{s} = 14 \text{ TeV}$ [41]), observation of triple Higgs production is
 1117 even farther in the future, and so is not considered here. However this may be interesting for
 1118 future work in a variety of Beyond the Standard Model scenarios (e.g. [42–44]).

1119

Chapter 3

1120

EXPERIMENTAL APPARATUS

1121 What machines must we build to examine the smallest pieces of the universe? The famous
 1122 equation $E = m$ provides that to create massive particles, we need to provide enough energy.
 1123 In order to give kinematic phase space to the types of processes that are examined in this
 1124 thesis (and many others besides), a system must be created in which there is enough energy
 1125 to (at bare minimum), overcome kinematic thresholds: if you want to search for HH decays,
 1126 you should have at least 250 GeV ($= 2 \times m_H$) to work with. It is not enough to simply induce
 1127 such processes, however. These processes need to be captured in some way, emitted energy
 1128 and particles must be characterized and identified, and in the end all of this information must
 1129 be put into a useful and useable form such that selections can be made, statistics can be run,
 1130 and a meaningful statement can be made about the universe. In this chapter, we describe the
 1131 machines behind the physics, namely the Large Hadron Collider and the ATLAS experiment.

1132 **3.1 The Large Hadron Collider**

1133 The Large Hadron Collider is a particle accelerator near Geneva, Switzerland. In broad scope,
 1134 it is a ring with a 27 kilometer circumference. Hadrons (usually protons or heavy ions) move
 1135 in two counter-circulating beams, which are made to collide at four collision points at various
 1136 points on the ring. These four collision points correspond to the four detectors placed around
 1137 the ring: two “general purpose” experiments: ATLAS and CMS; LHCb, focused primarily on
 1138 flavor physics; and ALICE, focused primarily on heavy ions.

1139 The focus of this thesis is proton-proton collisions at center of mass energy $\sqrt{s} = 13$ TeV.
 1140 The process to achieve such collisions proceeds as follows: first, an electric field strips hydrogen
 1141 of its electrons, creating protons. A linear accelerator, LINAC 2, accelerates protons to

₁₁₄₂ 50 MeV. The resulting beam is injected into the Proton Synchrotron Booster (PSB), which
₁₁₄₃ pushes the protons to 1.4 GeV, and then the Proton Synchrotron, which brings the beam to
₁₁₄₄ 25 GeV.

₁₁₄₅ Protons are then transferred to the Super Proton Synchrotron (SPS), which ramps up
₁₁₄₆ the energy to 450 GeV. Finally, the protons enter the LHC itself, bringing the beam up to
₁₁₄₇ 6.5 TeV [45].

₁₁₄₈ While there is, of course, much that goes into the Large Hadron Collider development and
₁₁₄₉ operation, perhaps two of the most fundamental ideas are (1) how are the beams directed
₁₁₅₀ and manipulated and (2) what do we mean when we say “protons are accelerated”. These
₁₁₅₁ questions both are directly answered by pieces of hardware, namely (1) magnets and (2)
₁₁₅₂ radiofrequency (RF) cavities.

₁₁₅₃ One of fundamental components of the LHC is a large set of superconducting niobium-
₁₁₅₄ titanium magnets. These are cooled by liquid helium to achieve superconducting temperatures,
₁₁₅₅ and there are several types with very specific purposes. The obvious first question with a
₁₁₅₆ circular accelerator is how to keep the particle beam moving around in that circle. This job
₁₁₅₇ is done via a set of dipole magnets placed around the *beam pipes*: the tubes containing the
₁₁₅₈ beam. These are designed such that the magnetic field in the center of the beam pipe runs
₁₁₅₉ perpendicular to the velocity of the charged particles, providing the necessary centripetal
₁₁₆₀ force for the synchrotron motion.

₁₁₆₁ A proton beam is not made of a single proton, however, but of many protons, grouped
₁₁₆₂ into a series of *bunches*. As all of these are positively charged, if unchecked, these bunches
₁₁₆₃ would become diffuse and break apart. What we want is a stable beam with tightly clustered
₁₁₆₄ protons to maximize the chance of a high energy collision. Such clustering is done via a series
₁₁₆₅ of quadropole magnets, with field distributed as in *TODO: grab image from General Exam*.
₁₁₆₆ Alternating sets of quadropoles provide the necessary forces for a tight, stable beam. While
₁₁₆₇ these are the two major components of the LHC magnet system, it is not the full story –
₁₁₆₈ higher order magnets are used to correct for small imperfections in the beam.

₁₁₆₉ 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. This oscillation creates 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

1194 Small Wheel), liquid argon calorimeter; upgrades in preparation for the High Luminosity
1195 LHC (HL-LHC).

1196 5. Run 3 (2021–2023?): Third run of the LHC, target center of mass energy $\sqrt{s} =$
1197 $13 - 14 \text{ TeV}$, total target luminosity 300 fb^{-1} .

1198 6. Long Shutdown 3 (LS3; 2024?–2026?): Further upgrades for the HL-LHC.

1199 7. Run 4, 5, ... (2026? onward): High Luminosity LHC – goal is to achieve instantaneous
1200 luminosities by a factor of five, massively enlarging available statistics for physics.
1201 Projected 3000 to 4000 fb^{-1} , > 20 times the full Run 2 ATLAS dataset.

1202 3.2 The ATLAS Experiment

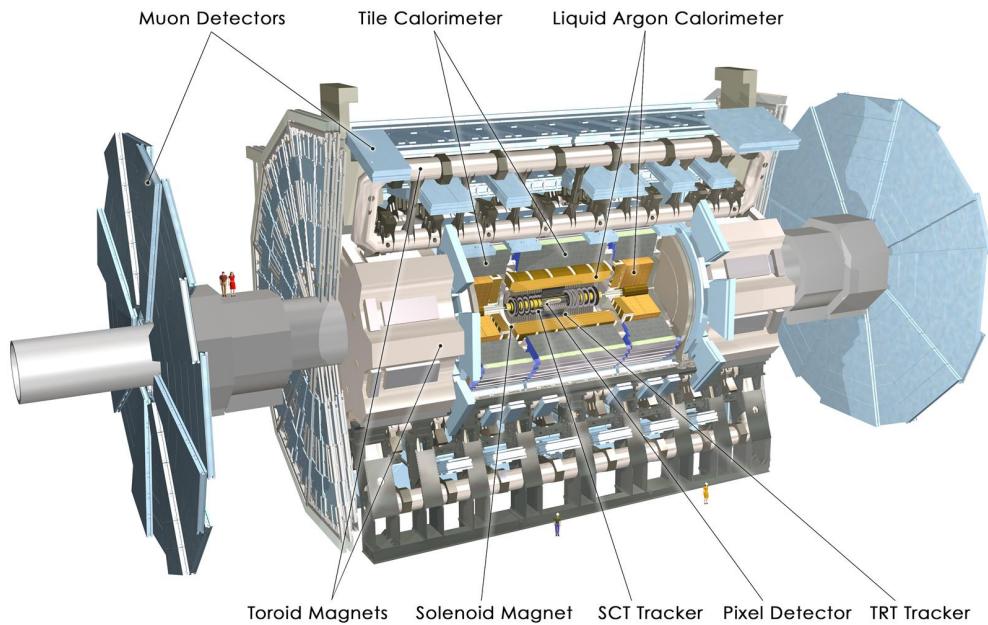


Figure 3.1: Diagram of the ATLAS detector [46]

1203 This thesis focuses on searches done with the ATLAS experiment. As mentioned, this is one

1204 of two “general purpose” experiments at the LHC, by which we mean there is a very large and
 1205 broad variety of physics done within the experimental collaboration. This broad physics focus
 1206 has a direct relation to the design of the ATLAS detector [47], pictured in Figure 3.1, which
 1207 is composed of a sophisticated set of subsystems designed to fully characterize the physics of
 1208 a given high energy particle collision. It consists of an inner tracking detector surrounded
 1209 by a thin superconducting solenoid, electromagnetic and hadronic calorimeters, and a muon
 1210 spectrometer incorporating three large superconducting toroidal magnets. The ATLAS
 1211 detector covers nearly the entire solid angle around the collision point, fully characterizing
 1212 the “visible” components of a collision and allowing for indirect sensitivity to particles that
 1213 do not interact with the detector (e.g. neutrinos) via “missing” energy (roughly momentum
 1214 balance). We will go through the design and physics contribution of each of the detector
 1215 components in the following. A schematic of how various particles interact with the detector
 1216 is shown in Figure 3.2.

1217 3.2.1 ATLAS Coordinate System

1218 Of relevance for the following discussion, as well as for the analysis presented in Chapters
 1219 6 through 10, is the ATLAS coordinate system. ATLAS uses a right-handed coordinate
 1220 system with its origin at the nominal interaction point (IP) in the center of the detector and
 1221 the z -axis along the beam pipe. The x -axis points from the IP to the center of the LHC
 1222 ring, and the y -axis points upwards. Cylindrical coordinates (r, ϕ) are used in the transverse
 1223 plane, ϕ being the azimuthal angle around the z -axis. The pseudorapidity is defined in
 1224 terms of the polar angle θ as $\eta = -\ln \tan(\theta/2)$. Angular distance is measured in units of
 1225 $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$. These coordinates are shown in Figure 3.3.

1226 3.2.2 Inner Detector

1227 The purpose of the inner detector is the reconstruction of the trajectory of charged particles,
 1228 called *tracking*. This is accomplished primarily through the collection of electrons displaced
 1229 when a charged particle passes through a tracking detector. By setting up multiple layers of

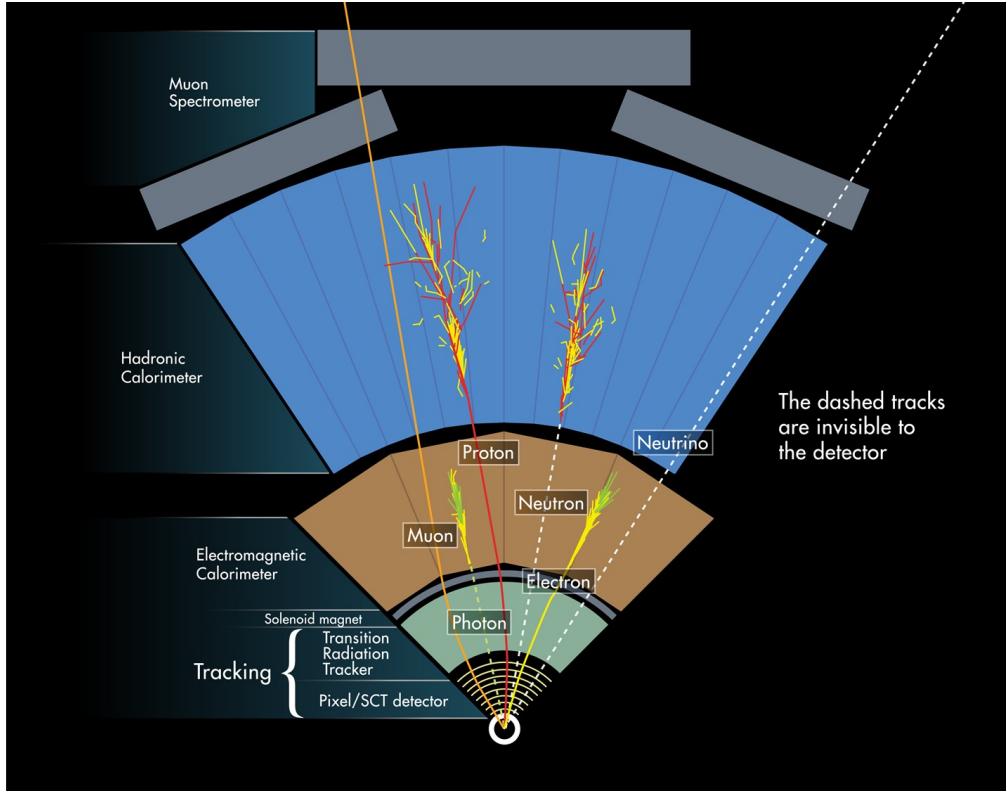


Figure 3.2: Cross section of the ATLAS detector showing how particles interact with various detector components [48]

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

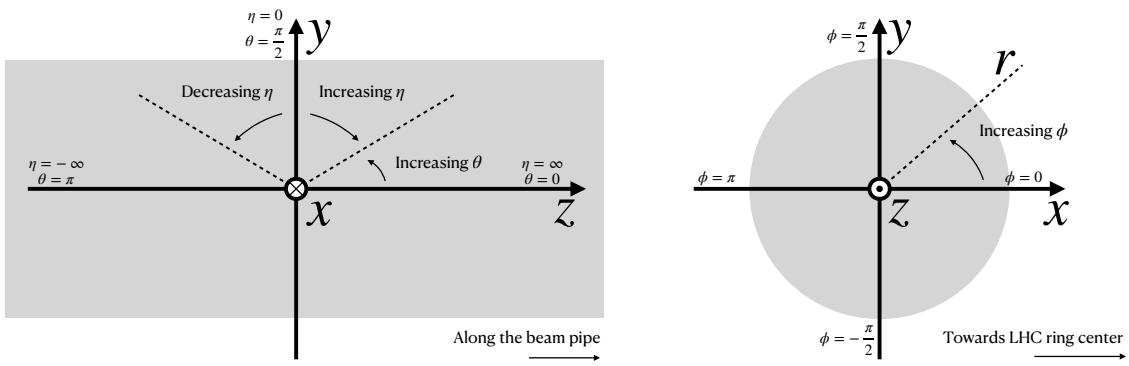


Figure 3.3: 2D projections of the ATLAS coordinate system

1240 (IBL) installed before Run 2 [49, 50]. This is very close to the interaction point with a
 1241 high degree of positional information, and is therefore very important for e.g. b -tagging (see
 1242 Chapter 5). It is followed by the silicon microstrip tracker (SCT), which usually provides
 1243 eight measurements per track. This is lower granularity, but similar in concept to the pixel
 1244 detector.

1245 Both of these silicon detectors are complemented by the transition radiation tracker
 1246 (TRT), which extends the radial track reconstruction within the range $|\eta| < 2.0$. This is
 1247 a different design, composed of *drift tubes*, i.e. straws filled with Xenon gas with a wire
 1248 in the center, but similarly collects electrons displaced by ionizing particles. In addition,
 1249 the TRT includes materials with widely varying indices of refraction, which leads to the
 1250 production of transition radiation, namely radiation produced by a charged particle passing
 1251 through an inhomogeneous medium. The energy loss on such a transition is proportional
 1252 to the Lorentz factor $\gamma = E/m$ – correspondingly, lighter particles (e.g. electrons) tend to
 1253 lose more energy and emit more photons compared to heavier particles (e.g. pions). In the
 1254 detector, this corresponds to a larger fraction of hits (typically 30 in total) above a given

1255 high energy-deposit threshold for electrons, providing particle identification information.

1256 *3.2.3 Calorimeter*

1257 Surrounding the inner detector in ATLAS is the calorimeter. The principle of the calorimeter
1258 is to completely absorb the energy of a produced particle in order to measure it. However,
1259 a pure block of absorber does not provide much information about the particle interaction
1260 with the material. The ATLAS calorimeter therefore has a *sampling calorimeter* structure,
1261 namely, layers of absorber interspersed with layers of sensitive material, giving the calorimeter
1262 “stopping power” while allowing detailed measurement of the resulting particle shower and
1263 corresponding deposited energy.

1264 The ATLAS calorimetersystem covers the pseudorapidity range $|\eta| < 4.9$, and is primarily
1265 composed of two components, an electromagnetic calorimeter, designed to measure particles
1266 which primarily interact via electromagnetism (e.g. photons and electrons), and a hadronic
1267 calorimeter, designed to measure particles which interact via the strong force (e.g. pions,
1268 other hadrons). We will return to the differences between these in a moment.

1269 In ATLAS, the electromagnetic calorimeter covers the region of $|\eta| < 3.2$, and uses
1270 lead for the absorbers and liquid-argon for the sensitive material. It is high granularity
1271 and, geometrically, has two components: the “barrel”, which covers the cylindrical body of
1272 the detector volume and the “endcap”, covering the ends. An additional thin liquid-argon
1273 presampler covers $|\eta| < 1.8$ to correct for energy loss in material upstream of the calorimeters.

1274 The hadronic calorimeter is composed of alternating steel and plastic scintillator tiles,
1275 segmented into three barrel structures within $|\eta| < 1.7$, in addition to two copper/liquid-argon
1276 endcap calorimeters.

1277 The solid angle coverage is completed with forward copper/liquid-argon and tungsten/liquid-
1278 argon calorimeter modules optimized for electromagnetic and hadronic energy measurements
1279 respectively.

1280 3.2.4 *Muon Spectrometer*

1281 While muons interact electromagnetically, they are around 200 times heavier than electrons
 1282 ($m_\mu = 106 \text{ MeV}$, while $m_e = 0.510 \text{ MeV}$). Therefore, electromagnetic interactions with
 1283 absorbers in the calorimeter are not sufficient to stop them, and, as they do not interact
 1284 via the strong force, hard scattering with nuclei is rare. A dedicated system for muon
 1285 measurements is therefore required.

1286 The muon spectrometer (MS) is the outermost layer of ATLAS and is designed for this
 1287 purpose. It is composed of three parts: a set of triggering chambers, which detect if there is
 1288 a muon and provide a coordinate measurement, in conjunction with high-precision tracking
 1289 chambers, which measure the deflection of muons in a magnetic field to measure muon
 1290 momentum, similar to the inner detector solenoid. The magnetic field is generated by the
 1291 superconducting air-core toroidal magnets, with a field integral between 2.0 and 6.0 T m
 1292 across most of the detector. The toroid magnetic field runs roughly in a circle in the $x - y$
 1293 plane around the beam line, leading to muon curvature along the z-axis.

1294 The precision tracking system covers the region $|\eta| < 2.7$ via three layers of monitored
 1295 drift tubes, and is complemented by cathode-strip chambers in the forward region, where the
 1296 background is highest. The muon trigger system covers the range $|\eta| < 2.4$ with resistive-plate
 1297 chambers in the barrel, and thin-gap chambers in the endcap regions.

1298 3.2.5 *Triggering*

1299 During a typical run of the LHC, there are roughly 1 billion collisions in ATLAS per second
 1300 (1 GHz), corresponding to a 40 MHz bunch crossing rate [51]. Saving the information from
 1301 all of them is not only unnecessary, but infeasible. The ATLAS trigger system provides a
 1302 sophisticated set of selections to filter the collision data and only keep those collision events
 1303 useful for downstream analysis.

1304 These events are selected by the first-level trigger system, which is implemented in custom
 1305 hardware, and accepts events at a rate below 100 kHz. Selections are then made by algorithms

1306 implemented in software in the high-level trigger [52], reducing this further, and, in the end,
1307 events are recorded to disk at much more manageable rate of about 1 kHz.

1308 An extensive set of ATLAS software [53] is open source, including the software used for
1309 real and simulated data reconstruction and analysis and that used in the trigger and data
1310 acquisition systems of the experiment.

1311 3.2.6 Particle Showers and the Calorimeter

1312 The design of the ATLAS detector is directly tied to the physics it is trying to detect. Of these,
1313 possibly the most non-trivial distinction is in the calorimeter design. It is therefore useful to
1314 discuss in more detail the various properties of electromagnetic and hadronic interactions
1315 with material, and how these correspond to the particle showers measured by the detector
1316 described above.

1317 Electromagnetic showers in ATLAS predominantly occur via bremsstrahlung, or “braking
1318 radiation”, and electron-positron pair production. This proceeds roughly as follows: an
1319 electron entering a material is deflected by the electromagnetic field of a heavy nucleus. This
1320 results in the radiation of a photon. That photon produces an electron-positron pair, and
1321 the process repeats, resulting in a shower structure. At each step, characterized by *radiation*
1322 *length*, X_0 , the number of particles approximately doubles and the average particle energy
1323 decreases by 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

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

1324 where Z is the nuclear charge. From a similar analysis of rates, we may see that the
1325 bremsstrahlung rate is inversely proportional to the square of the mass of the particle. This

₁₃₂₆ explains why muons do not shower in a similar way, as the rate of bremsstrahlung is suppressed
₁₃₂₇ 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

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

₁₃₂₈ such that for a 100 GeV shower in lead, $x \sim 13$. The radiation length for lead is around
₁₃₂₉ 0.56 cm, such that an electromagnetic shower could be expected to be captured within 10 cm
₁₃₃₀ of lead.

₁₃₃₁ Electromagnetic showers are therefore characterized by depositing much of their energy
₁₃₃₂ within a small region of space. As we show below (Chapter 4) though electromagnetic
₁₃₃₃ showering is not deterministic, the large number of particles and the restricted set of processes
₁₃₃₄ involved means that the shower development as a whole is very similar between individual
₁₃₃₅ electromagnetic showers of the same energy.

₁₃₃₆ For completeness, note as well that pair production dominates for photons of energy greater
₁₃₃₇ than around 10 MeV, whereas for lower energies (below around 1 MeV), the photoelectric
₁₃₃₈ effect, namely atomic photon absorption and electron emission, dominates.

₁₃₃₉ Hadronic showers are distinguished by the fact that they interact strongly with atomic
₁₃₄₀ nuclei. They are correspondingly more complex because (1) they involve a wider variety
₁₃₄₁ of processes than electromagnetic showers, and (2) these processes have a wide variety of
₁₃₄₂ associated length scales. Because these are heavier than electrons (e.g. protons and charged
₁₃₄₃ pions) bremsstrahlung is suppressed, but ionization interactions with the electrons will cause
₁₃₄₄ these particles to lose energy as they pass through the material. Hadronic showering occurs
₁₃₄₅ on interaction with atomic nuclei. This may lead to production of, e.g. both charged (π^\pm)
₁₃₄₆ and neutral (π^0) pions. The π^0 lifetime is much much shorter than that of the charged pions
₁₃₄₇ (around a factor of 10^8), and immediately decays to two photons, starting an electromagnetic

1348 shower, as described above. The longer lived π^\pm travel further in the detector before
1349 experiencing another strong interaction with more particles produced, also with varying
1350 lifetimes and decay properties.

1351 It is therefore immediately apparent that hadronic showers are more complex than
1352 electromagnetic ones (electromagnetic showers can be a subset of the hadronic!), and therefore
1353 much more variable from shower to shower. The length scales involved are also significantly
1354 larger due to the reliance on nuclear interactions, characterized by length λ_I , which is around
1355 17 cm for iron (used in the ATLAS hadronic calorimeter). This motivates the calorimeter
1356 design, and results in the properties demonstrated in Figure 3.2.

1357

Chapter 4

1358

SIMULATION

1359 Simulated physics samples are a core piece of the physics output of the Large Hadron
 1360 Collider, providing a map from a physics theory into what is observed in our detector. This
 1361 is crucial for searches for new physics, where simulation is necessary to describe what a given
 1362 signal model looks like, but also extremely valuable for describing the physics of the Standard
 1363 Model, providing detailed predictions of background processes for use in everything from
 1364 designing simple cuts to training multivariate discriminators. Broadly, simulation can be split
 1365 into two stages: *event generation*, in which physics theory is used to generate a description of
 1366 particles present after a proton-proton collision, and *detector simulation*, which passes this
 1367 particle description through a simulation of the detector material, providing a view of the
 1368 physics event as it would be seen in ATLAS data. Such simulation is often called Monte Carlo
 1369 in reference to the underlying mathematical framework, which relies on random sampling.

1370 **4.1 Event Generation**

1371 A variety of tools are used to simulate various aspects of event generation. One such aspect
 1372 is generation of the “hard scatter” event, i.e., two protons collide and some desired physics
 1373 process happens. In practice, this is not quite as simple as two quarks or gluons interacting.
 1374 Protons are composed of three “valence” quarks with various momenta interacting with each
 1375 other via exchange of gluons, but also a sea of virtual gluons which may decay into other
 1376 quarks. A hard scatter event is therefore characterized by the corresponding particle level
 1377 diagrams, but additionally by a set of *parton distribution functions* (PDFs), which describe
 1378 the probability to find constituent quarks or gluons at carrying various momenta at a given
 1379 energy scale (often written Q^2). Such PDFs are measured experimentally *TODO: cite* and

1380 the selection of a “PDF set” and a given physics process characterizes the hard scatter.
 1381 Depending on the model being considered and the particular theoretical constraints, processes
 1382 are often simulated at either leading (LO) or next to leading order (NLO), corresponding to
 1383 the order of the perturbative expansion (i.e. tree level or 1 loop diagrams). Various additional
 1384 tools are developed for such NLO calculations, including POWHEG Box v2 [54–56], which is
 1385 used for this thesis. MADGRAPH [57] is used in this thesis for leading order simulation.

1386 The hard scatter is not the only component of a given collider event, however. Incoming
 1387 and outgoing particles are themselves very energetic and may radiate particles along their
 1388 trajectory. In particular, gluons, which have a self-interaction term as described in Chapter 1,
 1389 may be radiated, which subsequently themselves radiate gluons or decay to quarks which can
 1390 also radiate gluons, in a whole mess of QCD that both contributes to the particle content
 1391 of a collider event and is not directly described by the hard scatter. This cascade, called a
 1392 *parton shower*, has a dedicated set of simulation tools. For this thesis, HERWIG 7 [58][59] and
 1393 PYTHIA 8 [60] are used, which interface with tools such as MADGRAPH for simulation.

1394 Due to color confinement (Chapter 1), quarks and gluons cannot be observed free particles,
 1395 but rather undergo a process called hadronization, in which they are grouped into colorless
 1396 hadrons (e.g. *mesons*, consisting of one quark and one anti-quark). In simulation, this is also
 1397 handled with tools such as HERWIG 7 or PYTHIA 8.

1398 The physics of b -quarks is quite important for a variety of searches for new physics and
 1399 measurements of the Standard Model, including this thesis work. Correspondingly, the decay
 1400 of “heavy flavor” particles (e.g. B and D mesons, containing b and c quarks respectively)
 1401 has been very well studied, and a dedicated simulation tool, EVTGEN [61], is used for such
 1402 processes.

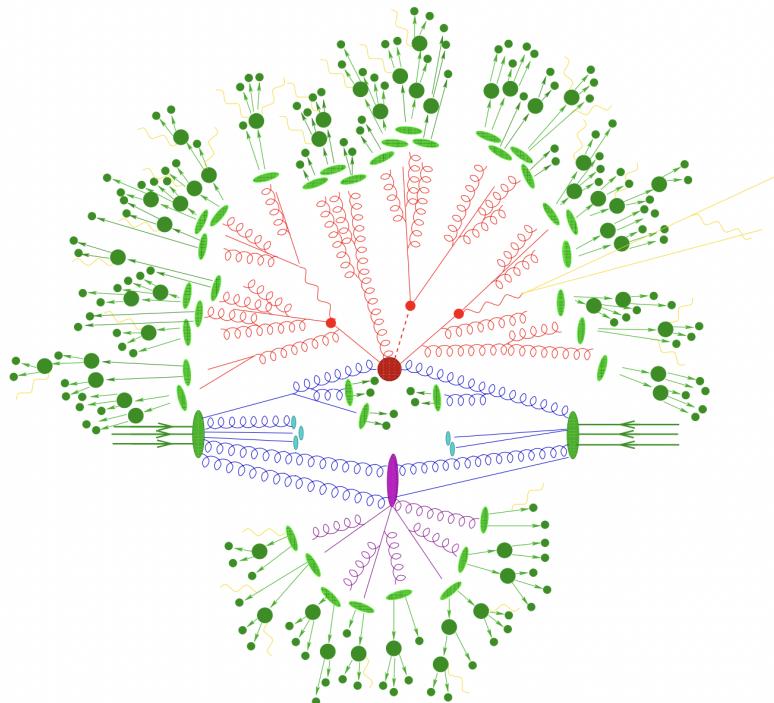


Figure 4.1: Schematic diagram of the Monte Carlo simulation of a hadron-hadron collision. The incoming hadrons are the green blobs with the arrows on the left and right, with the red blob in the center representing the hard scatter event, and the purple representing a secondary hard scatter. Radiation 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. [62]

1403 **4.2 Detector Simulation**

1404 Event generation provides a full and exact description of the particle content of a given
1405 collider event. This description is useful, but is an artifact of the simulation – for real physics
1406 events, we must rely on the information collected by sophisticated detectors (Chapter 3) to
1407 make statements about the physics content of collider events. The simulation of how particles
1408 interact with the physical detector and of the corresponding information that is collected is
1409 therefore a necessary step of physics simulation at the LHC. The design and components of
1410 the ATLAS detector are described in Chapter 3. Simulation of this detector quickly becomes
1411 complicated – there are a variety of different materials and sub-detectors, each with particular
1412 configurations and resolutions. Interactions of particles with the detector materials can cause
1413 showering, and such showers must be simulated and characterized.

1414 In ATLAS, the GEANT4 [63] simulation toolkit is used for detailed simulation of the
1415 ATLAS detector, often referred to as *full simulation*. The method can be thought of as
1416 proceeding step by step as a particle moves through the detector, simulating the interaction
1417 of the material at each stage, and following each branch of each resulting shower with a
1418 similarly detailed step by step simulation.

1419 This type of simulation is very computationally intensive, especially in the calorimeter,
1420 which has a high density of material, leading to an extremely large set of material interactions
1421 to simulate. There is correspondingly a large effort within ATLAS to develop techniques to
1422 decrease the computational load – these techniques will be of increasing importance for Run
1423 3 and the HL-LHC, which will have increased computational need due to the high complexity
1424 and large volume of collected physics events, along with the corresponding set of simulated
1425 physics events [64]. The divergence of the baseline computing model from the projected
1426 computing budget is shown in Figure 4.2.

1427 The fast simulation used for this thesis, AtlFast-II [66], is one such technique, which uses
1428 a parametrized simulation of the calorimeter, called FastCaloSim, in conjunction with full
1429 simulation of the inner detector, to achieve an order of magnitude speed up in simulation

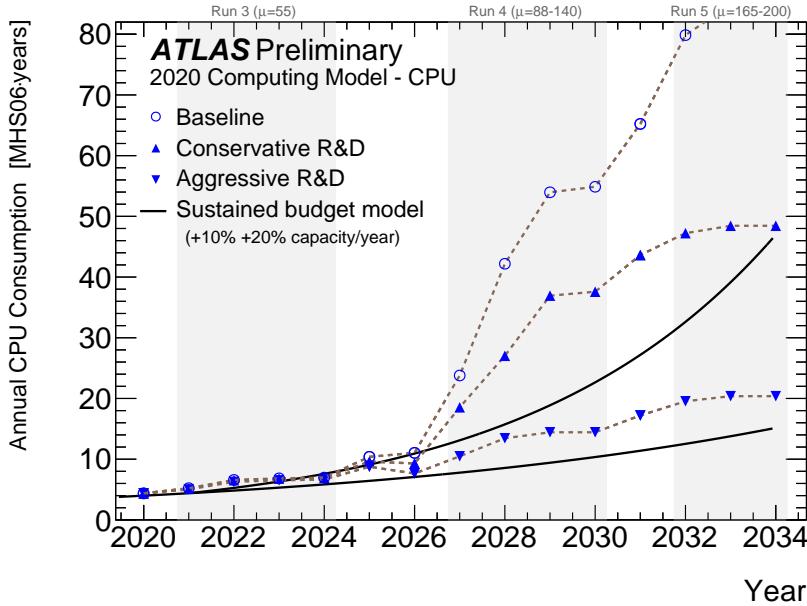


Figure 4.2: The projected ATLAS computational requirements for Run 3 and the HL-LHC relative to the projected computing budget. Aggressive R&D is required to keep resources within budget [65].

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.

1438 **4.3 Correlated Fluctuations in FastCaloSim**

1439 A variety of developments have been made to FastCaloSim, improving on the version used for
1440 AtlFast-II. This new fast calorimeter simulation [67] is largely based on two components: one
1441 which describes the *total energy* deposited in each calorimeter layer as a shower moves from
1442 the interaction point outward, and one which describes the *shape*, i.e., the pattern of energy
1443 deposits, of a shower in each respective calorimeter layer. Both methods are parametrizations
1444 of the full simulation, and therefore are considered to be performing well if they are able
1445 to reproduce corresponding full simulation distributions. Of course, directly sampling from
1446 a library of showers would identically reproduce such distributions – however a statistical
1447 sampling of various shower *properties* provides much more generality in the simulation.

1448 For the simulation of total energy in each given layer, the primary challenge is that such
1449 energy deposits are highly correlated. The new FastCaloSim thus relies on a technique called
1450 Principal Component Analysis (PCA) [68] to de-correlate the layers, aiding parametrization.

1451 The PCA chain transforms N energy inputs into N Gaussians and projects these Gaussians
1452 onto the eigenvectors of the corresponding covariance matrix. This results in N de-correlated
1453 components, as the eigenvectors are orthogonal. The component of the PCA decomposition
1454 with the largest corresponding eigenvalue is then used to define bins, in which showers
1455 demonstrate similar patterns of energy deposition across the calorimeter layers. To further
1456 de-correlate the inputs, the PCA chain is repeated on the showers within each such bin. This
1457 full process is reversed for the particle simulation. A full description of the method can be
1458 found in [67].

1459 Modeling of the lateral shower shape makes use of 2D histograms filled with GEANT4
1460 hit energies in each layer and PCA bin. Binned in polar $\alpha - R$ coordinates in a local plane
1461 tangential to the surface of the calorimeter system, these histograms represent the spatial
1462 distribution of energy deposits for a given particle shower. Such histograms are constructed
1463 for a number of GEANT4 events, and the histograms for each event are normalized to total
1464 energy deposited in the given layer. The average of these histograms is then taken (what is

¹⁴⁶⁵ called here the “average shape”).

¹⁴⁶⁶ In simulation, these average shape histograms are used as probability distributions, from
¹⁴⁶⁷ which a finite number of equal energy hits are drawn. This finite drawing of hits induces
¹⁴⁶⁸ a statistical fluctuation about the average shape which is tuned to match the expected
¹⁴⁶⁹ calorimeter sampling uncertainty.

¹⁴⁷⁰ As an example, the intrinsic resolution of the ATLAS Liquid Argon calorimeter has a
¹⁴⁷¹ sampling term of $\sigma_{\text{samp}} \approx 10\%/\sqrt{E}$ [69]. The number of hits to be drawn for each layer, $N_{\text{hits}}^{\text{layer}}$,
¹⁴⁷² is thus taken from a Poisson distribution with mean $1/\sigma_{\text{samp}}^2$, where the energy assigned to
¹⁴⁷³ 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
¹⁴⁷⁴ each bin in the average shape.

¹⁴⁷⁵ Figure 4.3 shows a comparison of energy and weta2 [70], defined as the energy weighted
¹⁴⁷⁶ lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons
¹⁴⁷⁷ simulated with the new FastCaloSim and with full GEANT4 simulation. The agreement is
¹⁴⁷⁸ 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.

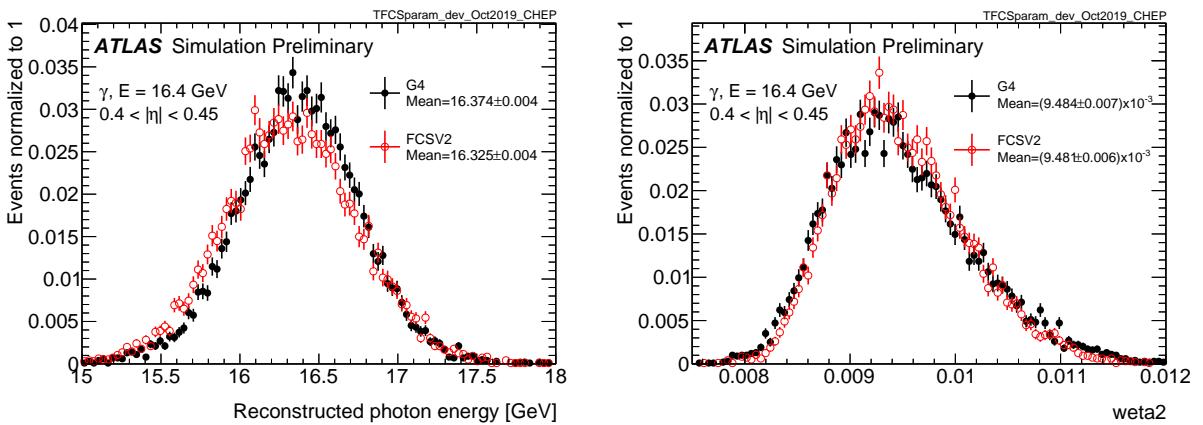


Figure 4.3: Energy and variable weta2, defined as the energy weighted lateral width of a shower in the second electromagnetic calorimeter layer, for 16 GeV photons with full simulation (G4) and FastCaloSimV2 (FCSV2) [67].

1480 *4.3.1 Fluctuation Modeling*

1481 Figure 4.4 shows the ratio of calorimeter cell energies for single GEANT4 photon and pion
 1482 events to the corresponding cell energies in their respective average shapes. While the photon
 1483 event is quite close to the corresponding average, the pion event shows a deviation from the
 1484 average which is much larger and has a non-trivial structure, reflecting the different natures
 1485 of electromagnetic and hadronic showering.

1486 While the shape parametrization described above is thus sufficient for describing electro-
 1487 magnetic showers, we will demonstrate below that it is not sufficient for describing hadronic
 1488 showers (Figures 4.7 and 4.8). We therefore present and validate methods to improve this
 1489 hadronic shower modeling. Such methods have been presented as well in [1].

1490 Two methods for modeling deviations from the average shape have been studied: (1)
 1491 a neural network based approach using a Variational Autoencoder (VAE) [71] and (2) a
 1492 map through cumulative distributions to an n -dimensional Gaussian. With both methods,
 1493 the shape simulation then proceeds as described in Section 4.3, with the drawing of hits
 1494 according to the average shape. However, these hits no longer have equal energy, but have
 1495 weights applied to increase or decrease their energy depending on their spatial position.
 1496 This application of weights is designed to mimic a realistic shower structure and to encode
 1497 correlations between energy deposits.

1498 Both methods are trained on ratios of energy in binned units called voxels. This voxelization
 1499 is performed in the same polar $\alpha - R$ coordinates as the average shape, with a 5 mm core in
 1500 R and 20 mm binning thereafter. There are a total of 8 α bins from 0 to 2π and 8 additional
 1501 R bins from 5 mm to 165 mm. The 5 mm core is filled with the average value of core voxels
 1502 across the 8 α bins when creating the parametrization. However, during simulation, each of
 1503 these 8 core bins is treated independently. The outputs of both methods mimic these energy
 1504 ratios and are used in the shape simulation as the weights described above. In contrast to
 1505 an approach based on, e.g., calorimeter cells, using voxels allows for flexibility in tuning the
 1506 binning used in creating the parametrization. Further, due to their relatively large size, using

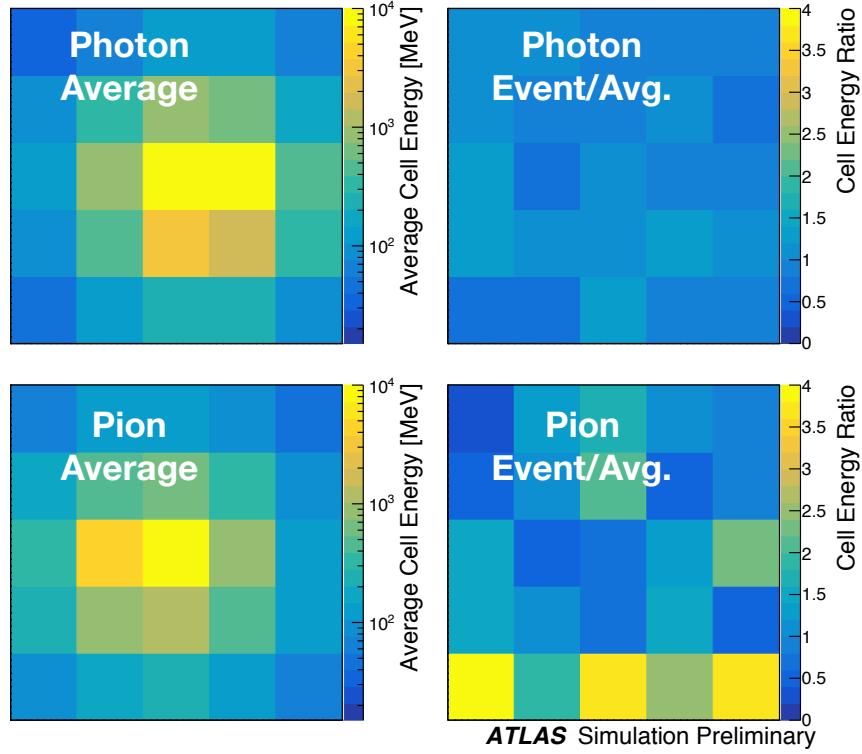


Figure 4.4: Example of photon and pion average shapes in 5×5 calorimeter cells. The left column shows the average shape over a sample of 10000 events, while the right column shows the energy ratio, in each cell, of single GEANT4 events with respect to this average. The photon ratios are all close to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

1507 calorimeter cells is subject to “edge effects”, where the splitting of energy between cells has a
 1508 non-trivial effect on the observed energy ratio. The binning used here is of the order of half
 1509 of a cell size, mitigating this effect.

1510 The Gaussian method operates by using cumulative distributions to map GEANT4 energy
 1511 ratios to a multidimensional Gaussian distribution. New events are generated by randomly
 1512 sampling from this Gaussian distribution.

1513 For the VAE method, a system of two linked neural networks is trained to generate events.

1514 The first “encoder” neural network maps input GEANT4 energy ratios to a lower dimensional
 1515 latent space. A second “decoder” neural network then samples from that latent space and
 1516 tries to reproduce the inputs. In simulation, events are generated by taking random samples
 1517 from the latent space and passing them through the trained decoder.

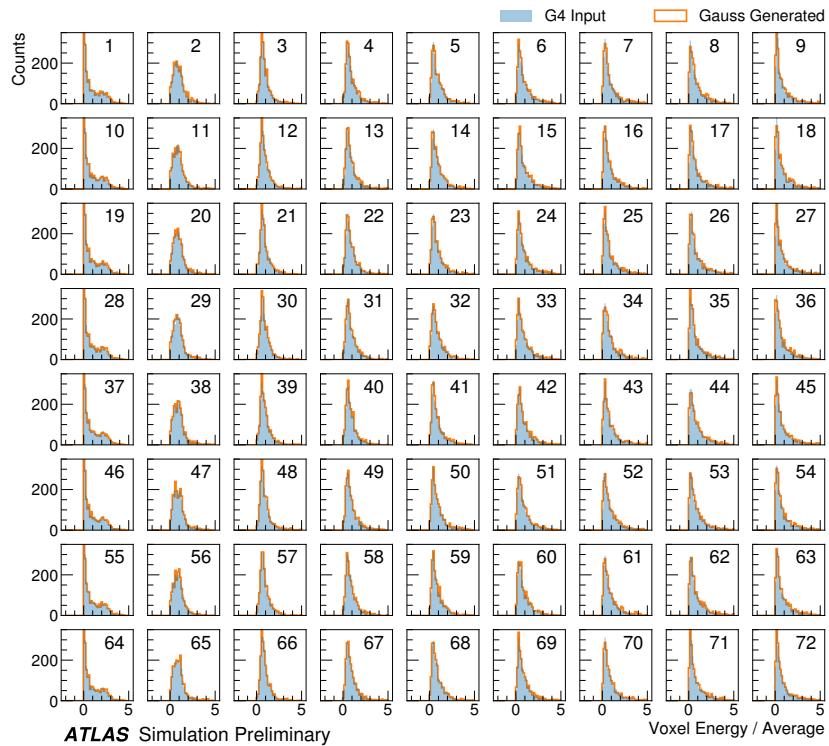


Figure 4.5: Distribution of the ratio of voxel energy in single events to the corresponding voxel energy in the average shape, with GEANT4 events in blue and Gaussian model events in orange, for 65 GeV central pions in EMB2. Moving top to bottom corresponds to increasing α , left to right corresponds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

1518 Figure 4.5 shows the distributions of input GEANT4 and Gaussian method generated
 1519 energy ratios in the grid of voxels. Figure 4.6 shows the correlation coefficient between the
 1520 center voxel from $\alpha = 0$ to $2\pi/8$ for input GEANT4 and the Gaussian and VAE fluctuation

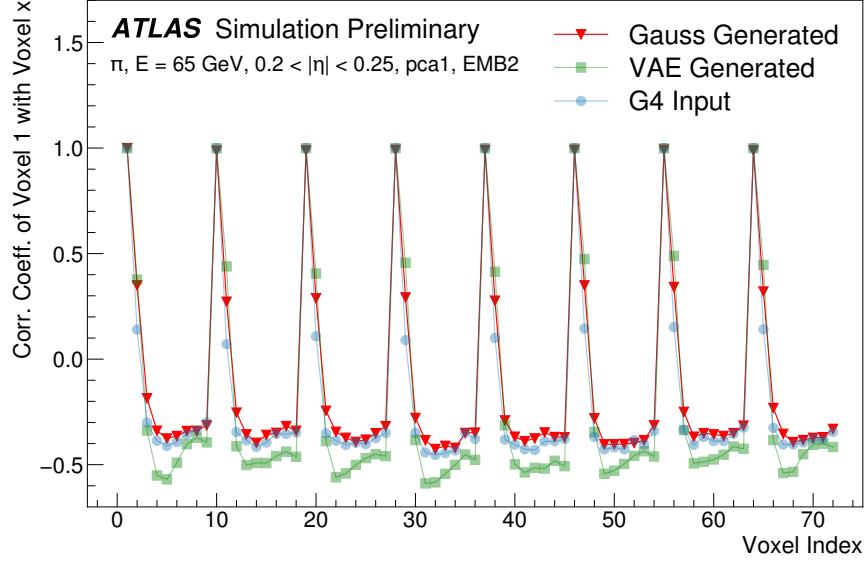


Figure 4.6: Correlation coefficient of ratios of voxel energy in single events to the corresponding voxel energy in the average shape, examined between the core bin from $\alpha = 0$ to $2\pi/8$ and each of the other voxels. The periodic structure represents the binning in α , and the increasing numbers 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.

1521 methods. Agreement is good throughout.

1522 Validation of the Gaussian and VAE fluctuation methods was performed within FastCaloSimV2.
 1523 Figure 4.7 shows the energy ratio of cells for a given simulation to the corresponding cells in
 1524 the average shape as a function of the distance from the shower center. The mean for all
 1525 simulation methods is expected to be around 1, with deviation from the average (the RMS
 1526 fluctuation) shown by the error bars. The Gaussian method RMS (red) and VAE method
 1527 RMS (green) both match the GEANT4 RMS (yellow) better than the case without correlated
 1528 fluctuations (blue) for a variety of energies, η points, and layers, often reproducing 80 – 100 %

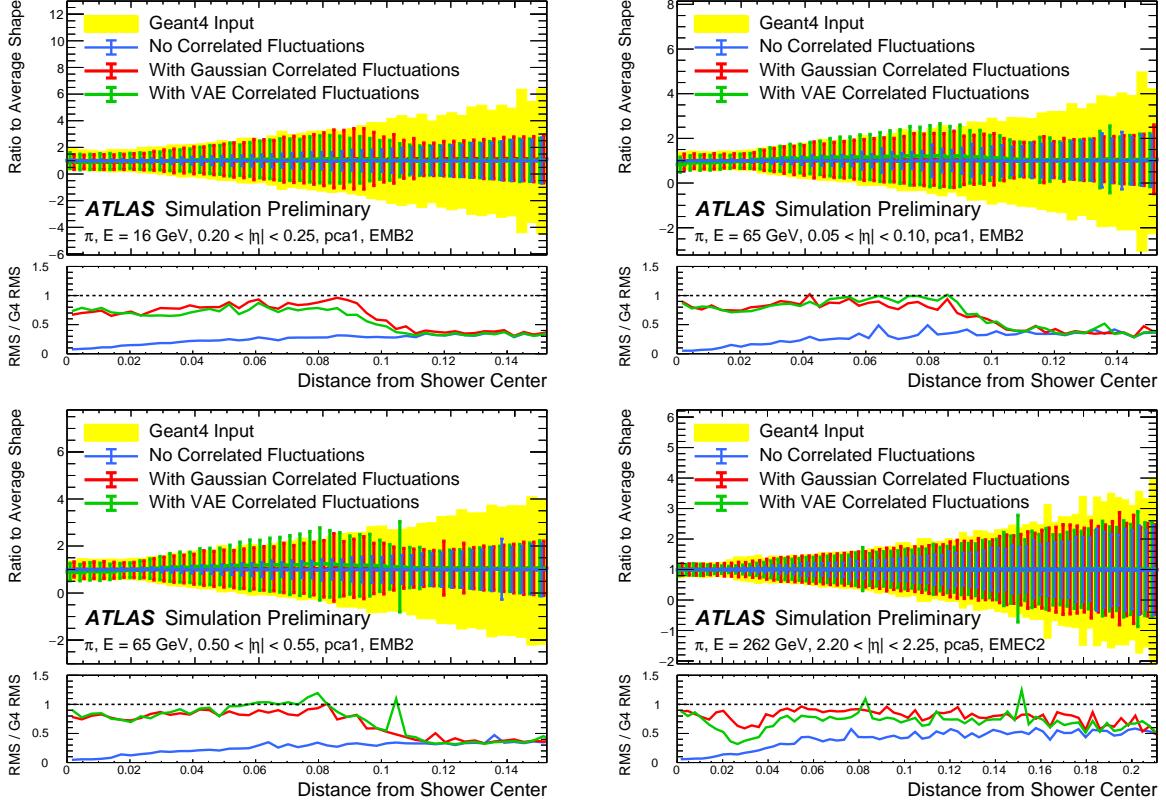


Figure 4.7: Comparison of the RMS fluctuations about the average shape with the Gaussian fluctuation model (red), the VAE fluctuation model (green), and without correlated fluctuations (blue) for a range of pion energies, η points, and layers.

1529 of the GEANT4 RMS magnitude, compared to the 5 – 30 % observed in the no correlated
1530 fluctuations case.

1531 Figure 4.8 shows the result of a simulation with full ATLAS reconstruction for 65 GeV
1532 central pions with the Gaussian fluctuation model. Here a *cluster* [72] is defined as a three-
1533 dimensional spatial grouping of calorimeter cells which are summed based on the input signals
1534 relative to their neighboring cells. The multiplicity, shape, and spatial distribution of such
1535 clusters provides a powerful insight on the structure of energy deposits in the calorimeter,
1536 and good performance in cluster variables is a promising step towards good performance

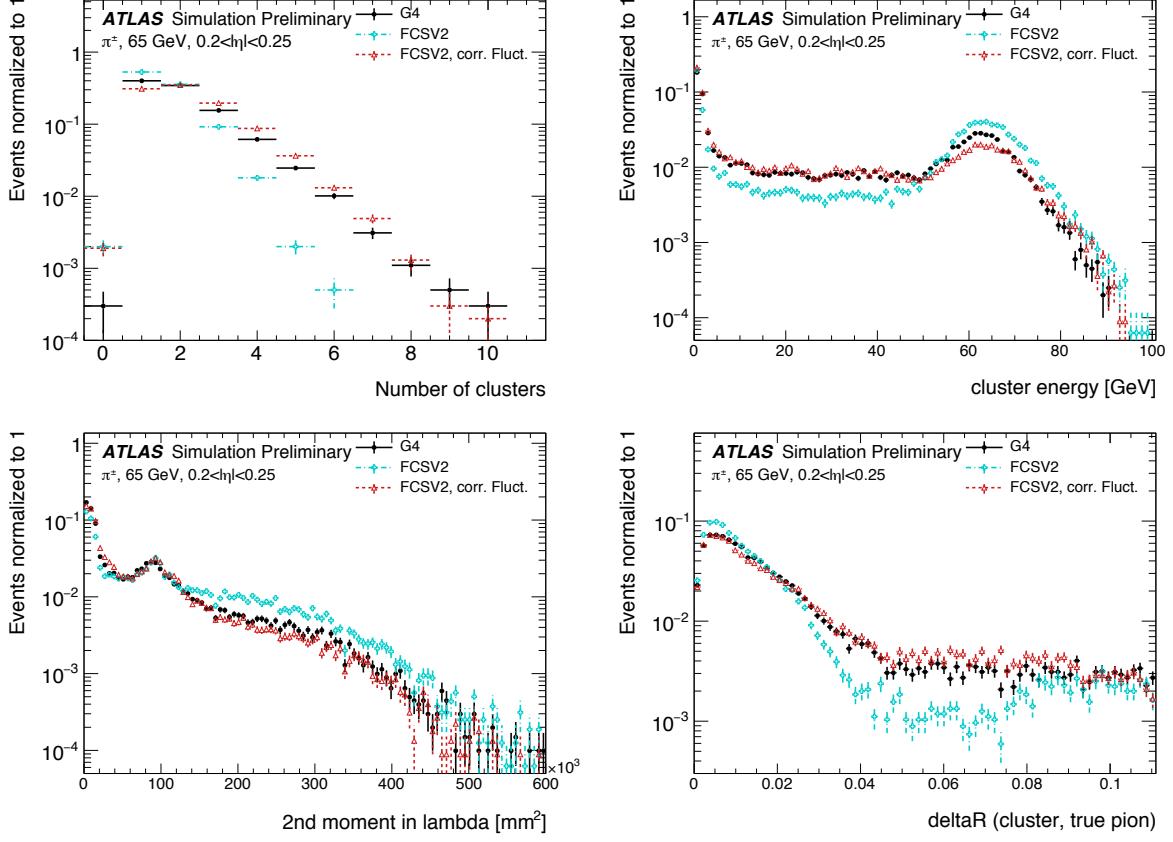


Figure 4.8: Comparison of the Gaussian fluctuation model to the default FCSV2 version and to G4 simulation, using pions of 65 GeV energy and $0.2 < |\eta| < 0.25$. Variables shown relate to calorimeter clusters, three-dimensional spatial groupings of cells [72] which provide powerful insight on the structure 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.

1537 in the modeling of jet substructure, as these clusters may themselves be summed to form
 1538 jets (see Chapter 5). The simulation with the Gaussian fluctuation model demonstrates
 1539 improved modeling of several of these cluster variables relative to baseline FastCaloSimV2,
 1540 reproducing the distributions of events simulated with GEANT4. These include number and
 1541 energy of clusters, the 2nd moment in lambda, ($\langle \lambda^2 \rangle$), which describes the square of the
 1542 longitudinal extension of a cluster, where λ is the distance of a cell from the shower center
 1543 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
 1544 ΔR , between the cluster and the true pion.

1545 The new fast calorimeter simulation is a crucial part of the future of simulation for the
 1546 ATLAS Experiment at the LHC. The per event simulation time of the full detector with
 1547 GEANT4, calculated over 100 $t\bar{t}$ events, is 228.9 s. Using FastCaloSim for the calorimeter
 1548 simulation reduces this to 26.6 s, of which FastCaloSim itself is only 0.015 s, with the majority
 1549 of the remaining simulation time due to GEANT4. Good physics modeling is achieved, and
 1550 the correlated fluctuations method shows good proof of concept improvement for the modeling
 1551 of hadronic showers.

1552 **4.4 Outlook**

1553 There has been significant effort in the community to develop a set of fast simulation tools,
 1554 with the use of machine learning methods at the forefront of such approaches (e.g. [73], [74]).
 1555 Most fast simulation approaches generally are based on parametrizations of fully simulated
 1556 events, but fall into two paradigms - a “by hand” simulation, which focuses on the modeling
 1557 of individual detector effects, or a fully parametrized simulation, in which a generative model
 1558 (e.g. a Generative Adversarial Network or Variational Autoencoder) is trained to directly
 1559 reproduce the input events. Both approaches can be extremely powerful, but each suffer from
 1560 certain drawbacks. The “by hand” approach offers the advantage of direct encoding of expert
 1561 knowledge – if an effect needs to be modeled, a new parametrization is introduced. However,
 1562 by the same token, it requires dedicated parametrizations for each effect. Fully parametrizing
 1563 the simulation with a generative model offloads this burden onto the network itself. However,

1564 by doing so, the ability to use expert knowledge is diminished – the network is required to
1565 learn all relevant effects.

1566 The method presented here represents an effort to step towards a hybrid between these two
1567 approaches, leveraging the power of machine learning techniques for individual parametriza-
1568 tions within the by hand framework. Such hybrid solutions have the potential to be extremely
1569 powerful, and further work on the development of these solutions is an interesting direction
1570 of future study.

1571

Chapter 5

1572

RECONSTRUCTION

1573 Chapter 3 discusses how a proton-proton collision may be captured by a physical detector
 1574 and turned into data that may be stored and analyzed. Chapter 4 discusses the simulation
 1575 of this same process. At this most basic level, however, the ATLAS detector is only a
 1576 machine for turning particles into a set of electrical signals, albeit in a very sophisticated,
 1577 physics motivated way. This chapter discusses the step of turning these electrical signals into
 1578 objects which may be identified with the underlying physics processes, and therefore used to
 1579 make statements about what occurred within a given collision event. This process is termed
 1580 *reconstruction*, and we will focus particularly on jets and flavor tagging, as the most relevant
 1581 pieces for this thesis work.

1582 **5.1 Jets**

1583 As discussed in Chapters 3 and 4, the production of particles with color charge from a
 1584 proton-proton interaction is complicated both by parton showering and by confinement: a
 1585 quark produced from a hard scatter is not seen as a quark, but rather, as a spray of particles
 1586 with a variety of hadrons in the final state, which subsequently shower upon interaction with
 1587 the calorimeter in a complicated way.

1588 For hard scatter electrons, photons, or muons on the other hand, the picture is much
 1589 clearer: there is no parton showering, and each has a distinct signature in the detector:
 1590 photons have no tracks and a very localized calorimeter shower, electrons are associated
 1591 with tracks and are similarly localized in the calorimeter, and muons are associated with
 1592 tracks, pass through the calorimeter due to their large mass, and leave signatures in the muon
 1593 spectrometer.

1594 Jets are a tool to deal with the messiness of quarks and gluons. The basic concept is to
 1595 group the multitude of particles produced by hadronization into a single object. Such an
 1596 object then has associated properties, including a four-vector, which may be identified with
 1597 the corresponding initial state particle. In practice a variety of information from the ATLAS
 1598 detector is used for such a reconstruction. The analysis considered in this thesis uses particle
 1599 flow jets [75], which combines information from both the tracker and the calorimeter, where
 1600 the combined objects may be identified with underlying particles. However, jets built from
 1601 clusters of calorimeter cells [76] as well as from charged particle tracks [77] have also been
 1602 used very effectively.

