©Copyright 2025
Charles Lewis

Searches for Higgs Boson Pair Production in the $b\bar{b}b\bar{b}$ Final State with the ATLAS Detector -or- May the 4b With You

Charles Lewis

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

Doctor of Philosophy

University of Washington

2025

Reading Committee:

Anna Goussiou, Chair

Shih-Chieh Hsu

Gordon Watts

Program Authorized to Offer Degree:
Physics

University of Washington

Abstract

Searches for Higgs Boson Pair Production in the $b\bar{b}b\bar{b}$ Final State with the ATLAS Detector -or- May the $4b$ With You

Charles Lewis

Chair of the Supervisory Committee:
Anna Goussiou
Physics

This thesis presents searches for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state using data collected by the ATLAS detector from proton-proton collisions at $\sqrt{s} = 13$ TeV during the entirety of LHC Run 2 and at $\sqrt{s} = 13.6$ TeV over a portion of Run 3. Three distinct analyses are detailed. The first analysis presents a search for scalar resonances in the asymmetric production of two scalar bosons using the full ATLAS Run 2 dataset. This Beyond-the-Standard-Model (BSM) scenario posits the production of a heavy Higgs-like resonance X via gluon-gluon fusion, which then decays into a Standard Model Higgs boson (H) and a lighter scalar (S). This study, representing the first ATLAS search in this decay topology, sets upper limits on the cross section for resonant scenarios with m_X between 200 GeV and 3 TeV and m_S from 70 GeV to 2.5 TeV. No significant excesses were observed. The resulting limits are competitive with a corresponding CMS search, which reported its highest excess with a local significance of 4.1σ (global 2.8σ) at $(m_X, m_S) = (700, 400)$ GeV. No such excess was found in this ATLAS search at a similar mass hypothesis of $(m_X, m_S) = (750, 400)$ GeV. The second analysis targets non-resonant Higgs boson pair production, probing both Standard Model HH production and potential deviations in the Higgs trilinear self-coupling using the Run 2 data. No significant evidence for Standard Model HH production is observed, but stringent upper limits on the cross section for Higgs boson pair production via gluon-gluon fusion and vector-boson fusion are established at 5.4 (8.1) times the Standard

Model prediction for observed (expected) yields. These results mark a substantial improvement in sensitivity relative to earlier Run 2 searches, positioning the $b\bar{b}b\bar{b}$ final state as the third most sensitive channel in the ATLAS Run 2 HH combination. Furthermore, cross section limits are set across a range of Higgs self-coupling values, parameterized by $\kappa_\lambda = \lambda_{HHH}/\lambda_{HHH}^{SM}$, with the observed interval $-3.9 \leq \kappa_\lambda \leq 11.1$ ($-4.6 \leq \kappa_\lambda \leq 10.8$ expected). The third analysis addresses non-resonant HH production through gluon-gluon fusion in the resolved topology, incorporating both advanced reconstruction and analytical techniques for the Run 2 dataset and extending to include partial Run 3 data from 2022 and 2023. While this study is ongoing, with further optimization and systematic studies underway, the core analysis workflow is established. For the most sensitive discriminant, the expected statistics-only upper limit on the HH production cross section is 2.80 times the SM prediction, reflecting a notable gain in sensitivity compared to the Run 2 stat-only result. This corresponds to an 18% improvement beyond what would be anticipated from the increase in luminosity alone. Cross section limits are again set as a function of κ_λ, which is constrained to $-2.86 \leq \kappa_\lambda \leq 10.02$ in the expected, stat-only scenario. Finally, this thesis also describes new methods and results for deriving dedicated jet calibrations and systematic uncertainties in the context of ATLAS fast calorimeter simulation.

TABLE OF CONTENTS

Page

List of Figures iv

Chapter 1: Introduction 1

Chapter 2: Theoretical Foundations 4
- 2.1 The Standard Model 4
- 2.2 Gauge Symmetry and the Higgs Mechanism 8
- 2.3 Experimental Verification of the Standard Model 11
- 2.4 Unanswered Questions 13

Chapter 3: The ATLAS Experiment 16
- 3.1 The Large Hadron Collider 16
- 3.2 The ATLAS Detector 25
- 3.3 Trigger and Data Storage 32
- 3.4 Processing of Simulated and Real Data 33

Chapter 4: Jet Calibration in Fast Simulation 36
- 4.1 Jet Definitions 36
- 4.2 AtlFast3 Simulation 42