1603 A variety of algorithms are used to associate detector level objects to a given jet. The
 1604 most commonly used in ATLAS is the anti- k_T algorithm [78], which is a successor to the
 1605 k_T algorithm, among others [79], and develops as follows. Both algorithms are sequential
 1606 recombination algorithms, which begin with the smallest distance, d_{ij} between considered
 1607 objects (e.g. particles or intermediate groupings of particles). If d_{ij} is less than a parameter
 1608 d_{iB} (B for “beam”) object i is combined with object j , the distance d_{ij} is recomputed, and
 1609 the process repeats. This proceeds until $d_{ij} \geq d_{iB}$, at which point the jet is “complete” and
 1610 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

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

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

1611 where p_{Ti} is the transverse momentum of object i , ΔR_{ij} is the angular distance between
 1612 objects i and j , R is a radius parameter, and p controls the tradeoff between the p_T and
 1613 angular distance terms. For the k_T algorithm $p = 1$; for the anti- k_T algorithm, $p = -1$. This
 1614 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

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

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

1615 Therefore, a single high p_T particle (p_{T1}) surrounded by a bunch of low p_T particles results in
 1616 a perfectly conical jet. This shape may change with the presence of other high momentum
 1617 particles, but the key feature of the dynamics is that the jet shape is determined by high p_T
 1618 objects due to the $\frac{1}{p_T}$ nature of this definition. In contrast, the k_T algorithm results in jets
 1619 influenced by low momentum particles, which results in a less regular shape. This property,
 1620 of regular jet shapes determined by high momentum objects, as well as demonstrated good
 1621 practical performance, makes the anti- k_T algorithm the favored jet algorithm in ATLAS.

1622 Because jets are composed of multiple objects, a useful property of jets is jet *substructure*,
 1623 that is, acknowledging that jets are composite objects, analyzing the structure of a given
 1624 jet to infer physics information. This leads to the use of *subjets*; that is, after running jet
 1625 clustering, often to create a “large-R”, $R = 1.0$ anti- k_T jet, a smaller radius jet clustering
 1626 algorithm is run within the jet. Subjets are often chosen using the k_T algorithm, which again
 1627 is sensitive to lower momentum particles, with $R = 0.2$ or 0.3 . For the boosted version of this
 1628 thesis analysis, such a strategy is used, in which the subjets are *variable radius* and depend
 1629 on the momentum of the (proto)jet in question. Beyond this thesis work, substructure is
 1630 used in a large variety of analyses, with a set of associated variables and tools developed for
 1631 exploiting this structure *TODO: Cite some?*.

1632 5.2 Flavor Tagging

1633 For this this thesis, the physics process being considered is $HH \rightarrow b\bar{b}b\bar{b}$. From the previous
 1634 section, we know that the standard practice is to identify these b quarks (or, rather, the
 1635 resulting B hadrons, due to confinement) with jets – in our case, these b -jets are $R=0.4$
 1636 anti- k_T particle flow jets (see Chapter 7). However, not all jets produced at the LHC are
 1637 from B hadrons; rather, there are a variety of different types of jets corresponding to different

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

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

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

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

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

1666 handles still apply – c -hadron lifetimes are between 0.5 and 1 ps, a factor of 2 smaller than B
1667 hadrons. Their mass is around 1.9 GeV, 2 to 3 times smaller than B hadrons, and c -hadrons
1668 contribute to an average of around 55 % of c -jet energy. Therefore, while the gap is slightly
1669 smaller, a distinction may still be made.

1670 The ATLAS flavor tagging framework [81] relies on developing a suite of “low-level”
1671 taggers, which use a variety of information about tracks and vertices as inputs. The output
1672 of these lower level taggers are then fed into a higher level tagger, which aggregates these
1673 results into a high level discriminant. Each of these taggers is described below.

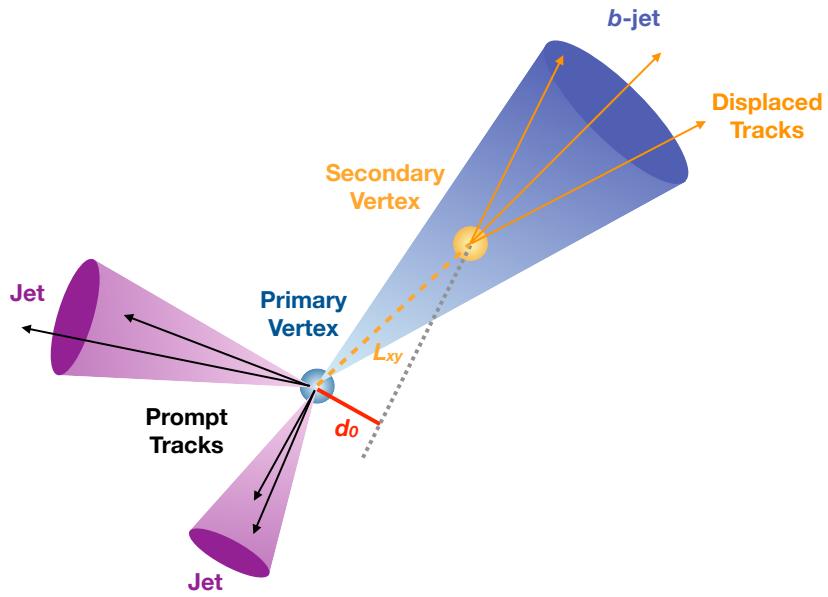


Figure 5.1: Illustration of an interaction producing two light jets and one b -jet in the transverse plane. While the light jets decay “promptly”, coinciding with the primary vertex of the proton-proton interaction, the longer lifetime of B hadrons leads to a secondary decay vertex, displaced from the 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. [80]

1674 5.2.1 IP2D/3D

1675 IP2D and IP3D are taggers based on the large track impact parameter (IP) nature of B
 1676 hadron decays. Both are based on histogram templates derived from Monte Carlo simulation,
 1677 which are used as probability density functions to construct log-likelihood discriminants.
 1678 IP2D incorporates just the transverse impact parameter information using 1D histogram
 1679 templates, whereas IP3D uses both transverse and longitudinal impact parameters in a 2D
 1680 template, which accounts for correlations. Importantly, these are *signed* impact parameters,
 1681 with sign based on the angle between the impact parameter and the considered jet – positive
 1682 impact parameters are consistent with a track extrapolation in front of the jet (angle between
 1683 impact parameter line and jet $< 90^\circ$), and therefore more consistent with tracks originating
 1684 from a displaced decay.

1685 Rather than using the impact parameters directly, an impact parameter *significance*
 1686 is used which incorporates an uncertainty on the impact parameter – tracks with a lower
 1687 uncertainty but the same impact parameter will contribute more in the calculation. This
 1688 signed significance is what is used to sample from the PDF templates, with the resulting
 1689 discriminants the sum of probability ratios between given jet hypotheses over tracks associated
 1690 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
 1691 p_{light} are the per-track probabilities. Similar discriminants are defined between b - and c -jets
 1692 and c and light jets. *TODO: show distributions?*

1693 5.2.2 SV1

1694 SV1 is an algorithm which aims to find a secondary vertex (SV) in a given jet. Operating
 1695 on all vertices associated with a considered jet, the algorithm discards tracks based on a
 1696 variety of cleaning requirements. It then proceeds to first construct all two-track vertices,
 1697 and then iterates over all the tracks involved in these two track vertices to try to fit a single
 1698 secondary vertex, which would then be identified with the secondary vertex from the b or c
 1699 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 [82], tries to reconstruct the full decay chain, including all involved vertices. This is structured around a Kalman filter formalism [83], 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.

1727 *5.2.4 RNNIP*

1728 The IP2D and IP3D algorithms rely on per-track probabilities, and the final discriminating
 1729 variables (and inputs to the higher level taggers) are sums (products) over these independently
 1730 considered quantities. In practice, however, the tracks are not independent – this is merely a
 1731 simplifying assumption to allow for the use of a binned likelihood, as treatment of all of the
 1732 interdependencies in such a framework quickly becomes intractable. To address this issue, a
 1733 recurrent neural network-based algorithm, RNNIP [84], is used, which takes as input a variety
 1734 of per-track variables, including the signed impact parameter significances (as in IP3D) as
 1735 well as track momentum fraction relative to the jet and ΔR between the track and the jet.
 1736 RNNs are sequence-based, and vectors of input variables corresponding to tracks for a given
 1737 jet are ordered by magnitude of transverse impact parameter significance and then passed
 1738 to the network, which outputs class probabilities corresponding to b-jet, c-jet, light-jet, and
 1739 τ -jet hypotheses. Such a procedure allows the network to learn interdependencies between
 1740 tracks, improving performance.

1741 *5.2.5 MV2 and DL1*

1742 Outputs from the above taggers are combined into high level taggers to aggregate all of the
 1743 information and improve discriminating power relative to the respective individual taggers as,
 1744 as shown in Figure 5.2. These high level taggers are primarily in two forms: MV2, which
 1745 uses a Boosted Decision Tree (BDT) for this aggregation, and DL1, which uses a deep neural
 1746 network. For the baseline versions of these taggers, only inputs from IP2D, IP3D, SV1, and
 1747 JetFitter are used. The tagger used for this thesis analysis, DL1r, additionally incorporates
 1748 RNNIP, demonstrating improved performance over the baseline DL1, as shown in Figure 5.3.
 1749 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.

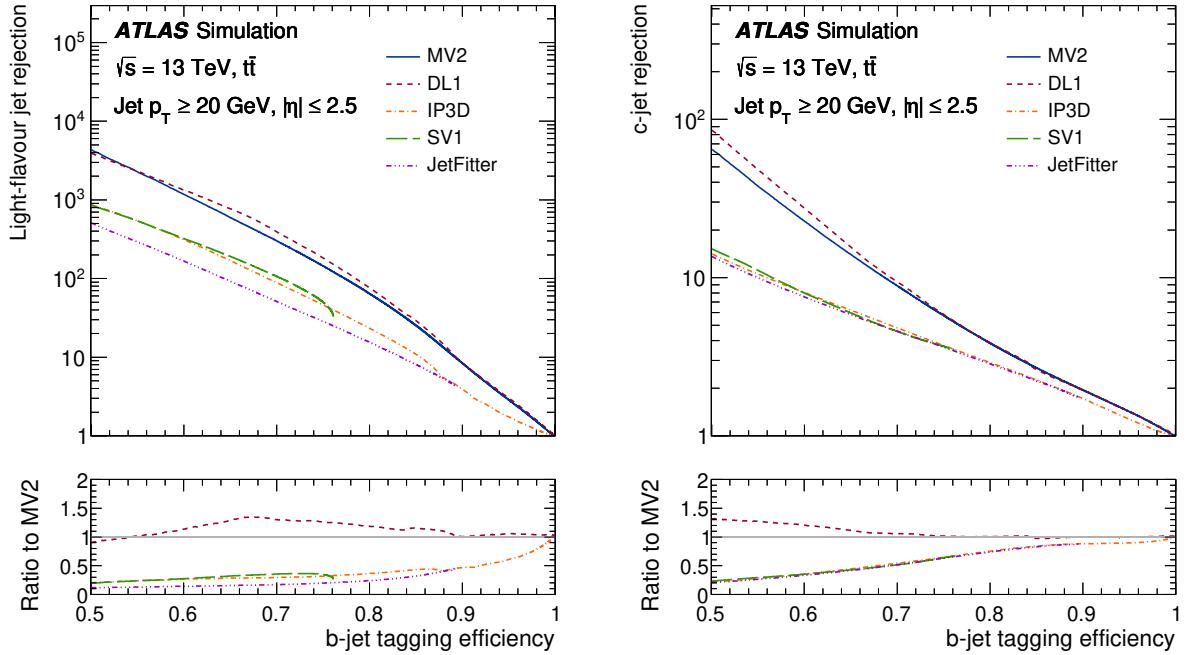


Figure 5.2: Performance of the various low and high level flavor tagging algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. The high level taggers demonstrate significantly better performance than any of the individual low level taggers, 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.

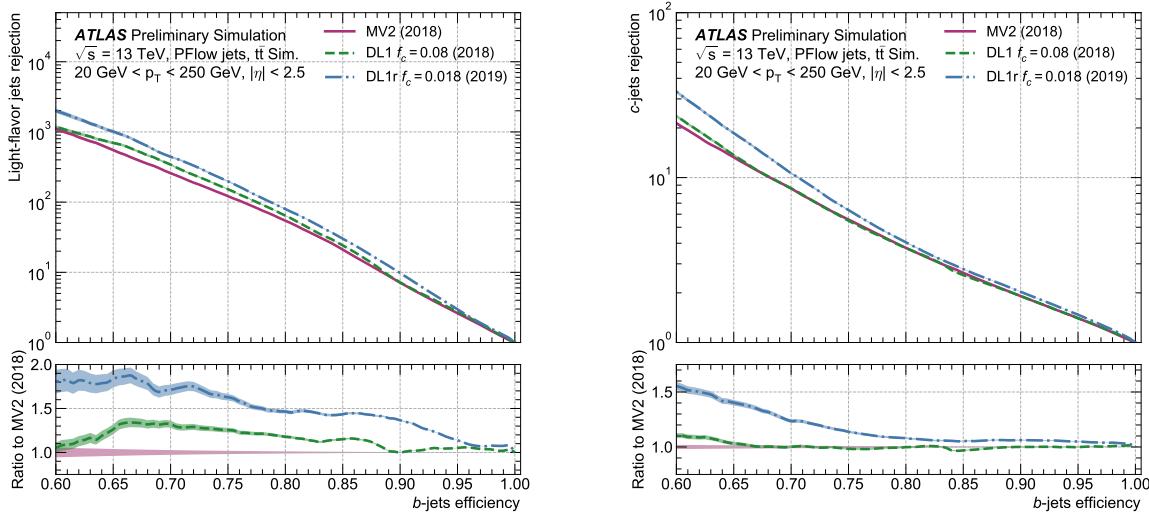


Figure 5.3: Performance of the MV2, DL1, and DL1r algorithms in $t\bar{t}$ simulation, demonstrating the tradeoff between b -jet efficiency and light and c -jet rejection. f_c controls the importance of c -jet rejection in the discriminating variable, and values shown have been optimized separately for each DL1 configuration. DL1r demonstrates a significant improvement in both light and c jet rejection over MV2 and DL1. [85]

1764 5.2.6 *Some Practical Notes*

1765 The b -tagging metrics presented in Figures 5.2 and 5.3 correspond to evaluating a tradeoff
1766 between b -jet efficiency and light jet and c -jet rejection. In this case, b -jet efficiency is defined
1767 such that, e.g. for a 77 % efficiency, 77 % of the real b -jets will be tagged as such. Somewhat
1768 counterintuitively, this means that a lower b -jet efficiency corresponds to a more aggressive
1769 (“tighter”) selection on the discriminating variable, while a higher b -jet efficiency corresponds
1770 to a less aggressive (“looser”) cut (100 % efficiency means no cut). Light and c jet efficiencies
1771 are defined similarly, with rejection defined as 1/ the corresponding efficiency.

1772 In ATLAS, the respective b -tagging efficiencies are used to define various b -tagging working
1773 points. The working point used for the nominal b -jet identification in this thesis is 77 % with
1774 DL1r. A loosened (less aggressive) selection at the 85 % working point is additionally used.
1775 See Chapter 7 for further details.

1776

Chapter 6

1777

SETTING UP THE $HH \rightarrow b\bar{b}b\bar{b}$ ANALYSIS

1778 The following chapters present two complementary searches for pair production of Higgs
 1779 bosons in the final state. Such searches are separated based on the signal models being
 1780 considered: resonant production, in which a new spin-0 or spin-2 particle is produced and
 1781 decays to two Standard Model Higgs bosons, and non-resonant production, which is sensitive
 1782 to the value of the Higgs self-coupling λ_{HHH} . Further information on the theory behind both
 1783 channels can be found in Chapter 2.

1784 While the searches face many similar challenges and proceed (in broad strokes) in a very
 1785 similar manner, separate optimizations are performed to maximize the respective sensitivities
 1786 for these two very different sets of signal hypotheses. More particularly, resonant signal
 1787 hypotheses are (1) very peaked in values of the mass of the HH candidate system near
 1788 the value of the resonance mass considered and (2) considered across a very broad range of
 1789 signal mass hypotheses. The resonant searches are therefore split into resolved and boosted
 1790 topologies based on Lorentz boost of the decay products, with the resolved channel as one of
 1791 the primary focuses of this thesis. Further, several analysis design decisions are made to allow
 1792 for sensitivity to a broad range of masses – in particular, though sensitivity is limited at lower
 1793 values of m_{HH} relative to other channels *TODO: Combination, bbyy* due to the challenging
 1794 background topology, retaining and properly reconstructing these low mass events allows the
 1795 $b\bar{b}b\bar{b}$ channel to retain sensitivity as low as the kinematic threshold at 250 GeV.

1796 In contrast, non-resonant signal hypotheses are quite broad in m_{HH} , and have a much
 1797 more limited mass range, with Standard Model production peaking near 400 GeV, and the
 1798 majority of the analysis sensitivity able to be captured with a resolved topology. Even for
 1799 Beyond the Standard Model signal hypotheses, which may have more events at low m_{HH} ,

1800 the non-resonant nature of the production allows the $b\bar{b}b\bar{b}$ channel to retain sensitivity while
 1801 discarding much of the challenging low mass background. Such freedom allows for decisions
 1802 which focus on improved background modeling for the middle to upper HH mass regime,
 1803 resulting in improved modeling and smaller uncertainties than would be obtained with a
 1804 more generic approach.

1805 Both searches are presented in the following, with emphasis on particular motivations for,
 1806 and consequences of, the various design decisions involved for each respective set of signal
 1807 hypotheses.

1808 The analyses improve upon previous work [2] in several notable ways. The resonant
 1809 search leverages a Boosted Decision Tree (BDT) based algorithm for the reconstruction of
 1810 the HH system from the jets considered for the analysis, offering an improved efficiency
 1811 of that reconstruction over a broad mass spectrum. The non-resonant adopts a different
 1812 approach, with a simplified algorithm based on the minimum angular distance (ΔR) between
 1813 jets in a reconstructed Higgs candidate. Such an approach very efficiently discards low mass
 1814 background events, resulting in an easier to estimate background with reduced systematic
 1815 uncertainties.

1816 A particular contribution of this thesis is the background estimation, which uses a novel,
 1817 neural network based approach to perform a data-driven estimation of the background, which
 1818 is dominated by QCD processes, for which a sufficient simulation is not available. This new
 1819 approach offers improved modeling over previous methods, as well as the ability to model
 1820 correlations between observables. While all aspects of the analysis of course contribute to
 1821 the final result, the author of this thesis wishes to emphasize that the background estimate,
 1822 with the corresponding uncertainties and all other associated decisions, really is the core of
 1823 the $HH \rightarrow b\bar{b}b\bar{b}$ analysis – the development of this procedure, and all associated decisions, is
 1824 similarly the core of this thesis work.

1825 ATLAS has performed a variety of searches in complementary decay channels as well,
 1826 notably for early Run 2 in the $b\bar{b}W^+W^-$ [86], $b\bar{b}\tau^+\tau^-$ [87], $W^+W^-W^+W^-$ [88], $b\bar{b}\gamma\gamma$ [89],
 1827 and $W^+W^-\gamma\gamma$ [90] final states, which were combined along with $b\bar{b}b\bar{b}$ in [26]. ATLAS has

1828 also released a variety of full Run 2 results, including boosted $b\bar{b}\tau^+\tau^-$ [91], VBF $b\bar{b}b\bar{b}$ [22],
1829 $b\bar{b}\ell\nu\ell\nu$ [92], and $b\bar{b}\gamma\gamma$ [93].

1830 CMS has also performed searches for resonant production of Higgs boson pairs in the
1831 $b\bar{b}b\bar{b}$ final state (among others) at $\sqrt{s} = 8$ TeV [94] and $\sqrt{s} = 13$ TeV [95]. CMS have also
1832 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
1833 in [96].

1834 This analysis also benefits from improvements to ATLAS jet reconstruction and calibration,
1835 and flavor tagging [81]. In particular, this analysis benefits from the introduction of particle
1836 flow jets [75]. These make use of tracking information to supplement calorimeter energy
1837 deposits, improving the angular and transverse momentum resolution of jets by better
1838 measuring these quantities for charged particles in those jets.

1839 The analysis also benefits from the new DL1r ATLAS flavor tagging algorithm. Whereas
1840 the flavor tagging algorithm used in the previous analysis (MV2) used a boosted decision tree
1841 (BDT) to combine the output of various low level algorithms, DL1r (and the baseline DL1
1842 algorithm) uses a deep neural network to do this combination. In addition to the low level
1843 algorithms used as inputs to MV2, DL1 includes a variety of additional variables used for
1844 c -tagging. DL1r further incorporates RNNIP, a recurrent neural network designed to identify
1845 b -jets using the impact parameters, kinematics, and quality information of the tracks in the
1846 jets, while also taking into account the correlations between the track features.

1847 The overall analysis sensitivity further benefits from a factor of ~ 4.6 increase in integrated
1848 luminosity.

1849 6.1 Data and Monte Carlo Simulation

1850 Both the resonant and non-resonant searches are performed on the full ATLAS Run 2 dataset,
1851 consisting of $\sqrt{s} = 13$ TeV proton-proton collision data taken from 2016 to 2018 inclusive.
1852 Data taken in 2015 is not used due to a lack of trigger jet matching information and b -jet

¹⁸⁵³ trigger scale factors¹. The integrated luminosity collected and usable in this analysis² was:

¹⁸⁵⁴ • 24.6 fb^{-1} in 2016

¹⁸⁵⁵ • 43.65 fb^{-1} in 2017

¹⁸⁵⁶ • 57.7 fb^{-1} in 2018

¹⁸⁵⁷ This gives a total integrated luminosity of 126 fb^{-1} . This is lower than the 139 fb^{-1} ATLAS
¹⁸⁵⁸ collected during Run 2 [98] due to the inefficiency described in footnote 2 as well as the
¹⁸⁵⁹ 3.2 fb^{-1} of 2015 data which is unused due to the trigger scale factor issue mentioned above.

¹⁸⁶⁰ In this analysis, Monte Carlo samples are used purely for modelling signal processes. The
¹⁸⁶¹ background is strongly dominated by events produced by QCD multijet processes, which are
¹⁸⁶² difficult to correctly model in simulation due to the complexity of the interactions involved
¹⁸⁶³ (including, e.g. non-perturbative effects), as well as the harsh requirement of four *b*-tagged
¹⁸⁶⁴ jets, which makes it difficult to collect sufficient Monte Carlo statistics. This necessitates the
¹⁸⁶⁵ use of a data-driven background modeling technique, which is described in Chapter 8.

¹⁸⁶⁶ The scalar resonance signal model is simulated at leading order in α_s using MADGRAPH
¹⁸⁶⁷ [57]. Hadronization and parton showering are done using HERWIG 7 [58][59] with EVTGEN [61],
¹⁸⁶⁸ and the nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a two Higgs
¹⁸⁶⁹ doublet model where the new neutral scalar is produced through gluon fusion and required
¹⁸⁷⁰ to decay to a pair of SM Higgs bosons. The heavy scalar is assigned a width much smaller
¹⁸⁷¹ than detector resolution, and the other 2HDM particles do not enter the calculation.

¹⁸⁷² Scalar samples are produced at resonance masses between 251 and 900 GeV and the
¹⁸⁷³ detector simulation is done using AtlFast-II [66]. In addition the samples at 400 GeV and
¹⁸⁷⁴ 900 GeV are also fully simulated to verify that the use of AtlFast-II is acceptable. For higher

¹These trigger scale factors account for differences in the performance of the *b*-tagging algorithms between simulation and data, with the jet matching providing a correspondence between the jets in the trigger decision and the jets in the offline analysis

²approximately 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 [97]

masses, as well as for the boosted analysis, samples are produced between 1000 and 5000 GeV, and the detector is fully simulated. As discussed in Chapter 4, an outstanding issue with AtlFast-II is the modeling of jet substructure. While such variables are not used for the resolved analysis, the boosted analysis begins at 900 GeV, motivating the different detector simulation in these two regimes.

The spin-2 resonance signal model is also simulated at LO in α_s using MADGRAPH. Hadronization and parton showering are done using PYTHIA 8 [60] with EVTGEN, and the nominal PDF is NNPDF 2.3 LO. In practice this is implemented as a Randall-Sundrum graviton with $c = 1.0$.

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

For the non-resonant search, samples are produced at values of $\kappa_\lambda = 1.0$ and 10.0, and are simulated using POWHEG Box v2 generator [54–56] at next-to-leading order (NLO), with full NLO corrections with finite top mass, using the PDF4LHC [99] parton distribution function (PDF) set. Parton showers and hadronization are simulated with PYTHIA 8.

Alternative ggF samples are simulated at NLO using POWHEG Box v2, but instead using HERWIG 7 [100] for parton showering and hadronization. The comparison between these two is used to assess an uncertainty on the parton showering.

6.2 Triggers and Object Definitions

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

- 2b + HT, which requires two b -tagged jets and a minimum value of H_T , defined to be the scalar sum of p_T across all jets in the event.
- 2b + 2j, which requires two b -tagged jets and two other jets matching some kinematic

1901 requirements

- 1902 • $2b + 1j$, which requires two b -tagged jets and one other jet matching some kinematic requirements

- 1904 • $1b$, which requires one b -tagged jet

1905 Due to minimal contributions from some of these triggers for the Standard Model non-resonant signal, a simplified strategy relying entirely on $2b + 1j$ and $2b + 2j$ triggers is used for the 1907 non-resonant search.

1908 While the use of multiple triggers is beneficial for analysis sensitivity, it comes with some 1909 complications. Namely, a set of scale factors must be assigned to simulated events to account 1910 for differences in trigger efficiency between real and simulated events. Because these scale 1911 factors may differ between triggers, the use of multiple triggers becomes complicated: an event 1912 may pass more than one trigger, while trigger scale factors are only provided for individual 1913 triggers.

1914 To simplify this calculation, a set of hierarchical offline selections is applied, closely 1915 mimicking the trigger selection. Based on these selections, events are sorted into categories 1916 (*trigger buckets*), after which the decision of a *single trigger* is checked.

1917 The resonant search applies such categorization in the following way, with selections 1918 considered in order:

- 1919 1. If the leading jet is b -tagged with $p_T > 325 \text{ GeV}$, the event is in the $1b$ trigger category.
- 1920 2. Otherwise, if the leading jet is not b -tagged, but has $p_T > 168.75 \text{ GeV}$, the event is in 1921 the $2b + 1j$ trigger category.
- 1922 3. If neither of the first two selections pass, if the scalar sum of jet p_T s, $H_T > 900 \text{ GeV}$, 1923 the event falls into the $2b + HT$ trigger category.

1924 4. Events that do not pass any of the above offline selections are in the $2b + 2j$ trigger
1925 category.

1926 Corresponding triggers are then checked in each category, and the final set of events consists
1927 of those events that pass the trigger decision in their respective categories.

1928 For the resonant search, the $2b + 1j$ and $2b + 2j$ triggers are the dominant categories,
1929 containing roughly 26 % and 49 % of spin-2 events, evaluated on MC16d samples with
1930 resonance masses between 300 and 1200 GeV. Notably, the $1b$ trigger efficiency is largest at
1931 high (> 1 TeV) resonance masses.

1932 For the non-resonant search, it was noted that the $1b$ trigger has minimal contribution,
1933 while the $2b + HT$ events are largely captured by the $2b + 2j$ trigger. Therefore, a simplified
1934 scheme is considered, with selections:

- 1935 1. If the 1st leading jet has $p_T > 170$ GeV and the 3rd leading jet has $p_T > 70$ GeV, the
1936 event is in the $2b + 1j$ trigger category.
- 1937 2. Otherwise, the event is in the $2b + 2j$ trigger category.

1938

Chapter 7

1939

ANALYSIS SELECTION

1940 7.1 Analysis Selection

1941 After the trigger selections of Section 6.2, a variety of selections on the analysis objects are
 1942 made, with the goal of (1) reconstructing a HH -like topology and (2) suppressing contributions
 1943 from background processes.

1944 Both analyses begin with a common pre-selection, requiring at least four $R = 0.4$ anti- k_T
 1945 jets with $|\eta| < 2.5$ and $p_T > 40 \text{ GeV}$. The $|\eta| < 2.5$ requirement is necessary for b -tagging
 1946 due to the coverage of the ATLAS tracking detector (see Chapter 3), while the $p_T > 40 \text{ GeV}$
 1947 requirement is motivated by the trigger thresholds. A low p_T category, which would include
 1948 events failing the analysis selection due to this p_T cut, was considered for the non-resonant
 1949 search, but was found to contribute minimal sensitivity. At least two of the jets passing this
 1950 pre-selection are required to be b -tagged, and additional b -tagging requirements are made to
 1951 define the following regions:

- 1952 • “2 b Region”: require exactly two b -tagged jets, used for background estimation
- 1953 • “4 b Region”: require at least (but possibly more) four b -tagged jets, used as a signal
 1954 region for both resonant and non-resonant searches

1955 The non-resonant analysis additionally defines two 3 b regions:

- 1956 • “3 b +1 loose Region”: require exactly three b -tagged jets which pass the 77 % b-tagging
 1957 working point (nominal) and one additional jet that fails the 77 % b-tagging working
 1958 point but passes the *looser* 85 % b-tagging working point. Used as a signal region for
 1959 the non-resonant search.

- 1960 • “3 b +1 fail Region”: complement of 3 b +1 loose. Require exactly three b -tagged jets
 1961 which pass the 77 % b-tagging working point, but require that none of the remaining jets
 1962 pass the 85 % b-tagging working point. Used as a validation region for the non-resonant
 1963 search.

1964 After these requirements, four jets are chosen, ranked first by b -tagging requirement and then
 1965 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
 1966 non-tagged jets; for the 4 b region, the jets are the four highest p_T b -tagged jets). To match
 1967 the topology of a $HH \rightarrow b\bar{b}b\bar{b}$ event, these four jets are then *paired* into *Higgs candidates*: the
 1968 four jets are split into two sets of two, and each of these pairs is used to define a reconstructed
 1969 object that is a proxy for a Higgs in a HH event.

1970 For four jets there are three possible pairings. For signal events, a correct pairing can be
 1971 identified (provided all necessary jets pass pre-selection) using the truth information of the
 1972 Monte Carlo simulation, and such information may be used to design/select an appropriate
 1973 pairing algorithm. This is only part of the story, however. The vast majority of the events in
 1974 data do *not* include a real HH decay (this is a search for a reason!), either because the event
 1975 originates from a background process (e.g. for 4 b events), or because the selection is not
 1976 designed to maximize the signal (e.g. 2 b events). As the pairing is part of the selection, it must
 1977 still be run on such events, such that various algorithms which achieve similar performance
 1978 in terms of pairing efficiency may have vastly different impacts in terms of the shape of the
 1979 background and the biases inherent in the background estimation procedure. The interplay
 1980 between these two facets of the pairing is an important part of the choices made for this
 1981 analysis.

A comparison of different shapes due to three different paring strategies is shown in Figure 7.1. The Boosted Decision Tree (BDT) pairing and min ΔR pairing are used for the analyses presented here, and are described in more detail below. The D_{HH} pairing was used for the

early Run 2 searches [2], and is based on minimizing the quantity

$$D_{HH} = \frac{|m_{H1} - \frac{120}{110}m_{H2}|}{\sqrt{1 + \left(\frac{120}{110}\right)^2}}, \quad (7.1)$$

corresponding to the the distance of the reconstructed Higgs candidate masses from a line running from $(0, 0)$ to the center of the signal region, $(120 \text{ GeV}, 110 \text{ GeV})$ in leading and subleading Higgs candidate masses, (m_{H1}, m_{H2}) . Note that while this achieves good pairing efficiency with respect to truth across a broad HH mass range, it significantly sculpts the mass plane (as seen in Figure 7.1), motivating the new approaches considered here.

7.1.1 Resonant Pairing Strategy

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

The boosted decision tree is also parameterized on the di-Higgs mass (m_{HH}) by providing this as an additional feature. Since the boosted decision tree is trained on correct and incorrect pairings in signal events, there will be exactly one correct pairing and two incorrect pairings in the training set for each m_{HH} value present in that set. As a result, this variable cannot, in itself, distinguish a correct pairing from an incorrect pairing, and therefore the inclusion of this variable simply serves to parameterize the BDT on m_{HH} ¹.

The boosted decision tree was trained on one quarter of the total AFII simulated scalar MC statistics, using the Gradient-based One Side Sampling (GOSS) algorithm which allows

¹That is, the conditions placed on the other variables by the BDT vary with m_{HH} .

2004 rapid training with very large datasets. A preselection was applied requiring events to have
 2005 four jets with a p_T of at least 35 GeV. Note that this is a looser requirement than the 40 GeV
 2006 used in the analysis selection, and is meant to increase the available statistics for events with
 2007 low m_{HH} and to ensure a better performance as a function of that variable. Events were also
 2008 required to have four distinct jets that could be geometrically matched (to within $\Delta R \leq 0.4$)
 2009 to the b -quarks. The events used to train the BDT were not included when the analysis was
 2010 run on these signal simulations. The boosted decision tree was constructed with the following
 2011 hyperparameters:

```

2012 min_data_in_leaf=50,
2013 num_leaves=180,
2014 learning_rate=0.01
  
```

2015 These hyperparameters were optimized using a Bayesian optimization procedure [101].
 2016 Three fold cross-validation was used to perform this optimization without the need for an
 2017 additional sample, while avoiding over-training on signal events.

2018 7.1.2 Non-resonant Pairing Strategy

2019 For the non-resonant analysis, a simpler pairing algorithm is used, which proceeds as follows:
 2020 in a given event, Higgs candidates for each possible pairing are sorted by the p_T of the vector
 2021 sum of constituent jets. The angular separation, ΔR is computed between jets in the each of
 2022 the leading Higgs candidates, and the pairing with the smallest separation (ΔR_{jj}) is selected.
 2023 This method will be referred to as $\min \Delta R$ in the following.

2024 The primary motivation for the use of this pairing in the non-resonant search is a *smooth*
 2025 *mass plane*: by efficiently discarding low mass events, $\min \Delta R$ removes the background peak
 2026 present in the resonant search while maintaining good pairing efficiency for the Standard
 2027 Model non-resonant signal. This facilitates a background estimate with small kinematic bias
 2028 – the region in which the background estimate is derived is more similar to the signal region.

2029 Along with discarding low mass background, there is a corresponding loss of low mass
 2030 signal. This predominantly impacts points away from the Standard Model (see Figure 7.2),

2031 but, because the $4b$ channel has the strongest contribution near the Standard Model and
 2032 because of the large low mass background present with other pairing methods, the impact on
 2033 analysis sensitivity is mitigated. The min ΔR pairing is thus adopted for the non-resonant
 2034 search.

2035 *7.1.3 Pairing Efficiencies*

2036 Though this is implicit in the above descriptions, an explicit examination of the pairing
 2037 efficiencies with respect to truth for the respective signal samples has been performed for both
 2038 min ΔR and the BDT pairing. Conceptually, for high invariant mass of the HH system, each
 2039 Higgs has a high p_T and the the b -jets corresponding to a given Higgs are more collimated.
 2040 In this case, angular information such as that exploited both directly by min ΔR and as
 2041 inputs in the BDT pairing may be expected to be a good discriminant for determining the
 2042 HH system. Indeed for resonance masses above 500 GeV, the pairing efficiency for both
 2043 algorithms is close to 100 %.

2044 For lower HH masses, the jets corresponding to a given Higgs are no longer as collimated,
 2045 such that min ΔR is no longer guaranteed to pick up the correct pairing (e.g. in a case when
 2046 the four jets involved are isotropic), and the pairing efficiency steadily gets worse as the HH
 2047 mass decreases. On resonant samples, e.g., the min ΔR efficiency drops below 80 % near
 2048 400 GeV. The additional information exploited by the BDT mitigates this somewhat, though
 2049 there is still a drop in efficiency at lower m_{HH} . Interestingly, the BDT pairing demonstrates
 2050 a rise in pairing efficiency near the threshold of 250 GeV, likely due to the limited kinematic
 2051 phase space for the HH system in this region.

2052 The examination of the pairing efficiency as a function of m_{HH} has a more direct cor-
 2053 respondence for resonant samples, but it of course applies to non-resonant samples as well,
 2054 resulting in the behavior shown in Figure 7.2. Note that the above statement that min ΔR
 2055 discards low mass events is a consequence of the reduced pairing efficiency at low mass – the
 2056 pairing algorithm itself does not make any cuts, but the mis-reconstruction of low mass signal
 2057 results in the reconstruction of Higgs candidates with masses away from 125 GeV, placing

2058 such events outside of the kinematic signal regions defined in Section 7.3.

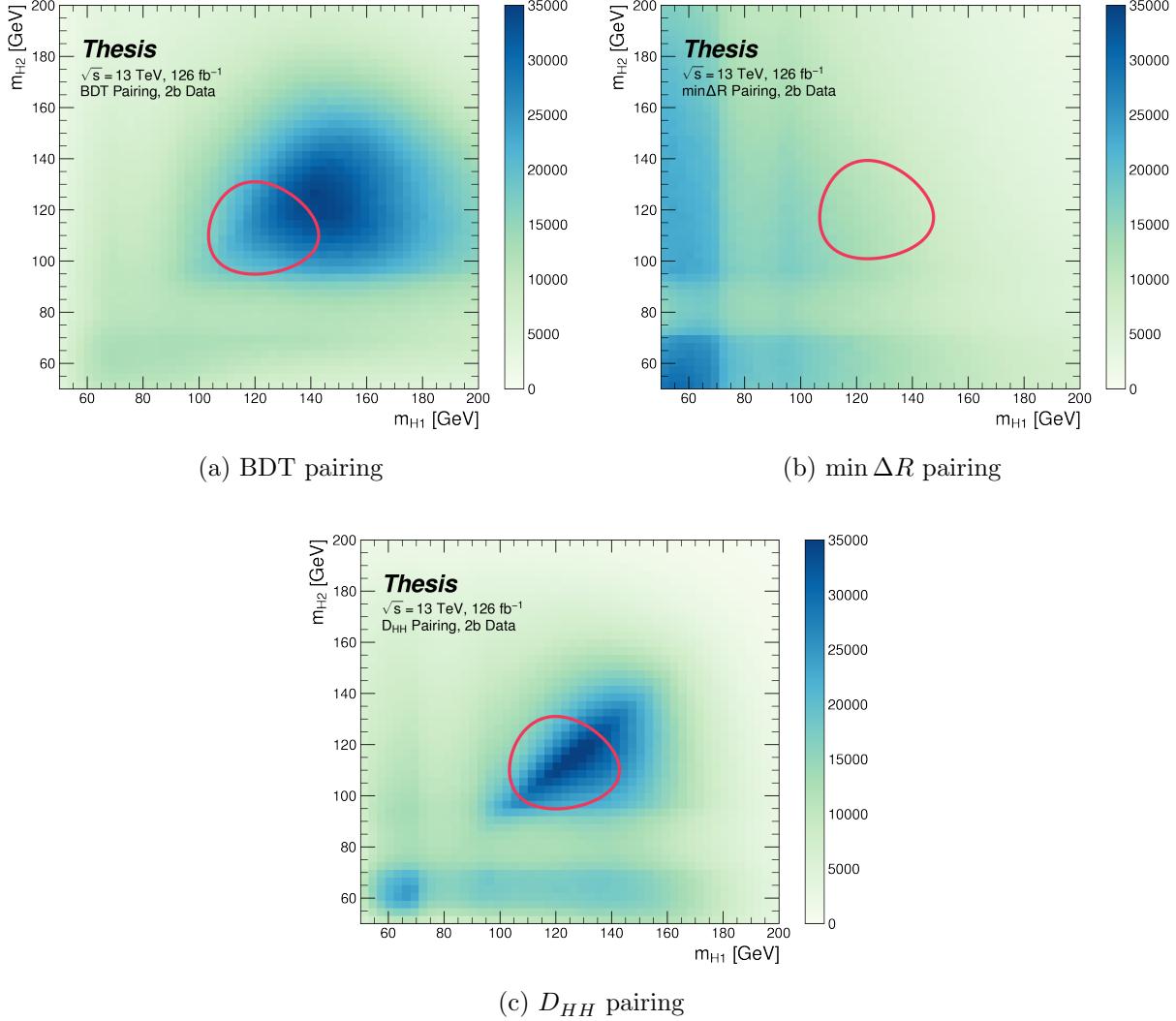


Figure 7.1: Comparison of m_{H1} vs m_{H2} planes for the full Run 2 2b dataset with different pairings. As evidenced, this choice significantly impacts where events fall in this plane, and therefore which events fall into the various kinematic regions defined in this plane (see Section 7.3). The signal regions for the resonant/early Run 2 analysis are shown for reference for the BDT and D_{HH} pairings, while the the min ΔR signal region shifted is 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 7.2.

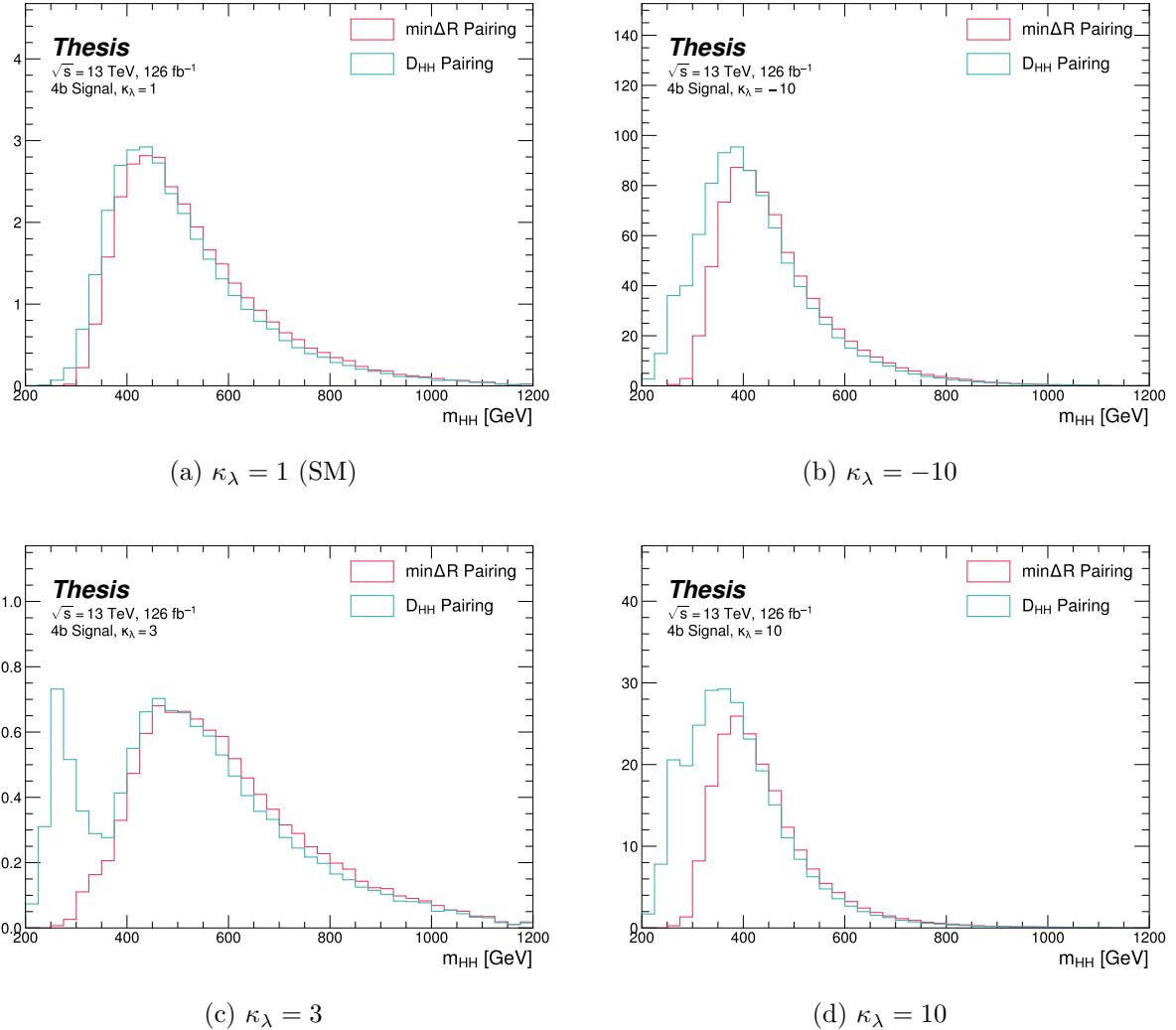


Figure 7.2: Comparison of signal distributions in the respective signal regions for the min ΔR and D_{HH} pairing for various values of the Higgs trilinear coupling. The distributions are quite similar at the Standard Model point, but for other variations, min ΔR does not pick up the low mass features.

2059 **7.2 Background Reduction and Top Veto**

2060 Choosing a pairing of the four b-tagged jets fully defines the di-Higgs candidate system used
2061 for each event in the remainder of the analysis chain. A requirement of $|\Delta\eta_{HH}| < 1.5$ on this
2062 di-Higgs candidate system mitigates QCD multijet background.

2063 In order to mitigate the hadronic $t\bar{t}$ background, a top veto is then applied, removing
2064 events consistent with a $t \rightarrow b(W \rightarrow q_1\bar{q}_2)$ decay.

2065 The jets in the event are separated into *HC jets* which are the four jets used to build the
2066 Higgs candidates, and *non-*HC jets**, the other jets (passing the p_T and $|\eta|$ requirements) in
2067 the event.

2068 W candidates are built by forming all possible pairs of all jets in each event. With n jets,
2069 there are $\binom{n}{2}$ such pairs. t candidates are then built by pairing each W candidate with each
2070 HC jet (for $4\binom{n}{2}$ combinations). Note that all jets in a t candidate must be distinct (i.e. a
2071 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

$$X_{Wt} = \sqrt{\left(\frac{m_W - 80.4 \text{ GeV}}{0.1 \cdot m_W}\right)^2 + \left(\frac{m_t - 172.5 \text{ GeV}}{0.1 \cdot m_t}\right)^2} \quad (7.2)$$

2072 is constructed for each combination.

2073 Events are then vetoed if the minimum X_{Wt} over all combinations is less than 1.5.

2074 The same definitions and procedures are used for both the resonant and non-resonant
2075 analyses. However, for the non-resonant search, the top candidates considered for X_{Wt} have
2076 the additional requirement that the jet used for the b is b -tagged. While this is identical to
2077 the resonant analysis by definition for $4b$ events, it does change the set of events considered in
2078 lower tag regions, in particular for the $2b$ events considered in the derivation of the background
2079 estimate. Such a change is found to reduce the impact of background systematics, an effect
2080 that is thought to be due to the shifting of $2b$ events to higher values of X_{Wt} (due to this
2081 more stringent requirement), where, e.g, the Standard Model signal peaks.

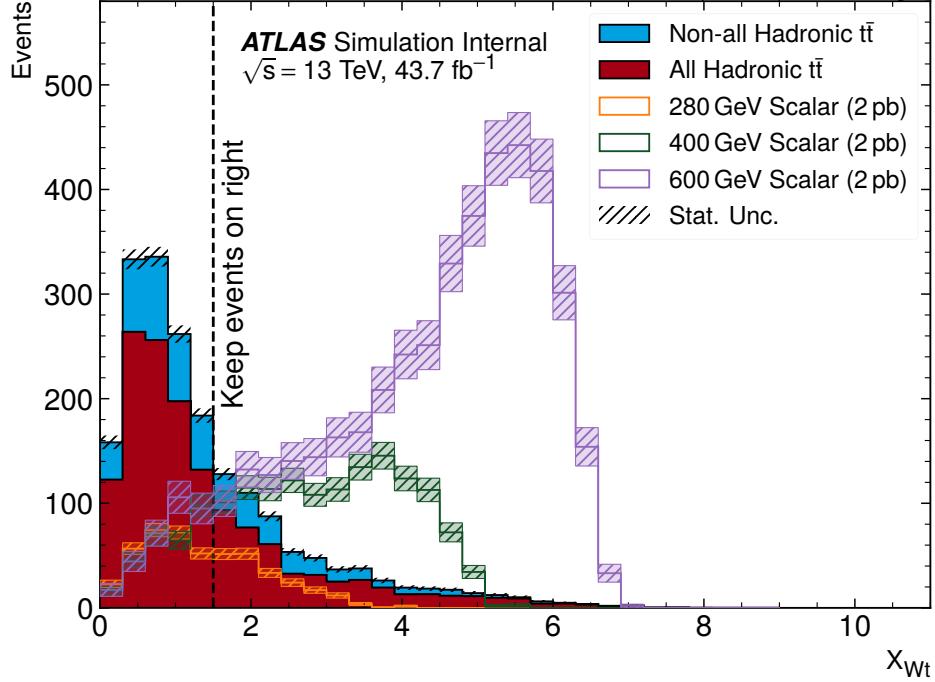


Figure 7.3: **Resonant search:** Illustration of the impact of the top veto on $t\bar{t}$ Monte Carlo for the resonant analysis, with representative scalar signals shown for reference. The cut value used is 1.5, shown in the dashed black, and events below this value are discarded. This top veto clearly removes the bulk of $t\bar{t}$ events, and the value of the cut is chosen to retain analysis sensitivity, particularly for low mass.

2082 The distribution of this variable is shown for $t\bar{t}$ Monte Carlo and representative signal
 2083 samples for the resonant and non-resonant 4 b signal regions in Figures 7.3 and 7.4 respectively,
 2084 with a line at the cut value of 1.5. Individual years are shown, but results are representative
 2085 across years. For the resonant analysis, the value of the cut is constrained by low mass
 2086 resonances, with the value of 1.5 chosen as a compromise between $t\bar{t}$ rejection and retaining
 2087 sensitivity for these signals. For the non-resonant, though e.g., the SM signal peaks at higher
 2088 values, a more aggressive cut on X_{Wt} was found to decrease analysis sensitivity, so the value
 2089 of 1.5 is kept.

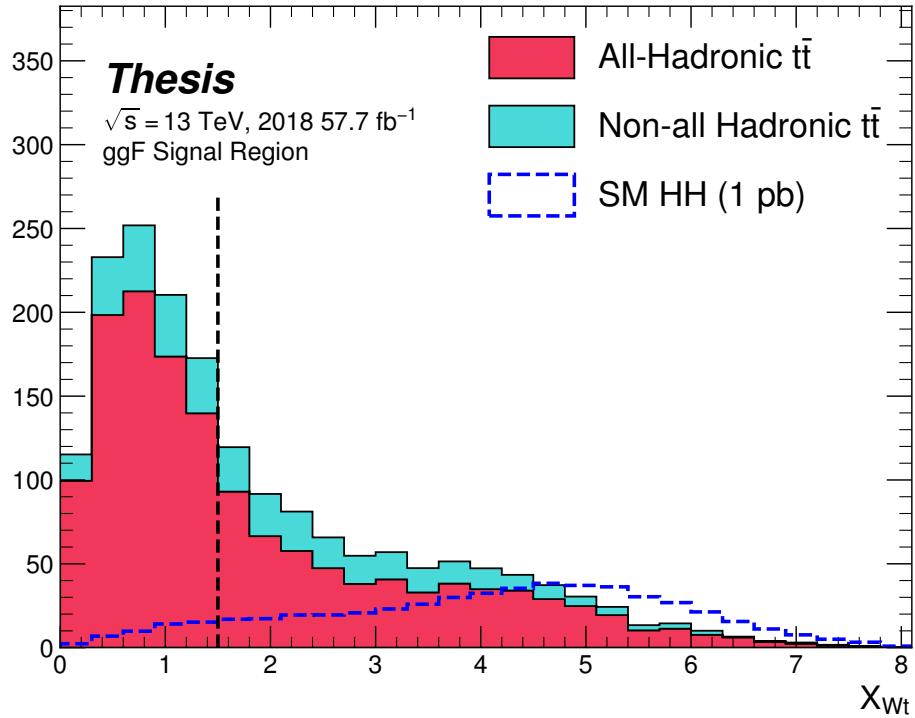


Figure 7.4: **Non-resonant search:** Illustration of the impact of the top veto on $t\bar{t}$ Monte Carlo for the non-resonant analysis, with the Standard Model signal shown for reference. The cut value used is 1.5, shown in the dashed black, and events below this value are discarded. This top veto clearly removes the bulk of $t\bar{t}$ events. While this plot may seem to motivate a more aggressive cut on X_{Wt} , increasing the value of the cut was found to reduce analysis sensitivity.

2090 **7.3 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:

$$X_{HH} = \sqrt{\left(\frac{m(H_1) - c_1}{0.1 \cdot m(H_1)}\right)^2 + \left(\frac{m(H_2) - c_2}{0.1 \cdot m(H_2)}\right)^2} \quad (7.3)$$

2091 with $m(H_1)$, $m(H_2)$ the leading and subleading Higgs candidate masses, c_1 and c_2 correspond
2092 to the center of the signal region, and the denominator provides a Higgs candidate mass
2093 dependent resolution of 10 %. For consistency with the HH decay hypothesis, c_1 and c_2
2094 are nominally (125 GeV, 125 GeV). However, these are allowed to vary due to energy loss,
2095 with specific values chosen described below. The selection of these values is one of several
2096 significant differences between the regions defined for the resonant and non-resonant search.
2097 We describe both below.

2098 **7.3.1 Resonant Kinematic Regions**

2099 For the resonant analysis, the signal region is centered at (120 GeV, 110 GeV) to account for
2100 energy loss leading to the Higgs masses being under-reconstructed. Note that leading and
2101 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

$$\sqrt{(m(H_1) - 1.03 \times 120 \text{ GeV})^2 + (m(H_2) - 1.03 \times 110 \text{ GeV})^2} < 30 \text{ GeV} \quad (7.4)$$

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

$$\sqrt{(m(H_1) - 1.05 \times 120 \text{ GeV})^2 + (m(H_2) - 1.05 \times 110 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.5)$$

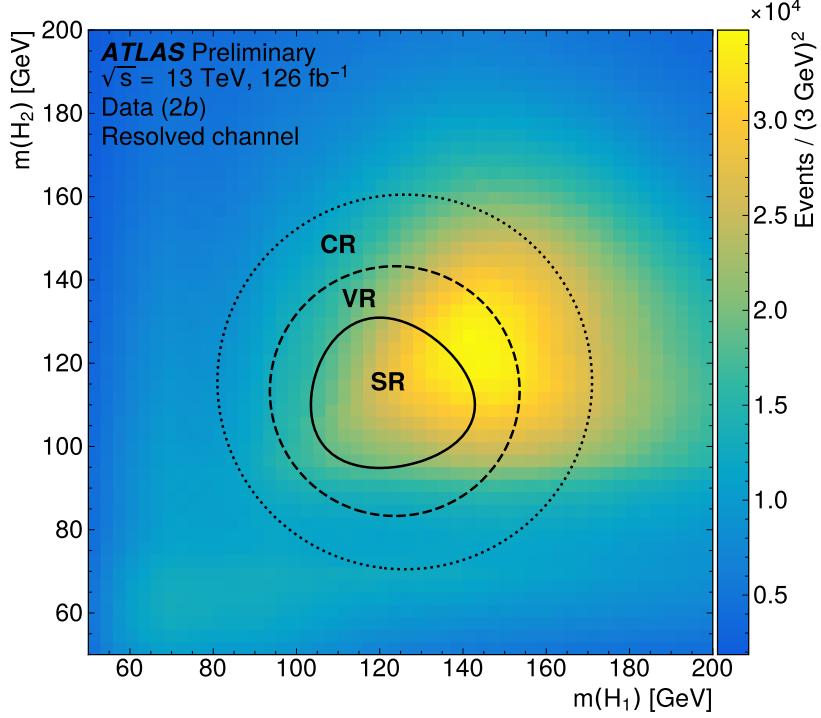


Figure 7.5: Regions used for the resonant search, shown on the $2b$ data mass plane. The outermost region (the “control region”) is used for derivation of the nominal background estimate. The innermost region is the signal region, where the signal extraction fit is performed. The region in between (the “validation region”) is used for the assessment of an uncertainty.

2102 For simplicity, the SR/VR/CR definitions from the early Run 2 paper [2] were chosen for
2103 the resonant analysis, and were found to be close to optimal. These regions are shown in
2104 Figure 7.5.

2105 7.3.2 Non-resonant Kinematic Regions

2106 For the non-resonant analysis the signal region is centered at (124 GeV, 117 GeV), corre-
2107 sponding to the means of *correctly paired* Standard Model signal events. The shape of the
2108 signal region (other than this change of center) was found to remain optimal.

2109 For the non-resonant search, leading and subleading Higgs candidates are defined according
 2110 to the *vector sum* of constituent jet p_T , more closely corresponding to the $1 \rightarrow 2$ decay
 2111 assumption behind the min ΔR pairing algorithm.

2112 Two areas for improvement were identified in the resonant analysis, which will be discussed
 2113 in more detail below: *signal contamination* of the validation region (which impacts the
 2114 uncertainty assessed due to extrapolation) and *large nuisance parameter pulls* for this
 2115 uncertainty, corresponding to a rough assumption that the validation region is closer to the
 2116 signal region in the mass plane, and so offers a better estimate of the signal region. Extensive
 2117 cross-checks were performed for the resonant search, which demonstrated minimal bias due
 2118 to the signal contamination and healthy behavior of the signal extraction fit, despite the
 2119 large pulls. However, these large pulls imply that the nominal estimate may be improved by
 2120 incorporating some of the information entering the definition of the extrapolation uncertainty.
 2121 Further, the resonant search benefits from a set of highly peaked signals, such that the
 2122 smooth nature of the background helps to mitigate signal contamination bias. With the
 2123 broad non-resonant signals, a bias due to signal contamination becomes more of a concern,
 2124 such that addressing this is highly motivated.

A redesign of the control and validation regions is therefore performed for the non-resonant analysis. The outer boundary defined by the shifted resonant control region:

$$\sqrt{(m(H_1) - 1.05 \times 124 \text{ GeV})^2 + (m(H_2) - 1.05 \times 117 \text{ GeV})^2} < 45 \text{ GeV} \quad (7.6)$$

2125 is kept, roughly corresponding to combining the regions used for the resonant analysis. In
 2126 order to assess the variation of the background estimate, two sets of regions are desired, so
 2127 this combined region is split into *quadrants*, that is, divided into four pieces along axes that
 2128 intersect with the signal region center. To avoid kinematic bias, quadrants on opposite sides
 2129 of the signal region are paired, with these pairs corresponding to the non-resonant control
 2130 and validation regions.

2131 The particular orientation of the regions is chosen such that region centers align with the
 2132 leading and subleading Higgs candidate masses, corresponding to a set of axes rotated at

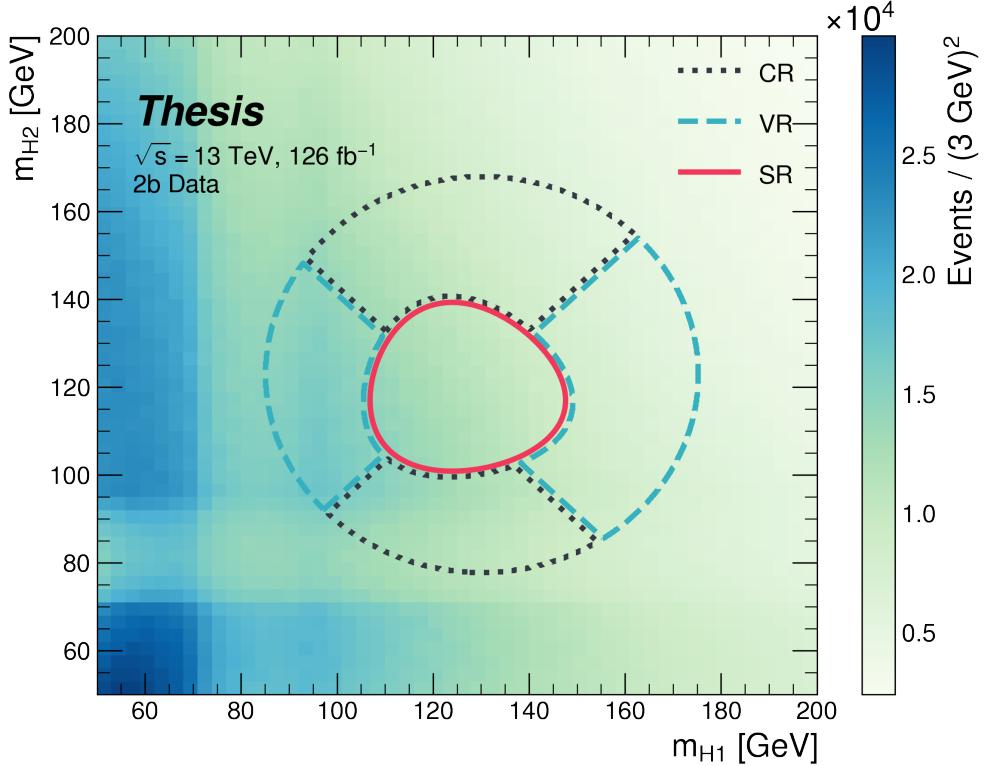


Figure 7.6: Regions used for the non-resonant search. The “top” and “bottom” quadrants together comprise the control region, in which the nominal background estimate is derived. The “left” and “right” quadrants together comprise the validation region, which is used to assess an uncertainty. The signal region, in the center, is where the signal extraction fit is performed.

²¹³³ 45° , with the “top” and “bottom” quadrants together comprising the control region, and the
²¹³⁴ other set (“left” and “right”) the validation region. These regions are shown in Figure 7.6

²¹³⁵ This design of regions includes a set of events closer to the signal region in the mass plane,
²¹³⁶ leveraging the assumption that these events are more similar to signal region events, while
²¹³⁷ also including events further away from the signal region, mitigating signal contamination.
²¹³⁸ This region selection is found to have good performance in alternate validation regions (see
²¹³⁹ Section 9.4).

2140 7.3.3 Discriminating Variable

2141 The discriminant used for the resonant analysis is *corrected* m_{HH} . This variable is calculated
 2142 by re-scaling the Higgs candidate four vectors such that each $m_H = 125$ GeV. These re-scaled
 2143 four-vectors are then summed, and their invariant mass is the corrected m_{HH} . These re-scaled
 2144 four-vectors are not used for any other purpose. The effect of this correction, which sharpens
 the m_{HH} peak and correctly centers it, is shown in Figure 7.7.

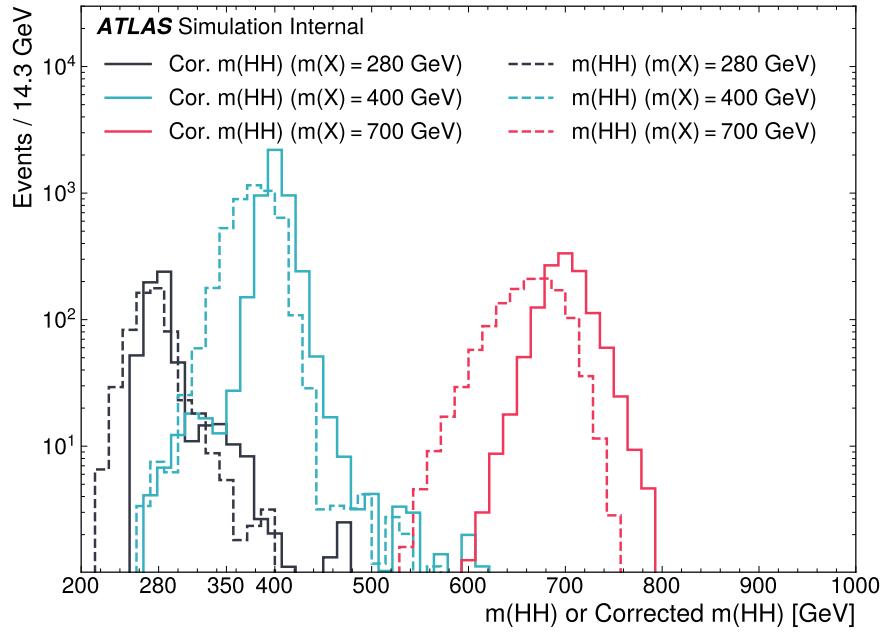


Figure 7.7: Impact of the m_{HH} correction on a range of spin-0 resonant signals. The corrected m_{HH} distributions (solid lines) are much sharper and more centered on the corresponding resonance masses than the uncorrected m_{HH} distributions (dashed).

2145

2146 For the non-resonant analysis, due to the broad nature of the signal in m_{HH} , such a
 2147 correction is not as motivated, and, indeed, is found to have very minimal impact. The
 2148 uncorrected m_{HH} (just referred to as m_{HH}) is therefore used as a discriminant. To maximize

sensitivity, the non-resonant analysis additionally uses two variables for categorization: $\Delta\eta_{HH}$, an angular variable which, along with m_{HH} , fully characterizes the HH system [102], and X_{HH} , the variable used for the signal region definition, which leverages the peaked structure of the signal in the $(m(H_1), m(H_2))$ plane to split the signal extraction fit into lower and higher purity regions (highest purity near $X_{HH} = 0$, the center of the signal region). Distributions of these variables are shown in *TODO: plots*. The categorization used for this thesis has been optimized to be 2×2 in these variables, with corresponding selections $0 \leq \Delta\eta_{HH} \leq 0.75$ and $0.75 \leq \Delta\eta_{HH} \leq 1.5$ for $\Delta\eta_{HH}$, and $0 \leq X_{HH} \leq 0.95$ and $0.95 \leq X_{HH} \leq 1.6$ for X_{HH} .

2157

Chapter 8

2158

BACKGROUND ESTIMATION

2159 After the event selection described above, there are two major backgrounds, QCD and $t\bar{t}$.
 2160 A very similar approach is used for both the resonant and the non-resonant analyses, with
 2161 some small modifications. This approach is notably fully data-driven, which is warranted due
 2162 to the flexibility of the estimation method, as well as the high relative proportion of QCD
 2163 background ($> 90\%$), and allows for the use of machine learning methods in the construction
 2164 of the background estimate. However, it sacrifices an explicit treatment of the $t\bar{t}$ component.
 2165 Performance of the background estimate on the $t\bar{t}$ component is checked explicitly, and
 2166 minimal impact due to $t\bar{t}$ mis-modeling is seen.

2167 Contributions of single Higgs processes and ZZ are found to be negligible, and the
 2168 Standard Model HH background is found to have no impact on the resonant search.

2169 The foundation of the background estimate lies in the derivation of a reweighting function
 2170 which matches the kinematics of events with exactly two b -tagged jets to those of events in
 2171 the higher tagged regions (events with three or four b -tagged jets). The reweighting function
 2172 and overall normalization are derived in the control region. Systematic bias of this estimate
 2173 is assessed in the validation region.

2174 For the resonant analysis, the systematic bias is a bias due to extrapolation: the validation
 2175 region lies between the control and signal regions. For the non-resonant analysis, the bias
 2176 instead comes from different possible interpolations of the signal region kinematics – given the
 2177 choice of nominal estimate, the validation region is a conceptually equivalent, but maximally
 2178 different, signal region estimate.

2179 **8.1 The Two Tag Region**

2180 Events in data with exactly two b-tagged jets are used for the data driven background estimate.
2181 The hypothesis here is that, due to the presence of multiple b -tagged jets, the kinematics of
2182 such events are similar to the kinematics of events in higher b-tagged regions (i.e. events
2183 with three and four b -tagged jets, respectively), and any differences can be corrected by a
2184 reweighting procedure. The region with three b -tagged jets is split into two b -tagging regions,
2185 as described in Section 7.1, with the $3b + 1$ loose region used as an additional signal region.
2186 The lower tagged $3b$ component ($3b + 1$ fail) is reserved for validation of the background
2187 modelling procedure. Events with fewer than two b -tagged jets are not used for this analysis,
2188 as they are relatively more different from the higher tag regions.

2189 The nominal event selection requires at least four jets in order to form Higgs candidates.
2190 For the four tag region, these are the four highest p_T b -tagged jets. For the three tag regions,
2191 these jets are the three b -tagged jets, plus the highest p_T jet satisfying a loosened b -tagging
2192 requirement. Similarly, and following the approach of the resonant analysis, the two tag region
2193 uses the two b -tagged jets and the two highest p_T non-tagged jets to form Higgs candidates.
2194 Combinatoric bias from selection of different numbers of b -tagged jets is corrected as a part
2195 of the kinematic reweighting procedure through the reweighting of the total number of jets in
2196 the event. In this way, the full event selection may be run on two tagged events.

2197 **8.2 Kinematic Reweighting**

2198 The set of two tagged data events is the fundamental piece of the data driven background
2199 estimate. However, kinematic differences from the four tag region exist and must be corrected
2200 in order for this estimate to be useful. Binned approaches based on ratios of histograms
2201 have been previously considered [2], [22], but are limited in their handling of correlations
2202 between variables and by the “curse of dimensionality”, i.e. the dataset becomes sparser and
2203 sparser in “reweighting space” as the number of dimensions in which to reweight increases,
2204 limiting the number of variables used for reweighting. This leads either to an unstable fit

2205 result (overfitting with finely grained bins) or a lower quality fit result (underfitting with
 2206 coarse bins).

2207 Note that even some machine learning methods such as Boosted Decision Trees (BDTs) [103],
 2208 may suffer from this curse of dimensionality, as the depth of each decision tree used is limited
 2209 by the available statistics after each set of corresponding selections (cf. binning in a more
 2210 sophisticated way), limiting the expressivity of the learned reweighting function.

2211 To solve these issues, a neural network based reweighting procedure is used here. This
 2212 is a truly multivariate approach, allowing for proper treatment of variable correlations. It
 2213 further overcomes the issues associated with binned approaches by learning the reweighting
 2214 function directly, allowing for greater sensitivity to local differences and helping to avoid the
 2215 curse of dimensionality.

2216 8.2.1 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

$$p_{2b}(x) \cdot w(x) = p_{4b}(x) \quad (8.1)$$

from which it follows that

$$w(x) = \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (8.2)$$

This falls into the domain of density ratio estimation, for which there are a variety of approaches. The method considered here is modified from [104, 105], and depends on a loss function of the form

$$\mathcal{L}(R(x)) = \mathbb{E}_{x \sim p_{2b}}[\sqrt{R(x)}] + \mathbb{E}_{x \sim p_{4b}}\left[\frac{1}{\sqrt{R(x)}}\right]. \quad (8.3)$$

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

$$\arg \min_R \mathcal{L}(R(x)) = \frac{p_{4b}(x)}{p_{2b}(x)} \quad (8.4)$$

2217 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

$$\mathcal{L}(Q(x)) = \mathbb{E}_{x \sim p_{2b}} [\sqrt{e^{Q(x)}}] + \mathbb{E}_{x \sim p_{4b}} \left[\frac{1}{\sqrt{e^{Q(x)}}} \right], \quad (8.5)$$

with solution

$$\arg \min_Q \mathcal{L}(Q(x)) = \log \frac{p_{4b}(x)}{p_{2b}(x)}. \quad (8.6)$$

2218 Taking the exponent then results in the desired reweighting function.

2219 Note that similar methods for density ratio estimation are available [106], e.g. from a
2220 more standard binary cross-entropy loss. However, these were found to perform no better
2221 than the formulation presented here.

2222 8.2.2 Variables and Results

2223 The neural network is trained on a variety of variables sensitive to two vs. four tag differences.
2224 To help bring out these differences, the natural logarithm of some of the variables with a
2225 large, local change is taken. The set of training variables used for the resonant analysis is

2226 1. $\log(p_T)$ of the 4th leading Higgs candidate jet

2227 2. $\log(p_T)$ of the 2nd leading Higgs candidate jet

2228 3. $\log(\Delta R)$ between the closest two Higgs candidate jets

2229 4. $\log(\Delta R)$ between the other two Higgs candidate jets

2230 5. Average absolute value of η across the four Higgs candidate jets

- 2231 6. $\log(p_T)$ of the di-Higgs system.
- 2232 7. ΔR between the two Higgs candidates
- 2233 8. $\Delta\phi$ between the jets in the leading Higgs candidate
- 2234 9. $\Delta\phi$ between the jets in the subleading Higgs candidate
- 2235 10. $\log(X_{Wt})$, where X_{Wt} is the variable used for the top veto
- 2236 11. Number of jets in the event.

- 2237 The non-resonant analysis uses an identical set of variables with two notable changes
- 2238 1. The definition of X_{Wt} differs from the resonant definition (as described in Section 7.2).
- 2239 2. An integer encoding of the two trigger categories is used as an input (variable which
2240 takes on the value 0 or 1 corresponding to each of the two categories). This was found
2241 to improve a mis-modeling near the tradeoff in m_{HH} of the two buckets.

2242 The neural network used for both resonant and non-resonant reweighting has three densely
2243 connected hidden layers of 50 nodes each with ReLU activation functions and a single node
2244 linear output. This configuration demonstrates good performance in the modelling of a variety
2245 of relevant variables, including m_{HH} , when compared to a range of networks of similar size.

2246 In practice, a given training of the reweighting neural network is subject to variation
2247 due to training statistics and initial conditions. An uncertainty is assigned to account for
2248 this (Chapter 9), which relies on training an ensemble of reweighting networks [107]. To
2249 increase the stability of the background estimate, the median of the predicted weight for each
2250 event is calculated across the ensemble. This median is then used as the nominal background
2251 estimate. This approach is indeed seen to be much more stable and to demonstrate a better
2252 overall performance than a single arbitrary training. Each ensemble used for this analysis
2253 consists of 100 neural networks, trained as described in Chapter 9.

2254 The training of the ensemble used for the nominal estimate is done in the kinematic
 2255 Control Region. The prediction of these networks in the Signal Region is then used for the
 2256 nominal background estimate. In addition, a separate ensemble of networks is trained in the
 2257 Validation Region. The difference between the prediction of the nominal estimate and the
 2258 estimate from the VR derived networks in the Signal Region is used to assign a systematic
 2259 uncertainty. Further details on this systematic uncertainty are discussed in Chapter 9. Note
 2260 that although the same procedure is used for both Control and Validation Region trained
 2261 networks, only the median estimate from the VR derived reweighting is used for assessing a
 2262 systematic – no additional “uncertainty on the uncertainty” from VR ensemble variation is
 2263 applied.

2264 Each reweighted estimate is normalized such that the reweighted $2b$ yield matches the $4b$
 2265 yield in the corresponding training region. Note that this applies to each of the networks used
 2266 in each ensemble, where the normalization factor is also subject to the procedure described
 2267 in Chapter 9. As the median over these normalized weights is not guaranteed to preserve this
 2268 normalization, a further correction is applied such that the $2b$ yield, after the median weights
 2269 are applied, matches the $4b$ yield in the corresponding training region. As no pre-processing
 2270 is applied to correct for the class imbalance between $2b$ and $4b$ events entering the training,
 2271 this ratio of number of $4b$ events ($n(4b)$) over number of $2b$ events ($n(2b)$) is folded into the
 2272 learned weights. Correspondingly, the set of normalization factors described above is near 1
 2273 and the learned weights are centered around $n(4b)/n(2b)$ (roughly 0.01 over the full dataset).
 2274 This normalization procedure applies for all instances of the reweighting (e.g. those used for
 2275 validations in Section 9.4), with appropriate substitutions of reweighting origin (here $2b$) and
 2276 reweighting target (here $4b$).

2277 Note that, due to different trigger and pileup selections during each year, the reweighting
 2278 is trained on each year separately. An approach of training all of the years together with
 2279 a one-hot encoding was explored, but was found to have minimal benefit over the split
 2280 years approach, and in fact to increase the systematic bias of the corresponding background
 2281 estimate. Because of this, and because trigger selections for each year significantly impact

the kinematics of each year, such that categorizing by year is expected to reflect groupings of kinematically similar events and to provide a meaningful degree of freedom in the signal extraction fit, the split-year approach is kept.

The control region closure for the 2018 dataset is shown for the resonant search in Figures 8.1 through 8.9 and for the non-resonant search in Figures 8.19 through 8.27 for 4*b* and Figures 8.37 through 8.45 for 3*b1l*. The impact of this control region derived reweighting on the validation region is shown in Figures 8.10 through 8.18 for the resonant search and Figures 8.28 through 8.36 for 4*b* and Figures 8.46 through 8.54 for 3*b1l* for the non-resonant search. 2018 is chosen because it is the largest subset of the data on which the year-by-year reweighting is trained. The other years are omitted here for brevity, but demonstrate very similar results. Generally good performance is seen, with some occasional mis-modeling. For the resonant search, this is most notable in the case of individual jet p_T . Such mis-modeling may be corrected by including the variables in the input set, but this was found to not improve the modeling of m_{HH} , and so is not done here. This mis-modeling is notable for the non-resonant search in the leading Higgs candidate jet p_T , and is a direct consequence of the trigger category input, which improves modeling of m_{HH} . Results are similar for other years, but are not included here for brevity.

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

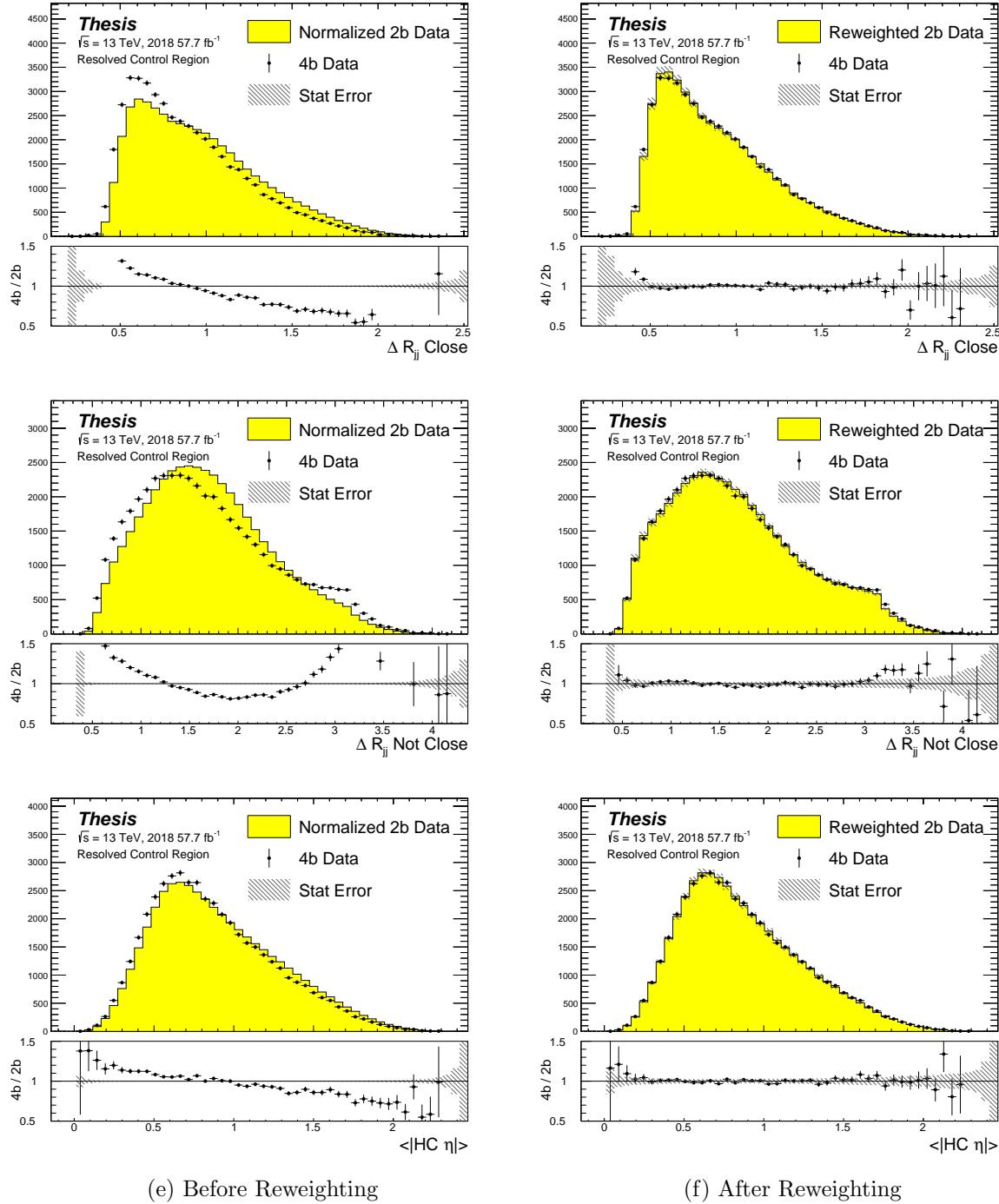


Figure 8.1: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 Control Region.

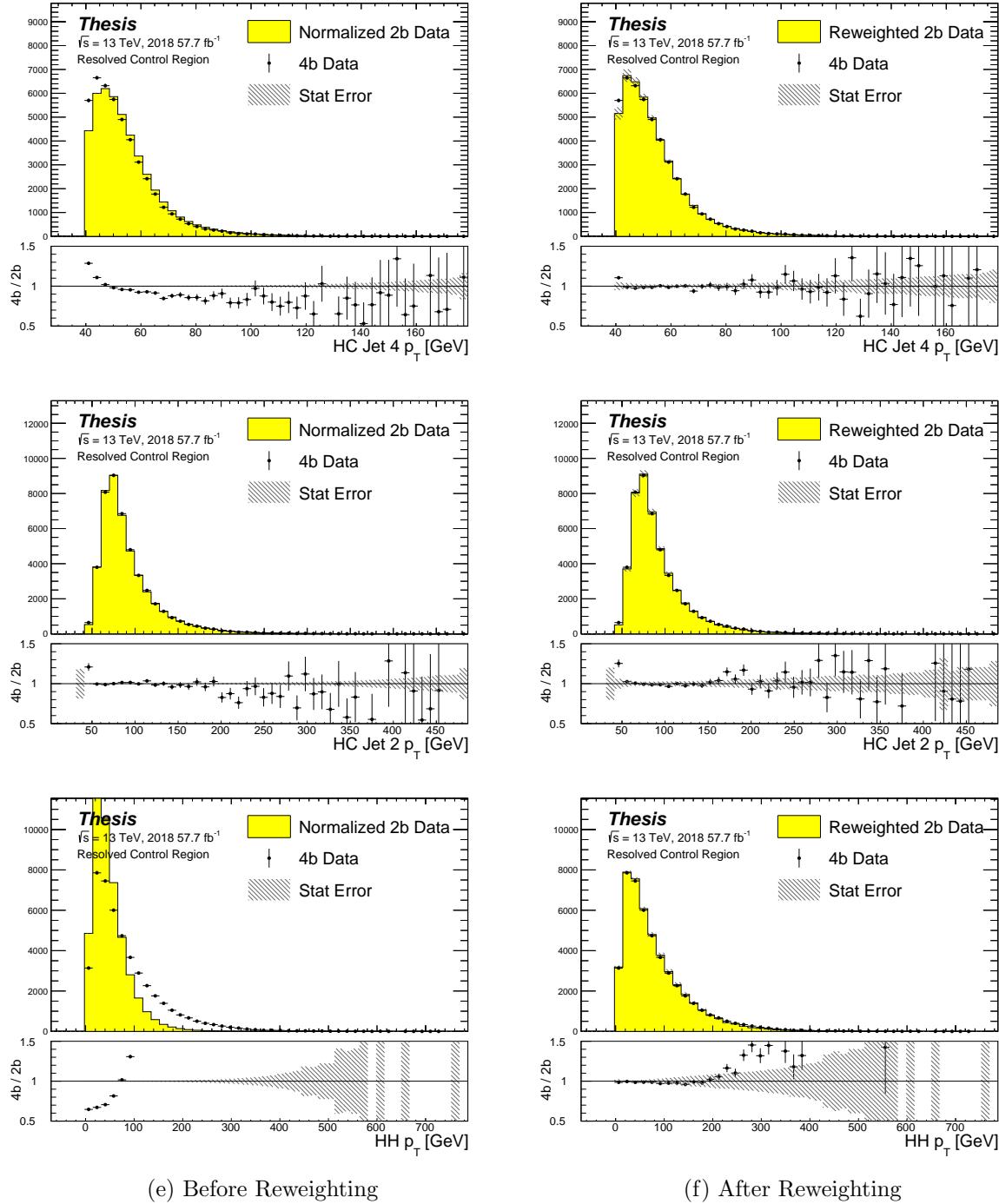


Figure 8.2: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Control Region.

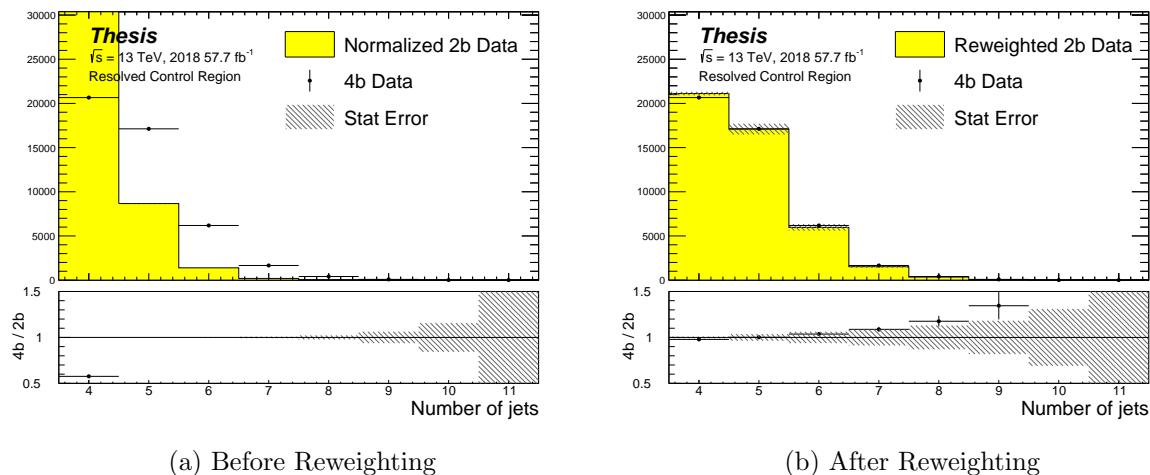


Figure 8.3: **Resonant Search:** Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

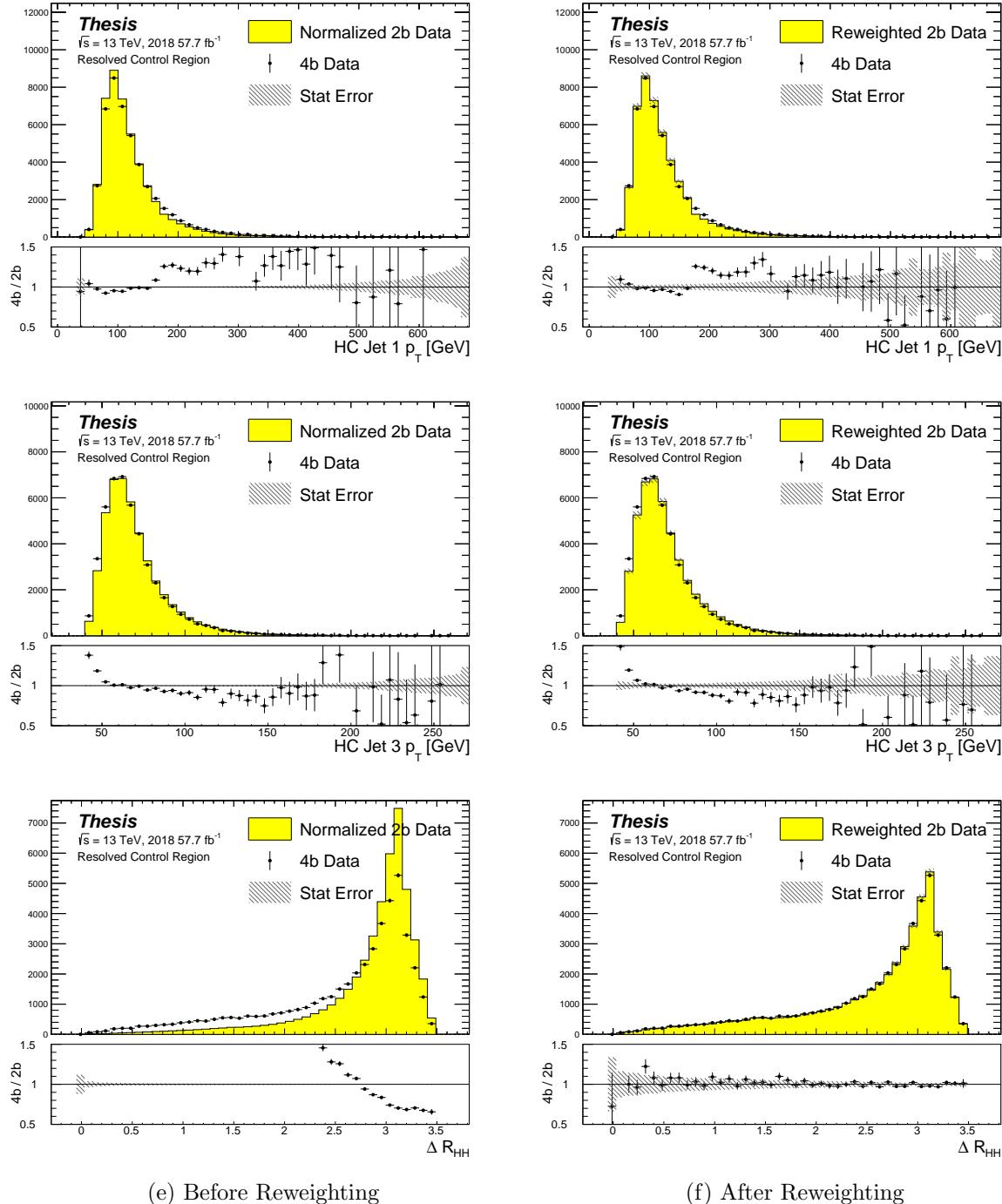


Figure 8.4: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region.

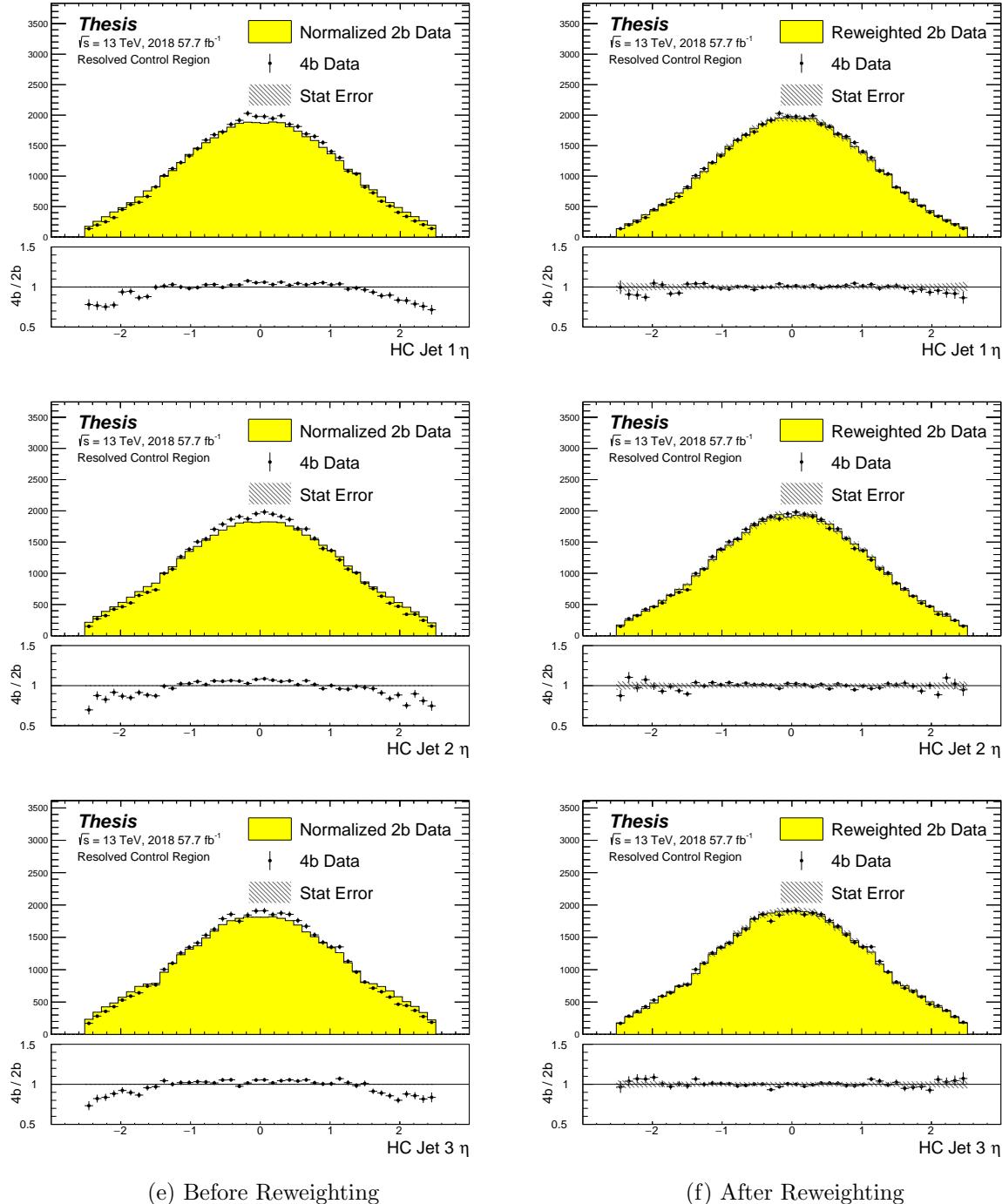


Figure 8.5: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 Control Region.

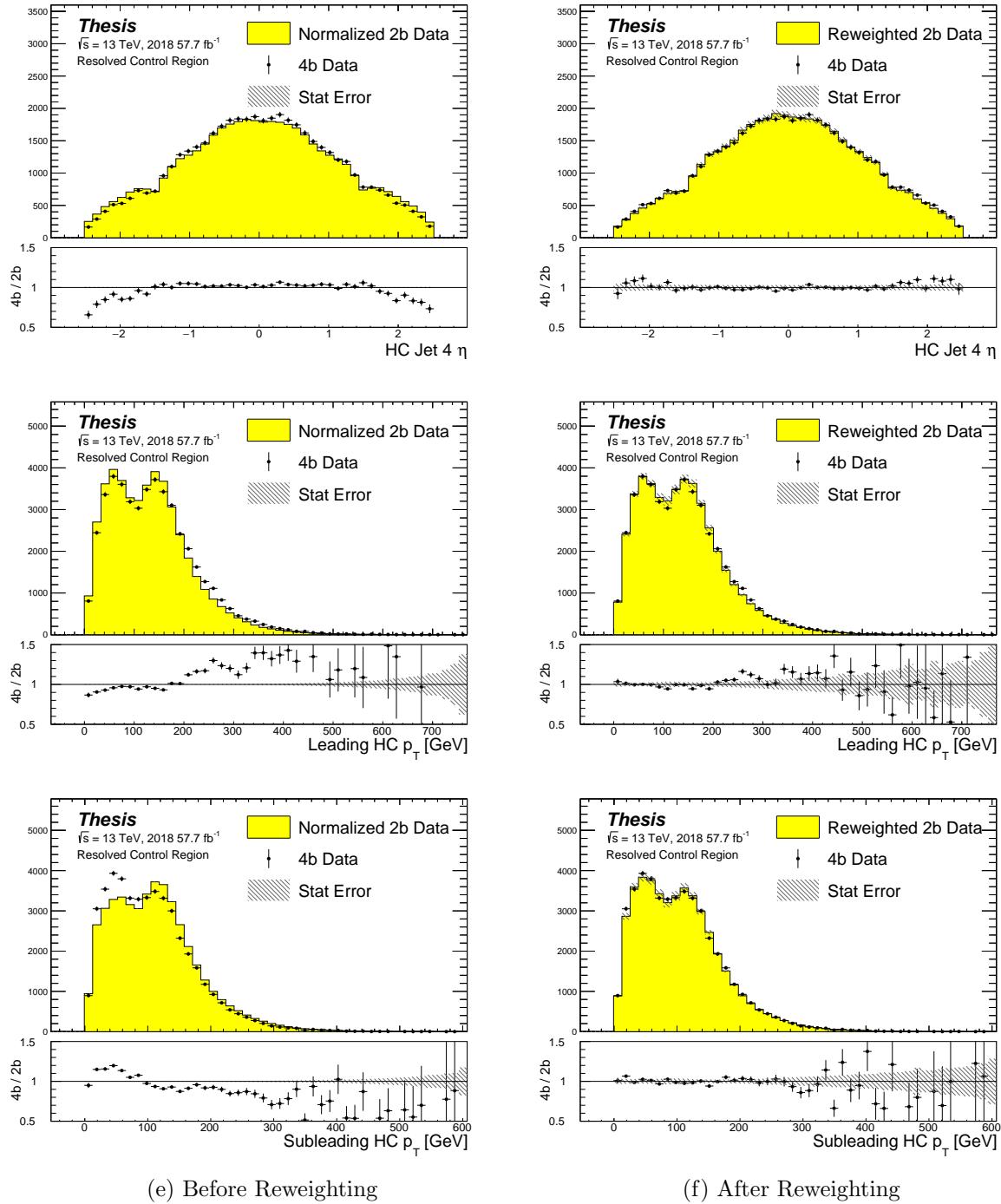


Figure 8.6: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region.

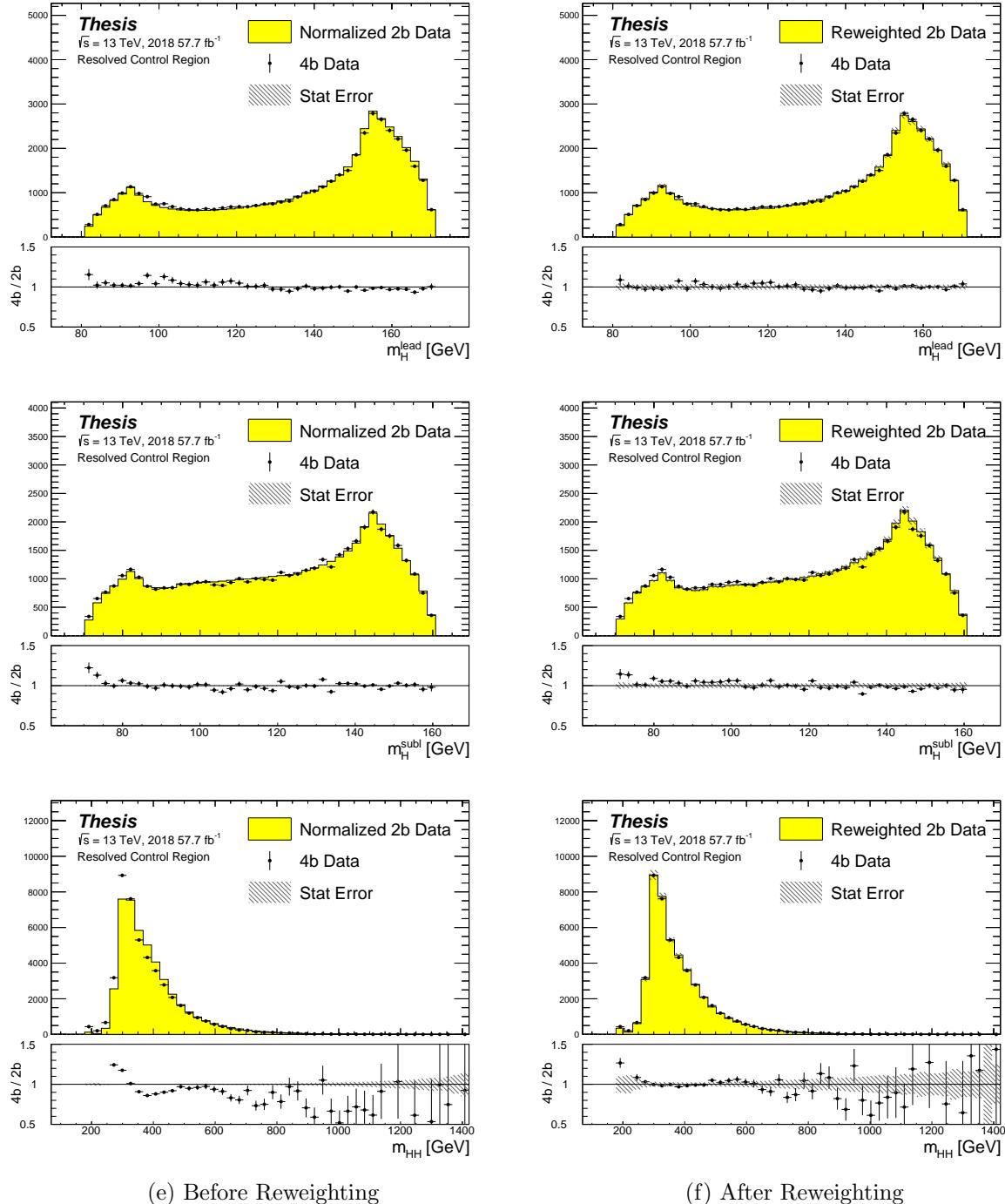


Figure 8.7: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Control Region.

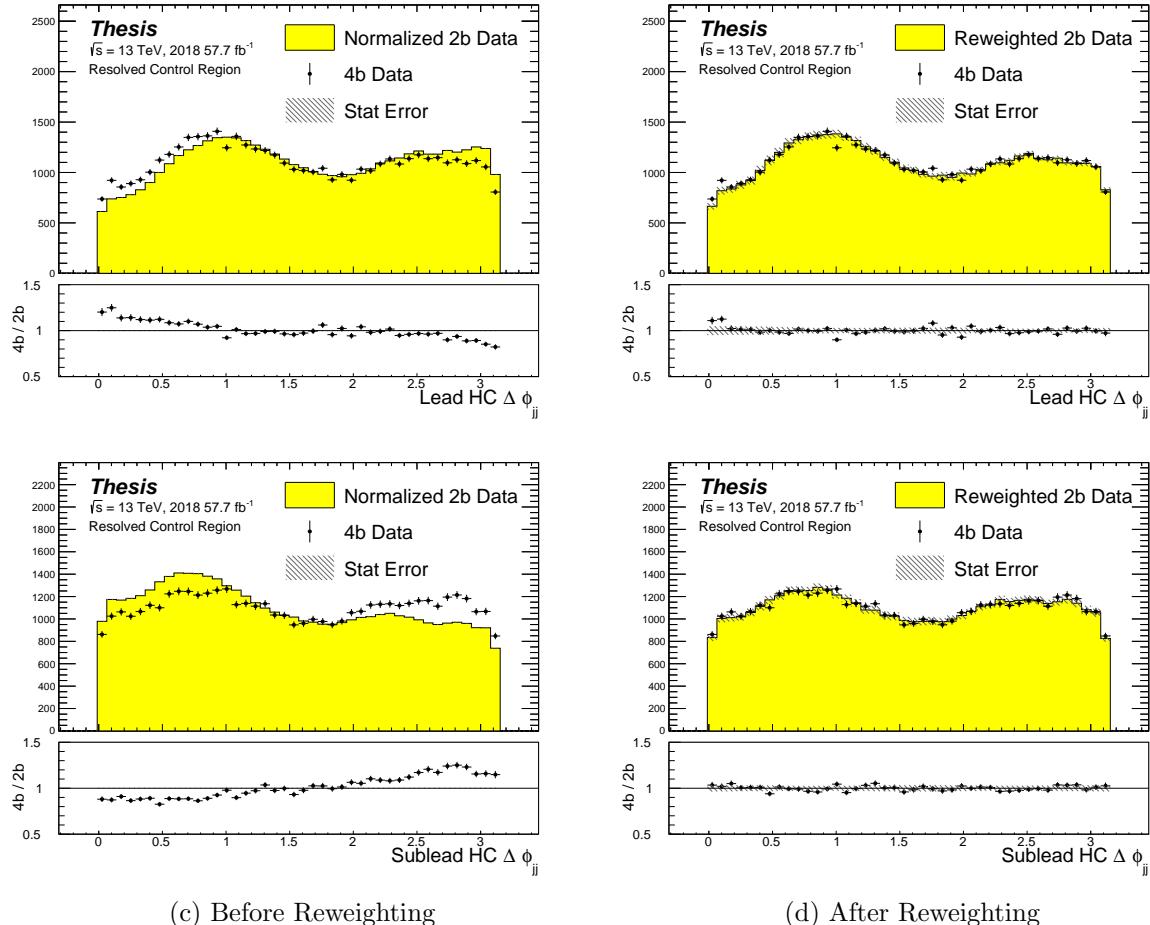


Figure 8.8: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Control Region.

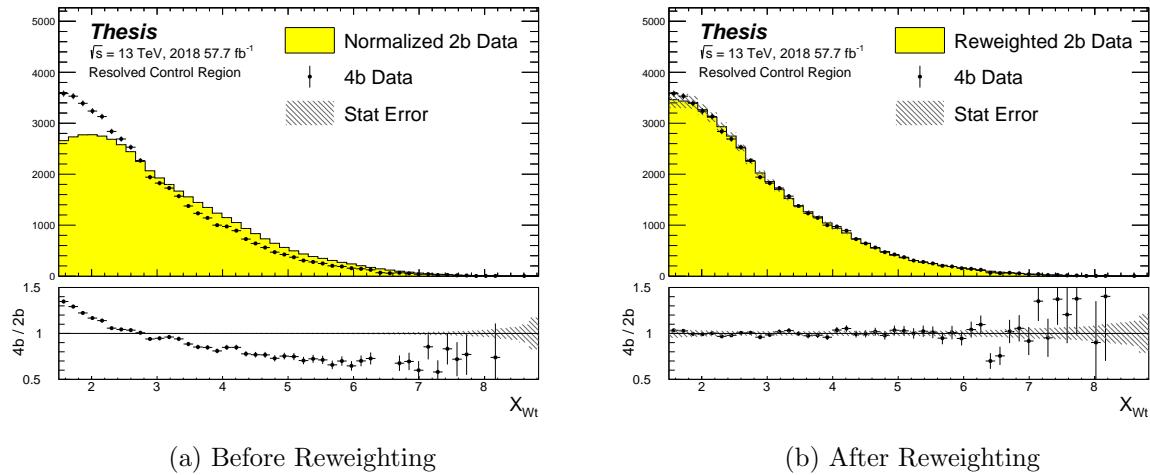


Figure 8.9: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 Control Region. Reweighting is done after the cut on this variable is applied

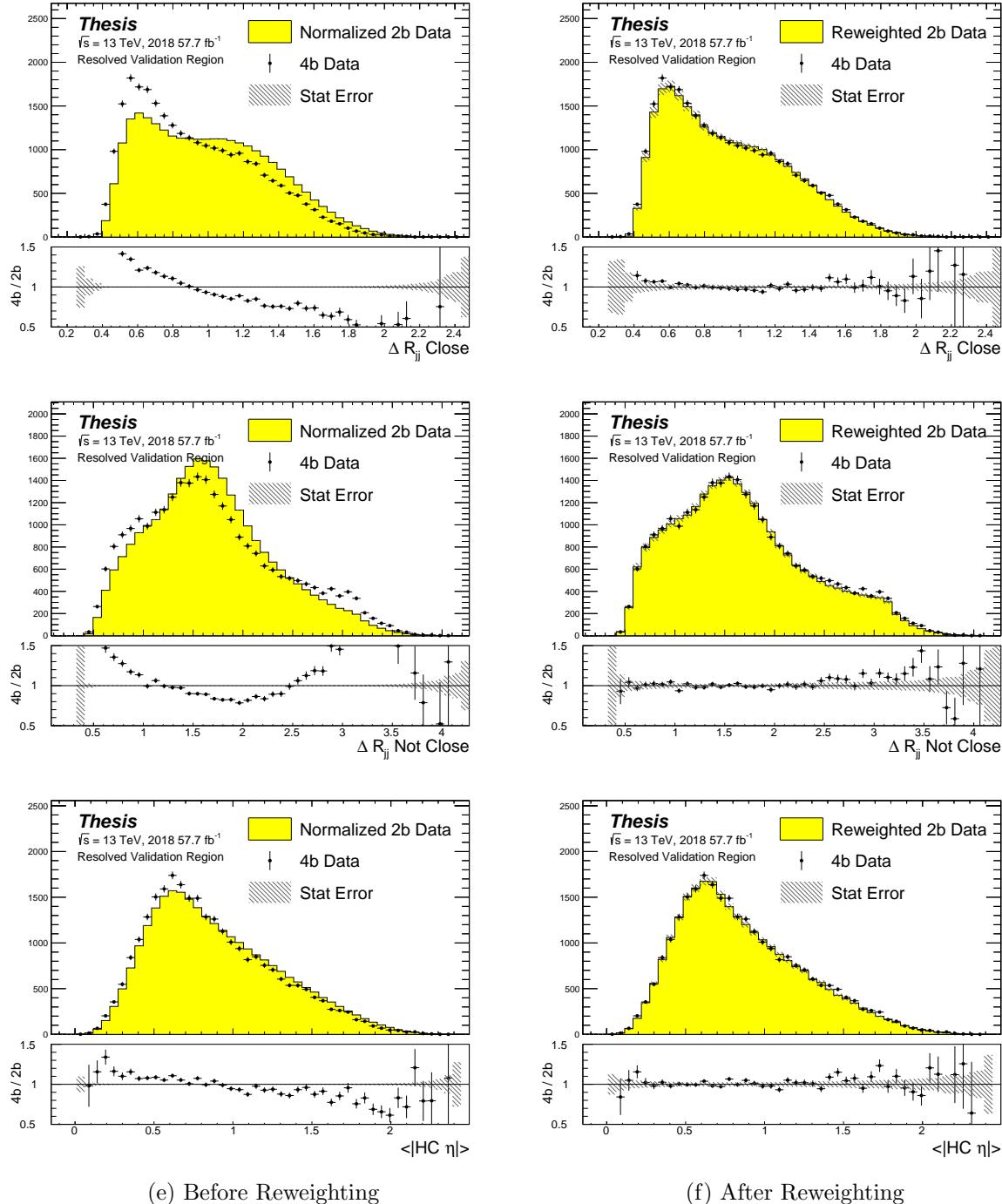


Figure 8.10: **Resonant Search:** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

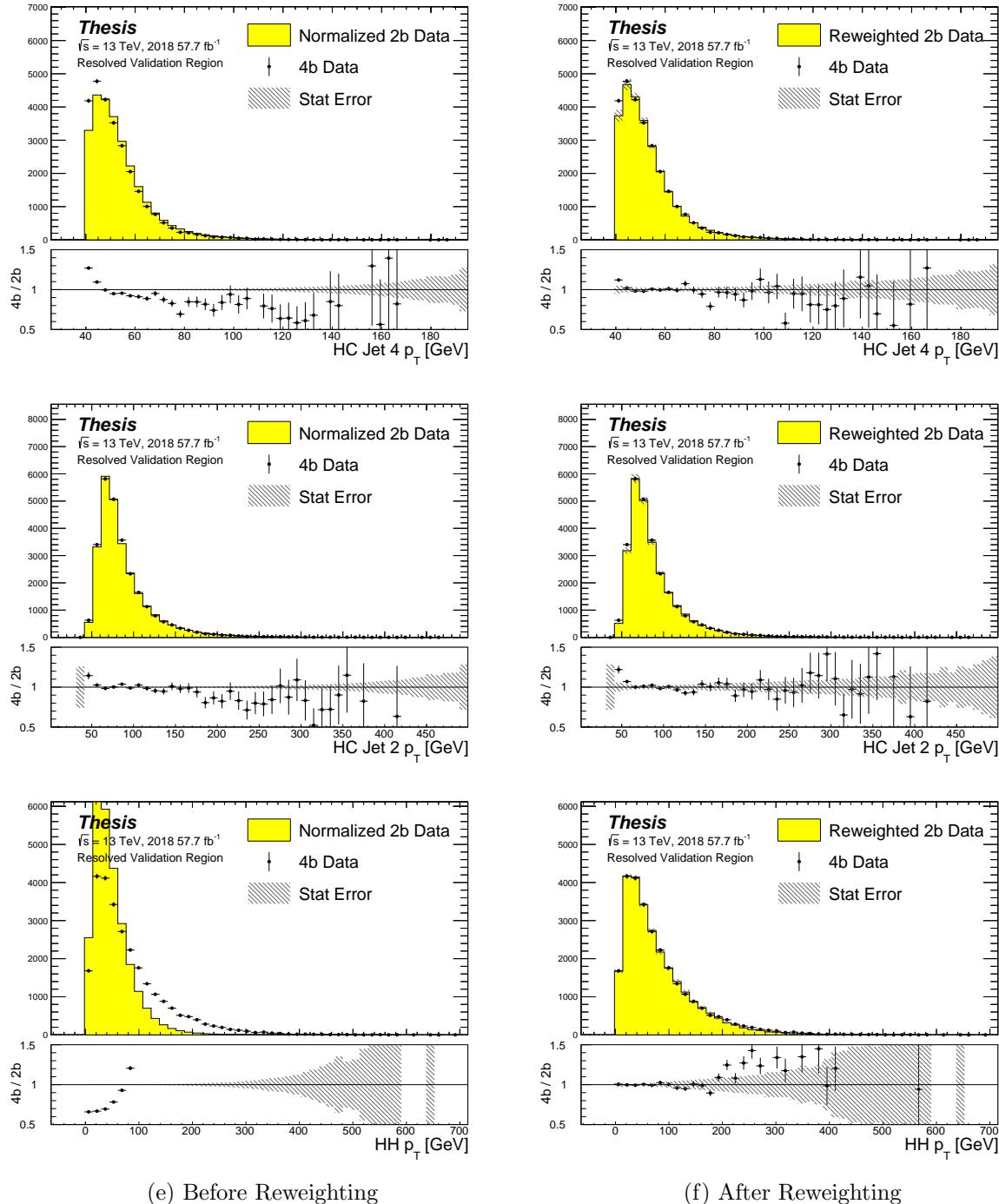


Figure 8.11: **Resonant Search:** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

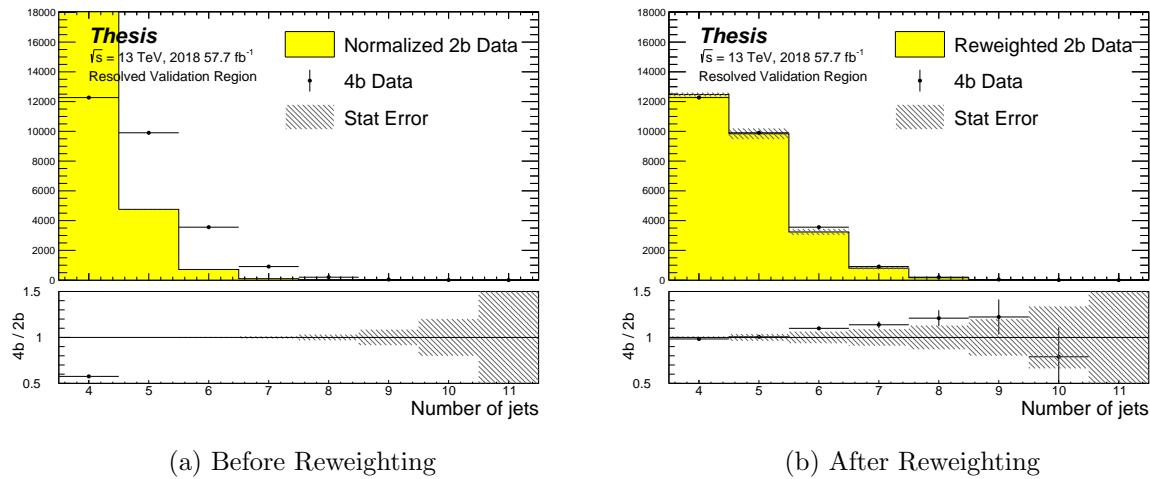


Figure 8.12: **Resonant Search:** Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