Chapter 5: Higgs Boson Pair Production 55
- 5.1 Theoretical Motivations 55
- 5.2 Production Modes 57
- 5.3 Experimental Results 64

Chapter 6: Common Aspects to $b\bar{b}b\bar{b}$ Analyses 70
- 6.1 b-Tagging 70
- 6.2 Analysis Topologies 74
- 6.3 Background Composition 76
- 6.4 Background Estimation Method 79

6.5 Systematic Uncertainties . . . 84
6.6 Limit Setting . . . 85

Chapter 7: Resonant $X \to SH \to b\bar{b}b\bar{b}$ With Run 2 Data . . . 88
7.1 Data and Monte Carlo Datasets . . . 88
7.2 Triggers . . . 92
7.3 Event Selection and Reconstruction . . . 95
7.4 Background Estimation . . . 111
7.5 Background Systematics . . . 119
7.6 Results . . . 129

Chapter 8: Non-resonant $HH \to b\bar{b}b\bar{b}$ With Run 2 Data . . . 142
8.1 Background Estimation . . . 142
8.2 Two-Dimensional Discriminants . . . 152
8.3 Results . . . 156
8.4 Areas for Improvement . . . 158

Chapter 9: Non-resonant $HH \to b\bar{b}b\bar{b}$ With Run 2 and Partial Run 3 Data . . . 161
9.1 Data and Monte Carlo Datasets . . . 161
9.2 Triggers . . . 162
9.3 Event Selection and Reconstruction . . . 164
9.4 Background Estimation . . . 168
9.5 Preliminary Results . . . 180

Chapter 10: Conclusion . . . 184
10.1 Summary . . . 184
10.2 Future Outlook . . . 185
10.3 Contributions . . . 189

Appendix A: Low-Mass Events in the $HH \to b\bar{b}b\bar{b}$ Analysis . . . 210
A.1 Motivation . . . 210
A.2 Pairing . . . 210
A.3 Potential Sensitivity Gain . . . 213
A.4 NN Reweighting . . . 214
A.5 Outlook . . . 216

Appendix B: BDT for Pairing in $X \to SH \to b\bar{b}b\bar{b}$ 218
B.1 BDT Pairing and Training 218
B.2 Reconstructed Mass Resonances 219

Appendix C: Comparison of Asymmetric and Symmetric Scalar Resonance Searches 223

LIST OF FIGURES

Figure Number Page

2.1 The fundamental particles of the Standard Model illustrating the different classifications and generations of matter and the force carriers that mediate their interactions [15]. 5

2.2 Distributions used in the Higgs boson discovery: (a) the four-lepton invariant mass $m_{4\ell}$ from the ATLAS analysis of $H \to ZZ^* \to 4\ell$, showing data compared to the background expectation and the signal prediction for a 125 GeV Higgs, combining $\sqrt{s} = 7$ TeV and 8 TeV data; (b) the diphoton invariant mass distribution from the CMS analysis of $H \to \gamma\gamma$, with events weighted by $S/(S+B)$ to enhance signal sensitivity [40, 41]. 13

2.3 SM Higgs branching ratios and their uncertainties for possible Higgs boson masses in the range from 90 - 200 GeV [42]. 14

3.1 A view of the LHC in its underground tunnel, alongside a cross sectional illustration of the beam pipe. The diagram shows the two separate beam pipes, which carry protons in opposite directions, and highlights the surrounding dipole magnets used to steer the beams. 18

3.2 A general diagram illustrating the path that protons (or lead in the case of the ion collisions) take to reach the ultra-relativistic velocities for colliding in the CERN accelerator complex. 19

3.3 Total integrated luminosity delivered by the LHC to ATLAS in Runs 1, 2, and 3 (to date) for the pp collisions. Each row corresponds to the total luminosity collected during the respective data-taking years: 2010–2012 (Run 1), 2015–2018 (Run 2), and 2022–2024 (Run 3 to date) [50]. 22

3.4 Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy pp collisions for each data-taking year. Note that data-taking for 2025 is ongoing [51]. 23

3.5 The typical coordinate system used in analyzing collisions at the LHC in 3.5a based on the four-vector of (p_T, η, ϕ, m). The pseudorapidity, η, is related to the polar angle in 3.5b. 24

3.6 A layout of the ATLAS subdetector systems in 3.6a with two people shown left for size. In 3.6b is myself standing in front of the ATLAS Detector during one of the winter shutdowns when cavern visits are allowed. 26