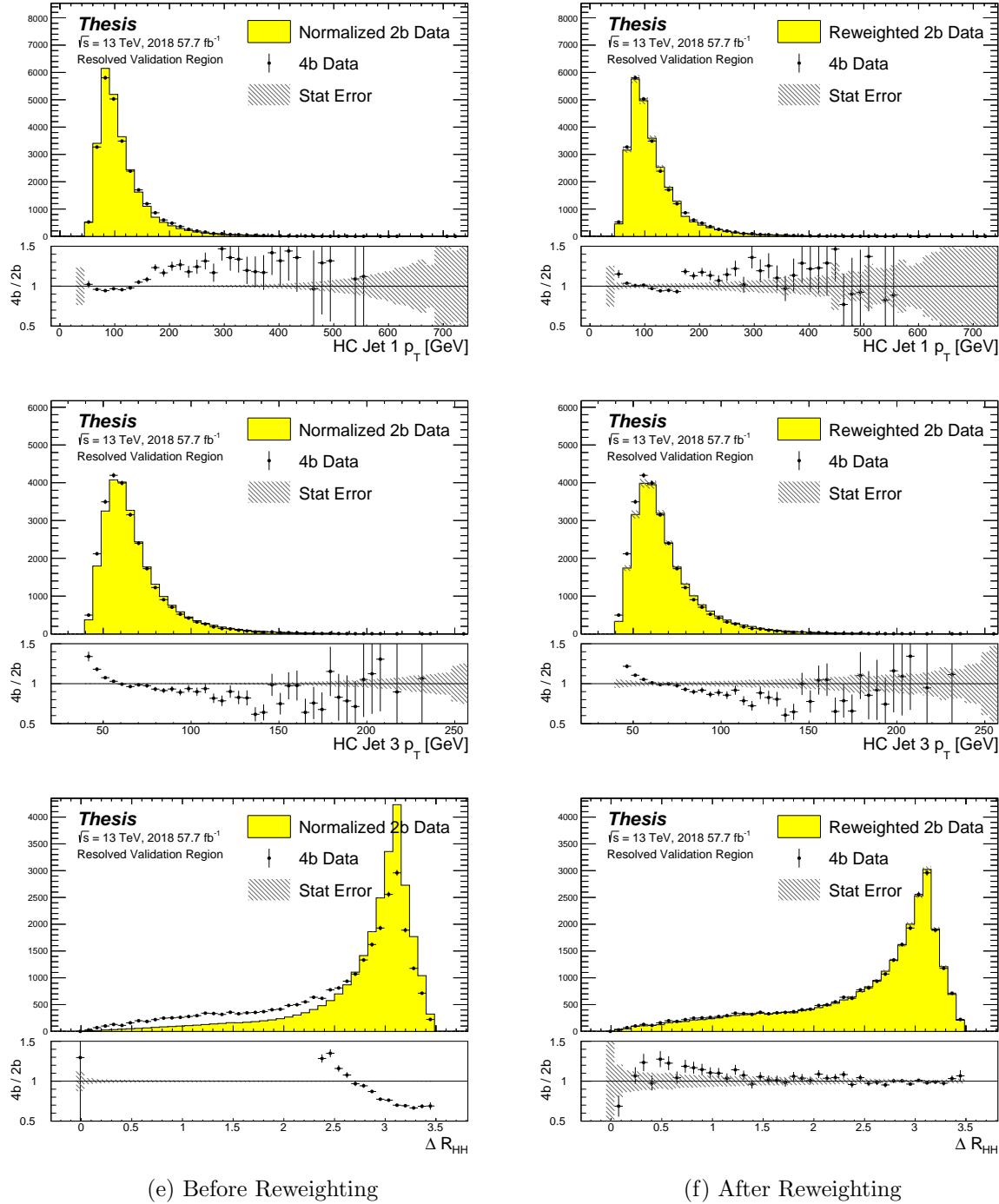


Figure 8.13: **Resonant Search:** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

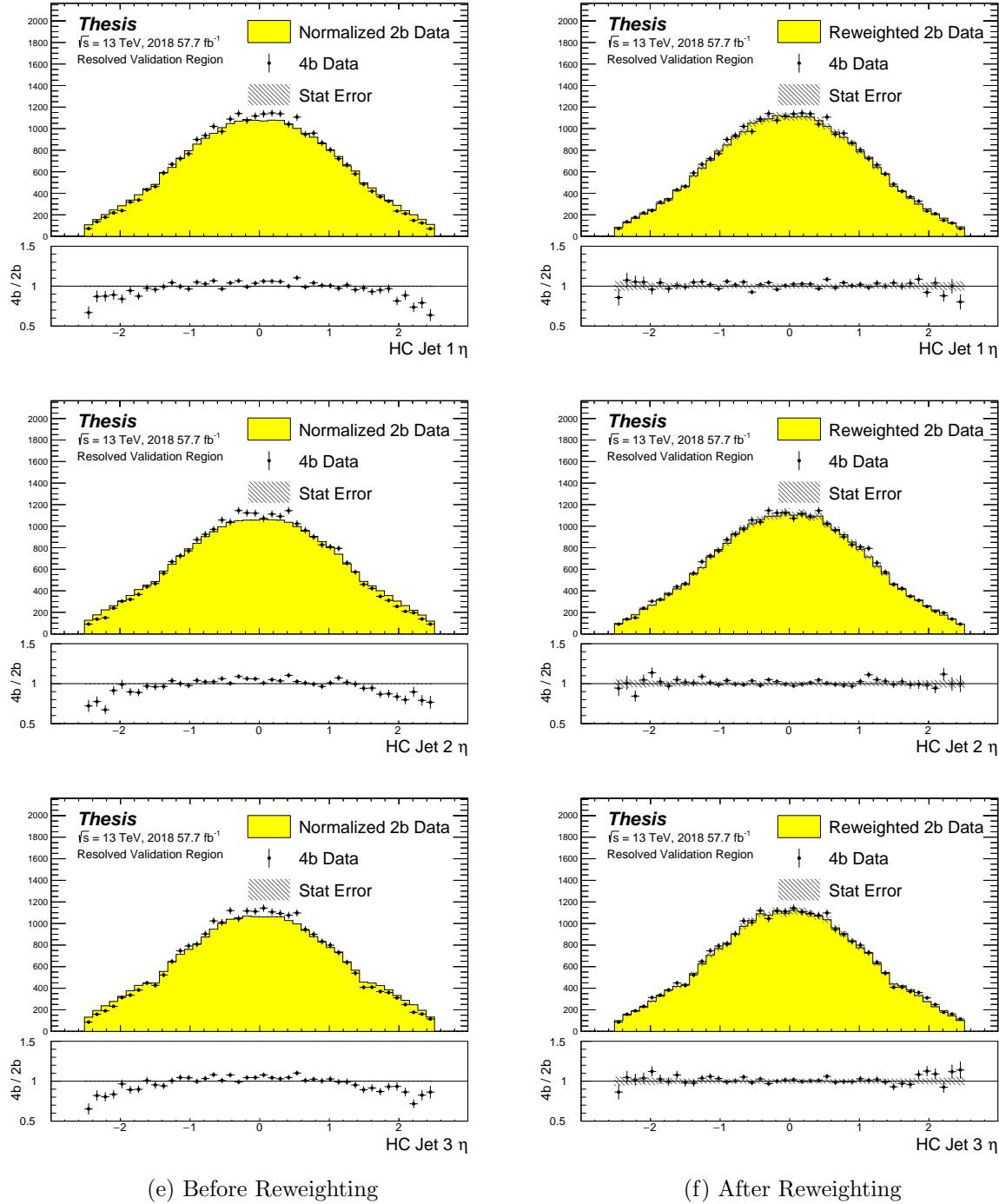


Figure 8.14: **Resonant Search:** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

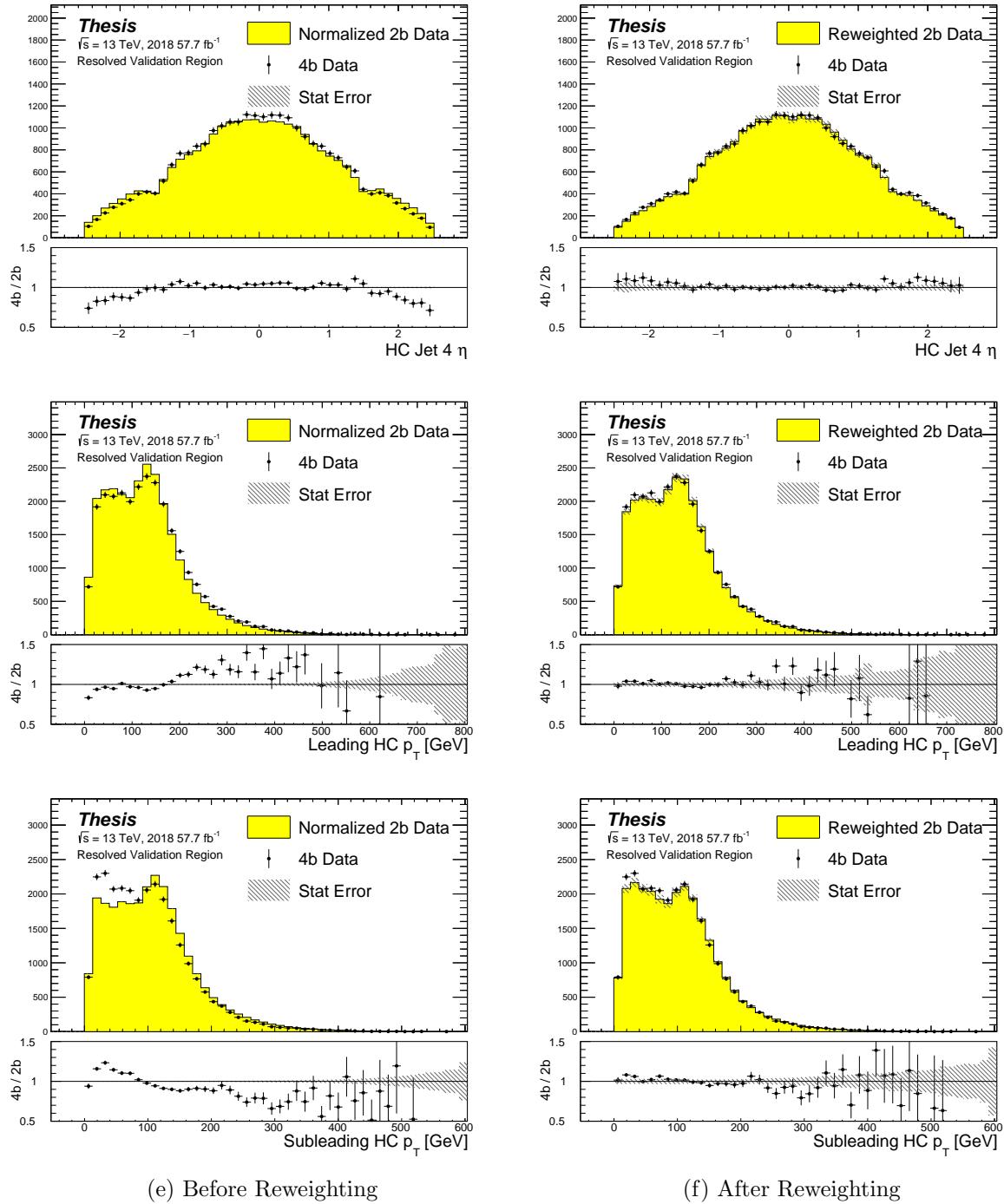


Figure 8.15: **Resonant Search:** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

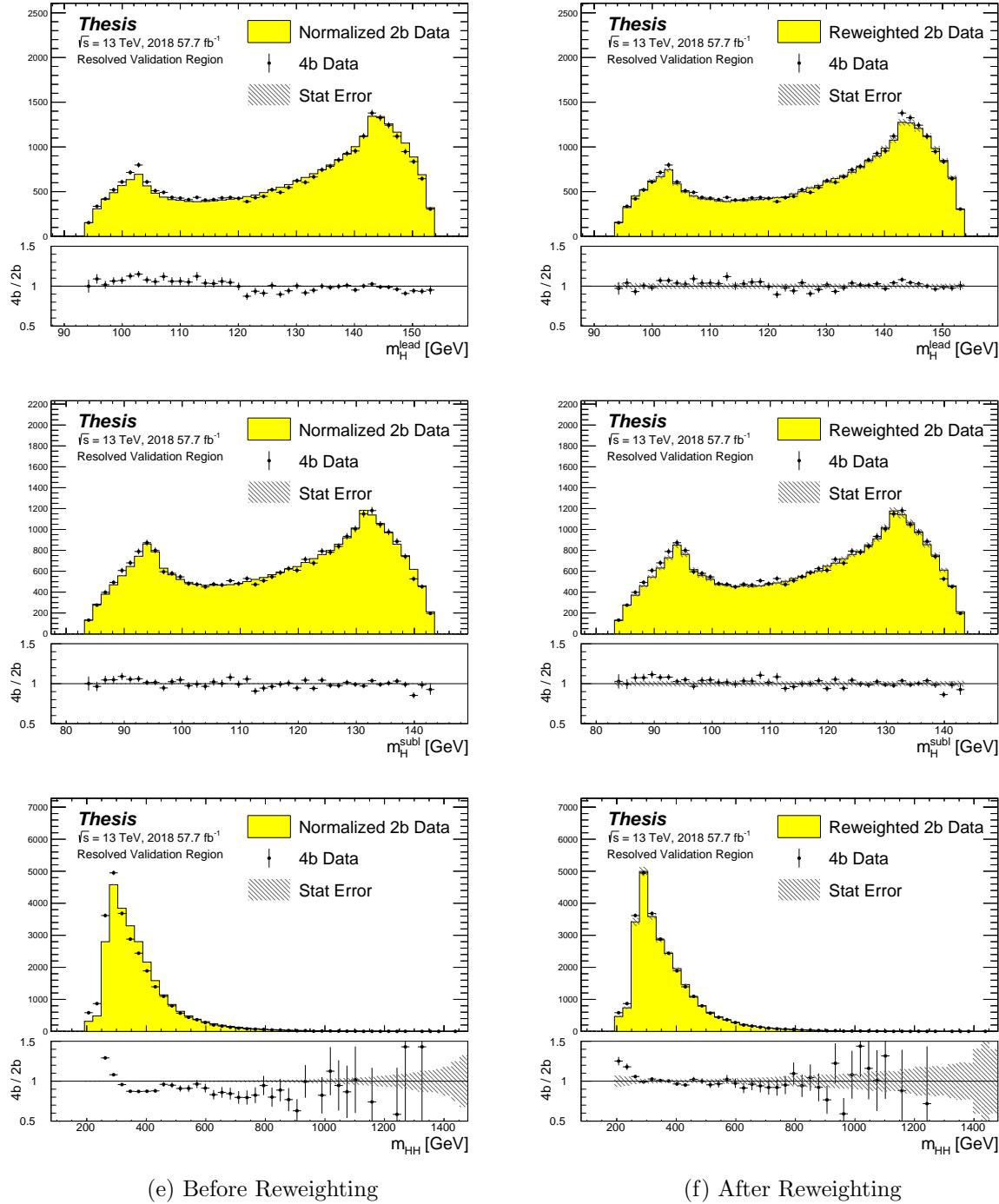


Figure 8.16: **Resonant Search:** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

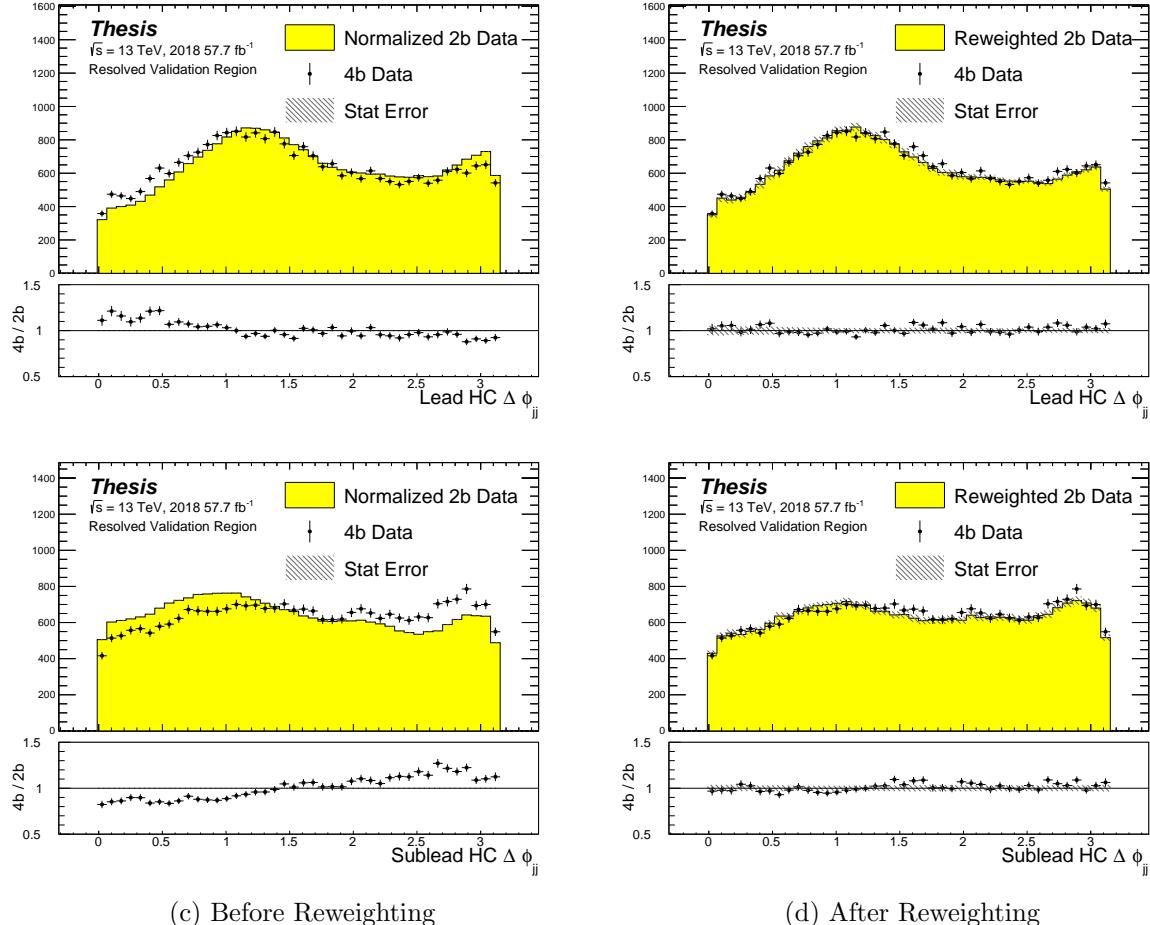


Figure 8.17: **Resonant Search:** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 Validation Region.

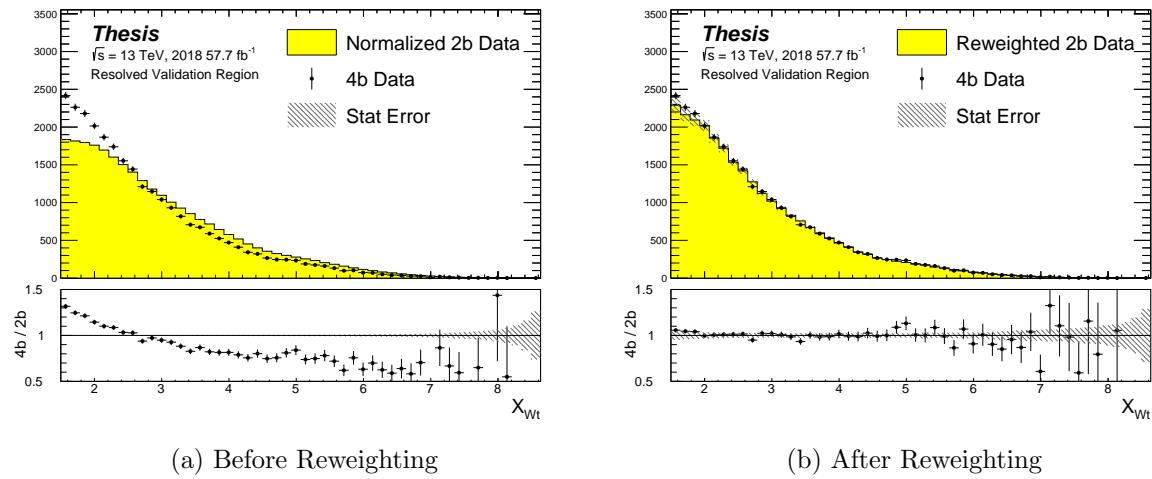


Figure 8.18: **Resonant Search:** Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 Validation Region. Reweighting is done after the cut on this variable is applied

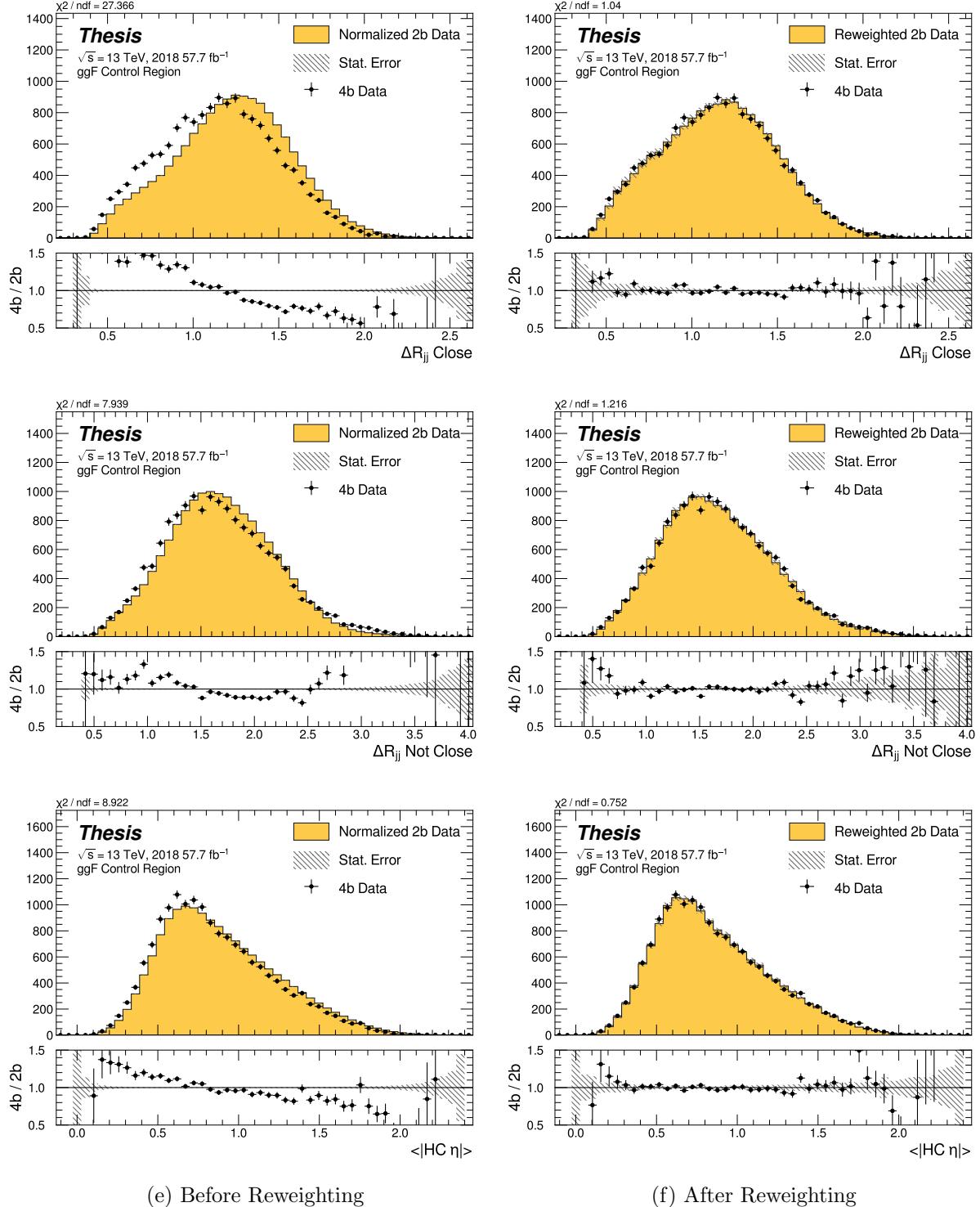


Figure 8.19: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

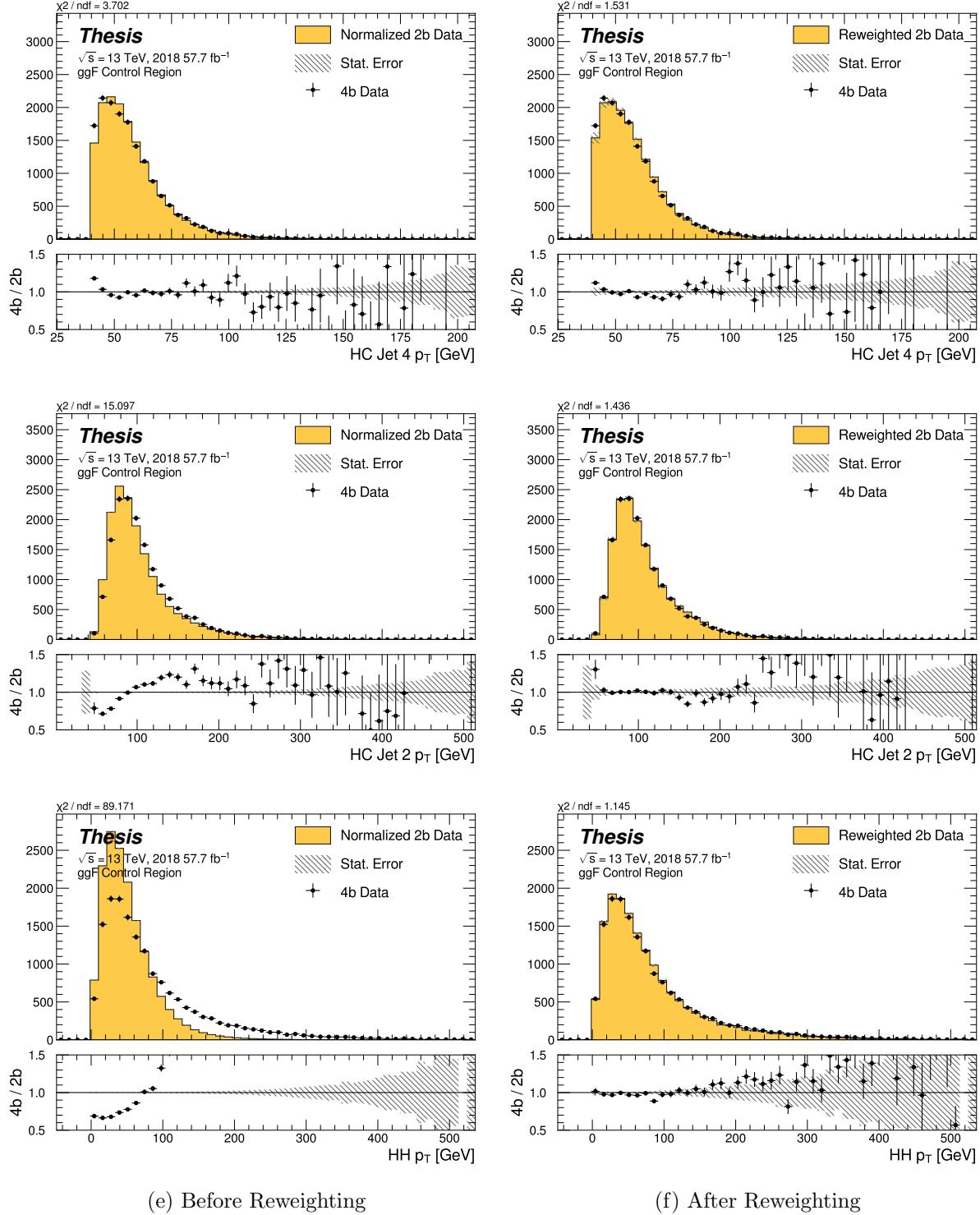


Figure 8.20: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

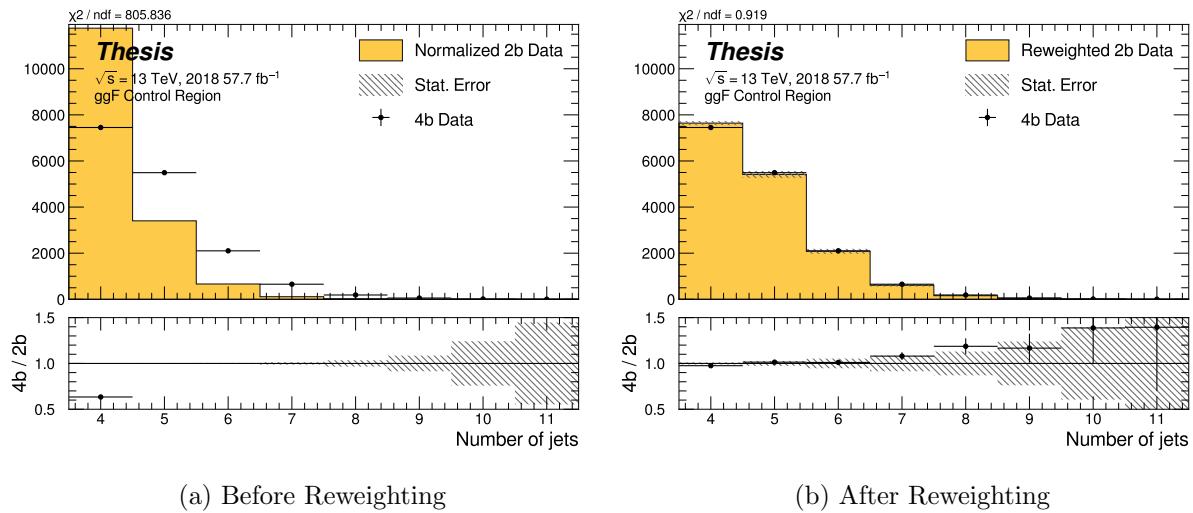


Figure 8.21: **Non-resonant Search (4b)**: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 4b Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

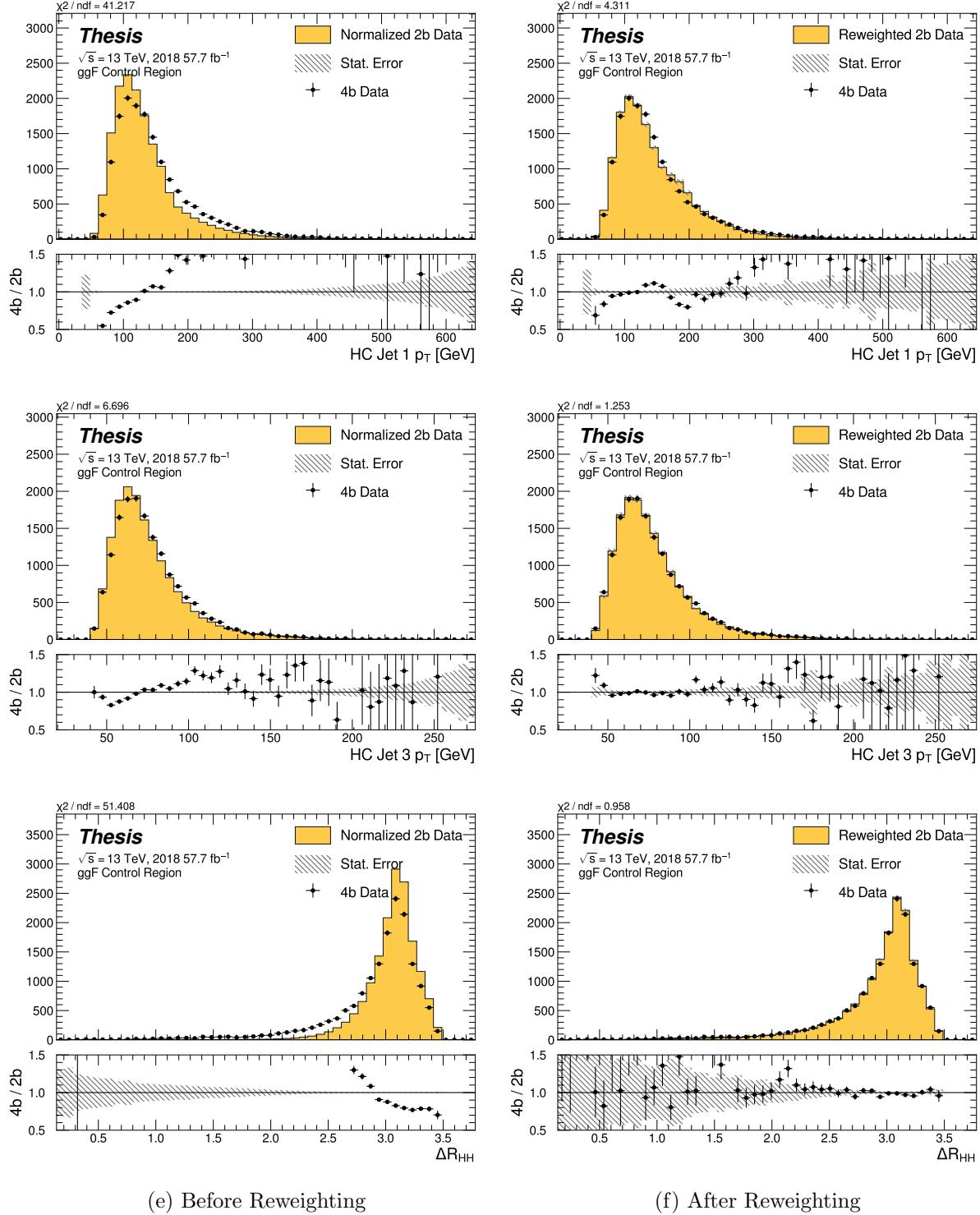


Figure 8.22: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

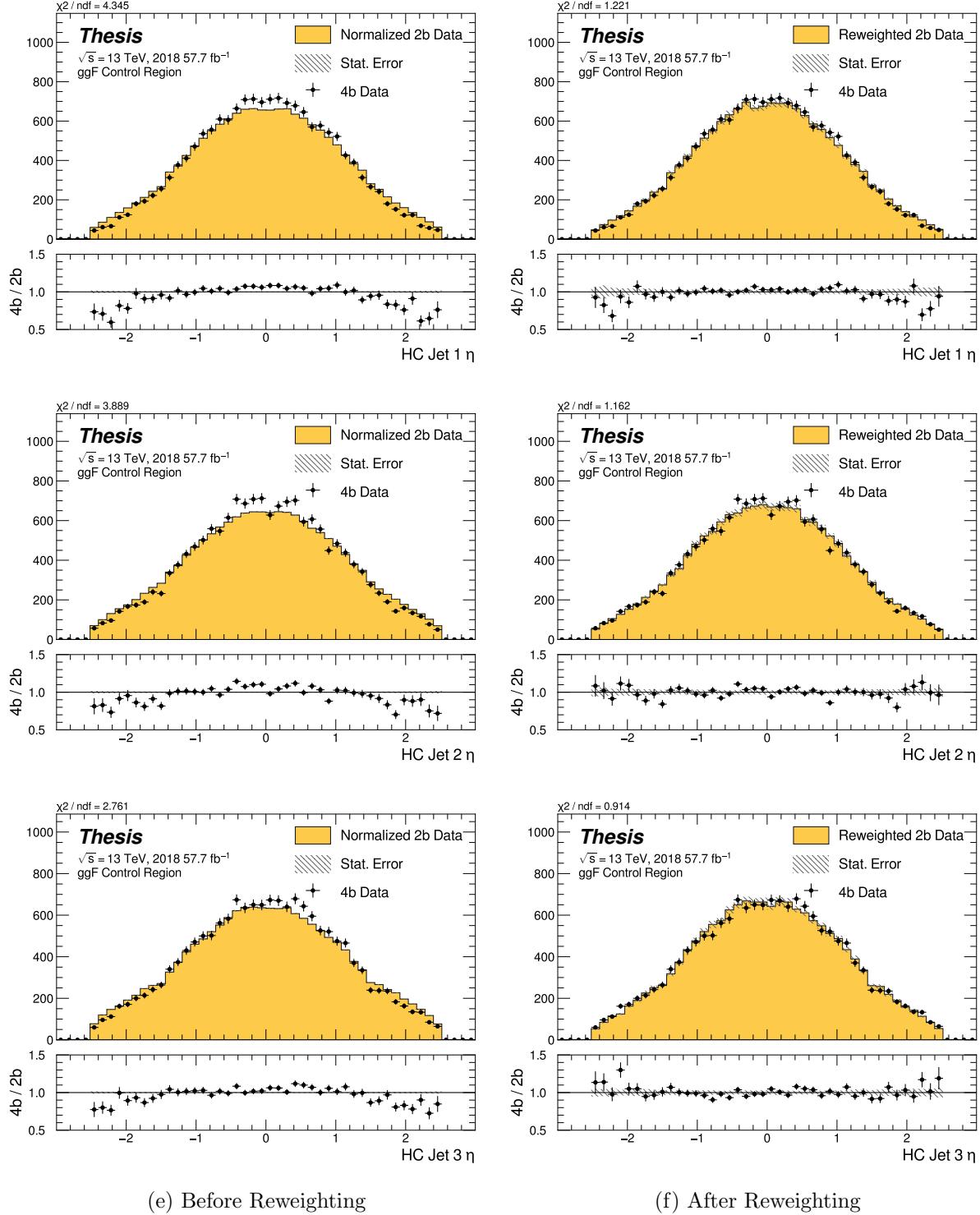


Figure 8.23: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

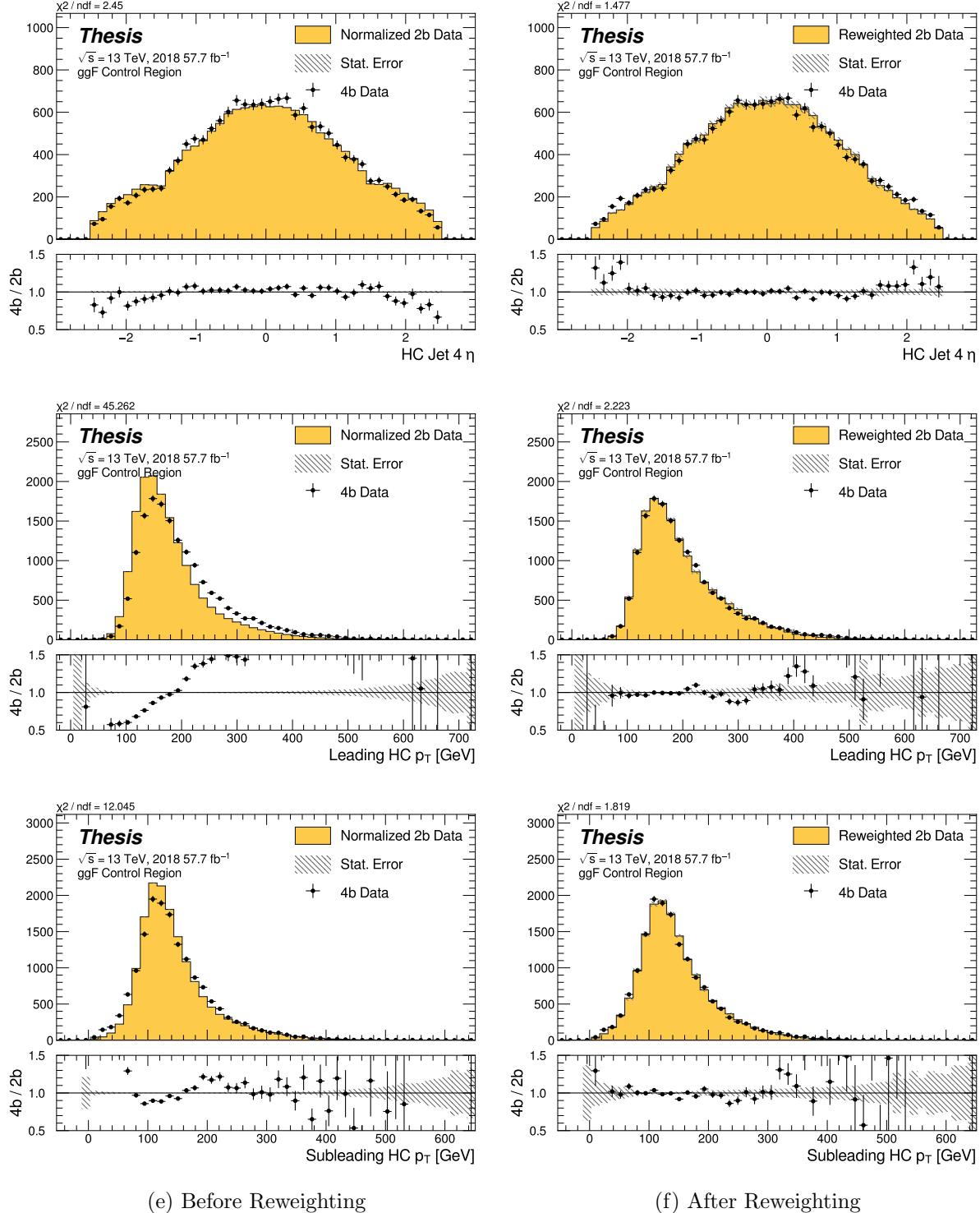


Figure 8.24: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

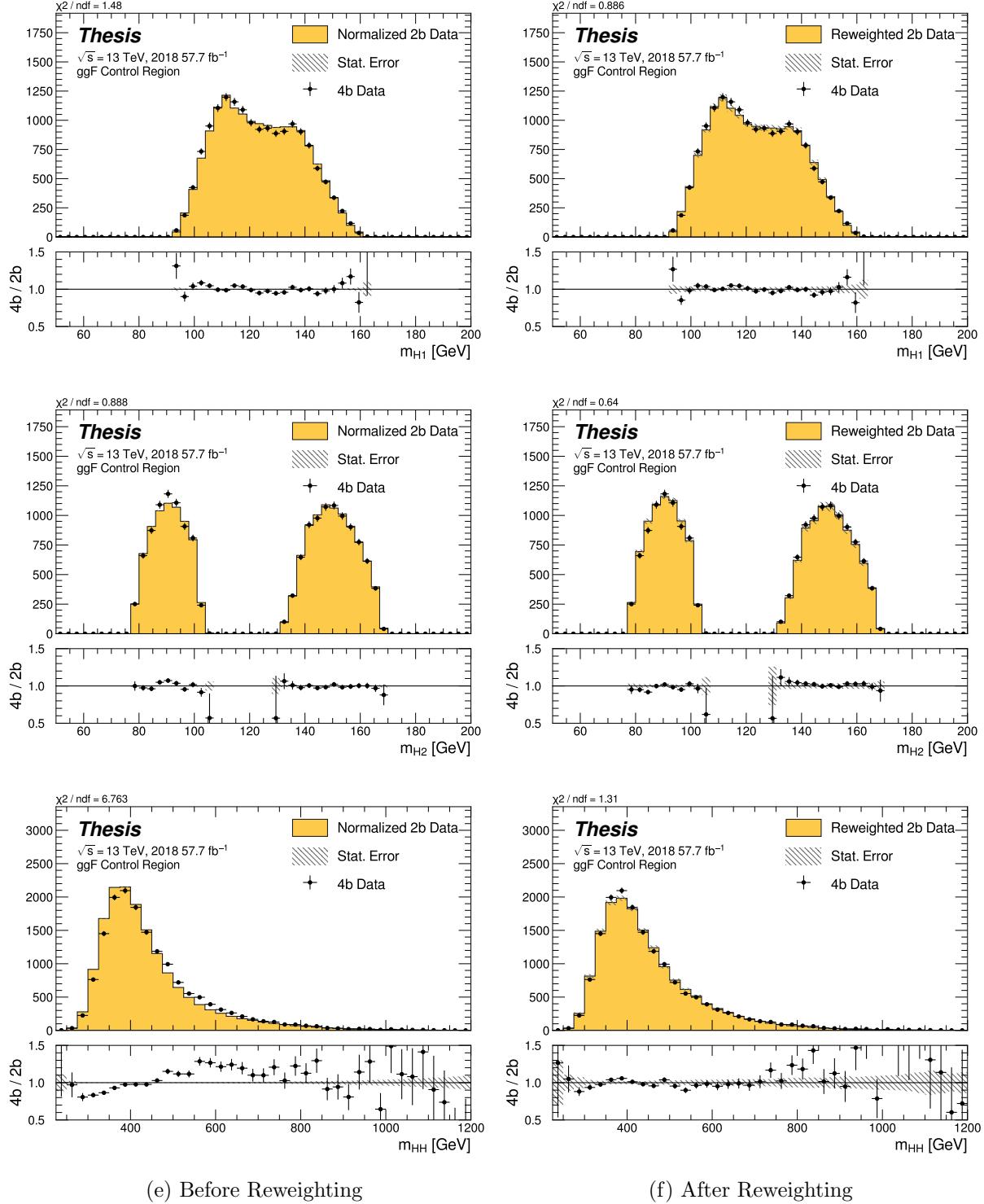


Figure 8.25: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

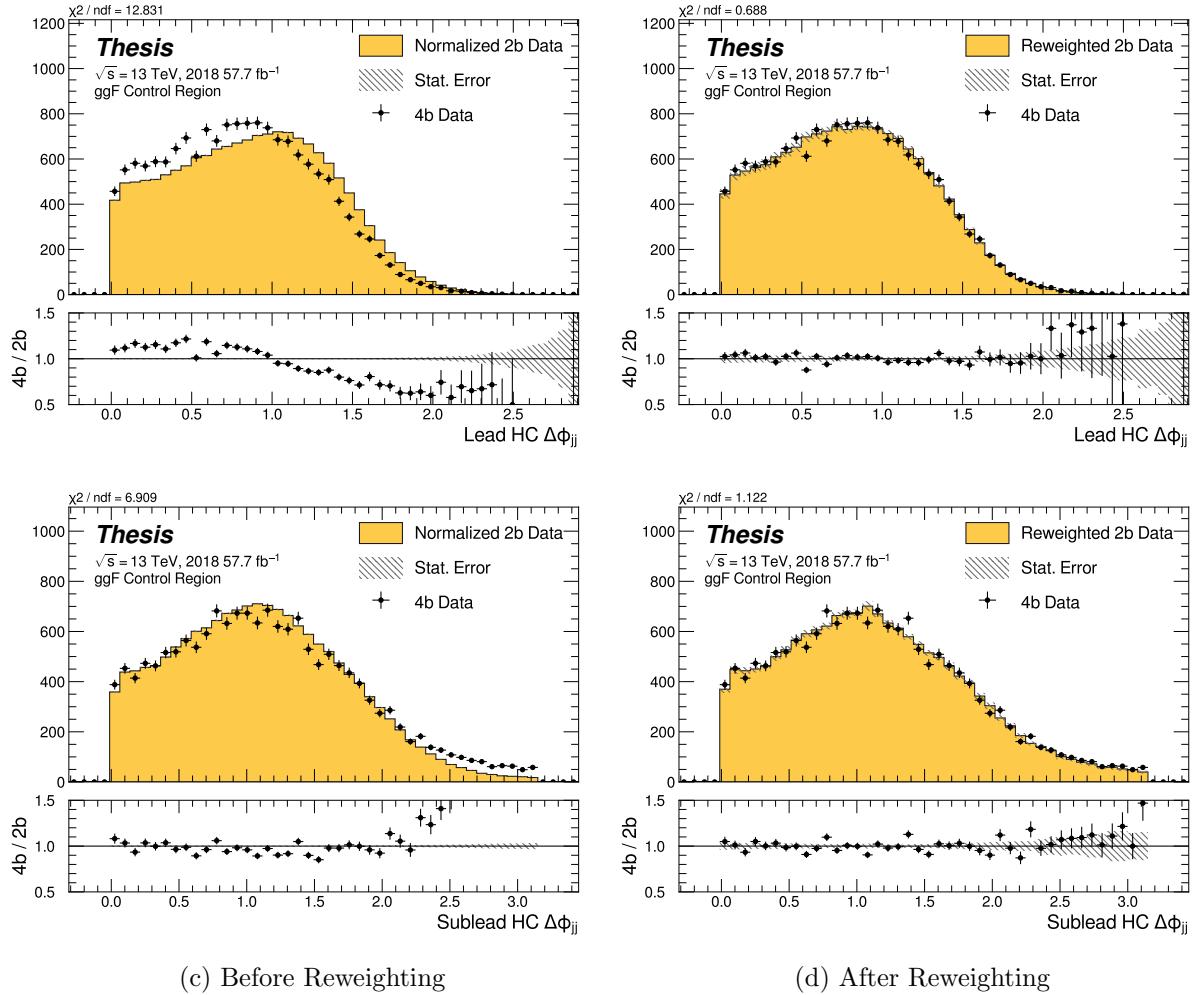


Figure 8.26: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Control Region.

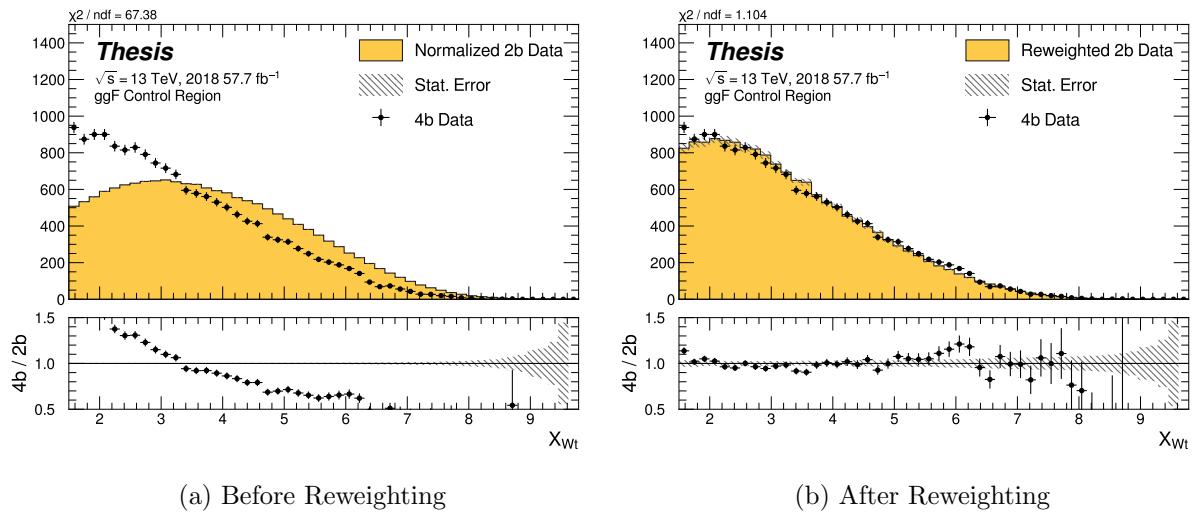


Figure 8.27: **Non-resonant Search (4b)**: Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 4b Control Region. Reweighting is done after the cut on this variable is applied.

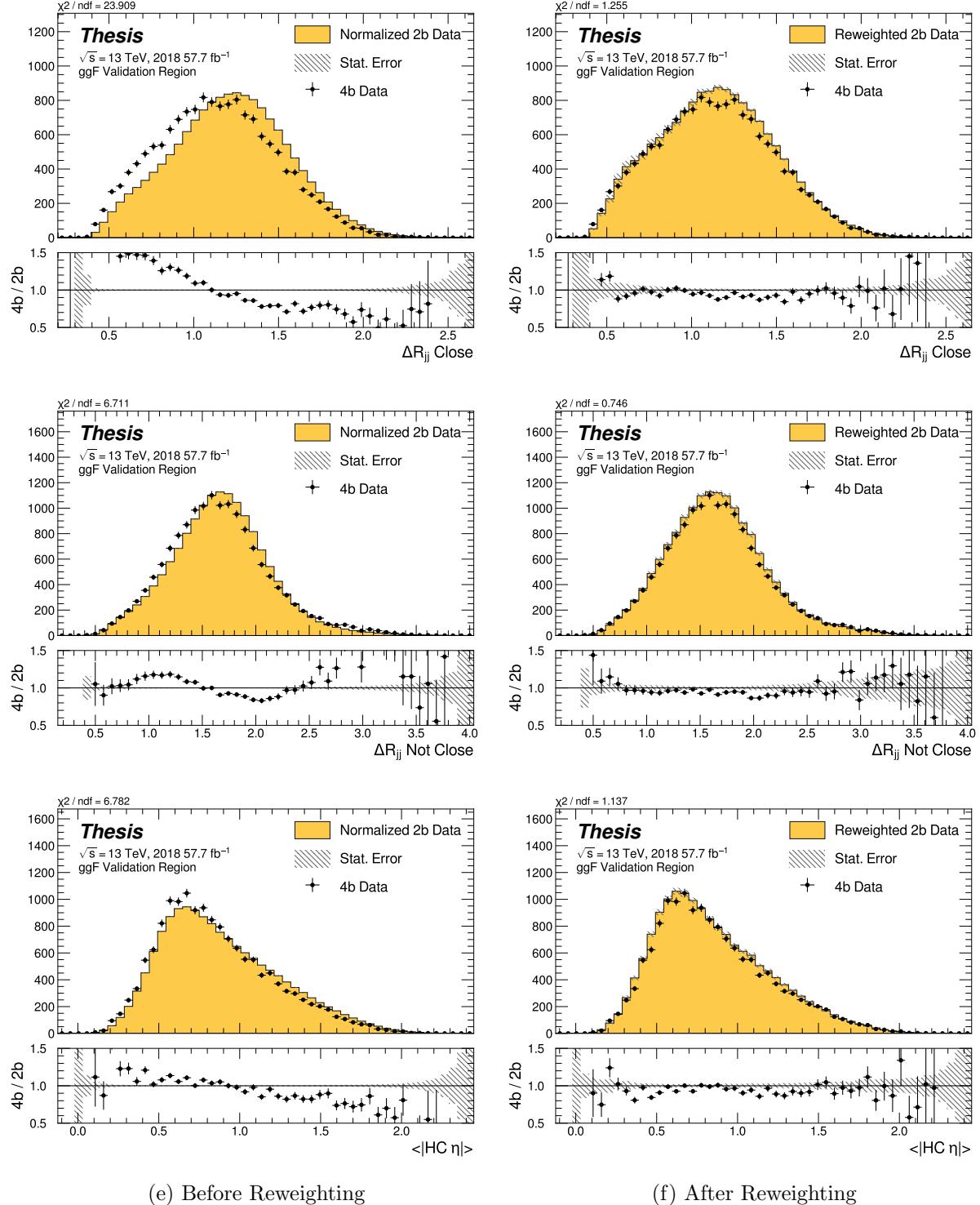


Figure 8.28: **Non-resonant Search (4b):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

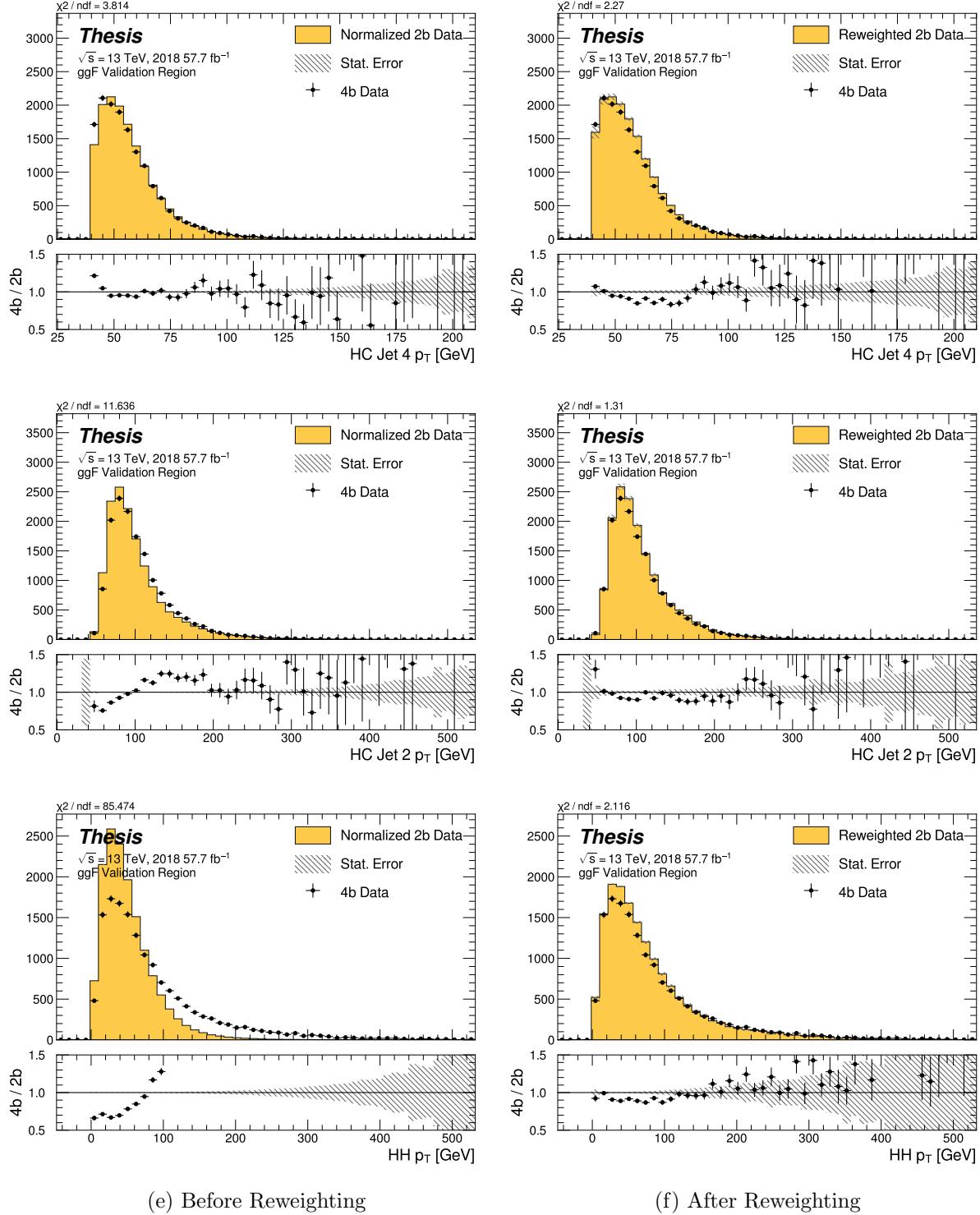


Figure 8.29: **Non-resonant Search (4b):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

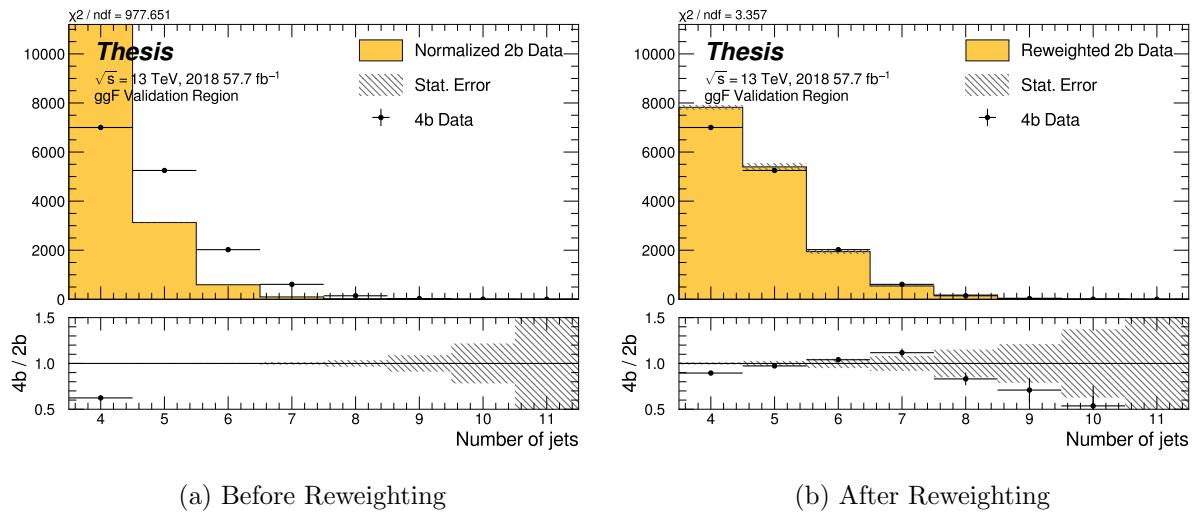


Figure 8.30: **Non-resonant Search (4b)**: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

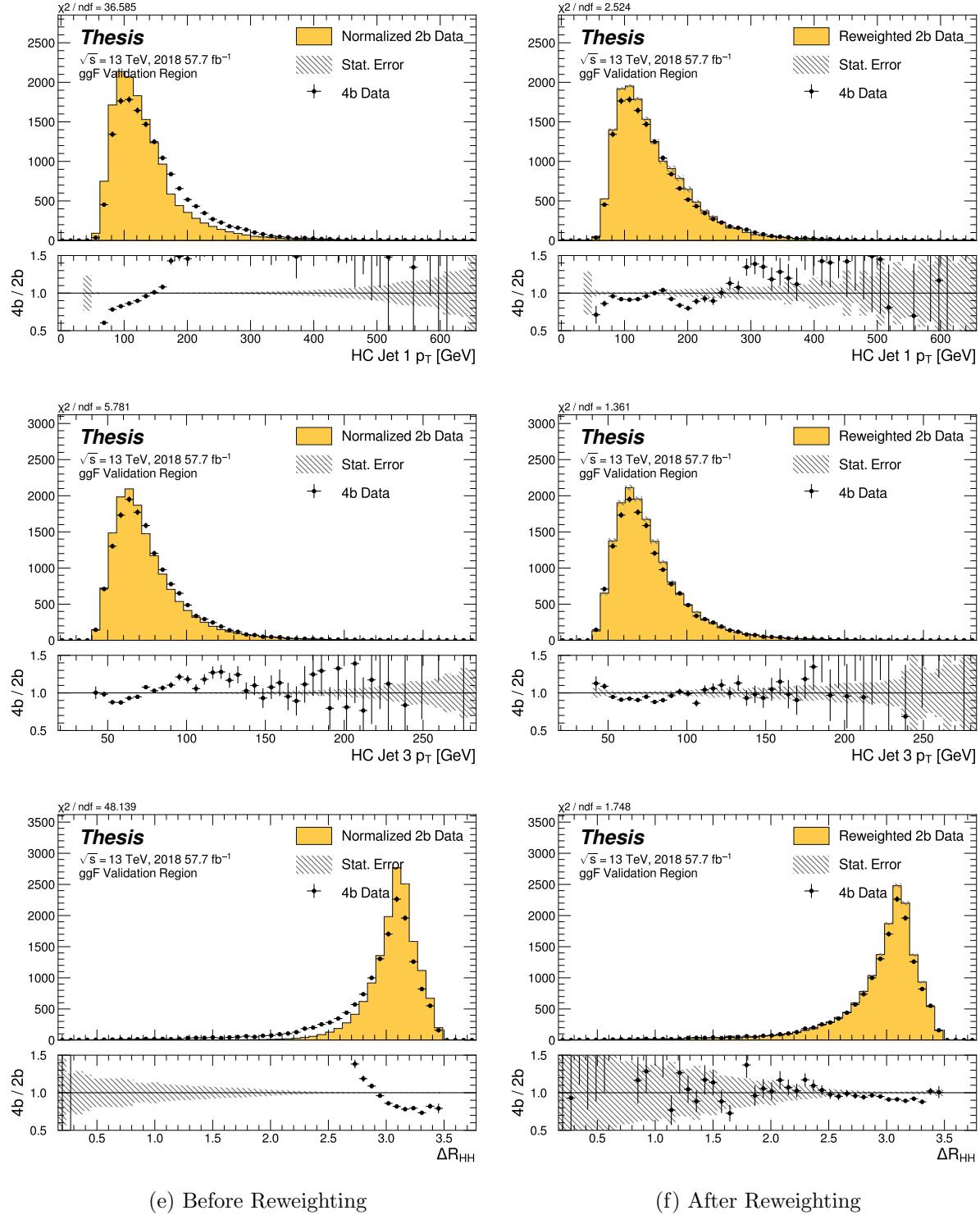


Figure 8.31: **Non-resonant Search (4b):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

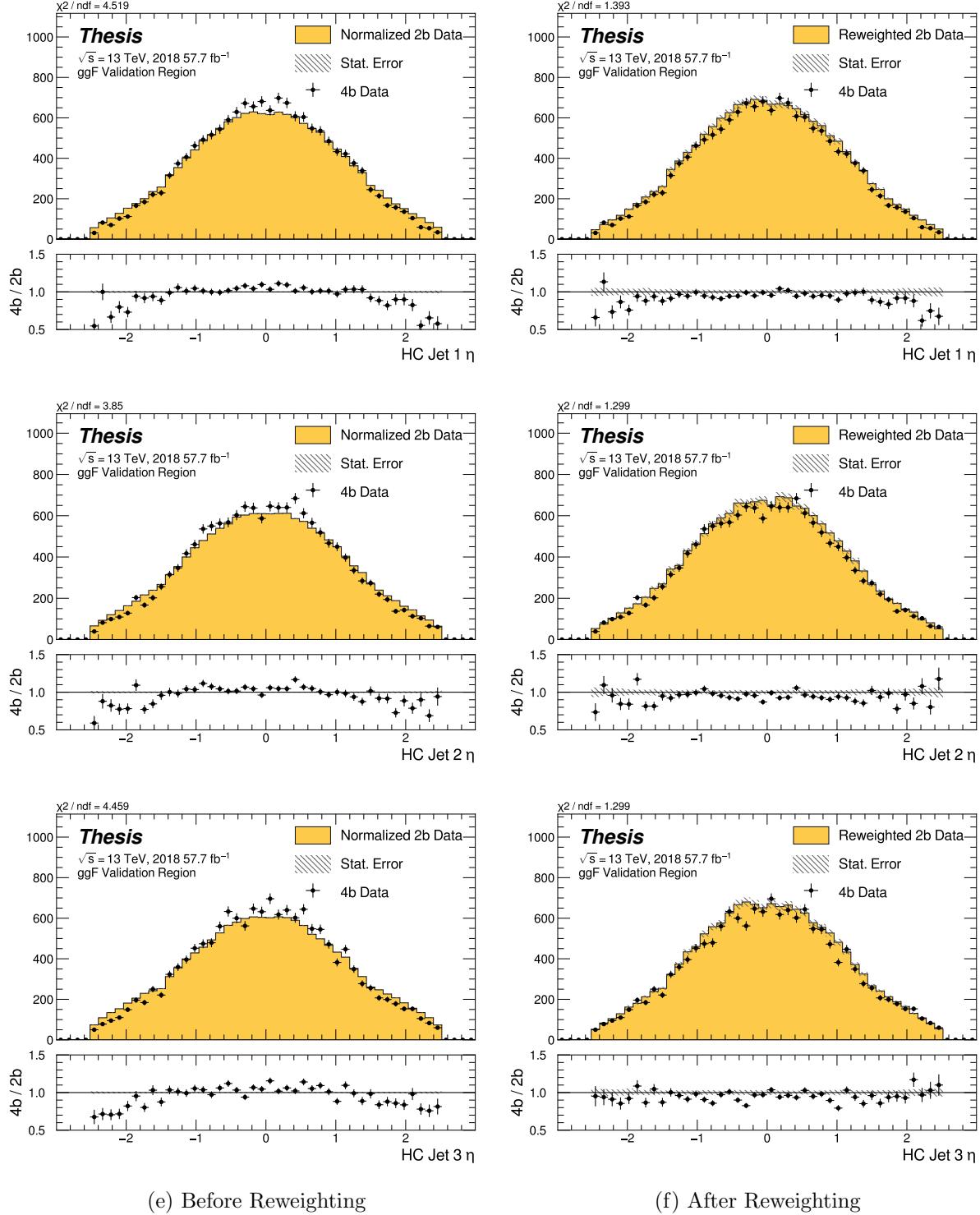


Figure 8.32: **Non-resonant Search (4b):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

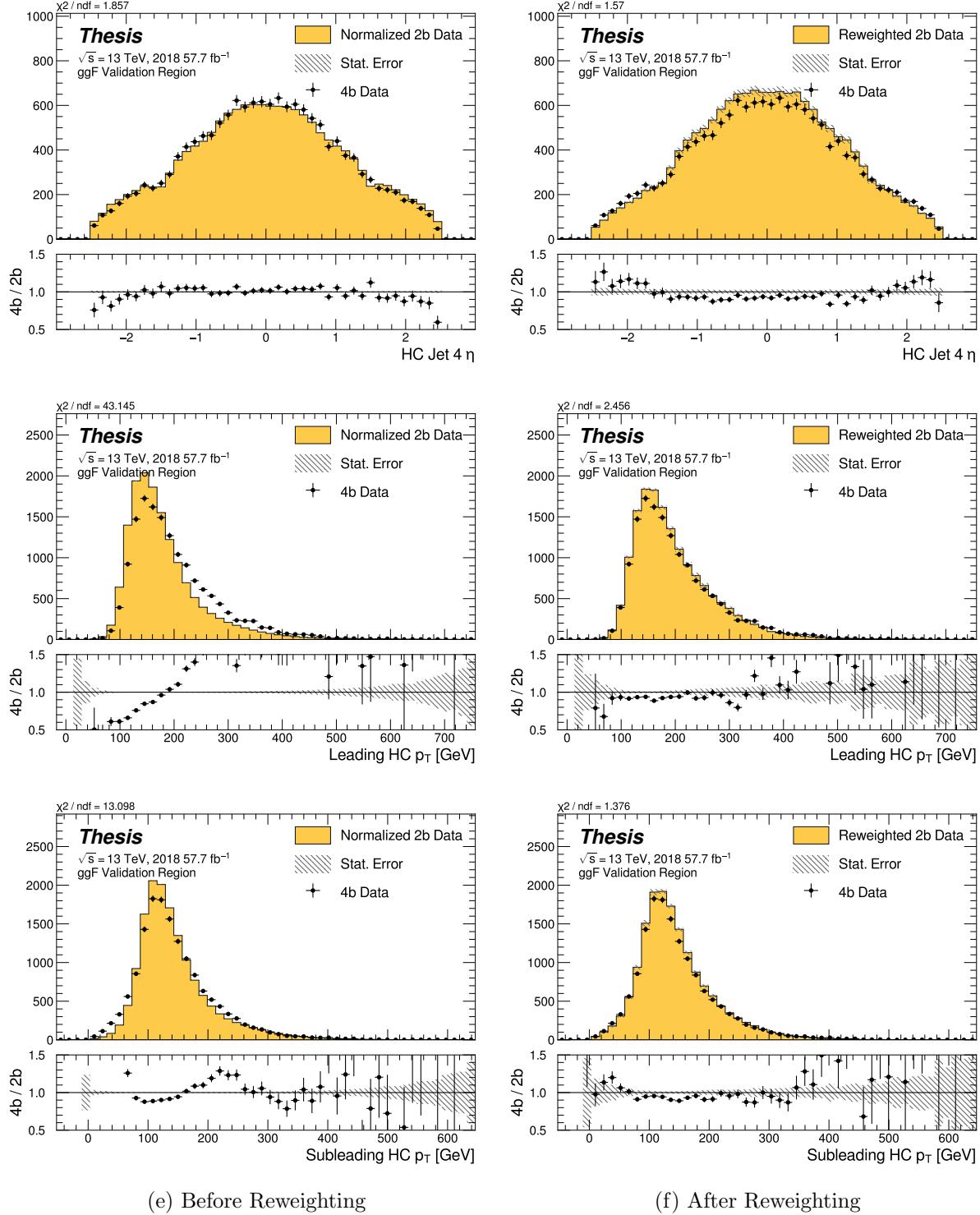


Figure 8.33: **Non-resonant Search (4b):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

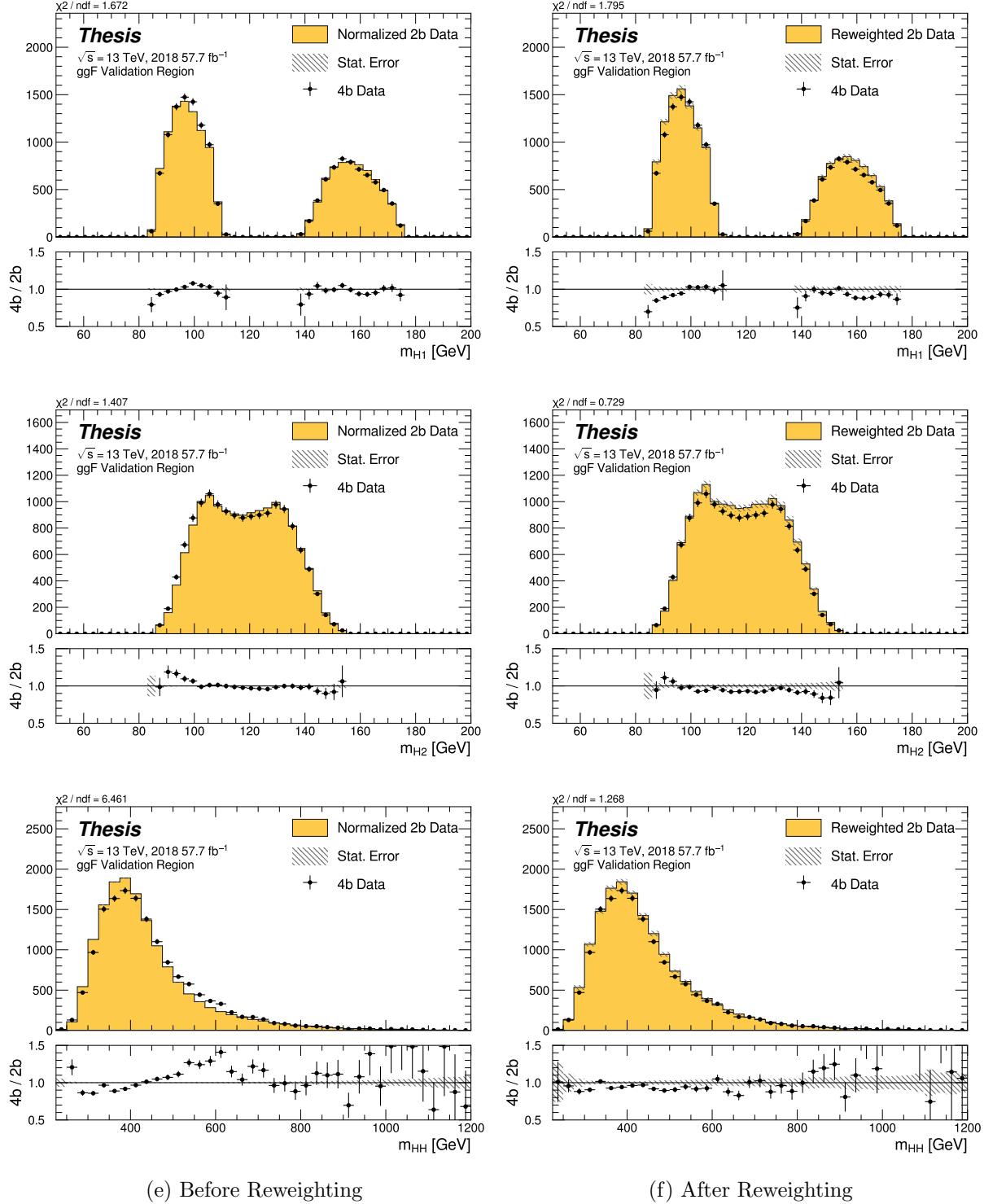


Figure 8.34: **Non-resonant Search (4b):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

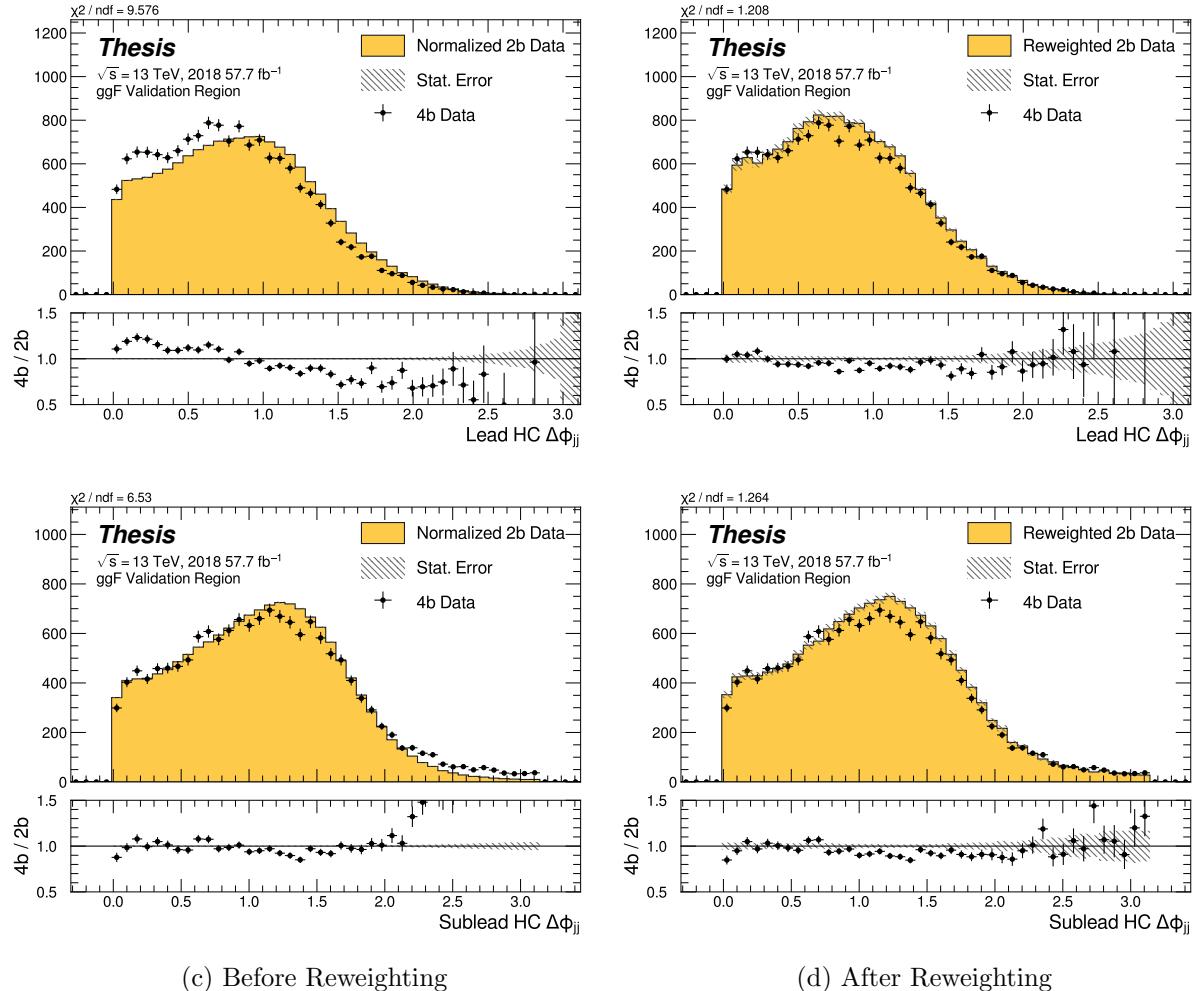


Figure 8.35: **Non-resonant Search (4b):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region.

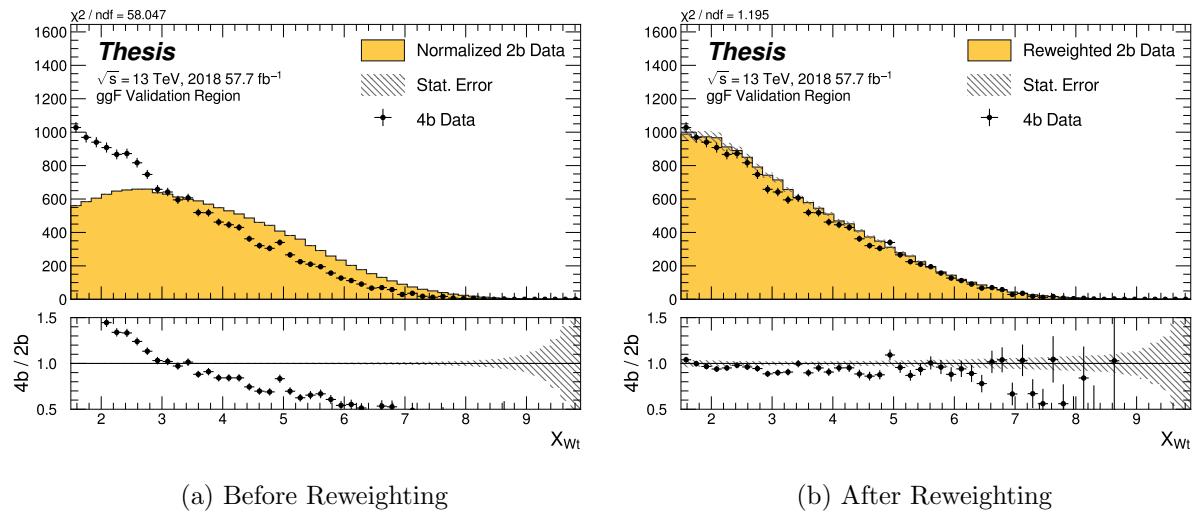


Figure 8.36: **Non-resonant Search (4b):** Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 4b Validation Region. Reweighting is done after the cut on this variable is applied.

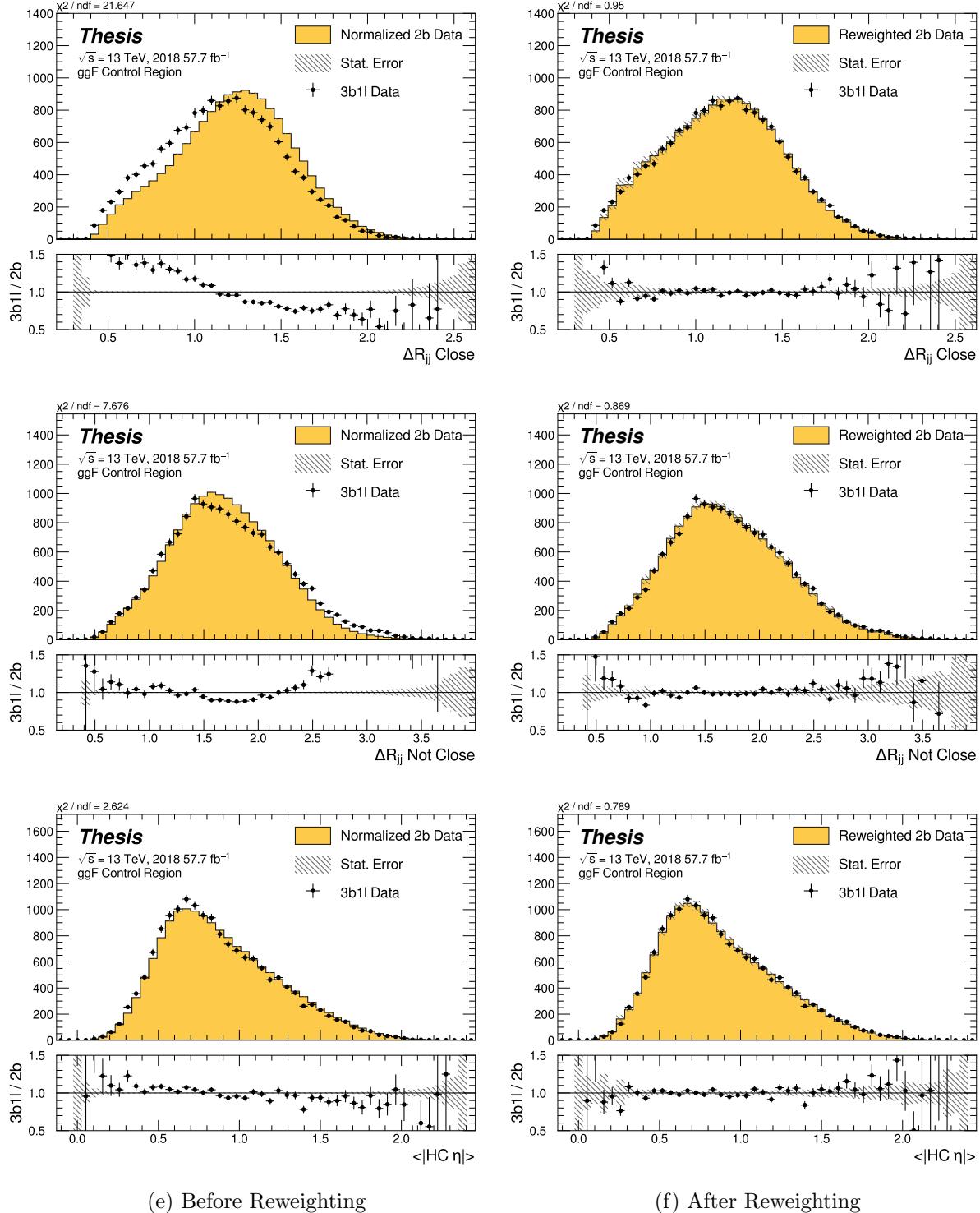


Figure 8.37: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

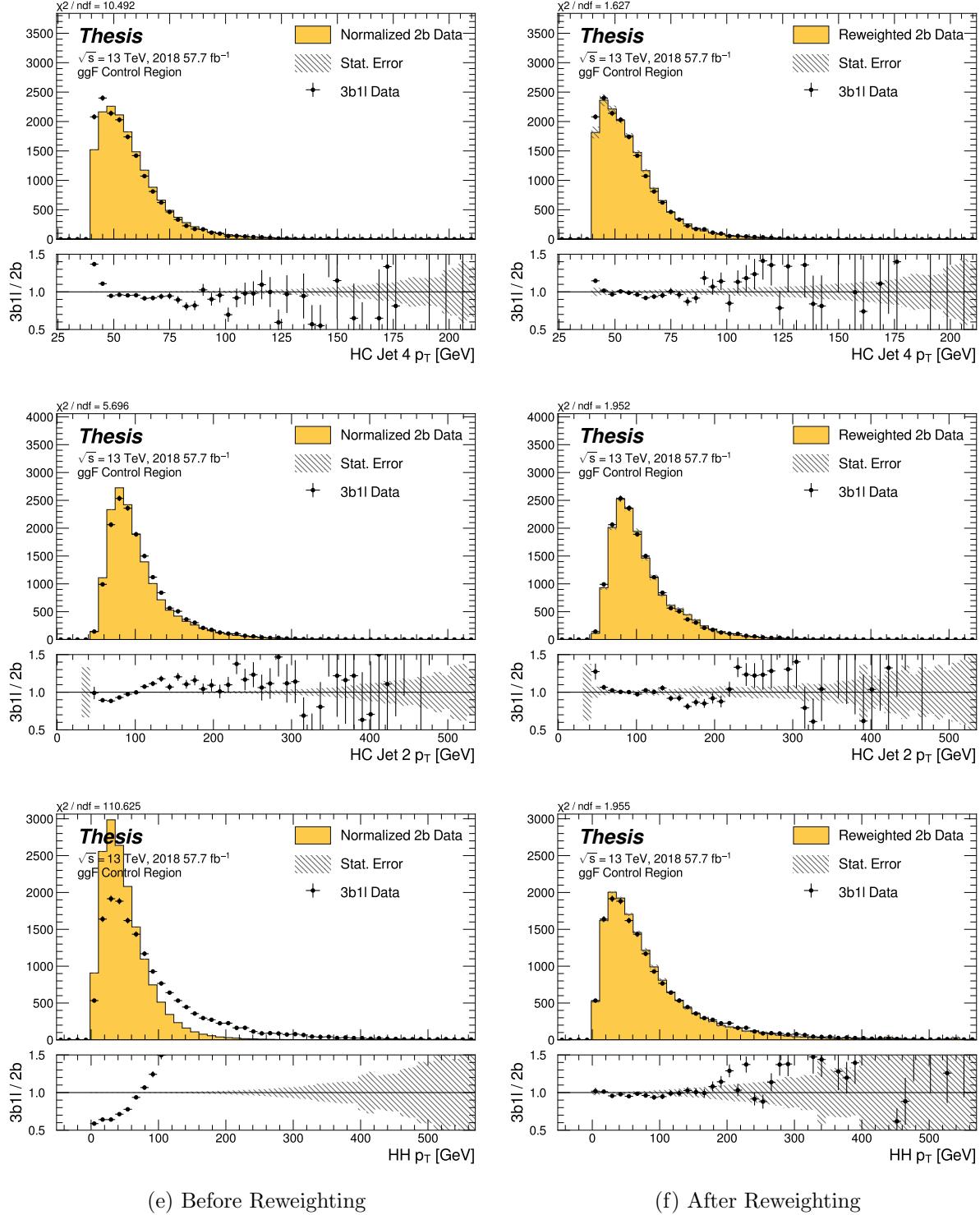


Figure 8.38: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

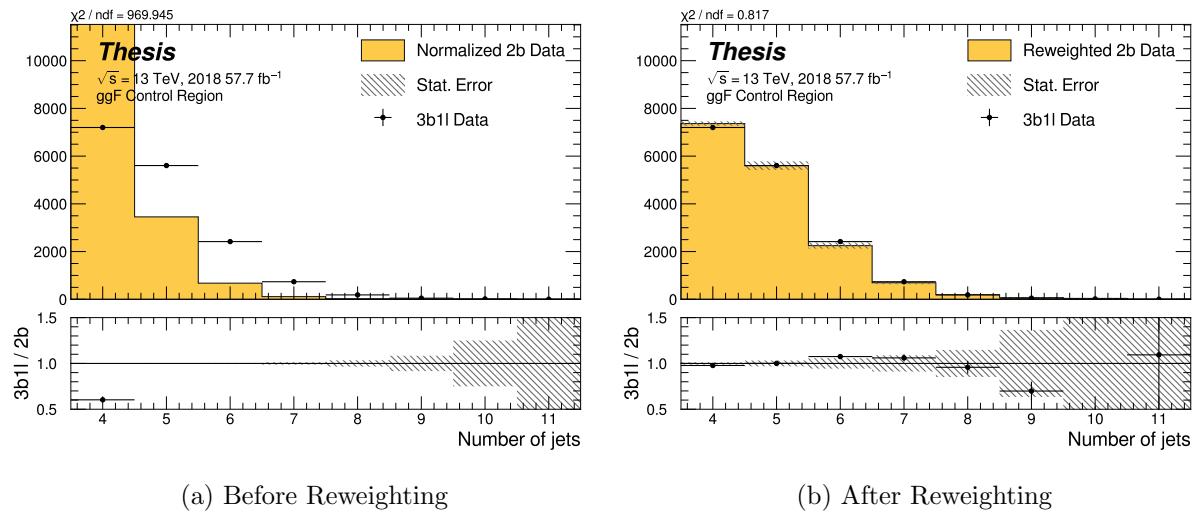


Figure 8.39: **Non-resonant Search (3b1l)**: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

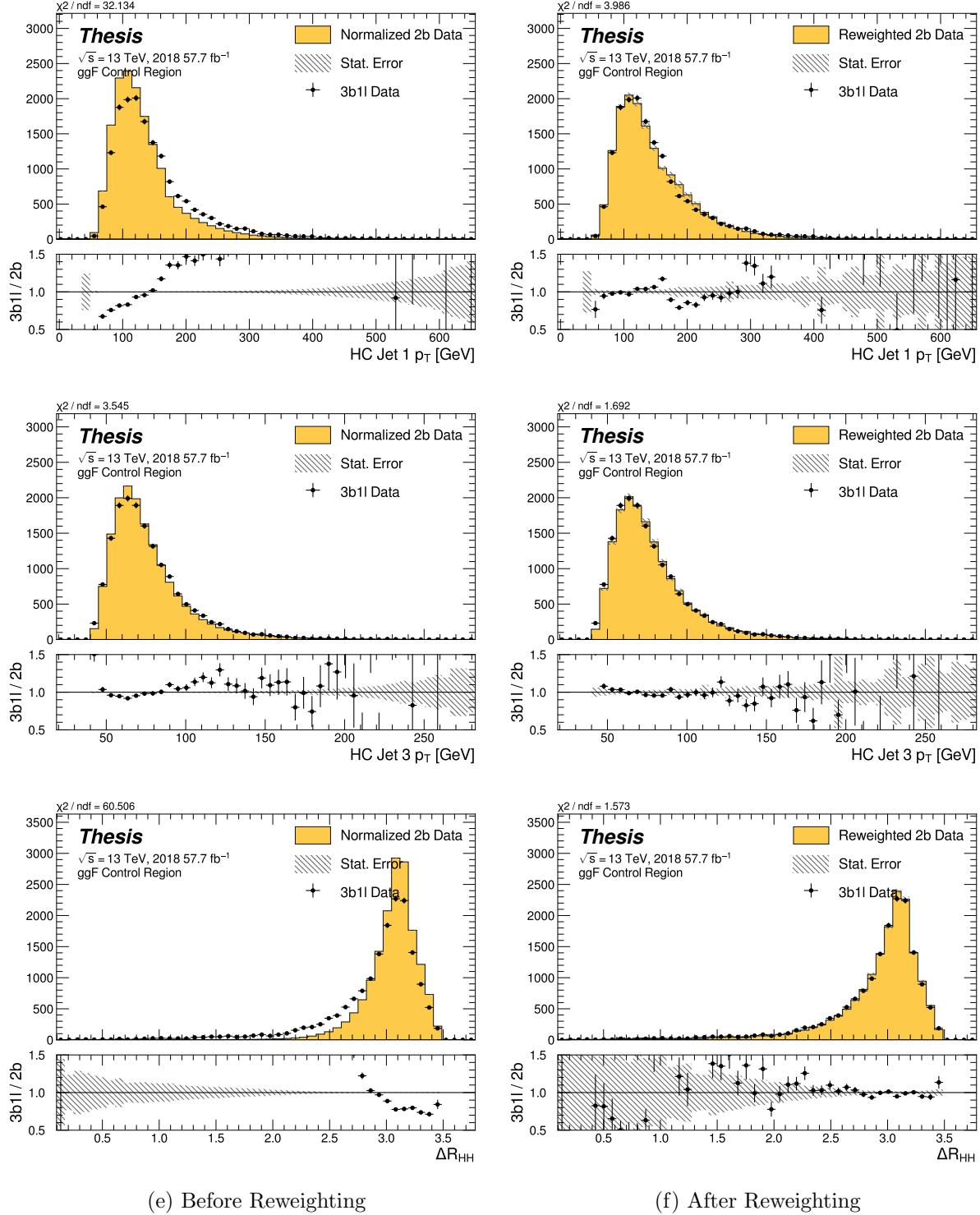


Figure 8.40: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

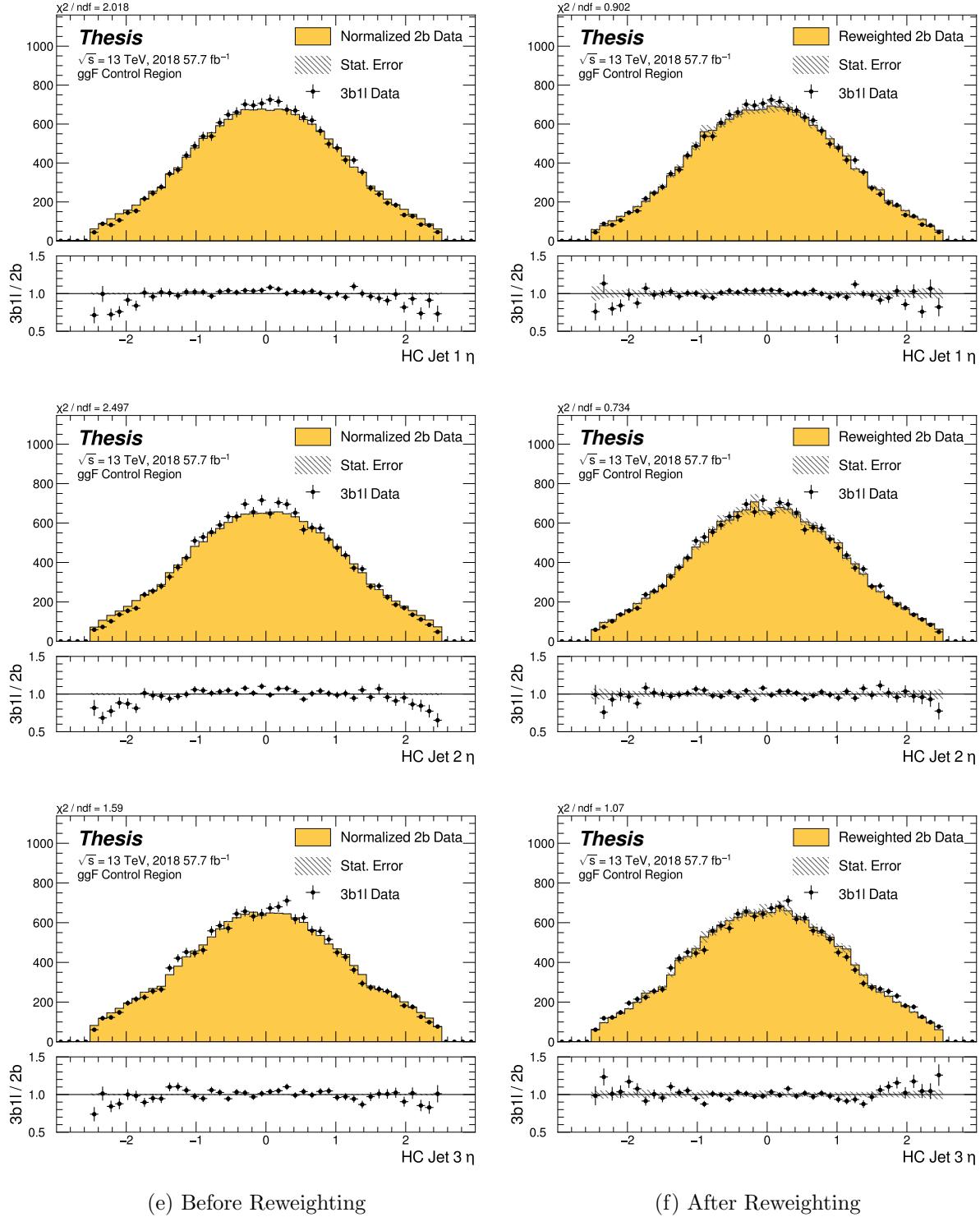


Figure 8.41: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

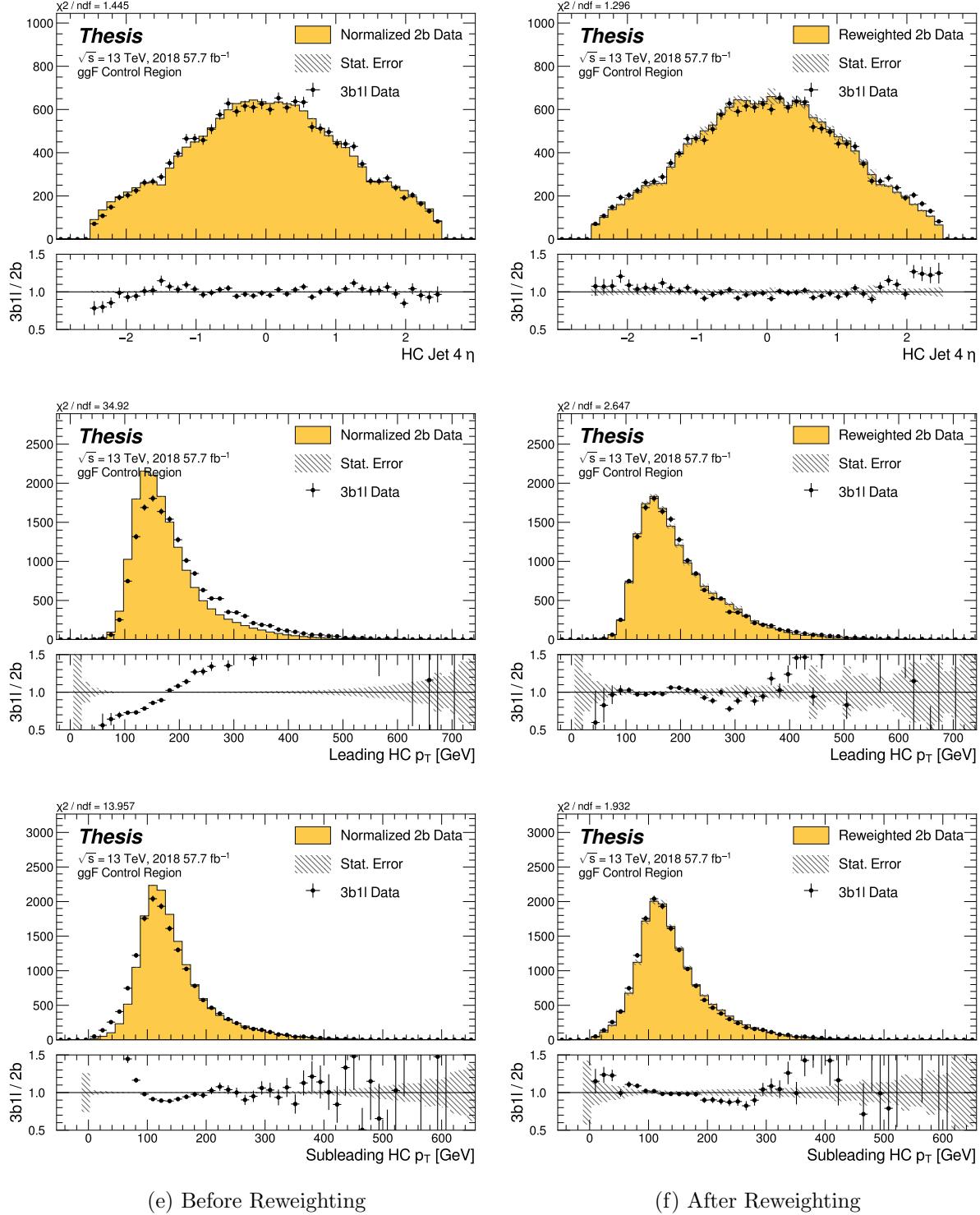


Figure 8.42: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

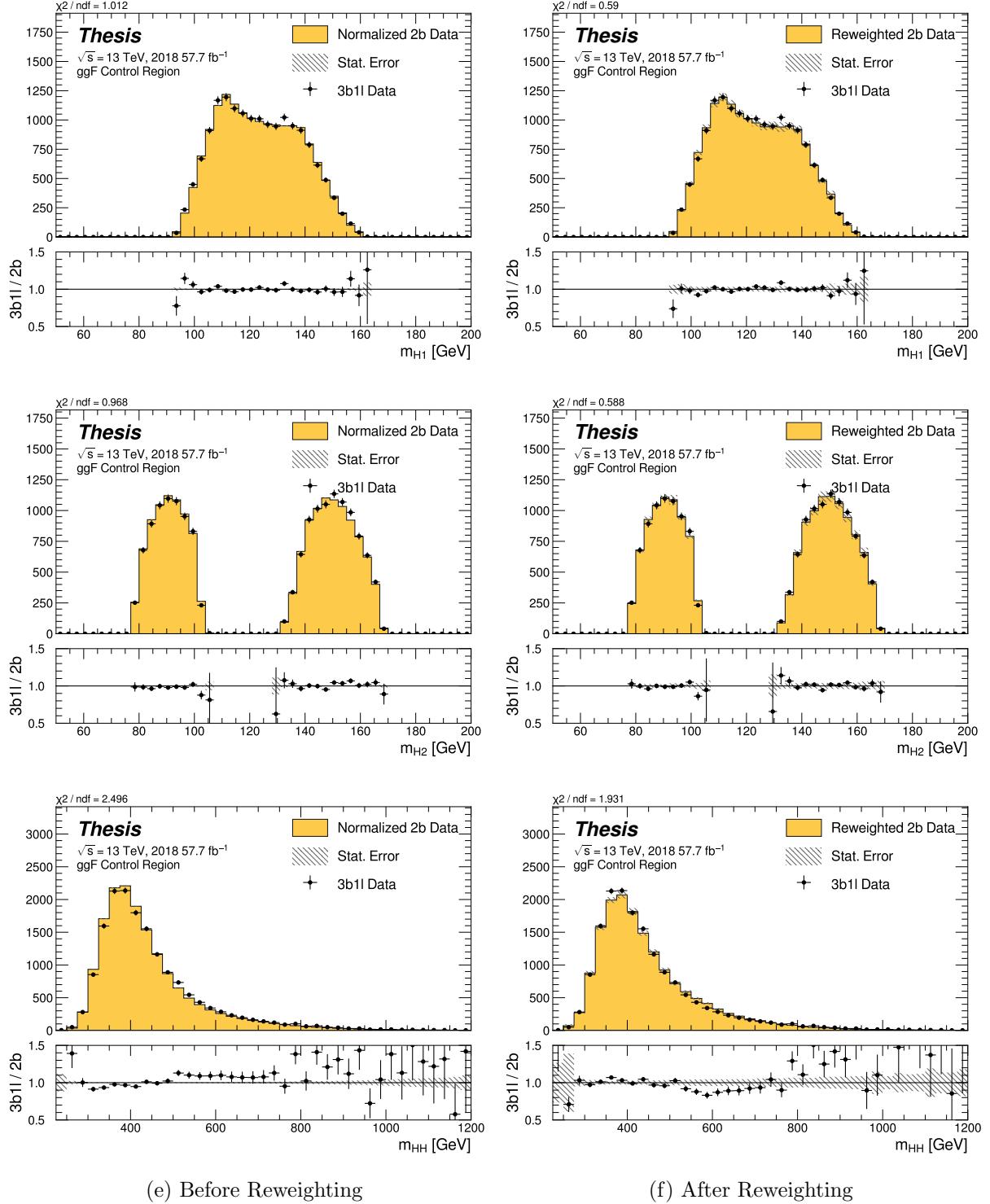


Figure 8.43: **Non-resonant Search (3b1l):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

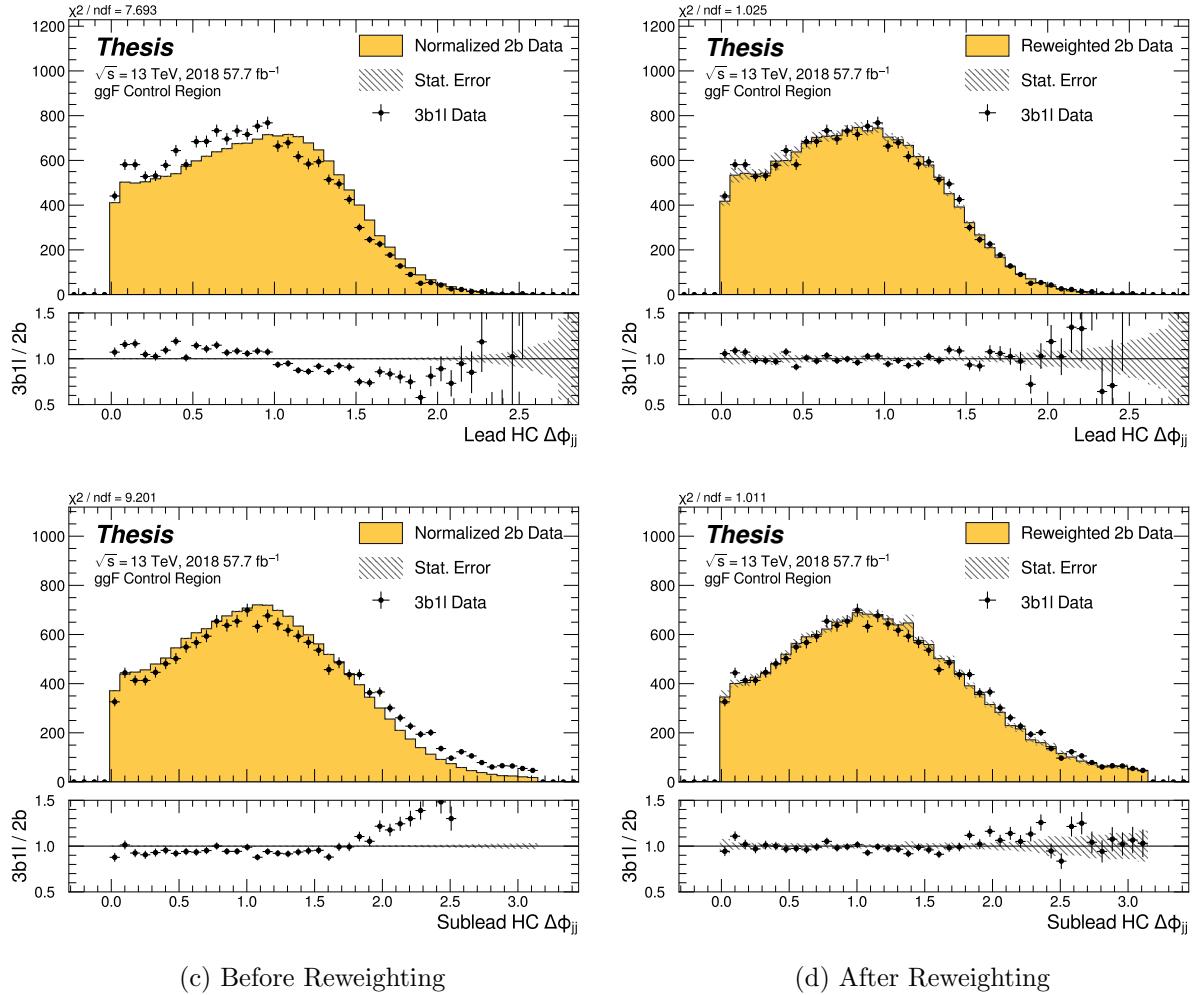


Figure 8.44: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region.

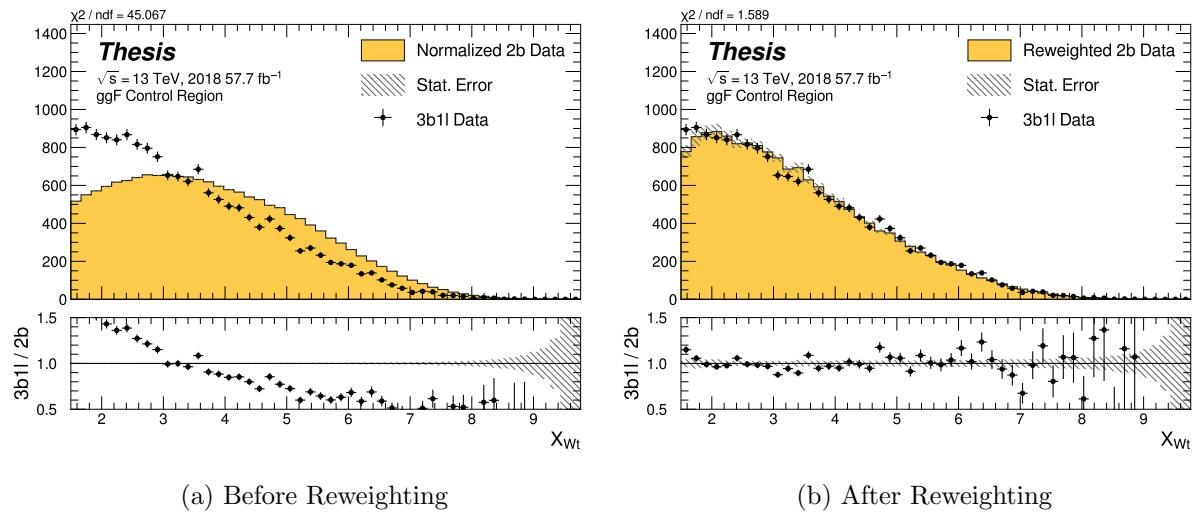


Figure 8.45: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 3b1l Control Region. Reweighting is done after the cut on this variable is applied.

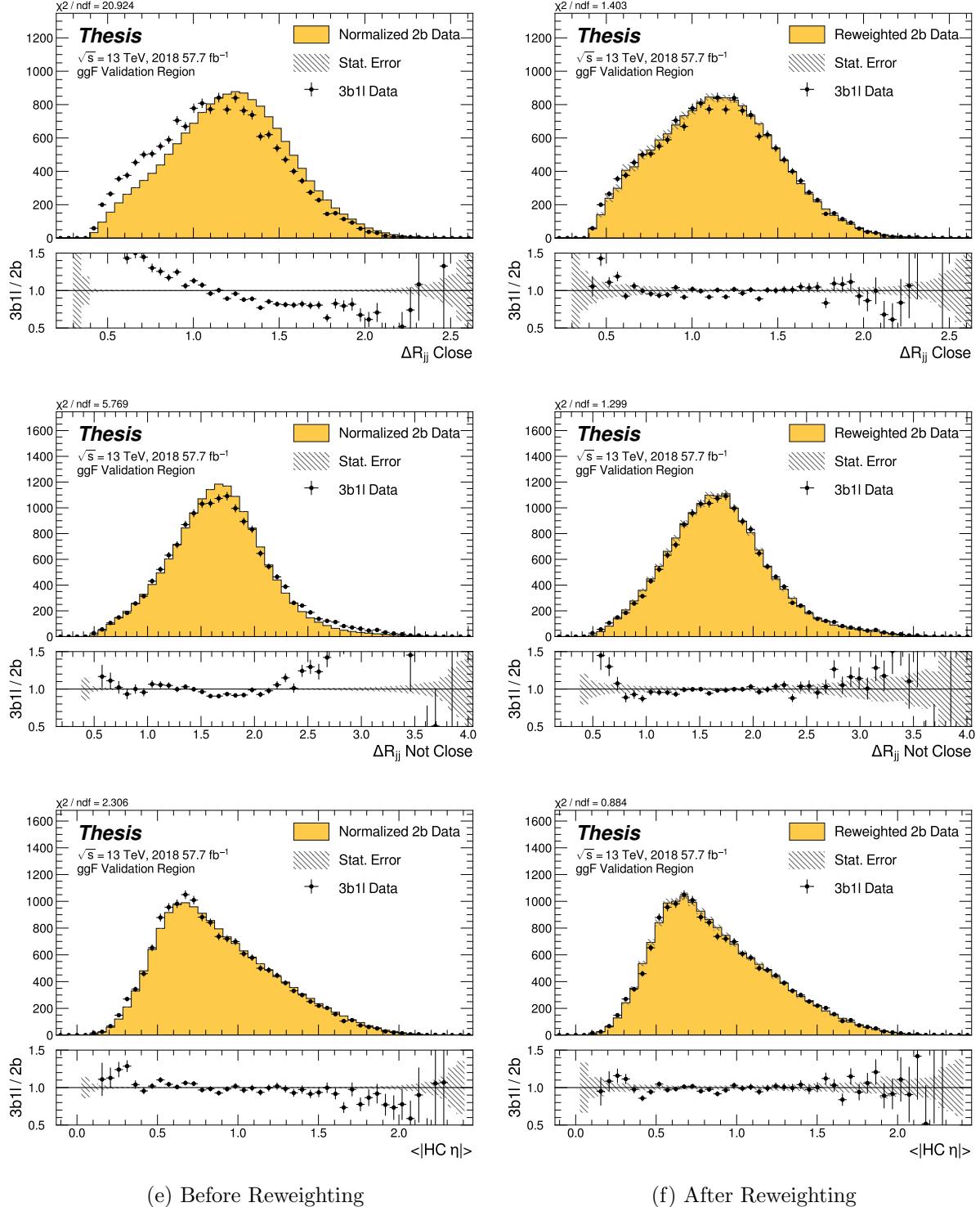


Figure 8.46: **Non-resonant Search (3b1l):** Distributions of ΔR between the closest Higgs Candidate jets, ΔR between the other two, and average absolute value of HC jet η before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

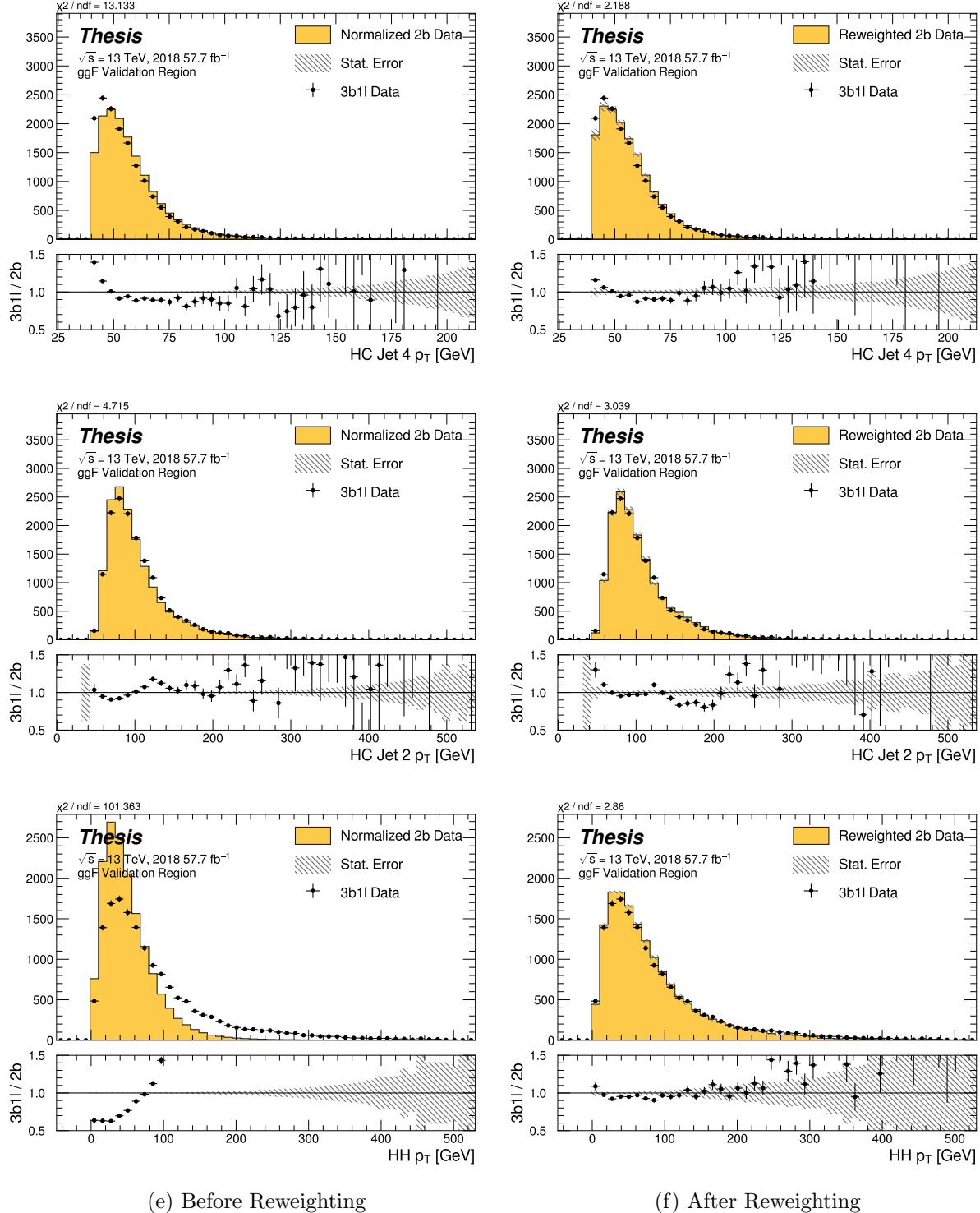


Figure 8.47: **Non-resonant Search (3b1l):** Distributions of p_T of the 2nd and 4th leading Higgs Candidate jets and the p_T of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

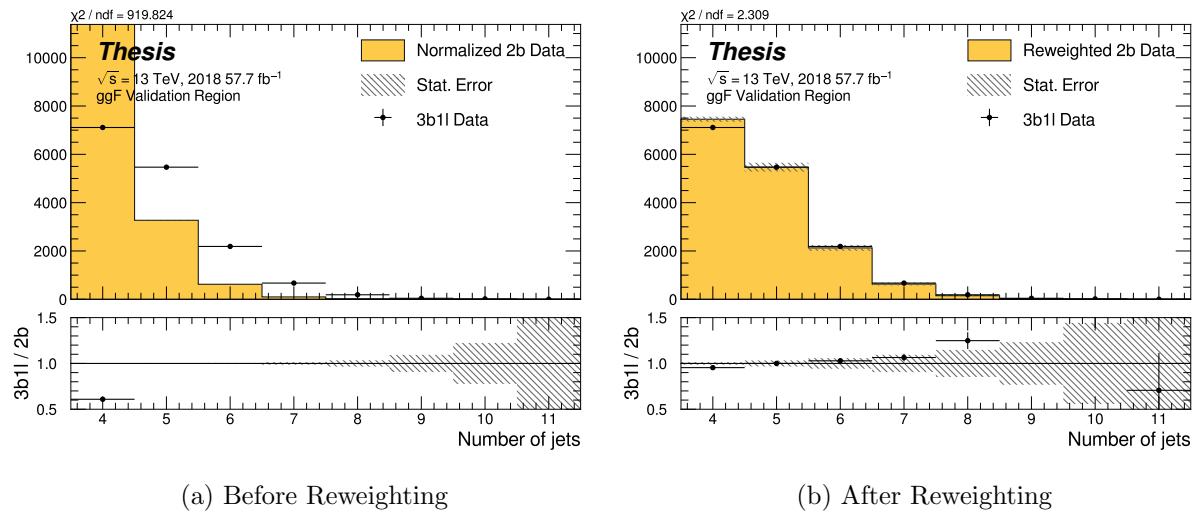


Figure 8.48: **Non-resonant Search (3b1l)**: Distributions of the number of jets before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. A minimum of 4 jets is required in each event in order to form Higgs candidates.