3.7 Overview of ATLAS tracking and magnet systems. (a) The Inner Detector (ID), composed of the Pixel Detector, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT), provides high-precision tracking of charged particles. (b) The eight superconducting coils of the Barrel Toroid magnet, part of the ATLAS toroidal system, generate a magnetic field used primarily to bend muon trajectories in the outer detector region. Together with the Central Solenoid, these systems are essential for tracking and identifying particles across a wide momentum range. 27
3.8 Illustrations of the two main calorimeter systems in the ATLAS detector. (a) View of the accordion-shaped structure of alternating absorber and readout layers of the Electromagnetic Calorimeter. (b) View of the Tile Hadronic Calorimeter, which consists of alternating layers of steel and plastic scintillator tiles. This system is optimized for measuring the energy of hadrons and forms the largest and most massive component of the calorimeter system. . . 29
3.9 ATLAS Muon Spectrometer overview. (a) Schematic of the spectrometer, showing the toroidal magnet system and layered drift chambers surrounding the calorimeters. (b) Diagram of the alignment system that ensures precise chamber positioning for accurate momentum measurements. 30
3.10 A 2D slice of the ATLAS Detector showing the different subdetector systems and how they function for object reconstruction in the events. The dashed lines indicate the neutrinos that pass through undetected. 31

4.1 The different stages of jet development going from the pp collision to the produced partons to stable particle jets to track jets and then the final calorimeter jets. Figure taken from [73]. 37
4.2 The average number of interactions per bunch crossing, $\langle\mu\rangle$, illustrating the pile-up density profiles for the data-taking years of Run 2 (4.2a) and Run 3 (4.2b) [50, 51]. 38
4.3 The mean CPU time per event simulated in Fast Simulation (AF3) and Full Simulation measured in standardized HS23 seconds for six different physics processes in 4.3a. Figure 4.3b show projected CPU consumption for the ATLAS experiment from 2020 to 2036 varies based on R&D investment [90, 91]. 43
4.4 Comparison of reconstructed small-R PFlow jet distributions in AF3 and FullSim for simulated Pythia dijet events. (a) The jet transverse momentum (p_T) distribution, showing that AF3 tends to overestimate p_T, particularly at high energies. (b) The jet pseudorapidity (η) distribution, highlighting discrepancies near calorimeter transition regions due to limitations in the parametric modeling. These examples illustrate key non-closure effects in AF3, motivating further improvements in jet reconstruction performance. . . 44

4.5 A two-dimensional slice of the ATLAS detector showing the calorimeter geometries at specific pseudorapidities along the beam axis. The poor modeling of jet η in AF3 tends to occur near the calorimeter boundaries [84]. 45

4.6 Jet calibration sequence for small-R jets in the ATLAS experiment. Starting from reconstructed jets, successive corrections are applied: pile-up subtraction, a residual pile-up correction, an absolute Monte Carlo–based calibration (also known as η-JES corrections), a Global Sequential Calibration to improve resolution and reduce flavor dependence, and a final in situ correction to align data with simulation [93]. 46

4.7 Residual pile-up correction factors for AF3-simulated jets as a function of $|\eta|$ using Pythia dijet samples from the MC20 campaign. (a) Correction factors α and β obtained from the first iteration, before applying any residual pile-up correction. (b) Correction factors recalculated after the initial correction is applied. Values below 0.03 across all $|\eta|$ regions indicate that pile-up effects have been effectively mitigated. 47

4.8 Jet energy response before and after applying the absolute JES calibration for small-R UFO jets in AF3 simulation. (a) Mean jet energy response, defined as $E_{\text{reco}}/E_{\text{truth}}$, as a function of jet energy and $|\eta|$ prior to calibration, showing significant deviations from unity across the detector. (b) Response after applying the JES correction function $C(E_{\text{reco}})$ obtained from numerical inversion and polynomial fitting. Closure is achieved in the central region ($|\eta| < 2.5$), indicating that the calibrated jet energies now accurately match the corresponding truth-level energies. 48

4.9 The jet p_T resolution for AF3 small-R UFO jets with the application of the global neural network calibration (GNNC) after the previous MC-JES calibration for the Pythia dijets in MC20 as shown for the η ranges of 0.0-0.2 in 4.9a, 0.2-0.7 in 4.9b, 0.7-1.3 in 4.9c, and 1.3-1.8 in 4.9d. All areas of the phase space see great improvement in p_T response and resolution. 49