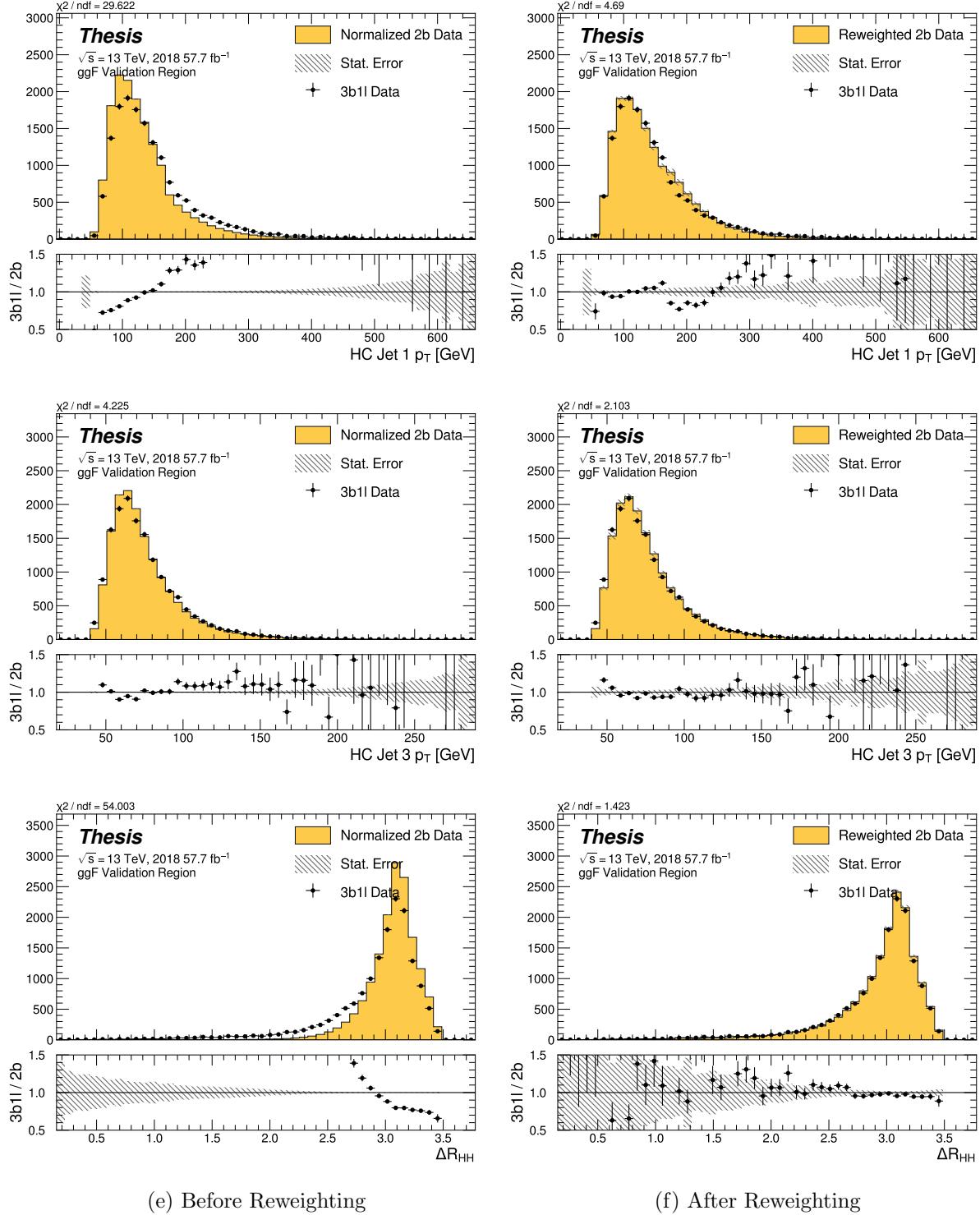


Figure 8.49: **Non-resonant Search (3b1l):** Distributions of p_T of the 1st and 3rd leading Higgs Candidate jets and ΔR between Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

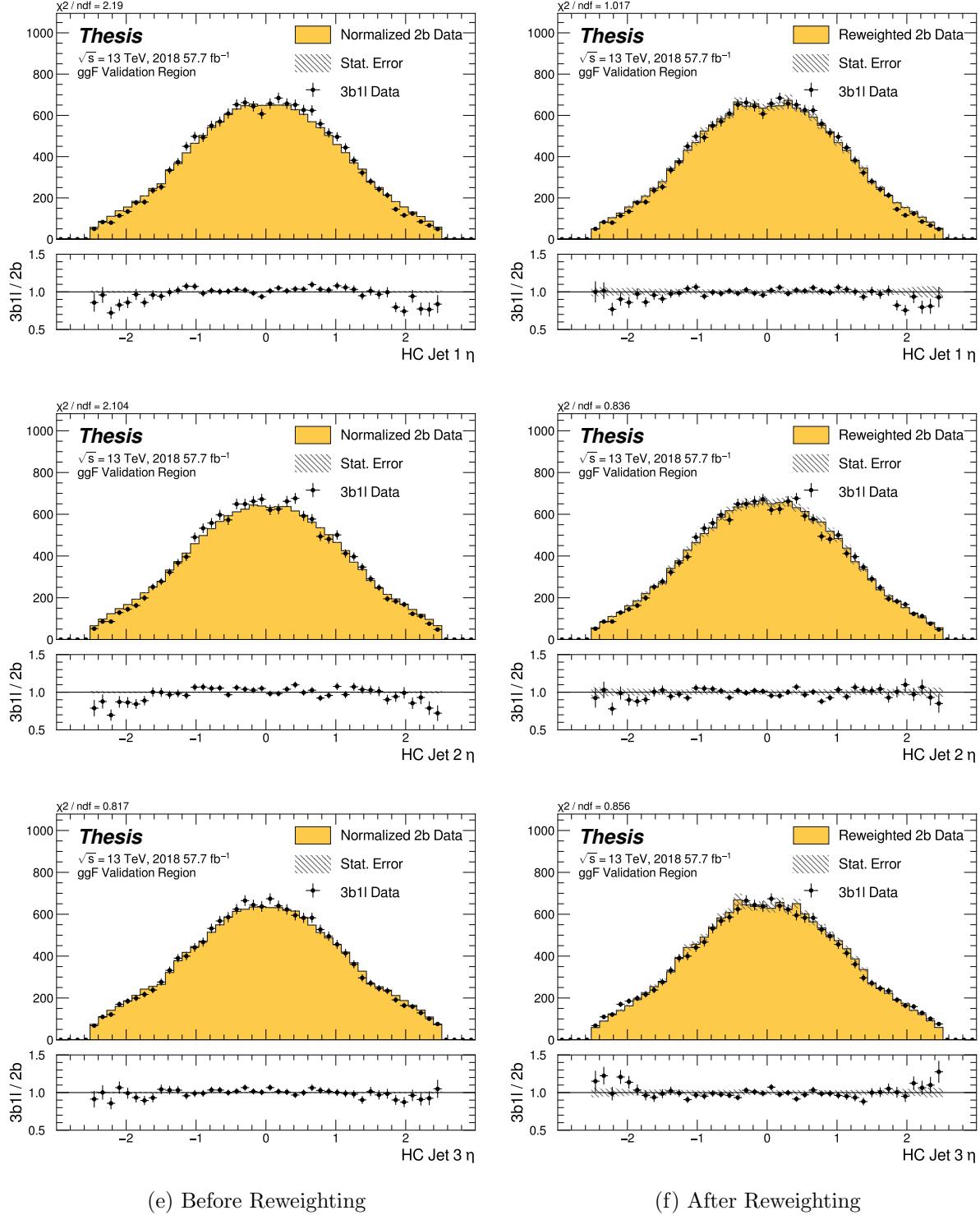


Figure 8.50: **Non-resonant Search (3b1l):** Distributions of η of the 1st, 2nd, and 3rd leading Higgs Candidate jets before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

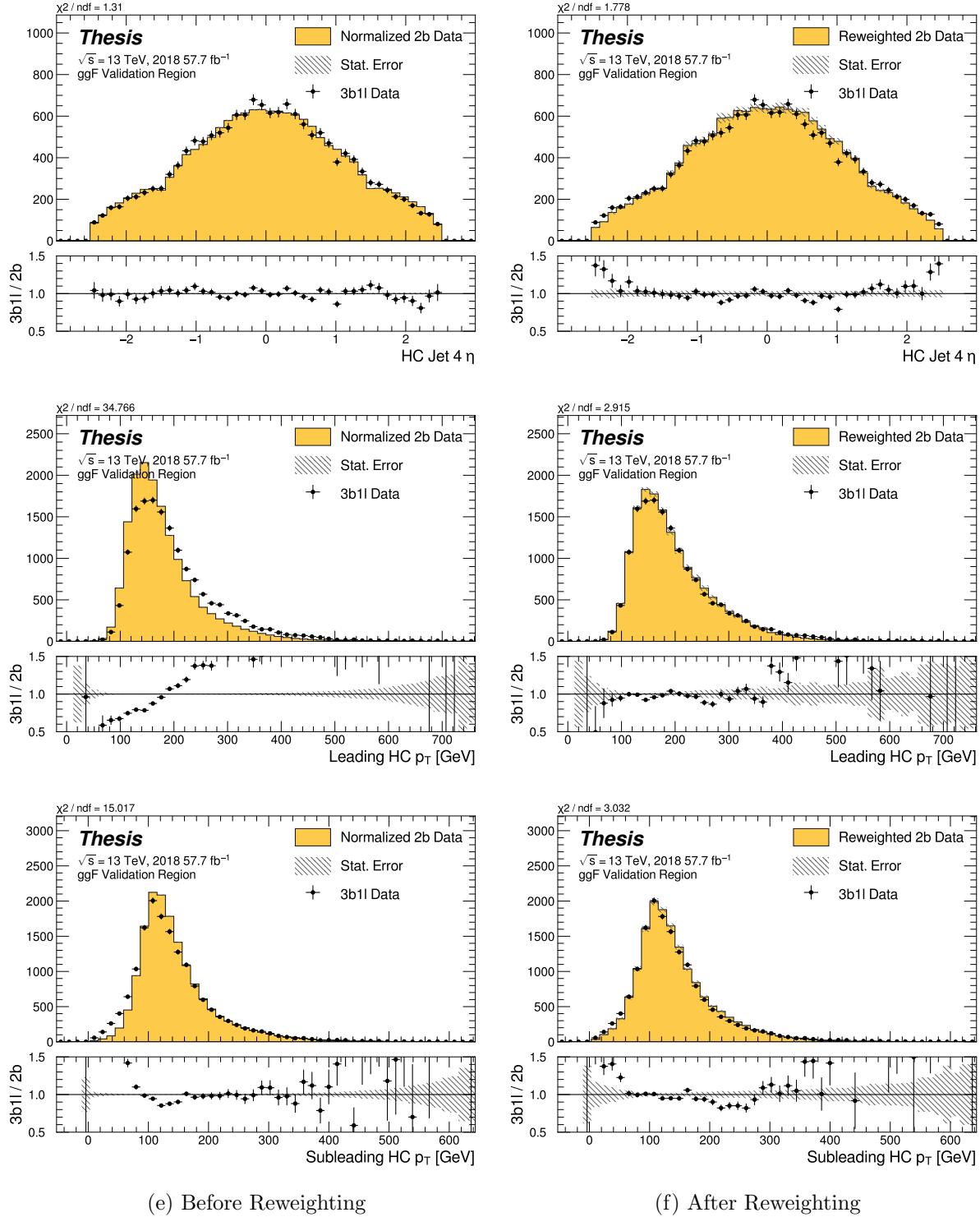


Figure 8.51: **Non-resonant Search (3b1l):** Distributions of η of the 4th leading Higgs Candidate jet and the p_T of the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

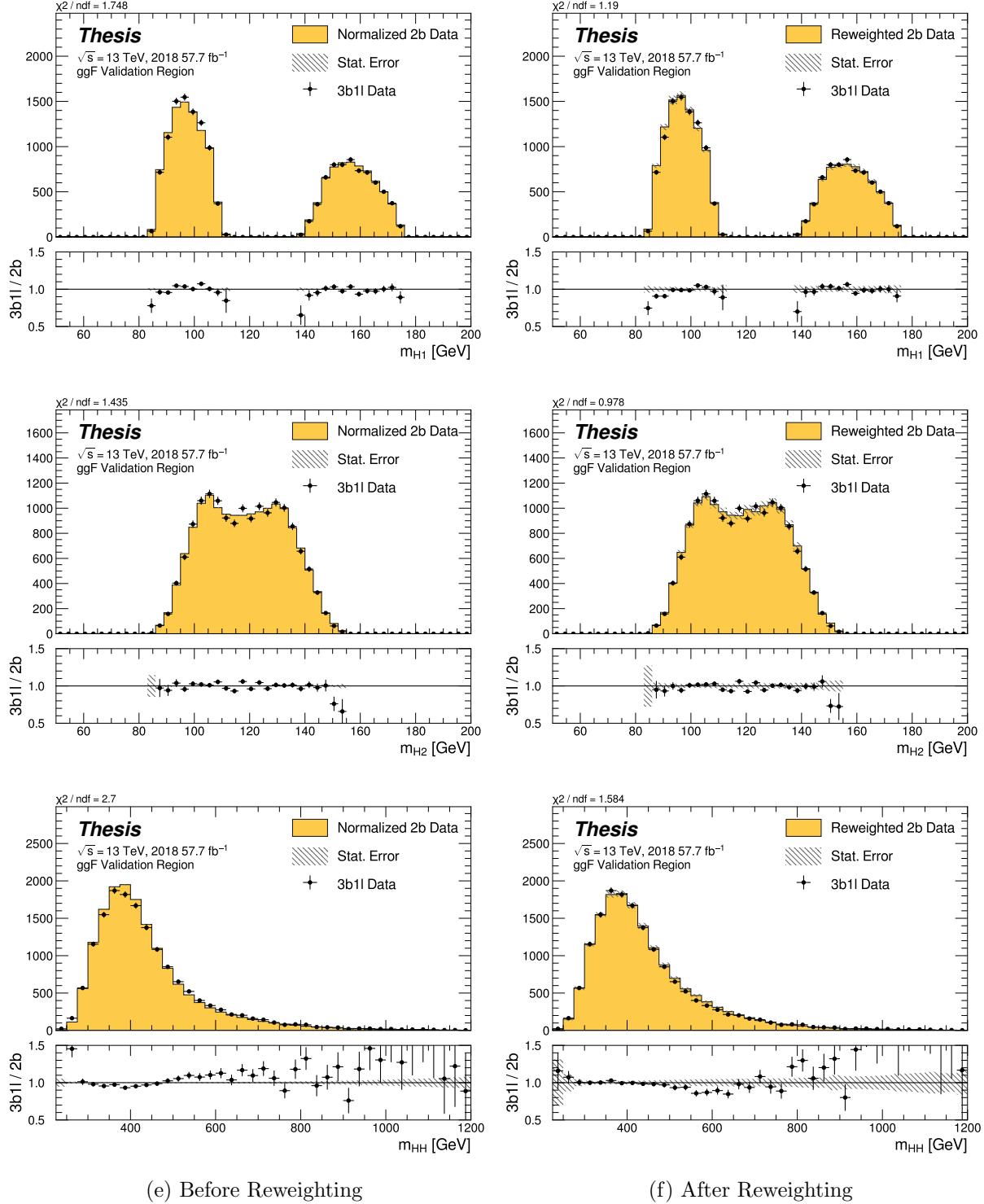


Figure 8.52: **Non-resonant Search (3b1l):** Distributions of mass of the leading and subleading Higgs candidates and of the di-Higgs system before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

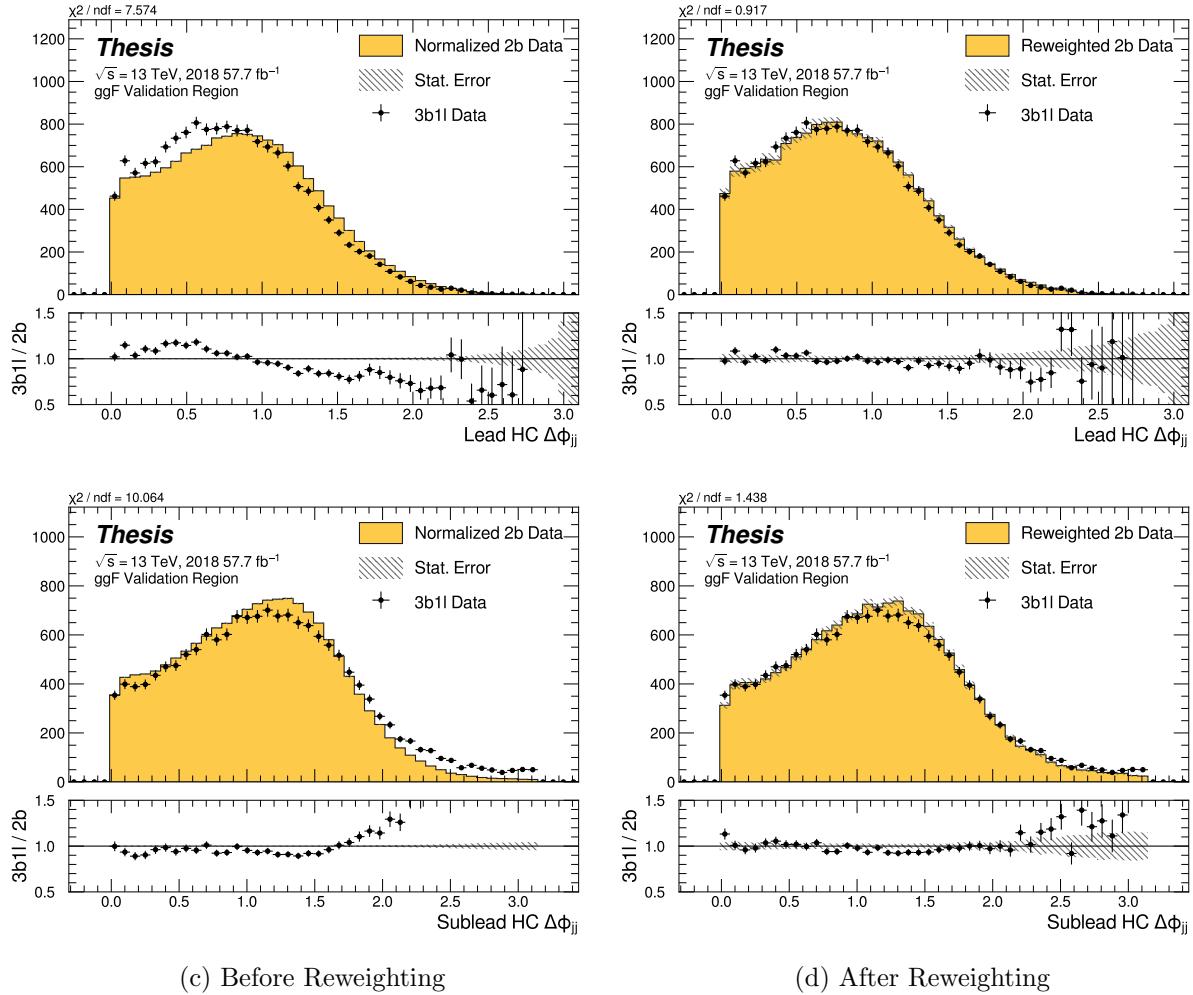


Figure 8.53: **Non-resonant Search (3b1l):** Distributions of $\Delta\phi$ between jets in the leading and subleading Higgs candidates before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region.

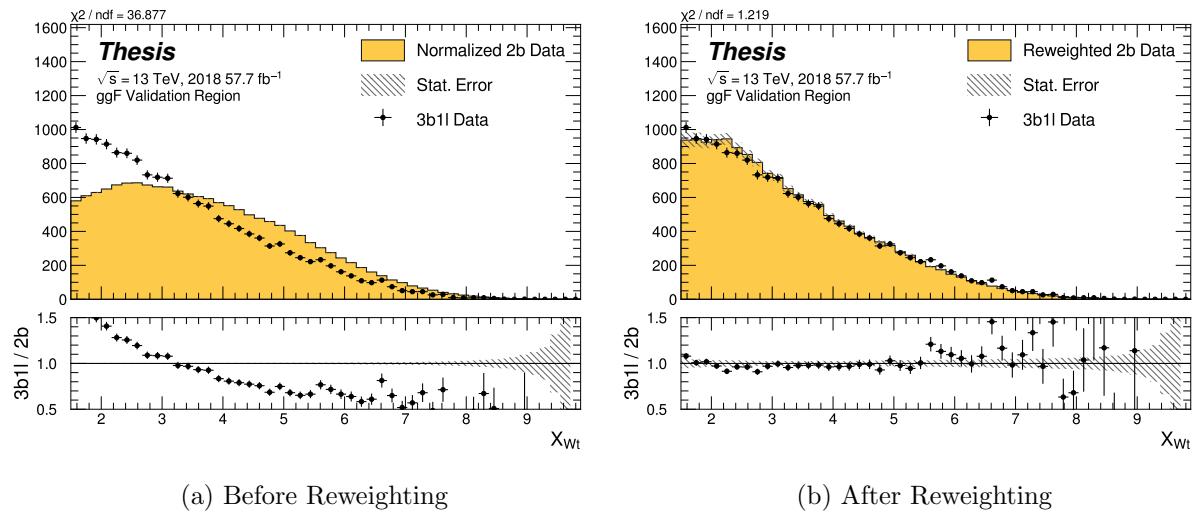


Figure 8.54: **Non-resonant Search (3b1l):** Distributions of the top veto variable, X_{Wt} , before (left) and after (right) CR derived reweighting for the 2018 3b1l Validation Region. Reweighting is done after the cut on this variable is applied.

2304

Chapter 9

2305

UNCERTAINTIES AND VALIDATION

2306 A variety of uncertainties are assigned to account for known biases in the underlying
 2307 methods, calibrations, and objects used for this analysis. The largest such uncertainty is
 2308 associated with the kinematic bias inherent in deriving the background estimate outside of
 2309 the signal region. However, a statistical biasing of this same estimate also has a significant
 2310 impact. Additionally, due to the use of Monte Carlo for signal modelling and b -tagging
 2311 calibration, uncertainties related to mis-modelings in simulation must also be accounted for.
 2312 Note that the results for the non-resonant analysis presented here are preliminary and only
 2313 include background systematic, such that the discussion of the signal systematics *only* applies
 2314 for the resonant search. However, these background systematics are expected to be by far
 2315 the dominant uncertainties.

2316 **9.1 Statistical Uncertainties and Bootstrapping**

2317 There are two components to the statistical error for the neural network background estimate.
 2318 The first is standard Poisson error, i.e., a given bin, i , in the background histogram has value
 2319 $n_i = \sum_{j \in i} w_j$, where w_j is the weight for an event j which falls in bin i . Standard techniques
 2320 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
 2321 when all w_j are equal to 1.

2322 However, this procedure does not take into account the statistical uncertainty on the
 2323 w_j due to the finite training dataset. Due to the large size difference between the two tag
 2324 and four tag datasets, it is the statistical uncertainty due to the four tag training data that
 2325 dominates that on the background. A standard method for estimating this uncertainty is the
 2326 bootstrap resampling technique [108]. Conceptually, a set of statistically equivalent sets is

2327 constructed by sampling with replacement from the original training set. The reweighting
 2328 network is then trained on each of these separately, resulting in a set of statistically equivalent
 2329 background estimates. Each of these sets is below referred to as a replica.

2330 In practice, as the original training set is large, the resampling procedure is able to
 2331 be simplified through the relation $\lim_{n \rightarrow \infty} \text{Binomial}(n, 1/n) = \text{Poisson}(1)$, which dictates that
 2332 sampling with replacement is approximately equivalent to applying a randomly distributed
 2333 integer weight to each event, drawn from a Poisson distribution with a mean of 1.

2334 Though the network configuration itself is the same for each bootstrap training, the
 2335 network initialization is allowed to vary. It should therefore be noted that the bootstrap
 2336 uncertainties implicitly capture the uncertainty due to this variation in addition to the
 2337 previously mentioned training set variation.

2338 The variation from this bootstrapping procedure is used to assign a bin-by-bin uncertainty
 2339 which is treated as a statistical uncertainty in the fit. Due to practical constraints, a
 2340 procedure for approximating the full bootstrap error band is developed which demonstrates
 2341 good agreement with the full bootstrap uncertainty. This procedure is described below.

2342 *9.1.1 Calculating the Bootstrap Error Band*

2343 The standard procedure to calculate the bootstrap uncertainty would proceed as follows: first,
 2344 each network trained on each bootstrap replica dataset would be used to produce a histogram
 2345 in the variable of interest. This would result in a set of replica histograms (e.g. for 100
 2346 bootstrap replicas, 100 histograms would be created). The nominal estimate would then be
 2347 the mean of bin values across these replica histograms, with errors set by the corresponding
 2348 standard deviation.

2349 In practice, such an approach is inflexible and demanding both in computation and in
 2350 storage, in so far as we would like to produce histograms in many variables, with a variety
 2351 of different cuts and binnings. This motivates a derivation based on event-level quantities.
 2352 However, due to non-trivial correlations between replica weights, simple linear propagation of
 2353 event weight variation is not correct.

2354 We therefore adopt an approach which has been empirically found to produce results
 2355 (for this analysis) in line with those produced by generating all of the histograms, as in the
 2356 standard procedure. This approach is described below. Note that, for robustness to outliers
 2357 and weight distribution asymmetry, the median and interquartile range (IQR) are used for
 2358 the central value and width respectively (as opposed to the mean and standard deviation).

2359 The components involved in the calculation have been mentioned in Chapter 8 and are as
 2360 follows:

- 2361 1. Replica weight (w_i): weight predicted for a given event by a network trained on replica
 dataset i .
- 2363 2. Replica norm (α_i): normalization factor for replica i . This normalizes the reweighting
 prediction of the network trained on replica dataset i to match the corresponding target
 yield.
- 2365 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

$$w_{med} \equiv \text{median}(\alpha_1 w_1, \dots, \alpha_{100} w_{100}) \quad (9.1)$$

- 2366 4. Normalization correction (α_{med}): normalization factor to match the predicted yield of
 the median weights (w_{med}) to the target yield in the training region.

2368 As mentioned in Chapter 8, the *nominal estimate* is constructed from the set of median
 2369 weights and the normalization correction, i.e. $\alpha_{med} \cdot w_{med}$.

2370 For the bootstrap error band, a “varied” histogram is then generated by applying, for
 2371 each event, a weight equal to the median weight (with no normalization correction) plus half
 2372 the interquartile range of the replica weights: $w_{varied} = w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})$.

2373 This varied histogram is scaled to match the yield of the nominal estimate. To account
 2374 for variation of the nominal estimate yield, a normalization variation is calculated from the

2375 interquartile range of the replica norms: $\frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100})$. This variation, multiplied into
2376 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:

$$H(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100})) \cdot \frac{Y(\alpha_{med} w_{med})}{Y(w_{med} + \frac{1}{2} \text{IQR}(w_1, \dots, w_{100}))} + \frac{1}{2} \text{IQR}(\alpha_1, \dots, \alpha_{100}) \cdot H(\alpha_{med} w_{med}) \quad (9.2)$$

2377 where the first term roughly describes the behavior of the bootstrap variation across the
2378 distribution of the variable of interest while the second term describes the normalization
2379 variation of the bootstrap replicas.

2380 The difference between the varied histogram and the nominal histogram is then taken to
2381 be the bootstrap statistical uncertainty on the nominal histogram.

2382 Figure 9.1 demonstrates how each of the components described above contribute to the
2383 uncertainty envelope for the non-resonant 2017 Control Region and compares this approximate
2384 band to the variation of histograms from individual bootstrap estimates. The error band
2385 constructed from the above procedure is seen to provide a good description of the bootstrap
2386 variation.

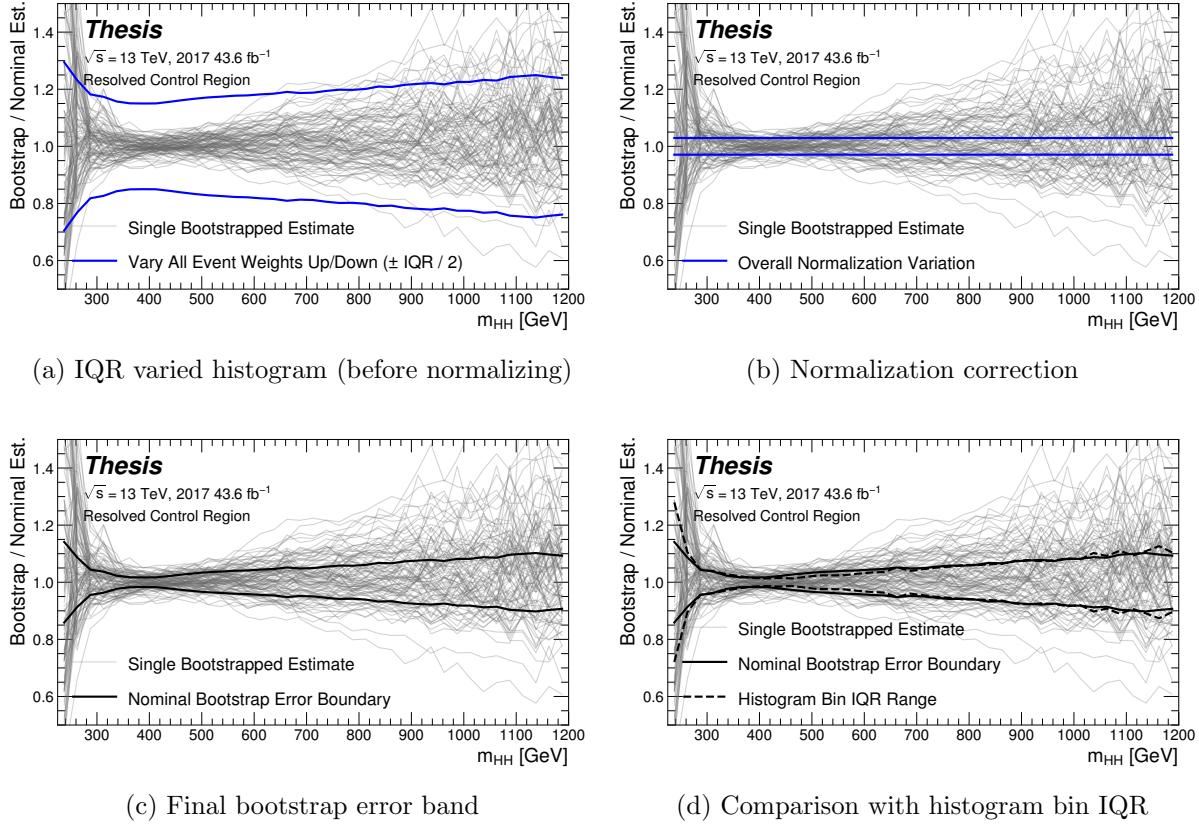


Figure 9.1: Illustration of the approximate bootstrap band procedure, shown as a ratio to the nominal estimate for the 2017 non-resonant background estimate. Each grey line is from the m_{HH} prediction for a single bootstrap training. Figure 9.1(a) shows the variation histograms constructed from median 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 9.1(b) demonstrates the applied normalization correction, and Figure 9.1(c) shows the final band (normalized Figure 9.1(a) + Figure 9.1(b)). Comparing this with the IQR variation for the prediction from each bootstrap in each bin in Figure 9.1(d), the approximate envelope describes a very similar variation.

2387 **9.2 Background Shape Uncertainties**

2388 To account for the systematic bias associated with deriving the reweighting function in the
2389 control region and extrapolating to the signal region, an alternative background model is
2390 derived in the validation region. Because of the fully data-driven nature of the background
2391 model, this is an uncertainty assessed on the full background. The alternative model and
2392 the baseline are consistent with the observed data in their training regions, and differences
2393 between the alternative and baseline models are used to define a shape uncertainty on the
2394 m_{HH} spectrum, with a two-sided uncertainty defined by symmetrizing the difference about
2395 the baseline.

2396 For the resonant analysis, this uncertainty is split into two components to allow for two
2397 independent variations of the m_{HH} spectrum: a low- H_T and a high- H_T component, where
2398 H_T is the scalar sum of the p_T of the four jets constituting the Higgs boson candidates, and
2399 serves as a proxy for m_{HH} , while avoiding introducing a sharp discontinuity. The boundary
2400 value is 300 GeV. The low- H_T shape uncertainty primarily affects the m_{HH} spectrum below
2401 400 GeV (close to the kinematic threshold) by up to around 5%, and the high- H_T uncertainty
2402 mainly m_{HH} above this by up to around 20% relative to nominal. These separate m_{HH}
2403 regimes are by design – the H_T split is introduced to prevent low mass bins from constraining
2404 the high mass uncertainty and vice-versa.

2405 This was the *status quo* shape uncertainty decomposition from the Early Run 2 analysis.
2406 A decomposition in terms of orthogonal polynomials, which would provide increased flexibility,
2407 was also evaluated. This study revealed that both decompositions are able to account for the
2408 systematic deviations between four tag data and the background estimate (evaluated in the
2409 kinematic validation region), and produce almost identical limits. The simpler *status quo*
2410 decomposition is therefore kept.

2411 For the non-resonant analysis, the quadrant nature of the background estimation leads to
2412 a natural breakdown of the nuisance parameters: quadrants are defined in the signal region
2413 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 9.2 shows an example of the variation in each H_T region for the 2018 resonant analysis. Figure 9.3 shows the example quadrant variation for the 2018 $4b$ non-resonant analysis.

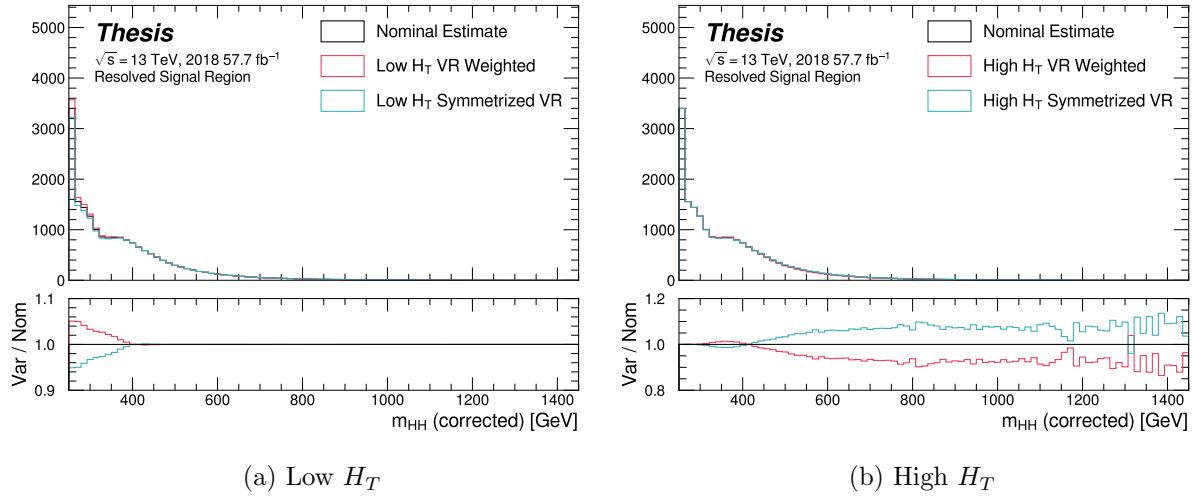
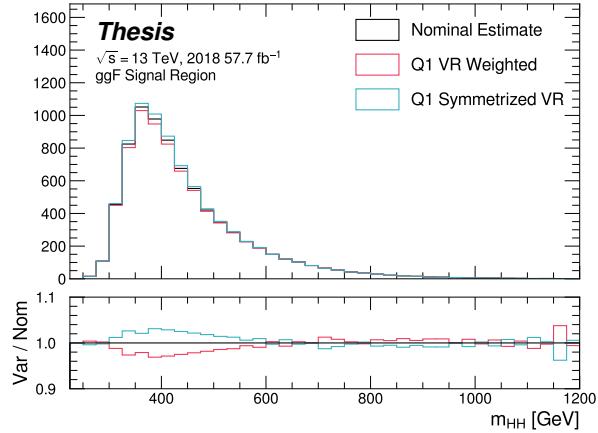
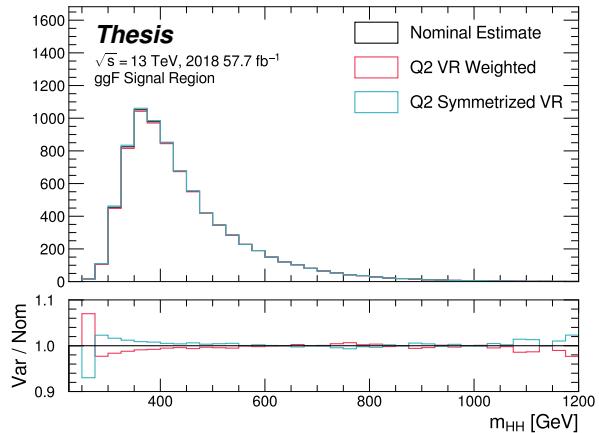


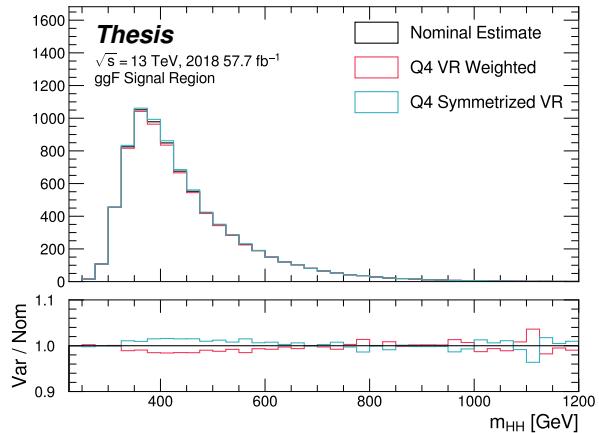
Figure 9.2: **Resonant Search:** Example of CR vs VR variation in each H_T region for 2018. The variation nicely factorizes into low and high mass components.



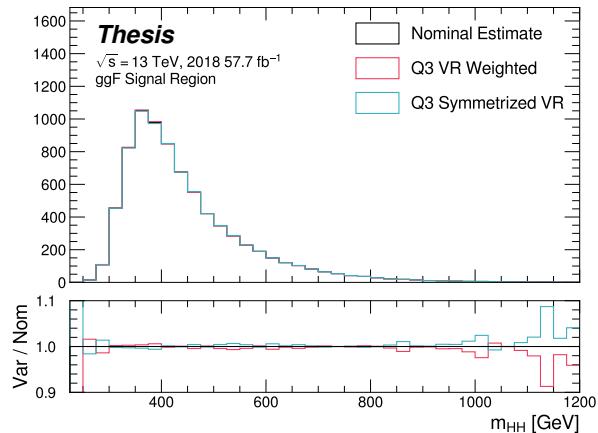
(a) Q1 (top)



(b) Q2 (left)



(c) Q4 (right)



(d) Q3 (bottom)

Figure 9.3: **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.

2420 **9.3 Signal Uncertainties**

2421 A variety of uncertainties are assessed on the signal Monte Carlo simulation. As the
2422 background estimate is fully data driven, such uncertainties are not needed for the background
2423 estimate. Note again that the results presented for the non-resonant search only include the
2424 background systematics described above.

2425 Detector modeling and reconstruction uncertainties account for differences between Monte
2426 Carlo simulation and real data due to mis-modeling of the detector as well as due to the
2427 different performance of algorithms on simulation compared to data. In this analysis they
2428 consist of uncertainties related to jet properties and uncertainties stemming from the flavor
2429 tagging procedure. The jet uncertainties are treated according to the prescription in [109] and
2430 are implemented as variations of the jet properties. These cover uncertainty in jet energy scale
2431 and resolution. Uncertainties in b -tagging efficiency are treated according to the prescription
2432 in Ref. [81] and implemented as scale factors applied to the Monte Carlo event weights. A
2433 systematic related to the PtReco b -jet energy correction has been studied in the $HH \rightarrow \gamma\gamma b\bar{b}$
2434 analysis [110] and found to be negligible compared to the other jet uncertainties. Following
2435 this example, such a systematic is therefore neglected here.

2436 Trigger uncertainties stem from imperfect knowledge of the ratio between the efficiency of
2437 a given trigger in data to its efficiency in Monte Carlo simulation. This ratio is applied as a
2438 scale factor to all simulated events, with the systematic variations produced by varying the
2439 scale factor up or down by one sigma. Such variations are evaluated based on measurements
2440 of per-jet online efficiencies for both jet reconstruction and b -tagging, and these are used to
2441 compute event-level uncertainties. These are then applied as overall weight variations on the
2442 simulated events.

2443 An uncertainty on the total integrated luminosity used in this analysis is also applied, and
2444 is measured to be 1.7% [98], obtained using the LUCID-2 detector for the primary luminosity
2445 measurements [111].

2446 A variety of theoretical uncertainties are also assessed on the signal. Such uncertainties

2447 are assessed by generating samples following the configuration of the baseline samples, but
 2448 with modifications to probe various aspects of the simulation. These include uncertainties in
 2449 the parton density functions (PDFs); uncertainties due to missing higher order terms in the
 2450 matrix elements; and uncertainties in the modelling of the underlying event, which includes
 2451 multi-parton interactions, of hadronic showers and of initial and final state radiation.

2452 Uncertainties due to modelling of the parton shower and the underlying event are eval-
 2453 uated by comparing results from using two different generators, namely HERWIG 7.1.3 and
 2454 PYTHIA 8.235. No significant dependence on the variable of interest, m_{HH} , is observed.
 2455 Therefore, a 5% flat systematic uncertainty is assigned to all signal samples, extracted from
 2456 the acceptance comparison for the full 4-tag selection.

2457 Uncertainties in the matrix element calculation are evaluated by varying the factorization
 2458 and renormalization scales used in the generator up and down by a factor of two, both
 2459 independently and simultaneously. This results in an effect smaller than 1% for all variations
 2460 and all masses; the impact of such uncertainties is therefore neglected.

2461 PDF uncertainties are evaluated using the PDF4LHC_NLO_MC set [99] by calculating
 2462 the signal acceptance for each PDF replica and taking the standard deviation. In all cases,
 2463 these uncertainties result in an effect smaller than 1% on the signal acceptance; therefore
 2464 these are also neglected.

2465 Theoretical uncertainties on the $H \rightarrow b\bar{b}$ branching ratio [112] are also included.

2466 **9.4 Background Validation**

2467 In addition to checking the performance of the background estimate in the control and
2468 validation regions, a variety of alternative selections are defined to allow for a full “dress
2469 rehearsal” of the background estimation procedure.

2470 Both the resonant and non-resonant analyses make use of a *reversed* $\Delta\eta$ region, in which
2471 the kinematic cut on $\Delta\eta_{HH}$ is reversed, so that events are required to have $\Delta\eta_{HH} > 1.5$.
2472 This is orthogonal to the nominal signal region and has minimal sensitivity, allowing for the
2473 comparison of the background estimate $4b$ data in the corresponding “signal region”. For
2474 this validation, a new reweighting is trained following nominal procedures, but entirely in the
2475 $\Delta\eta_{HH} > 1.5$ region.

2476 The non-resonant analysis additionally makes use of the $3b + 1$ fail region mentioned
2477 above, which again is orthogonal to the nominal signal regions and has minimal sensitivity.
2478 The reweighting in this case is between $2b$ and $3b + 1$ fail events rather than between $2b$
2479 and $3b + 1$ loose or $2b$ and $4b$. However, the kinematic selections of signal region events are
2480 otherwise identical, allowing for a complementary test of the background estimate.

2481 *TODO: Add shifted regions if they’re ready*

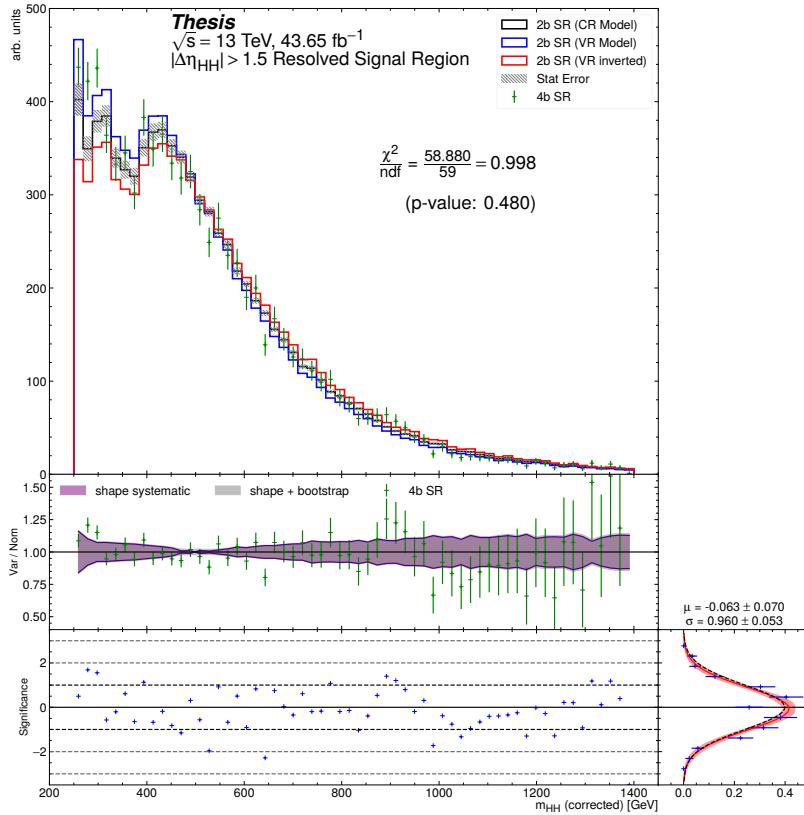


Figure 9.4: **Resonant Search:** Performance of the background estimation method in the resonant analysis reversed $\Delta\eta_{HH}$ kinematic signal region. A new background estimate is trained following nominal procedures entirely within the reversed $\Delta\eta_{HH}$ region, and the resulting model, 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.

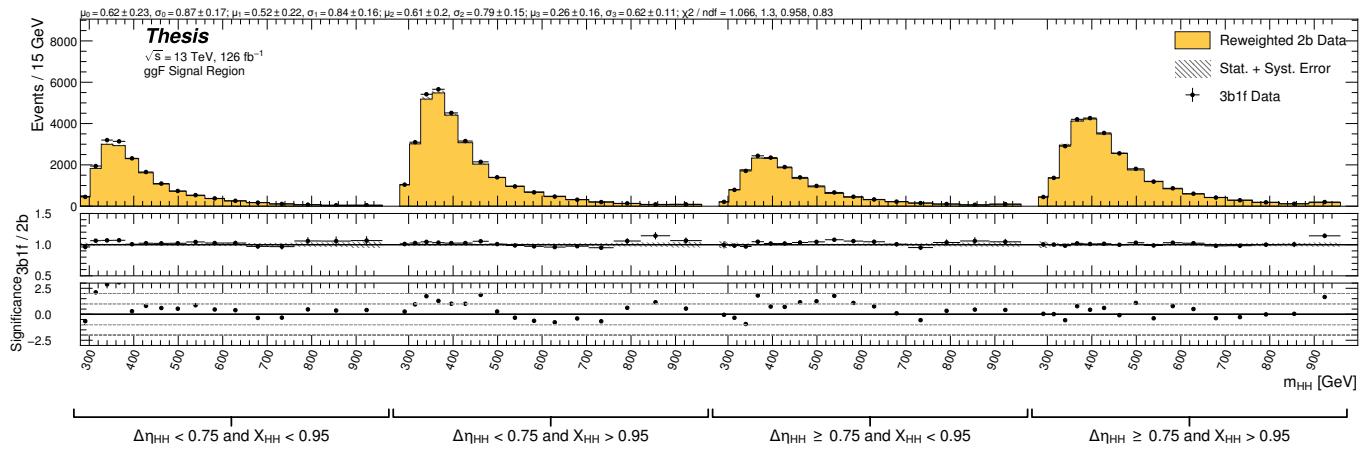


Figure 9.5: **Non-resonant Search:** Performance of the background estimation method in the $3b + 1$ fail validation region. A new background estimate is trained following nominal procedures but with a reweighting from $2b$ to $3b + 1$ fail events. Generally good agreement is seen, though there is some deviation at very low masses in the low $\Delta\eta_{HH}$ low X_{HH} category.

2482

Chapter 10

2483

RESULTS

2484 **10.1 m_{HH} Distributions**2485 **10.1.1 Resonant Search**

2486 The final discriminant used for the resonant search is corrected m_{HH} . Histogram binning
2487 was optimized for the resonant search to be 84 equal width bins from 250 GeV to 1450 GeV,
2488 corresponding to a bin width of 14.3 GeV, and overflow events (events above 1450 GeV) are
2489 included in the last bin. A demonstration of the performance of the reweighting on this
2490 distribution is shown in Figure 10.1 for the control region and Figure 10.2 for the validation
2491 region. A background-only profile likelihood fit is run for the distribution in the
2492 signal region, and results with spin-0 signals overlaid are shown in Figure 10.3. Note that the
2493 plots show the sum across all years, but the signal extraction fit and background estimate
2494 are run with the years separately. Agreement is generally good throughout.

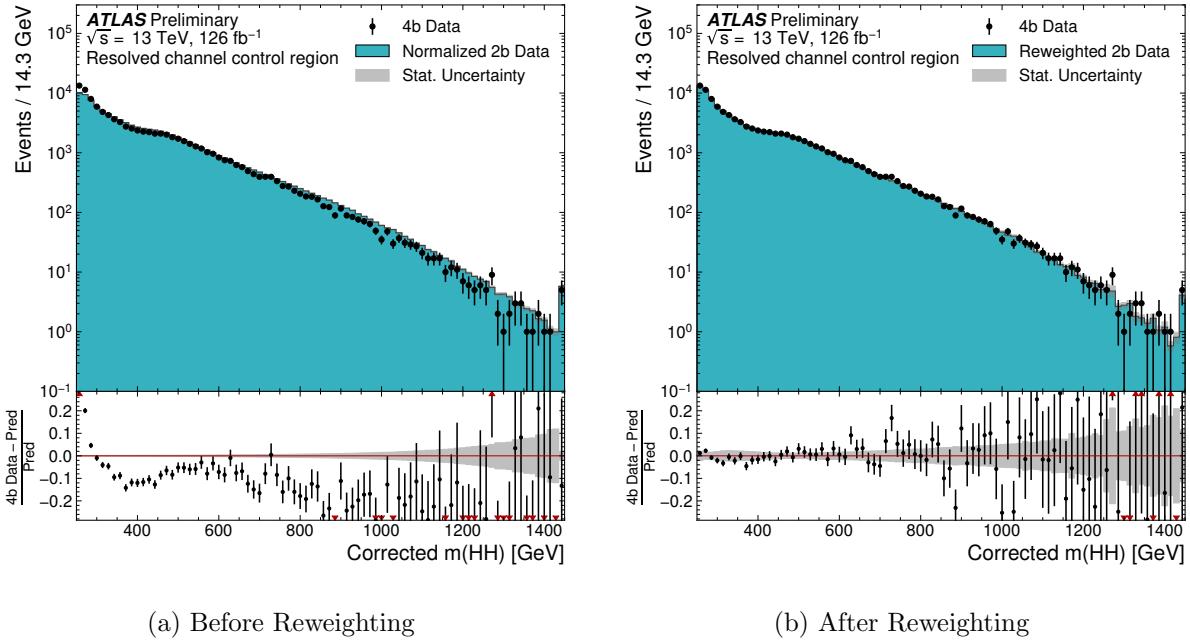


Figure 10.1: **Resonant Search:** Demonstration of the performance of the nominal reweighting in the control region on corrected m_{HH} , with Figure 10.1(a) showing $2b$ events normalized to the total $4b$ yield and Figure 10.1(b) applying the reweighting procedure. Agreement is much improved with the reweighting. Note that overall reweighted $2b$ yield agrees with $4b$ yield in the control region by construction.

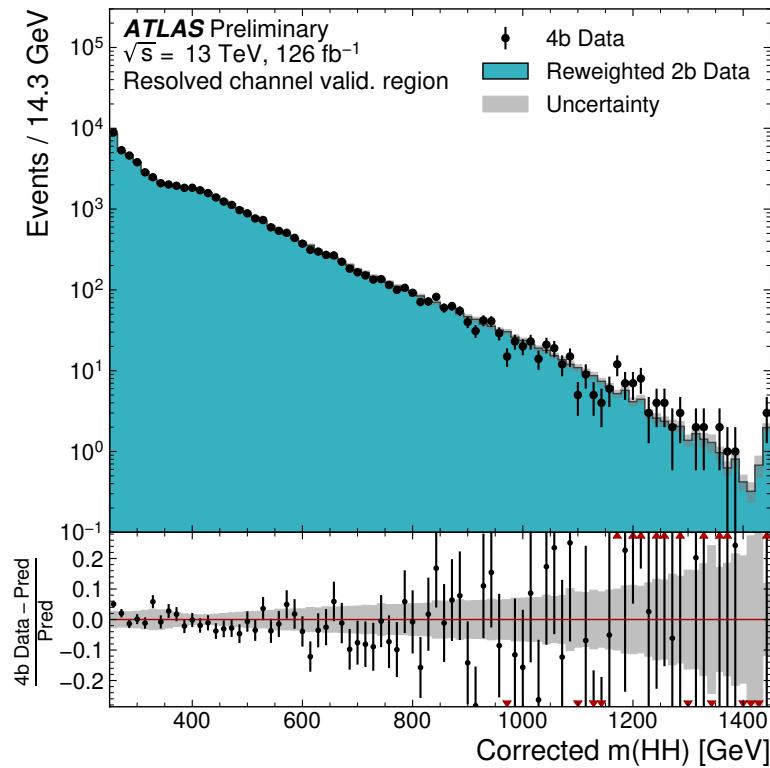


Figure 10.2: **Resonant Search:** Demonstration of the performance of the control region derived reweighting in the validation region on corrected m_{HH} . Agreement is generally good for this extrapolated estimate. Note that the uncertainty band includes the extrapolation systematic, which is defined by a reweighting trained in the validation region.