4.10 The jet p_T response distribution for AF3 small-R UFO jets with the application of the global neural network calibration (GNNC) after the previous MC-JES calibration for the Pythia dijets in MC20 as shown for the jets with 17 GeV $< p_T <$ 30 GeV and $|\eta| < 0.2$ in 4.10a, 30 GeV $< p_T <$ 50 GeV and $1.8 < |\eta| < 2.5$ in 4.10b, 300 GeV $< p_T <$ 400 GeV and $0.2 < |\eta| < 0.7$ in 4.10c, and 800 GeV $< p_T <$ 1000 GeV and $|\eta| < 0.2$ in 4.10d. All areas of the phase space see great improvement in p_T response and resolution. 50

4.11 Jet energy uncertainties as a function of jet η and p_T, derived by comparing FullSim and AF3 Pythia DiJet samples for large-R PFlow jets in the mc20 campaign. (a) Jet Energy Scale (JES) uncertainty, calculated according to Equation 4.2, indicates a general overestimation of the response in AF3 relative to FullSim. (b) Jet Energy Resolution (JER) uncertainty, defined in Equation 4.4, tends to increase with higher p_T at fixed η. Irregular features at high p_T and in forward regions are attributed to limited MC statistics. . . 53

4.12 AF3 JES non-closure uncertainties for PFlow jets in the central region ($0.0 < |\eta| < 0.2$). (a) JES uncertainties dominate for high-p_T jets above 200 GeV. (b) For jets with $80 < p_T < 100$ GeV, the uncertainties are particularly large in the forward region, highlighting AF3's degraded performance near subdetector boundaries, as discussed in Section 4.2. 54

4.13 Jet mass uncertainties as a function of jet mass and p_T for large-R PFlow jets in AF3 (mc20). (a) Jet Mass Scale (JMS) uncertainty, from Equation 4.2, remains generally small except at low jet mass and high p_T. (b) Jet Mass Resolution (JMR) uncertainty, from Equation 4.3, is largest at low p_T and high jet mass. Irregular features in the distributions reflect limited MC statistics in sparsely populated regions of phase space. 54

5.1 The Higgs potential, featured on the left, for $\mu^2 < 0$. Choosing any of the points at the bottom of the potential spontaneously breaks the rotational U(1) symmetry. Quantum corrections, featured on the right, can change the shape of the Higgs potential indicating whether the universe is stable or metastable. Here, the minimum of our vacuum is taken at $|\phi| = \frac{v}{\sqrt{2}}$. [98] . . . 55

5.2 The Feynman diagrams for the non-resonant Higgs boson pair production through gluon-gluon fusion (ggF) where the triangle diagram in 5.2a works through a top loop process including the Higgs self-coupling while the box diagram in 5.2b does not include the self-coupling. These diagrams destructively interfere [101, 102]. 58

5.3 The destructive interference between the triangle and box diagrams of the ggF Higgs boson pair production showing how the distribution for m_{HH}, the invariant mass of the di-Higgs system, changes depending on the Higgs self-coupling value. The green line in 5.3b where $\kappa_\lambda = 1$ indicates the distribution for the Standard Model prediction [103]. 58

5.4 Feynman diagrams illustrating the three main contributions to non-resonant Higgs pair production via vector boson fusion (VBF). The process proceeds through: 5.4a a diagram involving the Higgs self-coupling, proportional to κ_λ; 5.4b a diagram featuring a direct coupling of two Higgs bosons to vector bosons, governed by κ_{2V}; 5.4c a diagram with a single-Higgs coupling to vector bosons, proportional to κ_V [102]. 59

5.5 The branching ratios of the possible decay channels for Higgs pair production for pp collisions at $\sqrt{s} = 13$ TeV. The $b\bar{b}b\bar{b}$ decay channel has the largest branching ratio at 34% [105]. 61

5.6 (a) Expected values (dashed lines) of the test statistic $-2\ln\Lambda$ as a function of κ_λ, shown for the $b\bar{b}\gamma\gamma$ (purple), $b\bar{b}\tau^+\tau^-$ (green), multilepton (cyan), $b\bar{b}b\bar{b}$ (blue), and $b\bar{b}\ell\ell + E_T^{\text{miss}}$ (brown) channels, as well as their combination (black). Observed values from the combined data are indicated by the solid black line [120]. (b) Observed and expected 95% CL upper limits on the signal strength for inclusive ggF HH and VBF HH production from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}b\bar{b}$, multilepton, and $b\bar{b}\ell\ell + E_T^{\text{miss}}$ decay channels, and their combination. The expected limit and its $\pm1\sigma$ and $\pm2\sigma$ bands assume no HH production and profile all nuisance parameters to the observed data [120]. . 65

5.7 (a) The 95% CL upper limits on the inclusive HH cross section as a function of κ_λ when all other couplings are set to values predicted by the SM. The inner (green) band and the outer (yellow) band indicate the 68% and 95% CL intervals [129]. (b) Observed and expected 95% CL upper limits on the signal strength for inclusive ggF HH and VBF HH production from the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}W^+W^-$, multilepton, $\gamma\gamma\tau^+\tau^-$, $b\bar{b}ZZ, 4l$, and $W^+W^-\gamma\gamma$ decay channels, and their combination. The expected limit and its $\pm1\sigma$ and $\pm2\sigma$ bands assume no HH production and profile all nuisance parameters to the observed data [129]. 67

5.8 (a) Local p-value and (b) observed and expected upper limits at the 95% CL on the resonant Higgs boson pair production cross section as a function of the resonance mass m_X [109]. 68

5.9 (a) Observed 95% CL upper limits on the signal cross section times branching fraction for the $X \to SH \to b\bar{b}\gamma\gamma$ signal in the ATLAS Run 2 analysis, in the (mX, mS) plane with the points show where the limits were evaluated [132]. (b) Observed 95% CL upper limits on the signal cross section times branching fraction for the boosted $X \to YH \to b\bar{b}b\bar{b}$ signal in the CMS Run 2 analysis, in the (mX, mS) plane with the boundaries of the excluded regions of the NMSSM and TRSM included [133]. 69

6.1 A b-jet can be identified by tracks that originate from a displaced secondary vertex due to the relatively long lifetime of b-hadrons. These tracks have a large transverse impact parameter (d0), unlike those in light-flavour jets, which typically come directly from the primary vertex [140]. 71

6.2 The c-jet and light-jet rejections of the different ATLAS flavor tagging algorithms over time in the $t\bar{t}$ MC simulation in Figure 6.2a provided at the 70% efficiency b-tagging working point where the improvement is normalized to the DL1 tagger. The background rejection is also plotted against b-jet efficiency for the same samples in Figure 6.2b where the ratio for the different taggers is taken according to the DL1d tagger [141]. 73

6.3 Illustration of possible topological configurations for the $b\bar{b}b\bar{b}$ final state in the SH $\rightarrow 4b$ analysis. In the resolved topology, all four b-quarks hadronize into separate, individually reconstructed small-R jets. In the intermediate topology, one $b\bar{b}$ pair is sufficiently boosted that both quarks are reconstructed within a single large-R jet. In the boosted topology, both $b\bar{b}$ pairs are sufficiently boosted that they are reconstructed within two separate large-R jets. 75

6.4 There are six possible ways to reconstruct four small-R jets into two unique scalars and three possible ways to reconstruct them into two identical scalars. 76

6.5 (a) Flavor composition of the QCD background in the $4b$ signal region of the $HH \rightarrow b\bar{b}b\bar{b}$ analysis. (b) Bin-by-bin normalized version of (a), showing the fractional contribution of different jet flavors. 77

6.6 (a) Distribution of the top veto discriminant X_{Wt} for simulated $t\bar{t}$ background and $X \rightarrow SH \rightarrow b\bar{b}b\bar{b}$ signal hypotheses, showing reduced separation when m_S is near the top quark or W boson mass. (b) Reconstructed W boson and top quark candidate masses in the non-resonant $HH \rightarrow b\bar{b}b\bar{b}$ Control Region (2018 data), using the jet combination with minimum X_{Wt} (see Eq. 6.1). The ellipse marks the region excluded by the X_{Wt} cut. 78

6.7 A general illustration of the data-driven background estimation method employed in the $X \rightarrow SH \rightarrow b\bar{b}b\bar{b}$ analysis. Further technical details of the method and the results will follow in the subsequent chapters. 81

7.1 The Feynman diagram in (a) for the production of the heavy scalar resonance X from gluon-gluon fusion as it decays into a lighter scalar S and a Standard Model Higgs boson, where the S and Higgs bosons each decay into a pair of b-quarks. All signal mass pairs in (b) for the SH4b analysis are shown for the different channels in which they are reconstructed. The resolved channel has its own unique mass pairs analyzed and is also included in a combination scheme with the mixed channel for other mass pairs. See the SH4b Internal Note [145]. 89