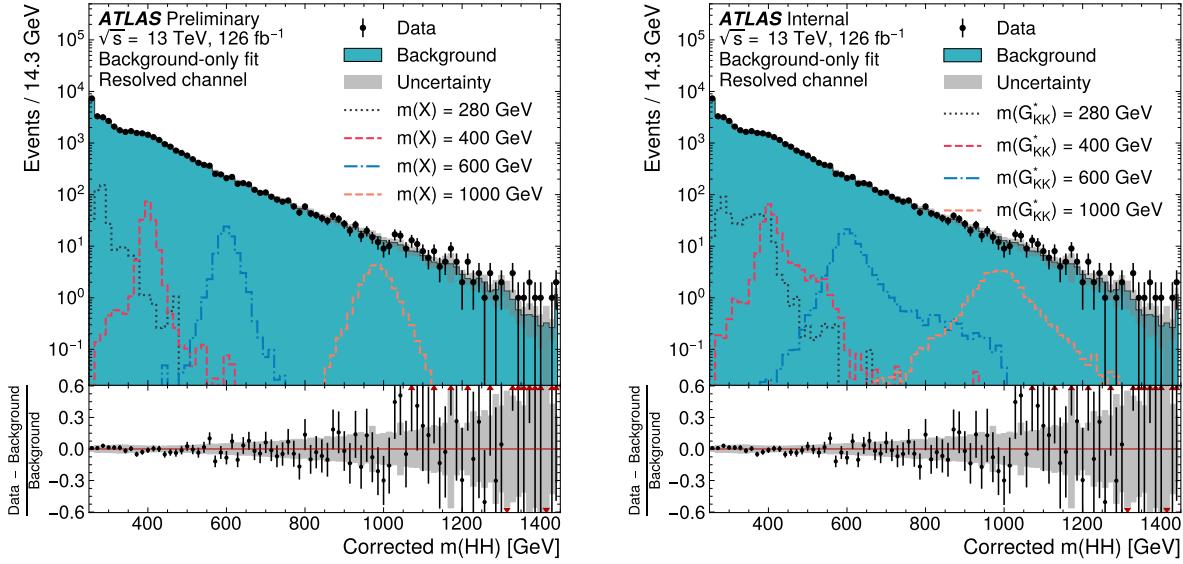


Figure 10.3: **Resonant Search:** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit. The left plot overlays a variety of representative spin-0 signals, while the right does the same for spin-2. The background and data are identical between the two. 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} . Note that the spin-2 signals are significantly wider than the spin-0 signals. Near the kinematic threshold of 250 GeV, this leads to, e.g., the double peaked structure of the 280 GeV signal, which is understood to be an effect of the limited kinematic phase space in this region.

2495 10.1.2 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

$$b_{low}^{i+1} = b_{low}^i + r \cdot b_{low}^i, \quad (10.1)$$

2496 where b_{low}^i is the low edge of bin i . The parameters used here are $r = 0.08$ over a range
2497 from 280 GeV to 975 GeV, and underflow and overflow are included in the initial and final
2498 bin contents respectively. m_{HH} with no correction is used as the final discriminant in each
2499 category.

2500 A demonstration of the performance of the reweighting on distributions unrolled across
2501 categories is shown in Figures 10.4 and 10.5 for the control region and Figures 10.6 and 10.7
2502 for the validation region. A background-only profile likelihood fit is run for the distribution in
2503 the signal region, and results with the Standard Model HH signal and $\kappa_\lambda = 6$ signal overlaid
2504 are shown for $4b$ in Figure 10.8 and $3b1l$ in Figure 10.9. Note that the plots show the sum
2505 across all years, but the signal extraction fit and background estimate are run with the years
2506 separately. All bins are normalized to represent a density of Events / 15 GeV.

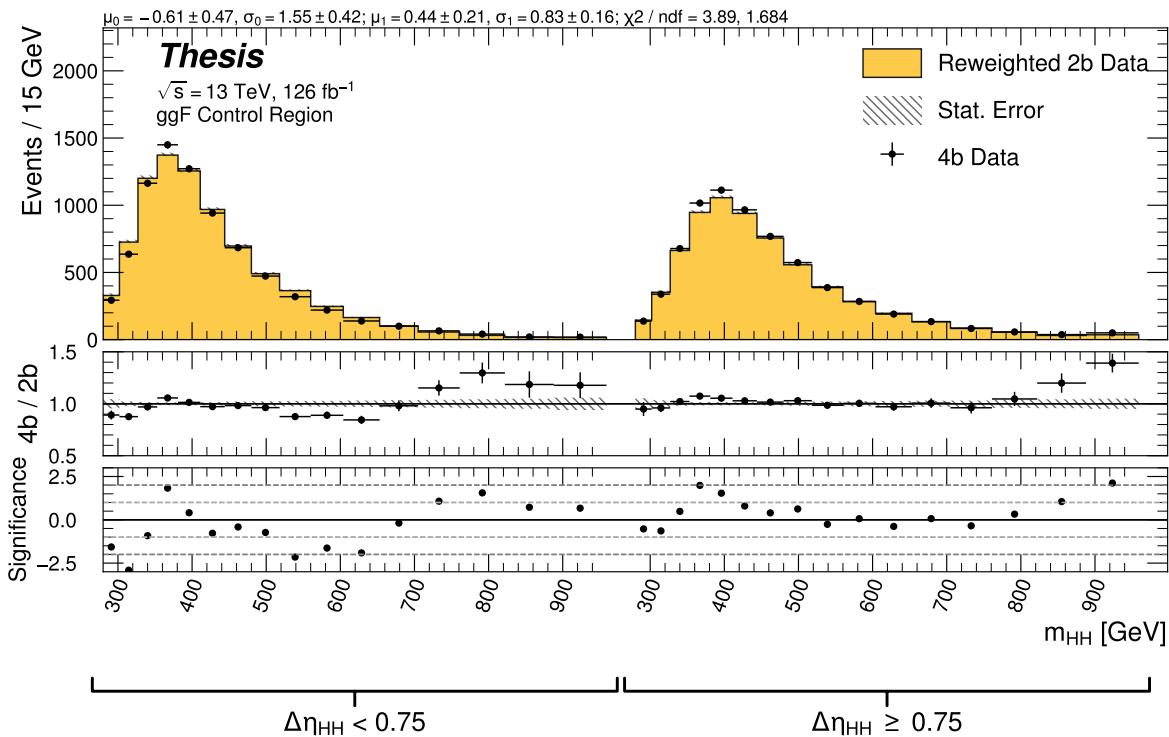


Figure 10.4: **Non-resonant Search (4b)**: Demonstration of the performance of the nominal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with some residual mis-modeling in the low $\Delta\eta_{HH}$ region near 600 GeV.

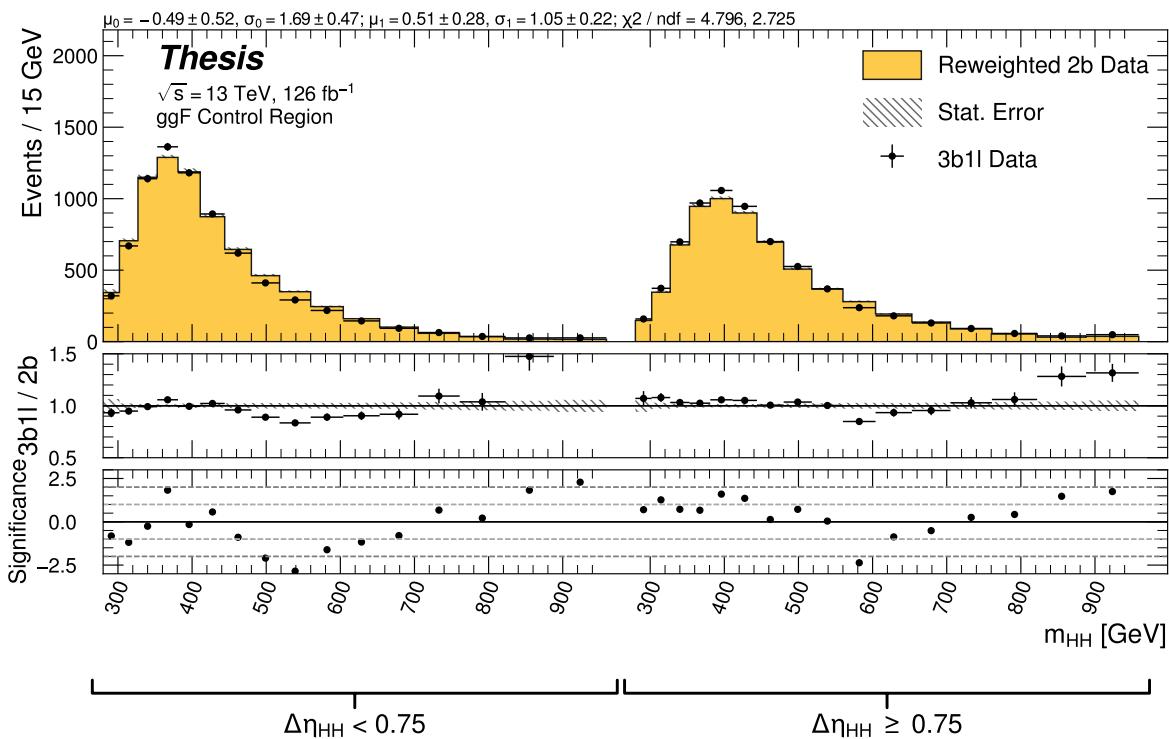


Figure 10.5: **Non-resonant Search (3b1l):** Demonstration of the performance of the nominal reweighting in the control region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. Closure is generally good, with similar conclusions as for the $4b$ region.

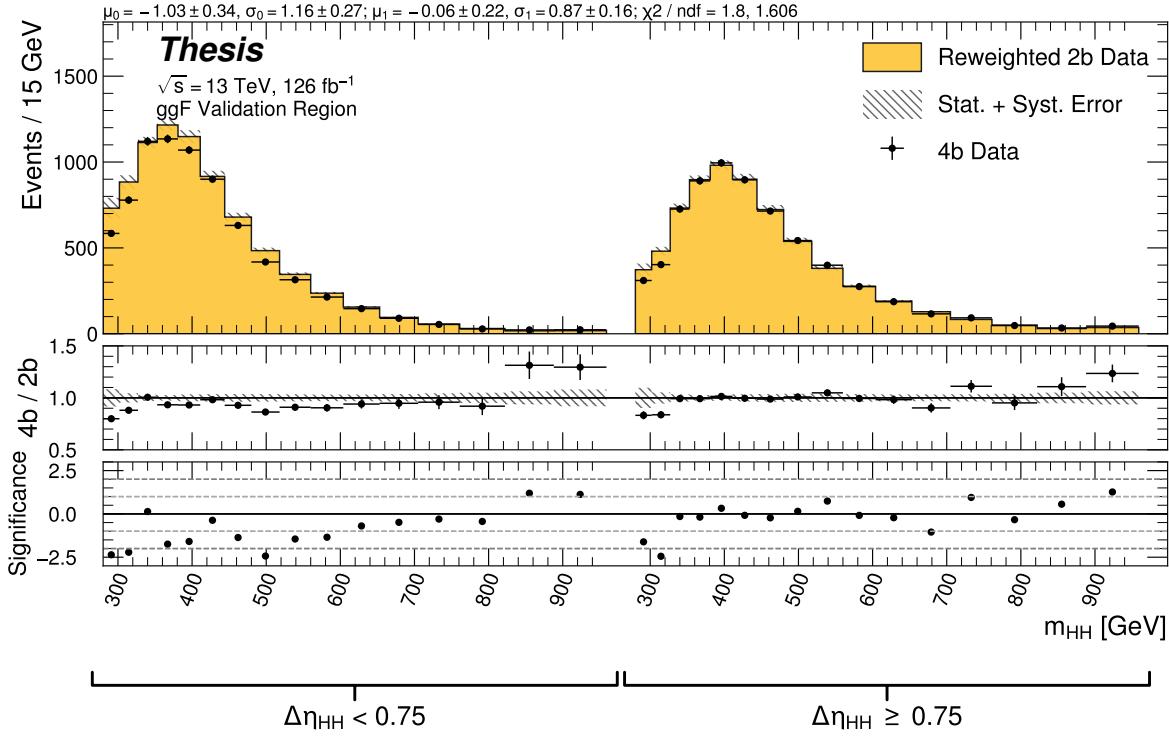


Figure 10.6: **Non-resonant Search (4b)**: Demonstration of the performance of the nominal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. The low $\Delta\eta_{HH}$ region is consistently overestimated, but, systematic uncertainties are defined via the difference between VR and CR estimates.

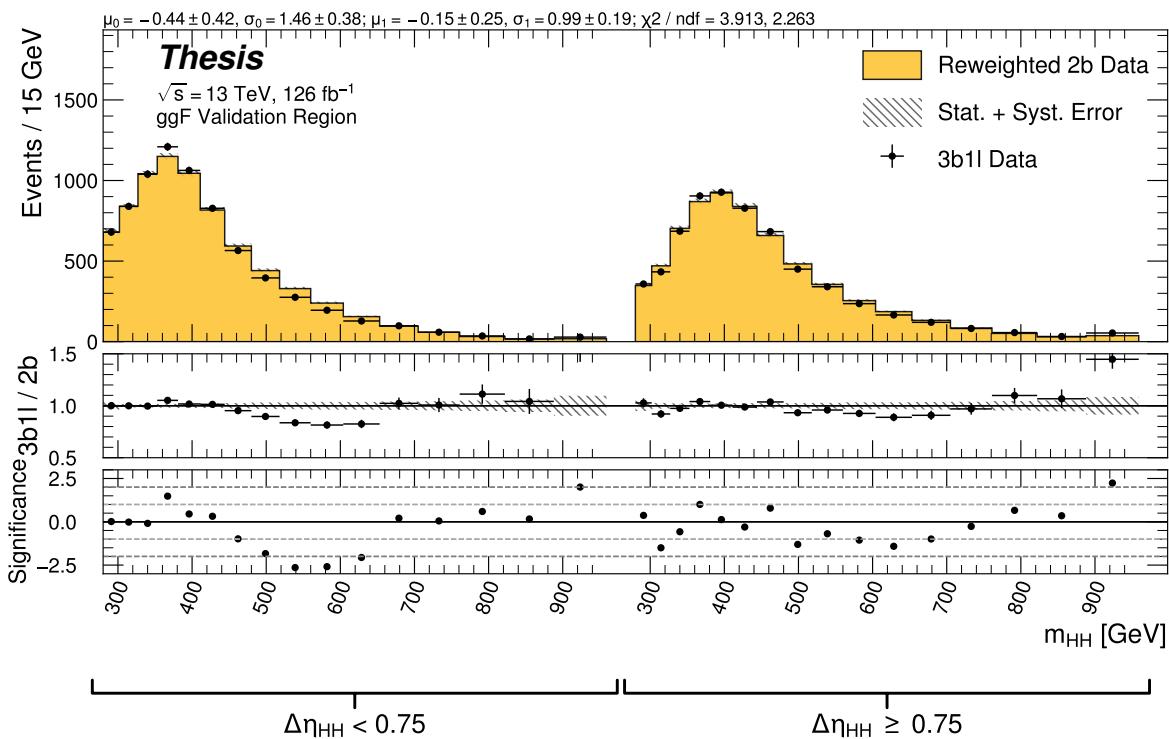


Figure 10.7: **Non-resonant Search (3b1l):** Demonstration of the performance of the nominal reweighting in the validation region on m_{HH} , split into the two $\Delta\eta_{HH}$ regions. A deficit is present near 600 GeV, but agreement is fairly good otherwise.

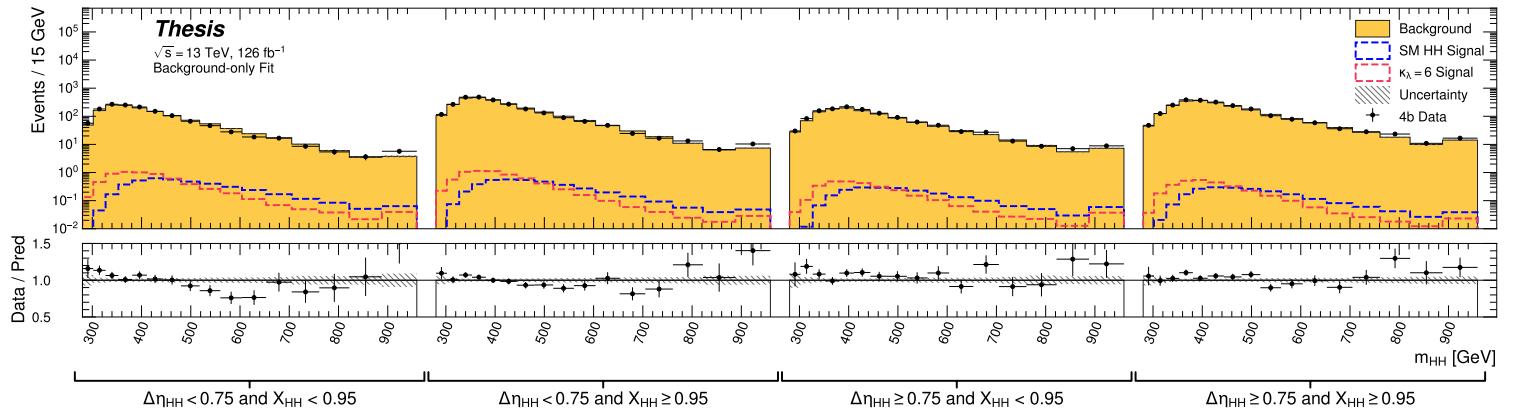


Figure 10.8: **Non-resonant Search (4b):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 4b channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Modeling is generally quite good near the Standard 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.

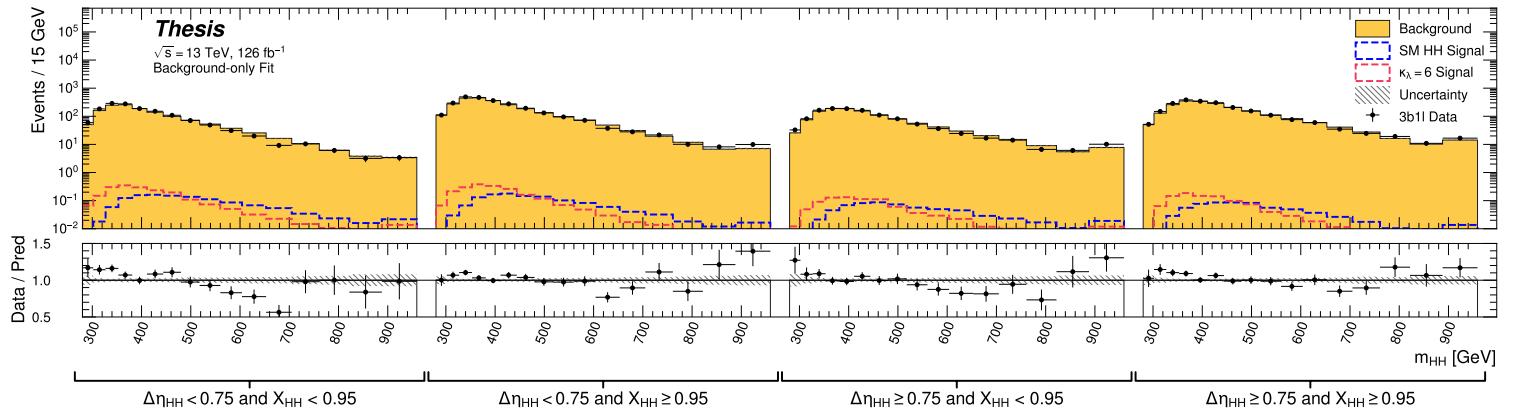


Figure 10.9: **Non-resonant Search (3b1l):** Signal region agreement of the background estimate and observed data after a background-only profile likelihood fit for the 3b1l channels, with Standard Model and $\kappa_\lambda = 6$ signal overlaid for reference. Conclusions are very similar to the 4b channels, with generally good modeling near the Standard Model peak, but disagreements at very low and high masses. A deficit is present near 600 GeV.

2507 **10.2 Statistical Analysis**

2508 The resonant analysis is used to set a 95% confidence level upper limit on the $pp \rightarrow X \rightarrow$
2509 $HH \rightarrow b\bar{b}b\bar{b}$ and $pp \rightarrow G_{KK}^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ cross-sections, while the non-resonant analysis
2510 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
2511 a variety of values of the trilinear Higgs coupling.

2512 The upper limit is extracted using the CL_s method [113]. The test statistic used is q_μ
2513 [114], where μ is the signal strength, and θ represents the nuisance parameters. A single
2514 hat represents the maximum likelihood estimate of a parameter, while $\hat{\theta}(x)$ represents the
2515 conditional maximum likelihood estimate of the nuisance parameters if the signal cross-section
2516 is fixed at x .

$$q_\mu = \begin{cases} -2 \ln \left(\frac{\mathcal{L}(\mu, \hat{\theta}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\theta})} \right) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} \quad (10.2)$$

2517 CL_s for some test value of μ is then defined by

$$CL_s = \frac{CL_{s+b}}{CL_b} = \frac{p(q_\mu \geq q_{\mu, \text{obs}} | s+b)}{p(q_\mu \geq q_{\mu, \text{obs}} | b)}, \quad (10.3)$$

2518 where the p -values are calculated in the asymptotic approximation [114], which is valid in
2519 the large sample limit.

2520 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 10.1: Limits on Standard Model $HH \rightarrow b\bar{b}b\bar{b}$ production, presented in units of the predicted Standard Model cross section. Results do not include signal systematics.

2521 10.3 Results

2522 Figure 10.10 shows the expected limit for the spin-0 and spin-2 resonant search. The resolved
 2523 channel covers the range between 251 and 1500 GeV and is combined with the boosted channel
 2524 between 900 and 1500 GeV. The boosted channel then extends to 5 TeV. All results use the
 2525 asymptotic approximation, though the validity of such an approximation for the boosted
 2526 results above 3 TeV is being studied. The most significant excess is seen for a signal mass of
 2527 1100 GeV, with local significance of 2.6σ for the spin-0 signal and 2.7σ for the spin-2 signal.
 2528 This is reduced to 1.0σ and 1.2σ globally.

2529 The spin-2 bulk Randall-Sundrum model with $k/\overline{M}_{\text{Pl}} = 1$ is excluded for graviton masses
 2530 between 298 and 1440 GeV.

2531 Preliminary results are presented here for the gluon-gluon fusion non-resonant search,
 2532 combining results from the $4b$ and $3b + 1l$ signal regions in the 2×2 category scheme in $\Delta\eta_{HH}$
 2533 and X_{HH} . These results will be further combined with a VBF channel as discussed, but this
 2534 is left for future work. Results shown here include background all background uncertainties,
 2535 but do not include signal systematics. Limits are set for κ_λ values from -20 to 20 . The cross
 2536 section limit for HH production is set at 140 fb (180 fb) observed (expected), corresponding
 2537 to an observed (expected) limit of 4.4 (5.9) times the Standard Model prediction (see Table
 2538 10.1). κ_λ is constrained to be within the range $-4.9 \leq \kappa_\lambda \leq 14.4$ observed ($-3.9 \leq \kappa_\lambda \leq 10.9$
 2539 expected). These results are shown in Figure 10.11.

2540 We note that this is a significant improvement over the early Run 2 result, which achieved
 2541 an observed (expected) limit of 12.9 (20.7) times the Standard Model prediction. The dataset

2542 is 4.6 times larger, and a naive scaling of the early Run 2 result (Poisson statistics \implies a factor
 2543 of $1/\sqrt{4.6}$) would predict an observed (expected) limit of 6.0 (9.7) times the Standard Model.
 2544 The result of 4.4 (5.9) observed (expected) presented here is therefore both an improvement
 2545 by a factor of 3 (3.5) over the previous result and also beats the statistical scaling by around
 2546 30 (40) %, demonstrating the impact of the various analysis improvements presented here.
 2547 We note again that these results do not include the complete set of uncertainties – however
 2548 we expect the addition of the remaining uncertainties to have no more than a few percent
 2549 impact.

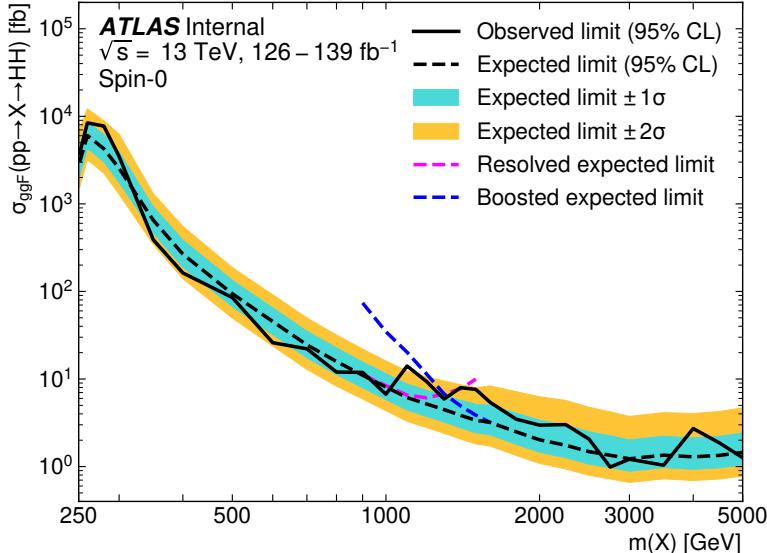
2550 The observed limits presented in Figure 10.11 are consistently above the 2σ band for
 2551 values of $\kappa_\lambda \geq 5$, peaking at a local significance of 3.8σ for $\kappa_\lambda = 6$. As this analysis is
 2552 optimized for points near the Standard Model, and as there is no excess present in more
 2553 sensitive channels in this same region (e.g. $HH \rightarrow bb\gamma\gamma$ *TODO: include comparison*), we do
 2554 not believe this is a real effect, but is rather due to a mis-modeling of the background at
 2555 low mass, where the min ΔR pairing has poor signal efficiency and the assumption of well
 2556 behaved background in the mass plane breaks down. This is consistent with the location of
 2557 the $\kappa_\lambda = 6$ signal in m_{HH} , as shown in Figures 10.8 and 10.9. It was considered, but not
 2558 implemented, for this analysis to impose a cut on m_{HH} near 350 or 400 GeV to avoid such a
 2559 low mass modeling issue.

2560 To check the impact of if we would have imposed such a cut, and to verify that the
 2561 excess is due to the low mass regime, we therefore run the same set of limits without the
 2562 low mass bins. In this case, we choose to simply drop the first few bins in m_{HH} such that
 2563 everything else, including the higher mass bin edges, is kept the same. Due to the variable
 2564 width binning, this corresponds to an m_{HH} cut of 381 GeV. The results of this check are
 2565 shown in Figure 10.12, and the corresponding limits for Standard Model HH are quoted in
 2566 Table 10.2. With the m_{HH} cut imposed, there is a slight degradation in the expected limits
 2567 for larger positive and negative values of κ_λ , but the points near the Standard Model are
 2568 nearly identical. Further, the observed excess is significantly reduced, with observed limits
 2569 for $\kappa_\lambda \geq 5$ now falling entirely within the expected 1σ band. Due to the preliminary nature

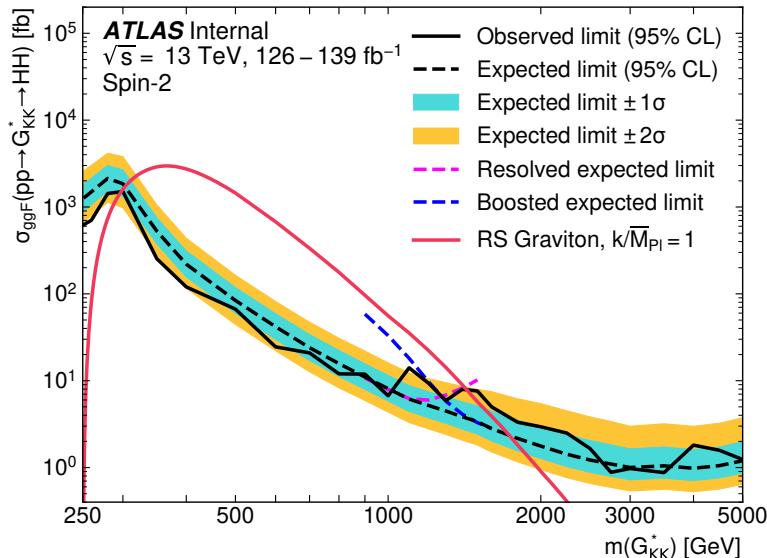
| Observed | -2σ | -1σ | Expected | $+1\sigma$ | $+2\sigma$ |
|------------|------------|------------|------------|------------|------------|
| 3.7 | 3.2 | 4.3 | 5.9 | 8.3 | 11.2 |

Table 10.2: Limits on Standard Model $HH \rightarrow b\bar{b}b\bar{b}$ production, presented in units of the predicted Standard Model cross section, corresponding to the $m_{HH} > 381$ GeV selection of Figure 10.12. Results do not include signal systematics. The deficit in the observed limit is larger than that of Table 10.1, but still within the 2σ band. There are only very minor differences in the expected limit band.

2570 of these results, further study is left for future work. However, we believe, in conjunction
2571 with the $HH \rightarrow bb\gamma\gamma$ results and our expectations about the difficulty of the background
2572 estimation at low mass, that this is demonstrative of a mis-modeling rather than a real excess.



(a)



(b)

Figure 10.10: 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 $k/\bar{M}_{\text{Pl}} = 1$ [25] (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. Note that all results use the asymptotic approximation, though the validity of this approximation for the boosted results above 3 TeV is being evaluated.

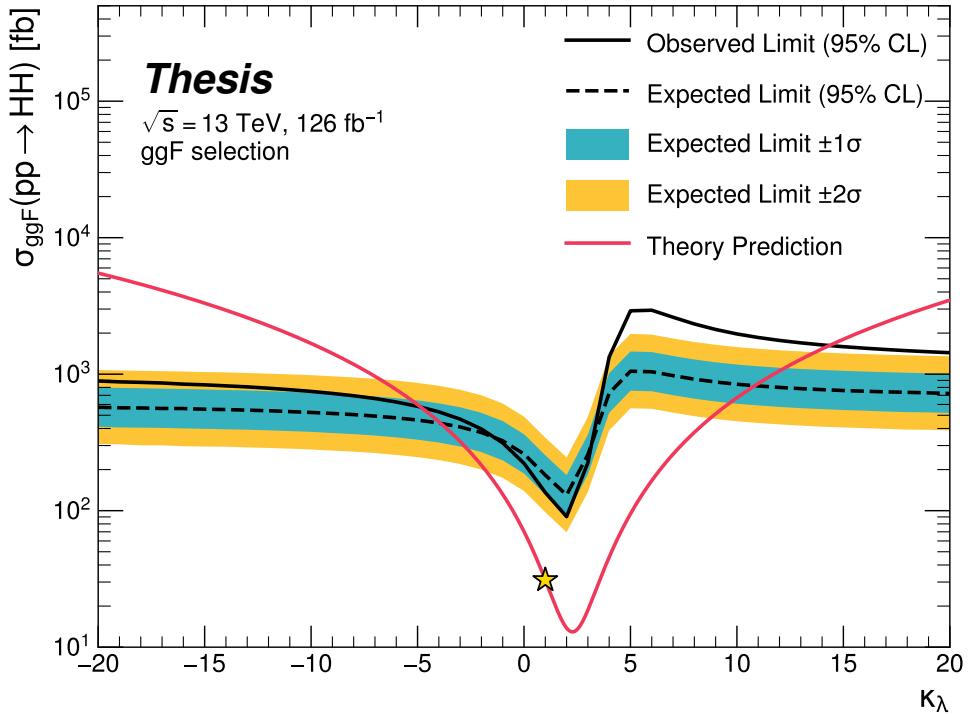


Figure 10.11: Expected (dashed black) and observed (solid black) 95% CL upper limits on the cross-section times branching ratio of non-resonant production for a range of values of the Higgs self-coupling, with the Standard Model value ($\kappa_\lambda = 1$) illustrated with a star. The $\pm 1\sigma$ 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. $HH \rightarrow bb\gamma\gamma$).

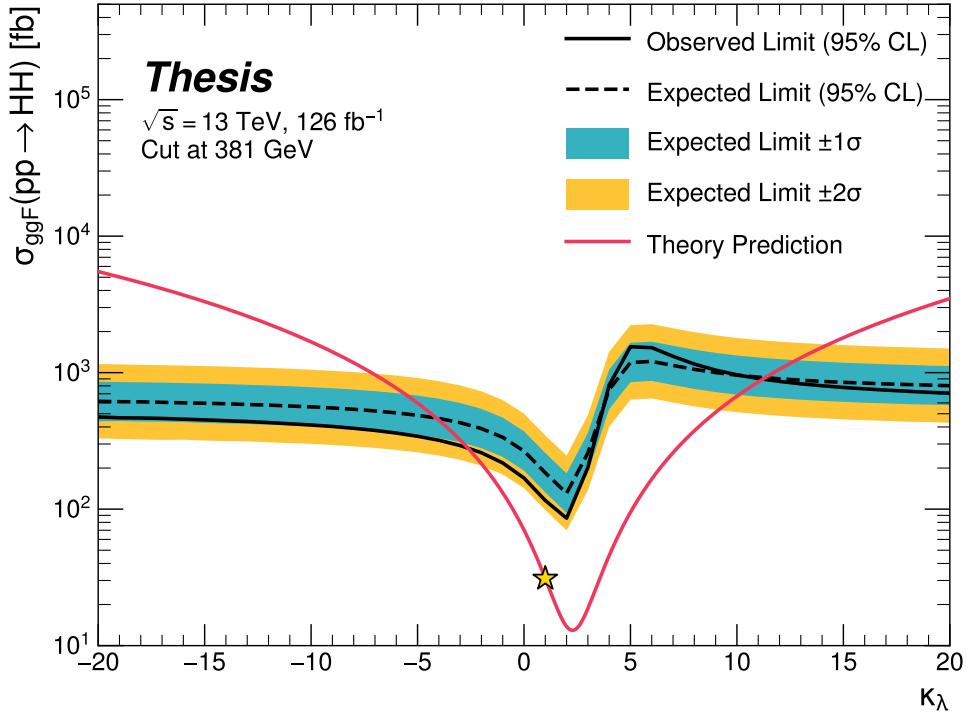


Figure 10.12: Limits including only events above 381 GeV in m_{HH} , to be compared with the limits in Figure 10.11. Such a cut is accomplished by dropping m_{HH} bins below 381 GeV, with the value of 381 GeV determined by the optimized variable width binning. All other aspects of the procedure and inputs are kept the same as in Figure 10.11. The excess at and above $\kappa_\lambda = 5$ is significantly reduced, demonstrating that such an excess is driven by low mass. Notably, there is minimal impact on the expected sensitivity with this m_{HH} cut.

2573

Chapter 11

2574

COMPARISONS WITH OTHER CHANNELS

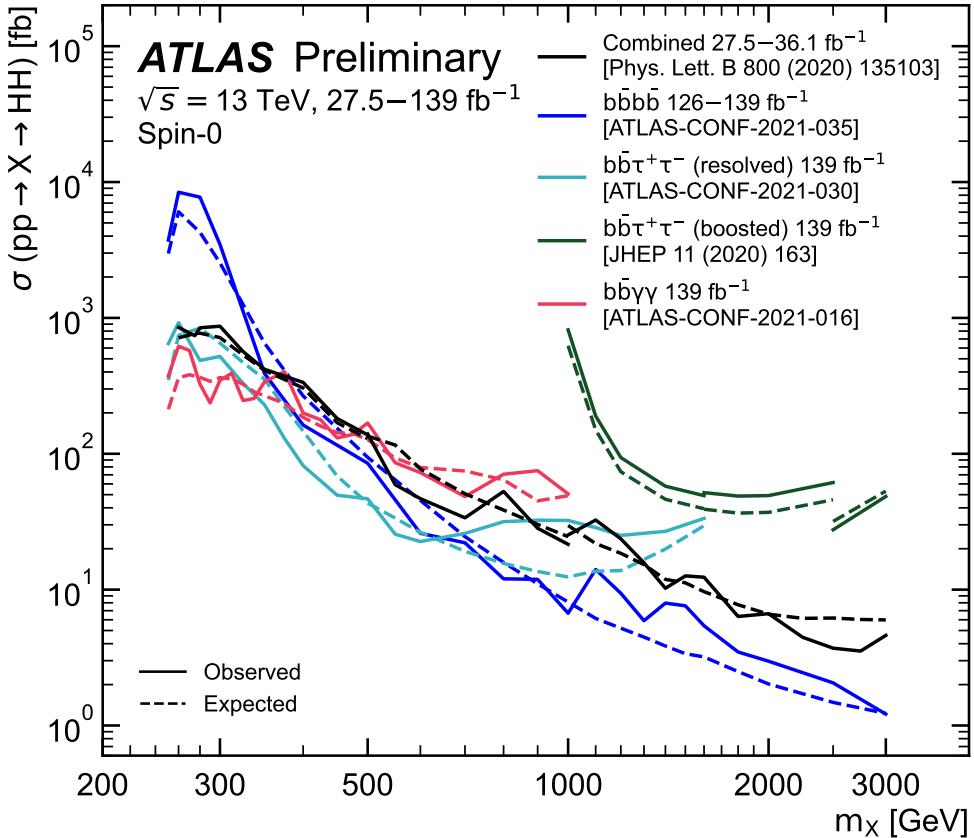


Figure 11.1: Comparison of full Run 2 ATLAS HH searches for spin-0 resonances. The $b\bar{b}b\bar{b}$ channel (blue) is compared with full Run 2 results from $b\bar{b}\tau^+\tau^-$ (both resolved and boosted) and $b\bar{b}\gamma\gamma$, as well as the combined early Run 2 results. The $b\bar{b}b\bar{b}$ channel has leading sensitivity above a mass of around 700 GeV, and is competitive with other channels across much of the mass range, demonstrating a strong contribution to the ATLAS HH experimental results. [115]

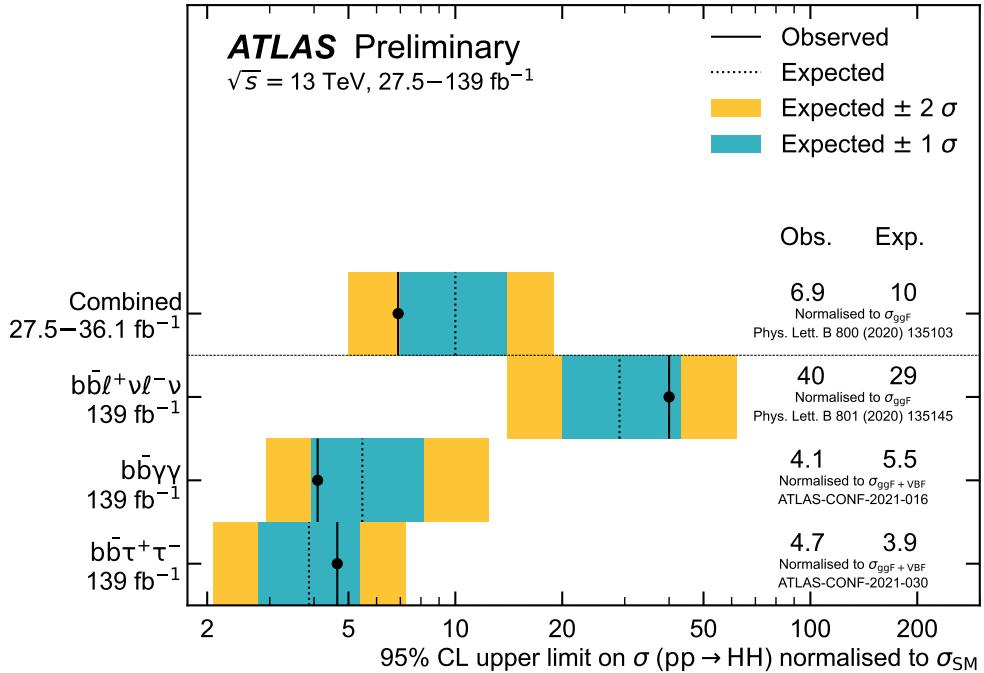


Figure 11.2: Comparison of full Run 2 ATLAS HH searches for Standard Model HH production. The preliminary results presented in this thesis are not yet included in these results. However, the results presented in Table 10.1 are quite competitive with the results from $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$, two of the ATLAS channels with leading sensitivity in the search for HH . Note that these results include signals produced via both gluon-gluon fusion (ggF) and vector boson fusion (VBF), and are normalized as such, while the results of this thesis only include (and are normalized to) ggF production [115]

2575

Chapter 12

2576

CONCLUSIONS

2577 This thesis has provided an overview of the Standard Model, with an emphasis on pair
2578 production of Higgs bosons and how this process may be used to both verify the Standard
2579 Model and to search for new physics. An overview of the Large Hadron Collider and the
2580 ATLAS detector has been provided, and the design and use of simulation infrastructure
2581 has been explained, including work to improve hadronic shower modeling in fast detector
2582 simulation. The translation of detector level information to analysis level information has
2583 been explained, with an emphasis on jets and the identification of B hadron decay. Finally,
2584 two searches for Higgs boson pair production have been presented, with a complete set of
2585 results for resonant production included, focusing on searches beyond the Standard Model,
2586 and a preliminary set of results for non-resonant production, targeting Standard Model
2587 production, with variations of the Higgs self-coupling. Two advanced techniques for the
2588 future of these analyses are further presented, along with proof-of-concept results.

2589

BIBLIOGRAPHY

- [1] S. Gasiorowski and H. Gray, *Fast Calorimeter Simulation in ATLAS*, EPJ Web Conf. **245** (2020) 02002. 7 p, URL: <https://cds.cern.ch/record/2756298> (cit. on pp. [xx](#), [56](#)).
- [2] ATLAS Collaboration, *Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final 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. [xxi](#), [77](#), [85](#), [95](#), [101](#)).
- [3] *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, tech. rep., All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATL-CONF-2021-035>: CERN, 2021, URL: <http://cds.cern.ch/record/2777861> (cit. on p. [xxi](#)).
- [4] A. Shmakov et al., *SPANet: Generalized Permutationless Set Assignment for Particle Physics using Symmetry Preserving Attention*, 2021, arXiv: [2106.03898 \[hep-ex\]](#) (cit. on p. [xxii](#)).
- [5] D. M. Reiman, J. Tamanas, J. X. Prochaska, and D. Ďurovčíková, *Fully probabilistic quasar continua predictions near Lyman- α with conditional neural spline flows*, 2020, arXiv: [2006.00615 \[astro-ph.CO\]](#) (cit. on p. [xxii](#)).
- [6] *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: 2021-07-23 (cit. on pp. [3](#), [27](#)).
- [7] J. D. Jackson, *Classical electrodynamics; 2nd ed.* Wiley, 1975 (cit. on p. [2](#)).

- 2611 [8] M. Thomson, *Modern particle physics*, Cambridge University Press, 2013, ISBN: 978-1-
2612 107-03426-6 (cit. on p. 2).
- 2613 [9] M. Srednicki, *Quantum Field Theory*, Cambridge Univ. Press, 2007 (cit. on p. 2).
- 2614 [10] M. Gell-Mann, *A schematic model of baryons and mesons*, Physics Letters **8** (1964)
2615 214, ISSN: 0031-9163, URL: <https://www.sciencedirect.com/science/article/pii/S0031916364920013> (cit. on p. 11).
- 2616 [11] G. Zweig, *An SU_3 model for strong interaction symmetry and its breaking; Version 1*,
2617 tech. rep., CERN, 1964, URL: <https://cds.cern.ch/record/352337> (cit. on p. 11).
- 2619 [12] O. W. Greenberg, *Spin and Unitary-Spin Independence in a Paraquark Model of
2620 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).
- 2621 [13] M. Y. Han and Y. Nambu, *Three-Triplet Model with Double $SU(3)$ Symmetry*, Phys.
2622 Rev. **139** (4B 1965) B1006, URL: <https://link.aps.org/doi/10.1103/PhysRev.139.B1006> (cit. on p. 11).
- 2625 [14] F. L. Wilson, *Fermi's Theory of Beta Decay*, American Journal of Physics **36** (1968)
2626 1150, eprint: <https://doi.org/10.1119/1.1974382>, URL: <https://doi.org/10.1119/1.1974382> (cit. on p. 14).
- 2628 [15] *The Nobel Prize in Physics 1979*, <https://www.nobelprize.org/prizes/physics/1979/summary/>, Accessed: 2021-07-23 (cit. on p. 15).
- 2630 [16] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*,
2631 Phys. Rev. Lett. **13** (9 1964) 321, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321> (cit. on p. 20).
- 2633 [17] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett.
2634 **13** (16 1964) 508, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>
2635 (cit. on p. 20).

- 2636 [18] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global Conservation Laws and*
 2637 *Massless Particles*, Phys. Rev. Lett. **13** (20 1964) 585, URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.585> (cit. on p. 20).
- 2639 [19] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
 2640 *Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
 2641 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214) (cit. on p. 20).
- 2642 [20] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS*
 2643 *experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv: [1207.7235 \[hep-ex\]](https://arxiv.org/abs/1207.7235)
 2644 (cit. on p. 20).
- 2645 [21] 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).
- 2647 [22] ATLAS Collaboration, *Search for the $HH \rightarrow b\bar{b}b\bar{b}$ process via vector-boson fusion*
 2648 *production using proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
 2649 JHEP **07** (2020) 108, arXiv: [2001.05178 \[hep-ex\]](https://arxiv.org/abs/2001.05178) (cit. on pp. 30, 78, 101), Erratum:
 2650 JHEP **01** (2021) 145.
- 2651 [23] G. Branco et al., *Theory and phenomenology of two-Higgs-doublet models*, Physics
 2652 Reports **516** (2012) 1, Theory and phenomenology of two-Higgs-doublet models,
 2653 ISSN: 0370-1573, URL: <http://www.sciencedirect.com/science/article/pii/S0370157312000695> (cit. on pp. 31, 32).
- 2655 [24] K. Agashe, H. Davoudiasl, G. Perez, and A. Soni, *Warped gravitons at the CERN*
 2656 *LHC and beyond*, Phys. Rev. D **76** (3 2007) 036006, URL: <https://link.aps.org/doi/10.1103/PhysRevD.76.036006> (cit. on pp. 31, 32).
- 2658 [25] A. Carvalho, *Gravity particles from Warped Extra Dimensions, predictions for LHC*,
 2659 2018, arXiv: [1404.0102 \[hep-ph\]](https://arxiv.org/abs/1404.0102) (cit. on pp. 31, 32, 189).

- 2660 [26] ATLAS Collaboration, *Combination of searches for Higgs boson pairs in pp collisions*
 2661 at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector, *Phys. Lett. B* **800** (2020) 135103, arXiv:
 2662 1906.02025 [hep-ex] (cit. on pp. 32, 35, 77).
- 2663 [27] P. D. Group et al., *Review of Particle Physics*, *Progress of Theoretical and Experimental*
 2664 *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. 34).
- 2667 [28] T. van Ritbergen and R. G. Stuart, *On the precise determination of the Fermi coupling*
 2668 *constant from the muon lifetime*, *Nuclear Physics B* **564** (2000) 343, ISSN: 0550-3213,
 2669 URL: [http://dx.doi.org/10.1016/S0550-3213\(99\)00572-6](http://dx.doi.org/10.1016/S0550-3213(99)00572-6) (cit. on p. 34).
- 2670 [29] S. Dawson, A. Ismail, and I. Low, *What's in the loop? The anatomy of double Higgs*
 2671 *production*, *Physical Review D* **91** (2015), ISSN: 1550-2368, URL: <http://dx.doi.org/10.1103/PhysRevD.91.115008> (cit. on p. 34).
- 2673 [30] A. M. Sirunyan et al., *Measurement of the top quark Yukawa coupling from $t\bar{t}$ kinematic*
 2674 *distributions in the dilepton final state in proton-proton collisions at $\sqrt{s}=13 \text{ TeV}$* ,
 2675 *Physical Review D* **102** (2020), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.102.092013> (cit. on p. 34).
- 2677 [31] S. Dawson, S. Dittmaier, and M. Spira, *Neutral Higgs-boson pair production at hadron*
 2678 *colliders: QCD corrections*, *Physical Review D* **58** (1998), ISSN: 1089-4918, URL:
 2679 <http://dx.doi.org/10.1103/PhysRevD.58.115012> (cit. on p. 36).
- 2680 [32] S. Borowka et al., *Higgs Boson Pair Production in Gluon Fusion at Next-to-Leading*
 2681 *Order with Full Top-Quark Mass Dependence*, *Physical Review Letters* **117** (2016),
 2682 ISSN: 1079-7114, URL: <http://dx.doi.org/10.1103/PhysRevLett.117.012001>
 2683 (cit. on p. 36).

- 2684 [33] J. Baglio et al., *Gluon fusion into Higgs pairs at NLO QCD and the top mass*
 2685 *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. 36).
- 2687 [34] D. de Florian and J. Mazzitelli, *Higgs Boson Pair Production at Next-to-Next-to-*
 2688 *Leading Order in QCD*, Physical Review Letters **111** (2013), ISSN: 1079-7114, URL:
 2689 <http://dx.doi.org/10.1103/PhysRevLett.111.201801> (cit. on p. 36).
- 2690 [35] D. Y. Shao, C. S. Li, H. T. Li, and J. Wang, *Threshold resummation effects in Higgs*
 2691 *boson pair production at the LHC*, Journal of High Energy Physics **2013** (2013), ISSN:
 2692 1029-8479, URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)169](http://dx.doi.org/10.1007/JHEP07(2013)169) (cit. on p. 36).
- 2693 [36] D. de Florian and J. Mazzitelli, *Higgs pair production at next-to-next-to-leading*
 2694 *logarithmic accuracy at the LHC*, 2015, arXiv: [1505.07122 \[hep-ph\]](https://arxiv.org/abs/1505.07122) (cit. on p. 36).
- 2695 [37] M. Grazzini et al., *Higgs boson pair production at NNLO with top quark mass effects*,
 2696 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. 36).
- 2698 [38] J. Baglio et al., *gg → HH: Combined uncertainties*, Physical Review D **103** (2021),
 2699 ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.056002> (cit. on
 2700 p. 36).
- 2701 [39] F. Dulat, A. Lazopoulos, and B. Mistlberger, *iHixs 2 — Inclusive Higgs cross sections*,
 2702 Computer Physics Communications **233** (2018) 243, ISSN: 0010-4655, URL: <http://dx.doi.org/10.1016/j.cpc.2018.06.025> (cit. on p. 36).
- 2704 [40] ATLAS Collaboration, *Measurement prospects of the pair production and self-coupling*
 2705 *of the Higgs boson with the ATLAS experiment at the HL-LHC*, ATL-PHYS-PUB-
 2706 2018-053, 2018, URL: <https://cds.cern.ch/record/2652727> (cit. on p. 36).
- 2707 [41] D. de Florian, I. Fabre, and J. Mazzitelli, *Triple Higgs production at hadron colliders*
 2708 *at NNLO in QCD*, Journal of High Energy Physics **2020** (2020), ISSN: 1029-8479, URL:
 2709 [http://dx.doi.org/10.1007/JHEP03\(2020\)155](http://dx.doi.org/10.1007/JHEP03(2020)155) (cit. on p. 36).

- 2710 [42] A. Papaefstathiou, T. Robens, and G. Tetlamatzi-Xolocotzi, *Triple Higgs boson*
2711 *production at the Large Hadron Collider with Two Real Singlet scalars*, Journal of
2712 High Energy Physics **2021** (2021), ISSN: 1029-8479, URL: [http://dx.doi.org/10.](http://dx.doi.org/10.1007/JHEP05(2021)193)
2713 (cit. on p. 36).
- 2714 [43] T. Robens, T. Stefaniak, and J. Wittbrodt, *Two-real-scalar-singlet extension of the*
2715 *SM: LHC phenomenology and benchmark scenarios*, The European Physical Journal C
2716 **80** (2020), ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-020-7655-x> (cit. on p. 36).
- 2718 [44] D. Egana-Ugrinovic, S. Homiller, and P. Meade, *Multi-Higgs boson production probes*
2719 *Higgs sector flavor*, Physical Review D **103** (2021), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.103.115005> (cit. on p. 36).
- 2721 [45] CERN, *CERN's accelerator complex*, <https://home.cern/science/accelerators/accelerator-complex>, Accessed: 27 July 2021 (cit. on p. 38).
- 2723 [46] J. Pequenao, “Computer generated image of the whole ATLAS detector”, 2008, URL:
2724 <https://cds.cern.ch/record/1095924> (cit. on p. 40).
- 2725 [47] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*,
2726 JINST **3** (2008) S08003 (cit. on p. 41).
- 2727 [48] J. Pequenao and P. Schaffner, “How ATLAS detects particles: diagram of particle
2728 paths in the detector”, 2013, URL: <https://cds.cern.ch/record/1505342> (cit. on
2729 p. 42).
- 2730 [49] ATLAS Collaboration, *ATLAS Insertable B-Layer: Technical Design Report*, ATLAS-
2731 TDR-19; CERN-LHCC-2010-013, 2010, URL: <https://cds.cern.ch/record/1291633> (cit. on p. 43), Addendum: ATLAS-TDR-19-ADD-1; CERN-LHCC-2012-009,
2732 2012, URL: <https://cds.cern.ch/record/1451888>.
- 2734 [50] B. Abbott et al., *Production and integration of the ATLAS Insertable B-Layer*, JINST
2735 **13** (2018) T05008, arXiv: [1803.00844 \[physics.ins-det\]](https://arxiv.org/abs/1803.00844) (cit. on p. 43).

- 2736 [51] A. Outreach, “ATLAS Fact Sheet : To raise awareness of the ATLAS detector and
 2737 collaboration on the LHC”, 2010, URL: <https://cds.cern.ch/record/1457044>
 2738 (cit. on p. 45).
- 2739 [52] ATLAS Collaboration, *Performance of the ATLAS trigger system in 2015*, *Eur. Phys.*
 2740 *J. C* **77** (2017) 317, arXiv: [1611.09661 \[hep-ex\]](https://arxiv.org/abs/1611.09661) (cit. on p. 46).
- 2741 [53] ATLAS Collaboration, *The ATLAS Collaboration Software and Firmware*, ATL-SOFT-
 2742 PUB-2021-001, 2021, URL: <https://cds.cern.ch/record/2767187> (cit. on p. 46).
- 2743 [54] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,
 2744 *JHEP* **11** (2004) 040, arXiv: [hep-ph/0409146 \[hep-ph\]](https://arxiv.org/abs/hep-ph/0409146) (cit. on pp. 50, 80).
- 2745 [55] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton*
 2746 *shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070, arXiv: [0709.2092](https://arxiv.org/abs/0709.2092)
 2747 [\[hep-ph\]](https://arxiv.org/abs/hep-ph/0709.2092) (cit. on pp. 50, 80).
- 2748 [56] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO*
 2749 *calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010)
 2750 043, arXiv: [1002.2581 \[hep-ph\]](https://arxiv.org/abs/1002.2581) (cit. on pp. 50, 80).
- 2751 [57] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*
 2752 *differential cross sections, and their matching to parton shower simulations*, *Journal of*
 2753 *High Energy Physics* **2014** (2014), URL: [https://doi.org/10.1007/jhep07\(2014\)](https://doi.org/10.1007/jhep07(2014)079)
 2754 079 (cit. on pp. 50, 79).
- 2755 [58] J. Bellm et al., *Herwig 7.0/Herwig++ 3.0 release note*, *The European Physical Journal*
 2756 *C* **76** (2016), URL: <https://doi.org/10.1140/epjc/s10052-016-4018-8> (cit. on
 2757 pp. 50, 79).
- 2758 [59] M. Bähr et al., *Herwig++ physics and manual*, *The European Physical Journal C*
 2759 **58** (2008) 639, URL: <https://doi.org/10.1140/epjc/s10052-008-0798-9> (cit. on
 2760 pp. 50, 79).

- 2761 [60] T. Sjöstrand et al., *An introduction to PYTHIA 8.2*, Computer Physics Communications **191** (2015) 159, ISSN: 0010-4655, URL: <https://www.sciencedirect.com/science/article/pii/S0010465515000442> (cit. on pp. 50, 80).
- 2762
2763
- 2764 [61] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462** (2001) 152 (cit. on pp. 50, 79).
- 2765
- 2766 [62] S. Höche, *Introduction to parton-shower event generators*, 2015, arXiv: [1411.4085](https://arxiv.org/abs/1411.4085) [hep-ph] (cit. on p. 51).
- 2767
- 2768 [63] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **506** (2003) 250, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/S0168900203013688> (cit. on p. 52).
- 2769
2770
2771
- 2772 [64] 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. 52).
- 2773
2774
- 2775 [65] A. Collaboration, *Computing and Software - Public Results*, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ComputingandSoftwarePublicResults>, Accessed: 27 July 2021 (cit. on p. 53).
- 2776
2777
- 2778 [66] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. C **70** (2010) 823, arXiv: [1005.4568](https://arxiv.org/abs/1005.4568) [physics.ins-det] (cit. on pp. 52, 79).
- 2779
- 2780 [67] ATLAS Collaboration, *The new Fast Calorimeter Simulation in ATLAS*, ATL-SOFT-PUB-2018-002, 2018, URL: <https://cds.cern.ch/record/2630434> (cit. on pp. 54, 55).
- 2781
2782
- 2783 [68] ROOT, *TPrincipal Class Reference*, <https://root.cern.ch/doc/master/classTPrincipal.html>, Accessed: 27 July 2021 (cit. on p. 54).
- 2784

- 2785 [69] M. Aharrouche et al., *Energy linearity and resolution of the ATLAS electromagnetic*
 2786 *barrel calorimeter in an electron test-beam*, Nuclear Instruments and Methods in
 2787 Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated
 2788 Equipment **568** (2006) 601, ISSN: 0168-9002, URL: <http://dx.doi.org/10.1016/j.nima.2006.07.053> (cit. on p. 55).
- 2790 [70] ATLAS Collaboration, *Evidence for prompt photon production in pp collisions at*
 2791 $\sqrt{s} = 7 \text{ TeV}$ *with the ATLAS detector*, ATLAS-CONF-2010-077, 2010, URL: <https://cds.cern.ch/record/1281368> (cit. on p. 55).
- 2793 [71] 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. 56).
- 2795 [72] G. Aad et al., *Topological cell clustering in the ATLAS calorimeters and its performance*
 2796 *in LHC Run 1*, The European Physical Journal C **77** (2017), ISSN: 1434-6052, URL:
 2797 <http://dx.doi.org/10.1140/epjc/s10052-017-5004-5> (cit. on pp. 60, 61).
- 2798 [73] A. Ghosh and A. Collaboration, *Deep generative models for fast shower simulation*
 2799 *in ATLAS*, Journal of Physics: Conference Series **1525** (2020) 012077, URL: <https://doi.org/10.1088/1742-6596/1525/1/012077> (cit. on p. 62).
- 2801 [74] R. Di Sipio, M. F. Giannelli, S. K. Haghight, and S. Palazzo, *DijetGAN: a Generative-*
 2802 *Adversarial Network approach for the simulation of QCD dijet events at the LHC*,
 2803 *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. 62).
- 2805 [75] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with*
 2806 *the ATLAS Detector*, Eur. Phys. J. C **77** (2017) 466, arXiv: [1703.10485 \[hep-ex\]](https://arxiv.org/abs/1703.10485)
 2807 (cit. on pp. 65, 78).
- 2808 [76] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and*
 2809 *its performance in LHC Run 1*, Eur. Phys. J. C **77** (2017) 490, arXiv: [1603.02934](https://arxiv.org/abs/1603.02934)
 2810 [hep-ex] (cit. on p. 65).

- 2811 [77] ATLAS Collaboration, *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet*
 2812 *Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*, ATL-PHYS-PUB-2017-010, 2017,
 2813 URL: <https://cds.cern.ch/record/2268678> (cit. on pp. 65, 211).
- 2814 [78] M. Cacciari, G. P. Salam, and G. Soyez, *The anti- k_T jet clustering algorithm*, Journal
 2815 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. 65).
- 2817 [79] G. P. Salam, *Towards jetography*, The European Physical Journal C **67** (2010) 637,
 2818 ISSN: 1434-6052, URL: <http://dx.doi.org/10.1140/epjc/s10052-010-1314-6>
 2819 (cit. on p. 65).
- 2820 [80] A. Collaboration, *Configuration and performance of the ATLAS b-jet triggers in Run*
 2821 *2*, 2021, arXiv: [2106.03584 \[hep-ex\]](https://arxiv.org/abs/2106.03584) (cit. on p. 69).
- 2822 [81] ATLAS Collaboration, *ATLAS b-jet identification performance and efficiency mea-*
 2823 *surement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*, Eur. Phys. J. C **79** (2019)
 2824 970, arXiv: [1907.05120 \[hep-ex\]](https://arxiv.org/abs/1907.05120) (cit. on pp. 68, 78, 169).
- 2825 [82] ATLAS Collaboration, *Topological b-hadron decay reconstruction and identification*
 2826 *of b-jets with the JetFitter package in the ATLAS experiment at the LHC*, ATL-
 2827 PHYS-PUB-2018-025, 2018, URL: <https://cds.cern.ch/record/2645405> (cit. on
 2828 p. 71).
- 2829 [83] R. Frühwirth, *Application of Kalman filtering to track and vertex fitting*, Nuclear
 2830 Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers,
 2831 Detectors and Associated Equipment **262** (1987) 444, ISSN: 0168-9002, URL: <https://www.sciencedirect.com/science/article/pii/0168900287908874> (cit. on
 2832 p. 71).
- 2834 [84] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent*
 2835 *Neural Networks at the ATLAS Experiment*, ATL-PHYS-PUB-2017-003, 2017, URL:
 2836 <https://cds.cern.ch/record/2255226> (cit. on p. 72).

- 2837 [85] *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
2838 p. 74).
- 2840 [86] ATLAS Collaboration, *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, **JHEP 04** (2019) 092, arXiv: [1811.04671 \[hep-ex\]](https://arxiv.org/abs/1811.04671) (cit. on p. 77).
- 2843 [87] ATLAS Collaboration, *A search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, **Phys. Rev. Lett.** **121** (2018) 191801, arXiv: [1808.00336 \[hep-ex\]](https://arxiv.org/abs/1808.00336) (cit. on p. 77), Erratum: **Phys. Rev. Lett.** **122** (2019) 089901.
- 2847 [88] ATLAS Collaboration, *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s} = 13 \text{ TeV}$* , **JHEP 05** (2019) 124, arXiv: [1811.11028 \[hep-ex\]](https://arxiv.org/abs/1811.11028) (cit. on p. 77).
- 2850 [89] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment*, **JHEP 11** (2018) 040, arXiv: [1807.04873 \[hep-ex\]](https://arxiv.org/abs/1807.04873) (cit. on p. 77).
- 2853 [90] ATLAS Collaboration, *Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector*, **Eur. Phys. J. C** **78** (2018) 1007, arXiv: [1807.08567 \[hep-ex\]](https://arxiv.org/abs/1807.08567) (cit. on p. 77).
- 2856 [91] ATLAS Collaboration, *Reconstruction and identification of boosted di- τ systems in a search for Higgs boson pairs using 13 TeV proton–proton collision data in ATLAS*, **JHEP 11** (2020) 163, arXiv: [2007.14811 \[hep-ex\]](https://arxiv.org/abs/2007.14811) (cit. on p. 78).
- 2859 [92] ATLAS Collaboration, *Search for non-resonant Higgs boson pair production in the $b\bar{b}\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13 \text{ TeV}$* , **Phys. Lett. B** **801** (2020) 135145, arXiv: [1908.06765 \[hep-ex\]](https://arxiv.org/abs/1908.06765) (cit. on p. 78).

- 2862 [93] ATLAS Collaboration, *Search for Higgs boson pair production in the two bottom quarks*
 2863 *plus two final state with in pp collisions at $\sqrt{s} = 13 \text{ TeV}$ with the ATLAS detector,*
 2864 ATLAS-CONF-2021-016, 2021, URL: <https://cds.cern.ch/record/2759683> (cit. on
 2865 p. 78).
- 2866 [94] CMS Collaboration, *Search for heavy resonances decaying to two Higgs bosons in final*
 2867 *states containing four b quarks*, Eur. Phys. J. C **76** (2016) 371, arXiv: [1602.08762](https://arxiv.org/abs/1602.08762)
 2868 [[hep-ex](#)] (cit. on p. 78).
- 2869 [95] CMS Collaboration, *Search for production of Higgs boson pairs in the four b quark*
 2870 *final state using large-area jets in proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$* , JHEP **01**
 2871 (2019) 040, arXiv: [1808.01473](https://arxiv.org/abs/1808.01473) [[hep-ex](#)] (cit. on p. 78).
- 2872 [96] CMS Collaboration, *Combination of Searches for Higgs Boson Pair Production in*
 2873 *Proton–Proton Collisions at $\sqrt{s} = 13 \text{ TeV}$* , Phys. Rev. Lett. **122** (2019) 121803, arXiv:
 2874 [1811.09689](https://arxiv.org/abs/1811.09689) [[hep-ex](#)] (cit. on p. 78).
- 2875 [97] G. Bartolini et al., *Performance of the ATLAS b-jet trigger in p p collisions at $\sqrt{s} =$*
 2876 *13 TeV* , tech. rep. ATL-COM-DAQ-2019-150, CERN, 2019, URL: <https://cds.cern.ch/record/2688819> (cit. on p. 79).
- 2878 [98] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 13 \text{ TeV}$*
 2879 *using the ATLAS detector at the LHC*, ATLAS-CONF-2019-021, 2019, URL: <https://cds.cern.ch/record/2677054> (cit. on pp. 79, 169).
- 2881 [99] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43**
 2882 (2016) 023001, arXiv: [1510.03865](https://arxiv.org/abs/1510.03865) [[hep-ph](#)] (cit. on pp. 80, 170).
- 2883 [100] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C**58** (2008) 639, arXiv:
 2884 [0803.0883](https://arxiv.org/abs/0803.0883) [[hep-ph](#)] (cit. on p. 80).
- 2885 [101] 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. 86).

- 2887 [102] A. Carvalho et al., *Higgs pair production: choosing benchmarks with cluster analysis*,
 2888 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. 99).
- 2890 [103] 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. 102).
- 2893 [104] G. V. Moustakides and K. Basioti, *Training Neural Networks for Likelihood/Density
 2894 Ratio Estimation*, 2019, arXiv: [1911.00405 \[eess.SP\]](https://arxiv.org/abs/1911.00405) (cit. on p. 102).
- 2895 [105] T. Kanamori, S. Hido, and M. Sugiyama, *A Least-Squares Approach to Direct Im-
 2896 portance Estimation*, J. Mach. Learn. Res. **10** (2009) 1391, ISSN: 1532-4435 (cit. on
 2897 p. 102).
- 2898 [106] A. Andreassen and B. Nachman, *Neural networks for full phase-space reweighting
 2899 and parameter tuning*, Physical Review D **101** (2020), ISSN: 2470-0029, URL: <http://dx.doi.org/10.1103/PhysRevD.101.091901> (cit. on p. 103).
- 2901 [107] S. Fort, H. Hu, and B. Lakshminarayanan, *Deep Ensembles: A Loss Landscape Per-
 2902 spective*, 2020, arXiv: [1912.02757 \[stat.ML\]](https://arxiv.org/abs/1912.02757) (cit. on p. 104).
- 2903 [108] B. Efron, *Bootstrap Methods: Another Look at the Jackknife*, Ann. Statist. **7** (1979) 1,
 2904 URL: <https://doi.org/10.1214/aos/1176344552> (cit. on p. 161).
- 2905 [109] ATLAS Collaboration, *Jet energy scale and resolution measured in proton–proton
 2906 collisions at $\sqrt{s} = 13\text{ TeV}$ with the ATLAS detector*, (2020), arXiv: [2007 . 02645 \[hep-ex\]](https://arxiv.org/abs/2007.02645) (cit. on p. 169).
- 2908 [110] J. Adelman et al., *Search for Higgs boson pair production in the $b\bar{b}\gamma\gamma$ final state with
 2909 the full Run 2 13 TeV pp collision data collected by the ATLAS Experiment Supporting
 2910 note.*, tech. rep., CERN, 2020, URL: <https://cds.cern.ch/record/2711865> (cit. on
 2911 p. 169).

- 2912 [111] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring*
 2913 in *ATLAS*, JINST **13** (2018) P07017. 33 p, URL: <https://cds.cern.ch/record/2633501> (cit. on p. 169).
- 2915 [112] D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature*
 2916 of the Higgs Sector, CERN Yellow Reports: Monographs, 869 pages, 295 figures, 248 ta-
 2917 bles and 1645 citations. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/>
 2918 CERN, 2016, URL: <https://cds.cern.ch/record/2227475> (cit. on p. 170).
- 2919 [113] A. L. Read, *Presentation of search results: the CLs technique*, Journal of Physics G:
 2920 Nuclear and Particle Physics **28** (2002) 2693, ISSN: 0954-3899 (cit. on p. 185).
- 2921 [114] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic formulae for likelihood-*
 2922 *based tests of new physics*, The European Physical Journal C **71** (2011), ISSN: 1434-6052
 2923 (cit. on p. 185).
- 2924 [115] *Summary of non-resonant and resonant Higgs boson pair searches from the AT-*
 2925 *LAS experiment*, tech. rep., All figures including auxiliary figures are available at
 2926 <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB->
 2927 2021-031: CERN, 2021, URL: <http://cds.cern.ch/record/2777013> (cit. on pp. 193,
 2928 194).
- 2929 [116] A. Sherstinsky, *Fundamentals of Recurrent Neural Network (RNN) and Long Short-*
 2930 *Term Memory (LSTM) network*, Physica D: Nonlinear Phenomena **404** (2020) 132306,
 2931 ISSN: 0167-2789, URL: <http://dx.doi.org/10.1016/j.physd.2019.132306> (cit. on
 2932 p. 215).
- 2933 [117] M. Zaheer et al., *Deep Sets*, 2018, arXiv: [1703.06114 \[cs.LG\]](https://arxiv.org/abs/1703.06114) (cit. on p. 215).
- 2934 [118] J. Zhou et al., *Graph Neural Networks: A Review of Methods and Applications*, 2021,
 2935 arXiv: [1812.08434 \[cs.LG\]](https://arxiv.org/abs/1812.08434) (cit. on p. 215).
- 2936 [119] A. Vaswani et al., *Attention Is All You Need*, 2017, arXiv: [1706.03762 \[cs.CL\]](https://arxiv.org/abs/1706.03762)
 2937 (cit. on p. 215).

- 2938 [120] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, <http://www.deeplearningbook.org>, MIT Press, 2016 (cit. on p. 216).
- 2939
- 2940 [121] I. Kobyzev, S. Prince, and M. Brubaker, *Normalizing Flows: An Introduction and*
2941 *Review of Current Methods*, IEEE Transactions on Pattern Analysis and Machine
2942 Intelligence (2020) 1, ISSN: 1939-3539, URL: <http://dx.doi.org/10.1109/TPAMI.2020.2992934> (cit. on p. 221).
- 2943
- 2944 [122] L. Dinh, J. Sohl-Dickstein, and S. Bengio, *Density estimation using Real NVP*, 2017,
2945 arXiv: [1605.08803 \[cs.LG\]](https://arxiv.org/abs/1605.08803) (cit. on p. 222).
- 2946 [123] C. Durkan, A. Bekasov, I. Murray, and G. Papamakarios, *Neural Spline Flows*, 2019,
2947 arXiv: [1906.04032 \[stat.ML\]](https://arxiv.org/abs/1906.04032) (cit. on p. 222).
- 2948 [124] F. Pedregosa et al., *Scikit-learn: Machine Learning in Python*, Journal of Machine
2949 Learning Research 12 (2011) 2825 (cit. on p. 224).

2950

Appendix A

2951

OVERVIEW OF OTHER $b\bar{b}b\bar{b}$ CHANNELS

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

2958

A.0.1 Resonant: Boosted Channel

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

The boosted analysis instead reconstructs H decays as large radius, $R = 1.0$ jets, with corresponding b -quarks identified with variable radius subjets, that is jets with a radius that scales as ρ/p_T , the p_T is that of the jet in question, and ρ is a fixed parameter, here chosen to be 30 GeV, which is optimized to maintain truth-level double b-labeling efficiency across the full range of Higgs jet p_T [77].

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

- 2970 • 4 b category: two b -tagged jets in each
- 2971 • 2 $b - 1$ category: two b -tagged jets in one, one in the other

- 2972 • $1b - 1$ category: one b -tagged jet in each

2973 The analysis then proceeds in each of these categories.

2974 The resolved and boosted channels are combined for resonance masses from 900 GeV to
2975 1.5 TeV inclusive. To keep the channels statistically independent, the boosted channel vetoes
2976 events passing the resolved analysis selection.

2977 A.0.2 *Non-resonant: VBF Channel*

2978 The vector boson fusion channel is only considered for the non-resonant search. While the
2979 sensitivity is in general much more limited than the gluon-gluon fusion analysis due to the
2980 much smaller production cross section, VBF is sensitive to a variety of Beyond the Standard
2981 Model physics, both complementary and orthogonal to the theoretical scope of gluon-gluon
2982 fusion.

2983 The VBF channel proceeds very similarly to the ggF, with the primary differences being
2984 the kinematic selections and the categorization, which are impacted by the presence of two
2985 *VBF jets*, resulting from the two initial state quarks. The ggF channel result presented here
2986 includes a veto on VBF events, such that if events pass the full VBF selection, they are not
2987 included in the set of events considered for the ggF result.

2988 Beginning with the assumption of four HH jets and two VBF jets, the VBF channel first
2989 requires an event to have a minimum six jets. The VBF jets are reconstructed as the two jets
2990 with the highest di-jet invariant mass, m_{jj} , out of the set of all non-tagged jets in the event.
2991 If no such pair exists (i.e., there are less than two non-tagged jets), the event is placed in the
2992 ggF channel. To reduce the number of background events, three cuts are then applied, VBF
2993 jets are required to have $\Delta\eta > 3$ and a combined invariant mass of $m_{jj} > 1000$ GeV. HH
2994 jets are identified as in the ggF channel, and the vector sum of the p_T of the HH and VBF
2995 jets is required to be less than 65 GeV. The remainder of the analysis proceeds similarly to
2996 the ggF channel, and events failing any stage of this selection are considered for ggF.

2997 Note that the background estimation for the VBF channel is inherited from the resonant

²⁹⁹⁸ and ggF analyses, a significant additional contribution of this thesis work.

2999

Appendix B

3000

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

3001 The searches presented in this thesis make use of a large suite of sophisticated techniques,
 3002 selected through careful study and validation. During this process, a variety of interesting
 3003 directions for the $HH \rightarrow b\bar{b}b\bar{b}$ analysis were explored by this thesis author, in collaboration
 3004 with a few others¹, but were not used due to a variety of constraints. We present two
 3005 such interesting directions here, with the hope of encouraging further exploration of these
 3006 techniques in future work.

3007 **B.1 pairAGraph: A New Method for Jet Pairing**

3008 As discussed in Chapter 7, one of the main problems to solve is the pairing of b -jets into
 3009 Higgs candidates. Figure 7.1 demonstrates that the choice of the pairing method, while
 3010 important for achieving good reconstruction of signal events, also significantly impacts the
 3011 structure of non- HH events, leading to various biases in the background estimate. Evaluation
 3012 of the pairing method therefore must take both of these factors into account. While we have
 3013 presented some advantages in respective contexts for the pairing methods considered here,
 3014 we of course would like to explore further improvements to this important component of the
 3015 analysis.

3016 To that end, we note that all of the pairing methods considered here share a common
 3017 feature: four jets are selected, and the pairing is some discrimination between the available
 3018 three pairings of these four jets. For the methods used in this analysis, the jet selection
 3019 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.

3020 tagged jets.

3021 With the advent of a variety of machine learning methods for dealing with a variable number
 3022 of inputs (e.g. recurrent neural networks [116], deep sets [117], graph neural networks [118],
 3023 and transformers [119]), a natural place to improve on the pairing is to consider more than
 3024 just four jets. The pairing and jet selection is then performed simultaneously, allowing for
 3025 the incorporation of more event information in the pairing decision and the incorporation of
 3026 jet correlation structure in the jet selection.

3027 In practice, the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets which pass the
 3028 kinematic preselection, and any gain from this additional freedom would come from events
 3029 with greater than or equal to five jets. However, this five jet topology is particularly exciting
 3030 for scenarios such as events with initial state radiation (ISR), in which the $HH \rightarrow 4b$ jets are
 3031 offset by a single jet with p_T similar in magnitude to that of the $HH \rightarrow 4b$ system. Such
 3032 events have explicit event level information which is not encoded with the inclusion of only
 3033 the $HH \rightarrow 4b$ jets, and are pathological if the ISR jet happens to pass b -tagging requirements.

3034 Additionally, with the use of lower tagged regions for background estimation and alternate
 3035 signal regions, this extra flexibility in jet selection may provide a very useful bias – since the
 3036 algorithm is trained on signal, the selected jets for the pairing will be the most “4b-like” jets
 3037 available in the considered set.

3038 For the studies considered here, a transformer [119] based architecture is used. This is
 3039 best visualized by considering the event as a graph with jets corresponding to nodes and edges
 3040 corresponding to potential connections – for this reason, we term this algorithm “pairAGraph”.
 3041 The approach is as follows: each jet, i , is represented by some vector of input variables, \vec{x}_i ,
 3042 in our case the four-vector information, (p_T, η, ϕ, E) of each jet, plus information on the
 3043 b -tagging decision. A multi-layer perceptron (MLP) is used to create a latent embedding,
 3044 $\mathbf{h}(\vec{x}_i)$, 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

$$\vec{z}_i = \sum_j w_{ij} \mathbf{h}(\vec{x}_j) \quad (\text{B.1})$$

3045 where j runs over all jets in the event (including $i = j$), the w_{ij} can be thought of as edge
 3046 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:

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

3047 where subscripts $1a$ and $1b$ correspond to the two jets in pair 1, $2a$ and $2b$ to the jets in pair
 3048 2 for a given pairing of four distinct jets.

3049 This similarity metric is calculated for all possible pairings, which are then passed through
 3050 a softmax [120] activation function, which compresses these scores to between 0 and 1 with
 3051 sum of 1, lending an interpretation as probability of each pairing.

3052 In training, the ground truth pairing is set by *truth matching* jets to the b -jets in the
 3053 HH signal simulation – a jet is considered to match if it is < 0.3 in ΔR away from a b -jet in
 3054 the simulation record. Given this ground truth, a cross-entropy loss *TODO: cite* is used on
 3055 the softmax outputs, and w_{ij} and \mathbf{h} are updated correspondingly. Training in such a way
 3056 corresponds to updating w_{ij} and \mathbf{h} to maximize the similarity metric for the correct pairing.

3057 In evaluation, the pairings with a higher score (and therefore higher softmax output)
 3058 given the trained h and w_{ij} therefore correspond to the pairings that are most “ HH -like”.
 3059 The maximum over these scores is therefore the pairing used as the predicted result from the
 3060 algorithm.

3061 Because the majority of $HH \rightarrow b\bar{b}b\bar{b}$ events have either four or five jets, it was found to
 3062 be sufficient to only consider a maximum of 5 jets. Consideration of more is in principle
 3063 possible, but the quickly expanding combinatorics leads to a rapidly more difficult problem.
 3064 The jets considered are the five leading jets in p_T . Notably, this set of jets may include jets
 3065 which are not b -tagged, even for the nominal 4 b region – therefore events with 4 b -tagged jets

3066 are not required to use all of them in the construction of Higgs candidates, in contrast to the
 3067 other algorithms used in this thesis.

3068 A comparison of the pairAGraph jet selection with the baseline selection used in Chapter
 3069 7 is considered in Table B.1 for the MC16a Standard Model non-resonant signal. As a
 3070 reminder, the baseline selection orders jets by p_T , selecting first the highest p_T b -tagged jets
 3071 (according to the b -tag region definition) and then the highest p_T non-tagged jets. The first
 3072 four jets in this ordering are used.

3073 For the comparison presented in Table B.1, only the leading five jets are considered in
 3074 applying both algorithms in order to compare results on more equal footing. The numbers
 3075 shown are the percent of the time that the correct jets are selected for the Higgs candidates
 3076 by each algorithm, given that the correct jets fall within these leading five jets, where “correct”
 3077 here means truth matched to the corresponding b -quarks. pairAGraph demonstrates a slight
 3078 improvement over the baseline for $4b$, which widens when considering lower b -tag categories.
 3079 Given that four b -quarks are present in all of these categories, this suggests that pairAGraph
 3080 is able to recover information in the case of, e.g., mis-tagged jets.

3081 Table B.2 compares the HH pairing accuracy of a few different pairing algorithms for
 3082 the Standard Model signal. Notably, pairAGraph demonstrates a higher pairing accuracy
 3083 immediately after paring, but all methods are quite comparable after the full analysis selection.
 3084

3085 As mentioned in Chapter 7, though the pairing is quite important for signal events, it also
 3086 must be applied to events in data, where the overwhelming majority of events do not contain
 3087 HH . Though in general, pairing methods select for an HH -like topology, the additional
 3088 flexibility of pairAGraph to choose which jets enter the candidate HH system provides an
 3089 additional handle to shape the kinematics of events in data. Examples of this impact are
 3090 seen in Figures B.1 and B.2, which compare the $2b$ and $4b$ distributions of p_T of the HH
 3091 candidate system between BDT pairing and pairAGraph pairing before and after reweighting.
 3092 HH p_T was chosen as it is a variable which demonstrates both a large difference between
 3093 $2b$ and $4b$ and a residual mis-modeling after reweighting. As can be seen in Figure B.1, the

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

Table B.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 7.1, 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 B.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.

3094 *2b* and *4b* distributions are more similar before reweighting with pairAGraph. Figure B.2
 3095 further shows that the residual mis-modeling after reweighting is reduced, along with the
 3096 corresponding uncertainty. While this is not fully conclusive, it provides a hint that the jets
 3097 chosen for the *2b* event *HH* candidate system may be more “*4b-like*” than the jets chosen
 3098 with the baseline selection.

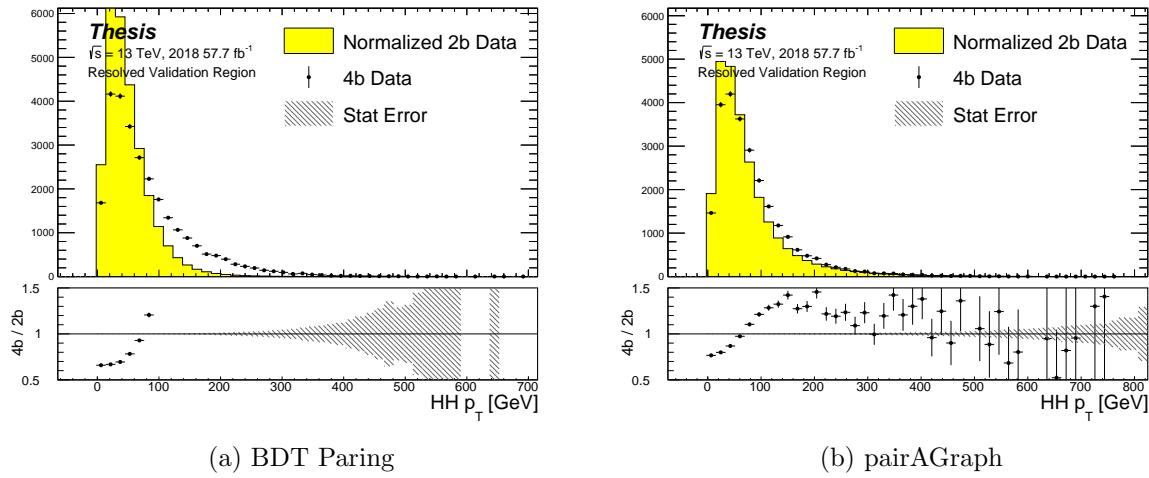


Figure B.1: Comparison of distributions of *HH* p_T in the 2018 resonant validation region before reweighting for BDT pairing (left) and pairAGraph (right). *HH* p_T is a variable with a large difference between *2b* and *4b*, but the relative shapes seem to be more similar for pairAGraph than for BDT paring, corresponding to the hypothesis that pairAGraph chooses more “*4b-like*” jets.

3099 **B.2 Background Estimation with Mass Plane Interpolation**

3100 The choice of a pairing algorithm that results in a smooth mass plane (such as $\min \Delta R$)
 3101 opens up a variety of options for the background estimation. While the method based on
 3102 reweighting of *2b* events used for this thesis performs well and has been extensively studied
 3103 and validated, it also relies on several assumptions. In particular, the reweighting is derived

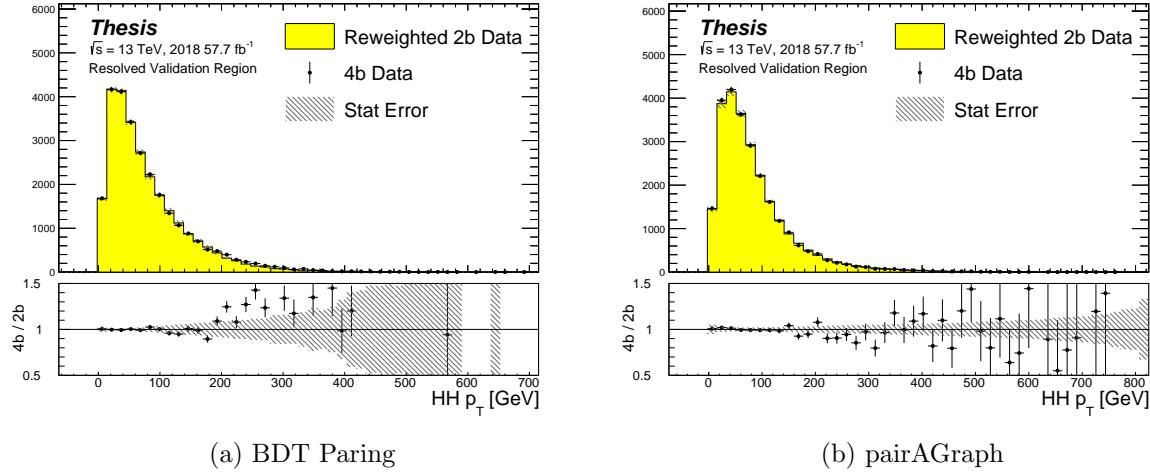


Figure B.2: Comparison of distributions of $HH p_T$ in the 2018 resonant validation region after reweighting for BDT pairing (left) and pairAGraph (right). $HH p_T$ is a variable with a large difference between $2b$ and $4b$, and the reweighted agreement in the high p_T tail is significantly improved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

3104 between e.g., $2b$ and $4b$ events *outside* of the signal region and then applied to $2b$ events *inside*
 3105 the signal region, with the assumption that the $2b$ to $4b$ transfer function will be sufficiently
 3106 similar in both regions of the mass plane. An uncertainty is assigned to account for the bias
 3107 due to this assumption, but the extrapolation in the mass plane is never explicitly treated in
 3108 the nominal estimate. While the approach of reweighting $2b$ events within the signal region
 3109 does have the advantage of incorporating explicit signal region information (that is, the $2b$
 3110 signal region events), the importance of the extrapolation bias motivates consideration of
 3111 a method that operates within the $4b$ mass plane. This additionally removes the reliance
 3112 on lower b -tagging regions, allowing for the use of, e.g. $3b$ triggers, and future-proofing the
 3113 analysis against trigger bandwidth constraints in the low tag regions.

3114 The pairAGraph pairing method discussed in the previous section was developed concur-

3115 rently with these studies and demonstrates good properties for an interpolated estimate (as
3116 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:

$$p(\vec{x}, m_{H1}, m_{H2}) = p(\vec{x}|m_{H1}, m_{H2})p(m_{H1}, m_{H2}) \quad (\text{B.3})$$

3117 by the chain rule of probability. This means that the full dynamics of events in the HH mass
3118 plane may be described by (1) the conditional probability of considered variables \vec{x} , given
3119 values of m_{H1} and m_{H2} , and (2) the density of the mass plane itself.

3120 We present here an approach which uses normalizing flows [121] to model the conditional
3121 probabilities of events in the mass plane and Gaussian processes to model the mass plane
3122 density. These models are trained in a region around, but not including, the signal region,
3123 and the trained models are then used to construct an *interpolated* estimate of the signal
3124 region kinematics. This approach therefore explicitly treats event behavior within the mass
3125 plane, avoiding the concerns associated with a reweighted estimate. Validation of such a
3126 method, as well as assessing of closure and biases of the method, may be done in alternate
3127 b -tagging or kinematic regions, notably the now unused $2b$ region, results of which are shown
3128 below.

3129 B.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

$$p_X(x) = p_Z(f(x)) \left| \det \left(\frac{d(f(x))}{dx} \right) \right| \quad (\text{B.4})$$

3130 where $\left(\frac{d(f(x))}{dx} \right)$ is the Jacobian of f at x .

3131 The problem of normalizing flows then reduces to (1) choosing sets of f which are both
3132 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 [122]), 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 [123] 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,

$$f_j(x_i) = \frac{a_{ij}x_i^2 + b_{ij}x_{ij} + c_{ij}}{d_{ij}x_i^2 + e_{ij}x_i + f_{ij}} \quad (\text{B.5})$$

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.

3156 In the context of the $HH \rightarrow 4b$ analysis, a neural spline flow is used to model the four
 3157 vector information of each Higgs candidate, conditional on their respective masses. The
 3158 resulting flow is therefore five dimensional, with inputs $x = (p_{T,H1}, p_{T,H2}, \eta_{H1}, \eta_{H2}, \Delta\phi_{HH})$,
 3159 where the ATLAS ϕ symmetry has been encoded by assuming $\phi_{H1} = 0$. Conditional variables
 3160 m_{H1} and m_{H2} are not modeled by the flow, but “come along for the ride”. A standard normal
 3161 distribution in 5 dimensions is used for the underlying prior. Modeling of the four vectors
 3162 was chosen in order to reduce bias from modeling m_{HH} directly.

3163 The trained flow model then gives a model for $p(x|m_{H1}, m_{H2})$ which may be sampled
 3164 from to reconstruct distributions of HH kinematics given values of m_{H1} and m_{H2} .

3165 *B.2.2 Gaussian Processes*

3166 The second piece of this background estimate is the modeling of the mass plane density,
 3167 $p(m_{H1}, m_{H2})$. This is done using Gaussian process regression – note that a similar procedure
 3168 is used to define a systematic in the boosted $4b$ analysis. Generally, Gaussian processes
 3169 are a collection of random variables in which every finite collection of said variables is
 3170 distributed according to a multivariate normal distribution. For the context of Gaussian
 3171 process regression, what we consider is a Gaussian process over function space, that is, for a
 3172 collection of points, x_1, \dots, x_N , the space of corresponding function values, $(f(x_1), \dots, f(x_N))$
 3173 is Gaussian process distributed, that is, described by an N dimensional normal distribution
 3174 with mean μ , covariance matrix Σ .

3175 For a single point, this would correspond to a function space described entirely by a
 3176 normal distribution, with various samples from that distribution yielding various candidate
 3177 functions. For multiple points, a covariance matrix describes the relationship between each
 3178 pair of points – correspondingly, it is represented via a *kernel function*, $K(x, x')$. As, in
 3179 practice, μ may always be set to 0 via a centering of the data, the kernel function fully defines
 3180 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

$$\begin{pmatrix} \vec{y}_1 \\ \vec{y}_2 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} K(X_1, X_1) & K(X_1, X_2) \\ K(X_1, X_2) & K(X_2, X_2) \end{pmatrix} \right) \quad (\text{B.6})$$

3181 where we have used that the kernel function is symmetric and assumed prior mean 0.

By standard conditioning properties of Gaussian distributions,

$$\vec{y}_2 | \vec{y}_1 \sim \mathcal{N}(K(X_2, X_1)K(X_1, X_1)^{-1}\vec{y}_1, K(X_2, X_2) - K(X_2, X_1)K(X_1, X_1)^{-1}K(X_1, X_2)) \quad (\text{B.7})$$

3182 which is the sampling distribution for a Gaussian process given kernel K . In practice, the
3183 mean of this sampling distribution is used as the function estimate, with an uncertainty from
3184 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

$$K(x, x') = \exp \left(-\frac{d(x, x')^2}{2l^2} \right) \quad (\text{B.8})$$

3185 where $d(\cdot, \cdot)$ is the Euclidean distance and $l > 0$ is a length scale parameter. Conceptually, as
3186 distances $d(x, x')$ increase relative to the chosen length scale, the kernel smoothly dies off –
3187 further away points influence each other less.

3188 Coming back to our case of the mass plane, the procedure runs as follows:

- 3189 1. A binned 2d histogram of the blinded mass plane is created in a window around the
3190 “standard” analysis regions. Bins which have any overlap with the signal region are
3191 excluded.
- 3192 2. A Gaussian process is trained using the bin centers, values as training points. The
3193 scikit-learn implementation [124] is used, with RBF kernel with anisotropic length scale
3194 (l is dimension 2). The length scale is initialized to $(50, 50)$ to cover the signal region,

3195 and optimized by minimizing the negative log-marginal likelihood on the training data,
 3196 $-\log p(\vec{y}|\theta)$. Training data is centered and scaled to mean 0, variance 1, and a statistical
 3197 error is included in the fit.

3198 3. The Gaussian process is then used to predict the density $p(m_{H1}, m_{H2})$ in the signal
 3199 region. This may then be sampled from via an inverse transform sampling to generate
 3200 values (m_{H1}, m_{H2}) according to the density (specifically, according to the mean of the
 3201 Gaussian process posterior). Though in principle the Gaussian process sampling is not
 3202 limited to bin centers, this is kept for simplicity, with a uniform smearing applied within
 3203 each sampled bin to approximate the continuous estimate, namely, if a bin is sampled
 3204 from, the returned value is drawn uniformly at random within the sampled bin.

4. The sampling in the previous step can be arbitrary – to set the overall normalization,
 a Monte Carlo sampling of the Gaussian process is done to approximate the relative
 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

$$n_{in} = \frac{n_{out}}{f_{out}} \cdot f_{in}. \quad (\text{B.9})$$

3205 Note that the Monte Carlo sampling procedure is simply a set of samples of the Gaussian
 3206 process from uniformly random values of m_{H1}, m_{H2} , and is the most convenient approach
 3207 given the irregular shape of the signal region.

3208 This procedure results in a generated set of predicted m_{H1}, m_{H2} values for signal region
 3209 background events, along with an overall yield prediction.

3210 B.2.3 The Full Prediction

3211 Given the normalizing flow parametrization of $p(x|m_{H1}, m_{H2})$ and the Gaussian process
 3212 generation of $(m_{H1}, m_{H2}) \sim p(m_{H1}, m_{H2})$ and prediction of the signal region yield, all of the

3213 pieces are in place to construct an interpolation background estimate. Namely

- 3214 1. Gaussian process sampled (m_{H1}, m_{H2}) values are provided to the normalizing flow to
 - 3215 predict the other variables for the Higgs candidate four-vectors. These are used to
 - 3216 construct the HH system (notably m_{HH}).
- 3217 2. These final distributions are normalized according to the predicted background yield.

3218 *B.2.4 Results*

3219 All of the following results use the pairAGraph pairing algorithm, and reweighted results use
3220 the region definitions from the resonant analysis.

3221 The Gaussian process sampling procedure is trained on a small fraction (0.01) of $2b$ data
3222 to mimic the available $4b$ statistics. This fraction of $2b$ data is blinded, and the prediction of
3223 the estimate trained on this blinded region may then be compared to real $2b$ data in the signal
3224 region. The predictions for signal region m_{H1} and m_{H2} individually are shown in Figure B.3,
3225 and the resulting mass planes are compared in Figure B.4. Good agreement is seen.

3226 The $4b$ region is kept blinded for this work, meaning that a direct comparison of the
3227 Gaussian process estimate in the $4b$ signal region is not done. However, a Gaussian process is
3228 trained on the blinded $4b$ region and compared to the corresponding reweighted $2b$ estimate,
3229 trained per the nominal procedures from the analyses above. The predictions for signal
3230 region m_{H1} and m_{H2} individually are shown in Figure B.5, compared to both the control and
3231 validation region derived reweighting estimates, and the resulting signal region mass planes
3232 are compared in Figure B.6. The estimates are seen to be compatible.

3233 The Gaussian process estimate may then be used as an input to the normalizing flow
3234 estimate to form a complete background estimate. Figure B.7 shows such an estimate for the
3235 subsampled $2b$ signal region. Results for the prediction of the normalizing flow with inputs of
3236 real $2b$ signal region m_{H1} and m_{H2} are compared to the results of using Gaussian process
3237 predicted m_{H1} and m_{H2} , and are seen to be consistent, demonstrating the above closure of

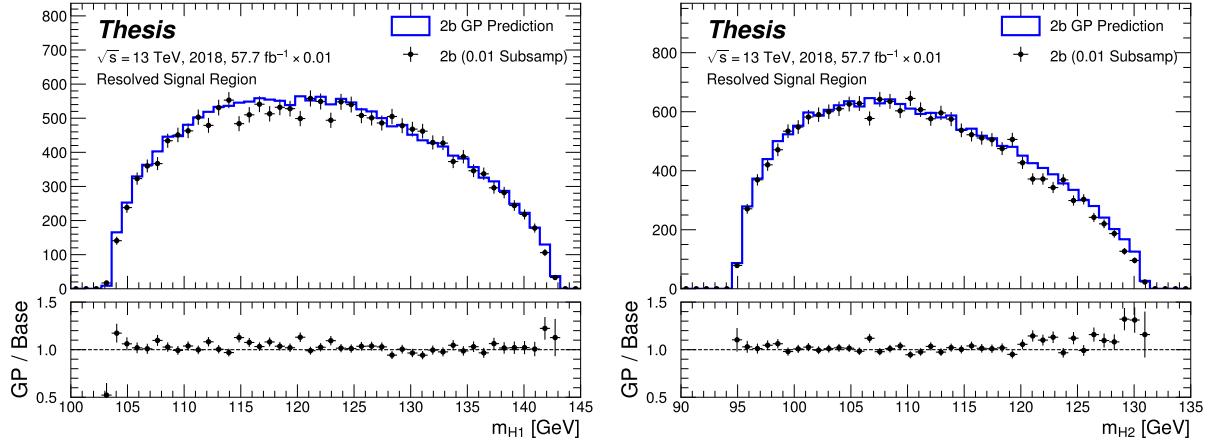


Figure B.3: Gaussian process sampling prediction of marginals m_{H_1} and m_{H_2} for $2b$ signal region events compared to real $2b$ signal region events for the 2018 dataset. Good agreement is seen. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics.

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

Figure B.8 demonstrates the application of this process to the $4b$ region, closely following how such an estimate would be used in the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. As the $4b$ signal region is kept blinded for these studies, no direct evaluation is made, but results are compared to a resonant control region derived reweighting. Both signal region predictions are seen to be comparable, though there are some systematic differences. However, only the nominal estimates are compared here, with assessment of uncertainties on the interpolated estimate left for future work.

3247 B.2.5 Outstanding Points

While good performance is demonstrated from the nominal interpolated background estimate, various uncertainties must be assigned according to the various stages of the estimate. These

3250 notably include

3251 • Assessing a statistical uncertainty on the normalizing flow training (cf. bootstrap
3252 uncertainty).

3253 • Propagation of the Gaussian process uncertainty through the sampling procedure.

3254 • Validation of the resulting estimate and assessment of necessary systematic uncertainties
3255 (e.g. from validation region non-closure).

3256 These are all quite tractable, but some, especially the choice of an appropriate systematic
3257 uncertainty, are certainly not obvious and require detailed study. In this respect, the
3258 reweighting validation work of the non-resonant analysis is certainly quite useful as a starting
3259 place in terms of the available regions and their correspondence to the nominal $4b$ signal
3260 region.

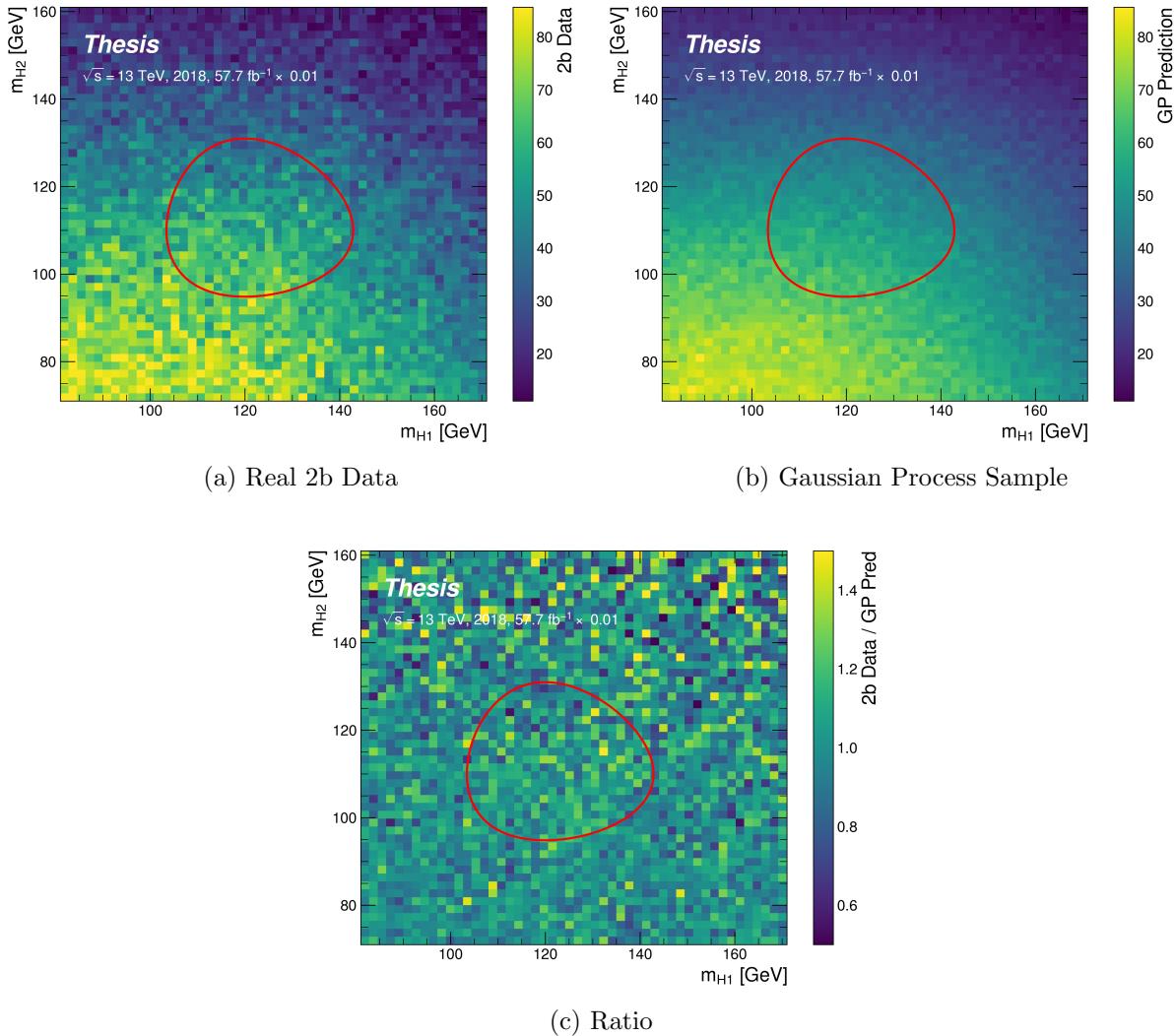


Figure B.4: Gaussian process sampling prediction for the mass plane compared to the real $2b$ dataset for 2018. Only a small fraction (0.01) of the $2b$ dataset is used for both training and this final comparison to mimic $4b$ statistics. Good agreement is seen.

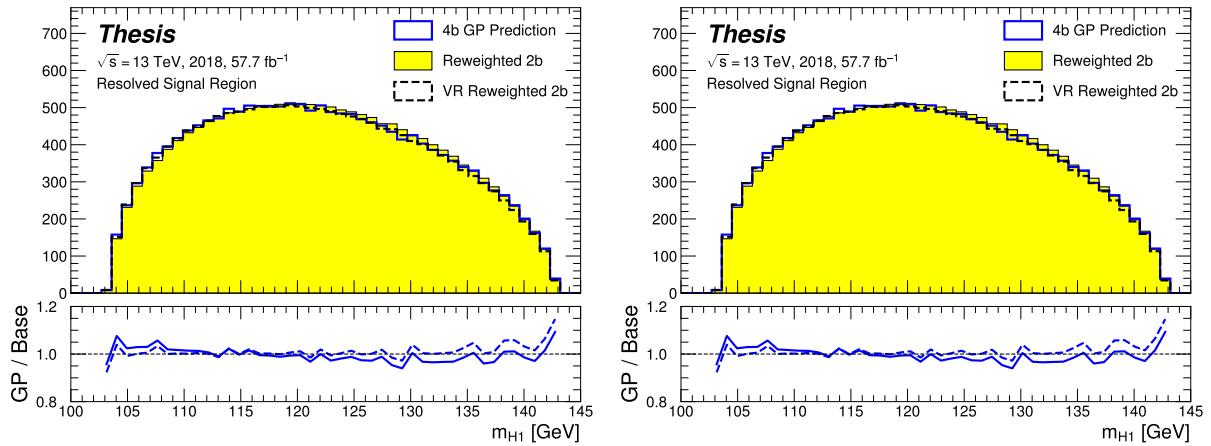


Figure B.5: Gaussian process sampling prediction of marginals m_{H1} and m_{H2} for 4b signal region events compared to both control and validation reweighting predictions. While there are some differences, the estimates are compatible.

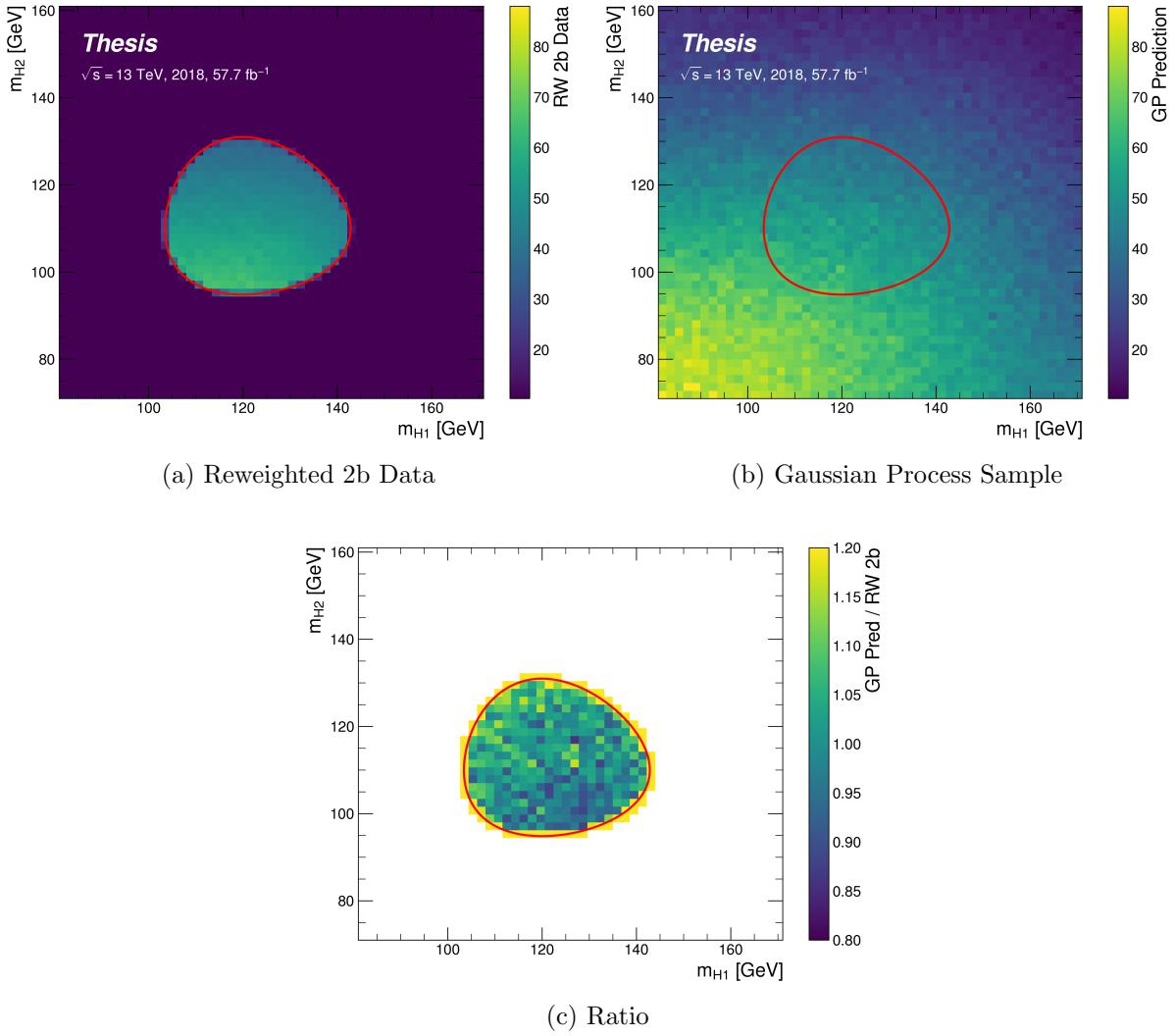


Figure B.6: Gaussian process sampling prediction for the $4b$ mass plane compared to the reweighted $2b$ estimate in the signal region. Both estimates are compatible.

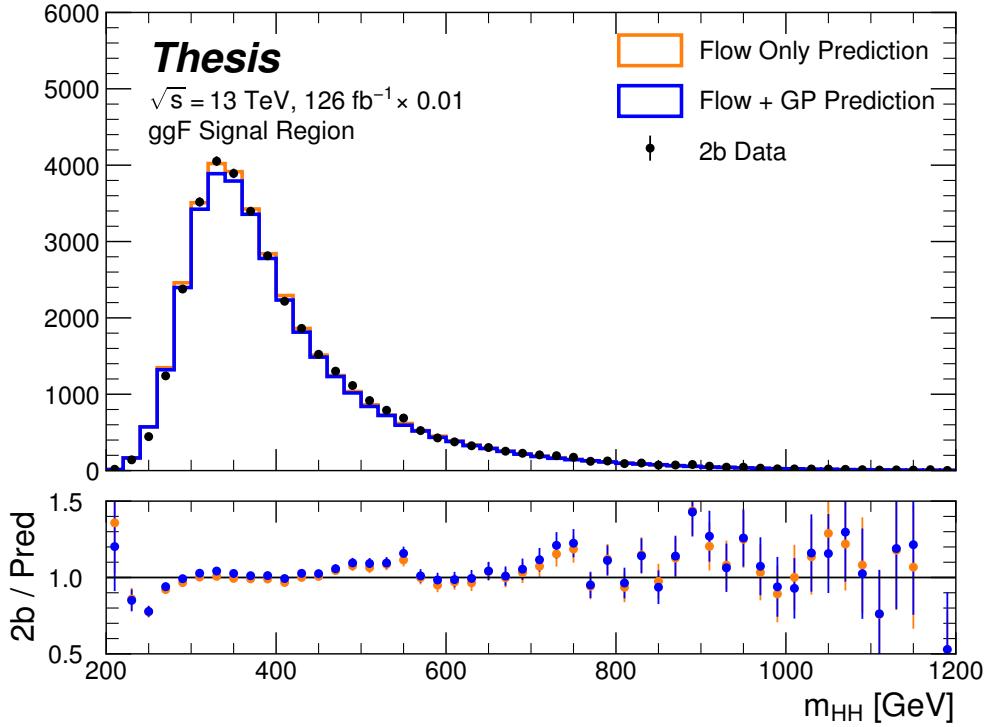


Figure B.7: Comparison of the interpolation background estimate with real 2b data in the signal region. Only 1 % of 2b data is used in order to mimic 4b statistics, and results are presented here summed across years. The “Flow Only” prediction uses samples of actual 2b signal region data for the input 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.

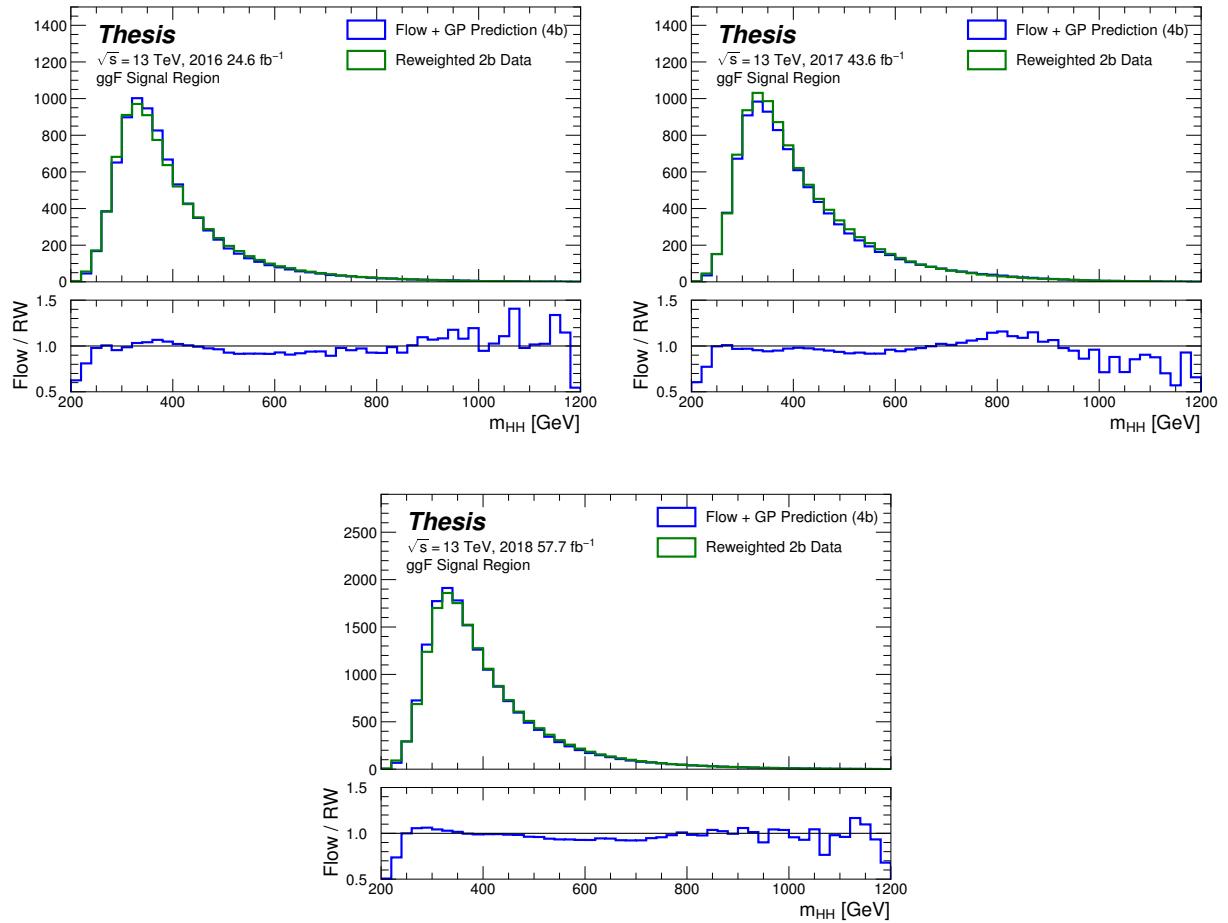


Figure B.8: Comparison of the interpolation background estimate in the $4b$ signal region with the control region derived reweighted 2b estimate, shown for each year individually. Results are generally similar, within around 10 %.