7.2 Truth-level distributions from the signal MC validation: (a) mass and (b) transverse momentum of the scalar boson S in addition to (c) $|\Delta\eta_{SH}|$ and (d) the transverse momentum of the Higgs boson for all signal mass pairs at $m_X = 3$ TeV. These checks were used to verify correct event generation and kinematic coverage for the $X \rightarrow SH \rightarrow bbbb$ analysis. 91

7.3 The nominal trigger bucket strategy (Bucket Strategy 1) employed that uses both **2b+1j** and **2b+2j** triggers and results in the highest efficiency across the phase space of the resolved channel. See the SH4b Internal Note [145]. . . 93

7.4 (a) Total trigger efficiency of Bucket Strategy 1 as a function of m_X for $m_S = 70$ GeV, showing the contribution from each individual bucket. Bucket 1 provides higher efficiency at large m_X values, while Bucket 2 is more efficient at lower m_X values. (b) Total efficiency of Bucket Strategy 1 for all signal mass pairs considered up to $m_X = 3000$ GeV. See the SH4b Internal Note [145]. 94

7.5 Schematic representation of the event selection criteria in the resolved channel of the SH4b analysis. Events are required to contain at least four small-R ($R = 0.4$) jets satisfying the specified kinematic and tagging requirements. . . 95

7.6 Assignment of signal mass points to the training groups of the NN (a) and the BDT (b). The compressed group consists of signal points where the mass difference between X and $S+H$ is small, resulting in S and H being produced nearly at rest. The non-compressed group in the BDT and the separated group in the NN contain signal points where both scalars are significantly boosted. The NN includes an additional group, referred to as the middle group, representing mass points whose kinematics are intermediate between the compressed and separated groups. 97

7.7 Distributions of key NN training inputs for the separated group: (a) $\Delta\eta_{jj}$, (b) $\Delta\phi_{jj}$, and (c) ΔR_{jj}, for S and H candidates. The signal corresponds to correctly paired jets, while the background consists of incorrect pairings. The histograms are normalized to account for the fact that there are twice as many incorrect pairings as correct pairings per event. 100

7.8 Reconstructed mass distributions using the NN pairing for $m_X = 400$ GeV: (a) m_H and (b) m_S. Clear mass peaks are observed, with a small excess near 125 GeV in (b) due to events where jets are correctly paired but S and H assignments are swapped. 100

7.9 Ratio of statistical-only 95% confidence level upper limits for $X \to SH \to b\bar{b}b\bar{b}$ between the NN and BDT pairing methods, evaluated using resolved signal samples. The NN pairing shows slightly better sensitivity for some low-mass points, while the BDT provides a substantial improvement at higher masses. 101

7.10 Distribution of ΔR between the two b-quarks from the decays of the scalar S (dashed) and the SM Higgs boson (solid) for signal samples with $m_X = 750$ GeV and m_S as indicated in each panel. See the SH4b Internal Note [145]. . 102

7.11 Illustration of the assignment of S and H candidates to the two b-jet pairs based on their angular separation. The pair with smaller ΔR is assigned to the lighter scalar boson for a given signal mass pair. 103

7.12 The distribution of the top veto discriminant, X_{Wt} with signal samples of (m_X GeV,m_S GeV): (300,70), (400,70), (400,170) and the simulated $t\bar{t}$. Requiring that $X_{Wt} > 1.5$ as an additional kinematic cut would result in a significant loss in signal events for these samples despite the background reduction. 105

7.13 Acceptance × efficiency plots obtained for the trigger, 4b selections, QCD background rejection ($\Delta\eta_{SH}$) selections in SRs defined by $X_{SH} < 1.6$, when using the BDT-paired jets for $m_X = 1000$ GeV and (a) $m_S = 400$ GeV, (b) $m_S = 500$ GeV, (c) $m_S = 750$ GeV. See the SH4b Internal Note [145]. 105

7.14 Example reconstructed mass distributions for selected signal points. The pink solid line shows the median after the 4b selection, with the pink dashed line indicating the same for the initial selection. The blue solid line marks the innermost 68% quantile, while the blue dashed line shows its median. The black dashed line represents the average median across all signals with the same m_S. Orange indicates BDT-paired events with the $\Delta\eta_{SH} < 1.5$ selection, while the green filled area shows the 1D signal region after the X_{SH} selection. a displays m_S for $m_X = 750$ GeV, $m_S = 400$ GeV, and b displays m_H for the same mass point. See the SH4b Internal Note [145]. . . . 107

7.15 Mass plane and kinematic regions for 2018 data, using the BDT pairing in the non-compressed region for a selected number of m'_S mass points considered in the resolved analysis. a shows the m_H-m_S mass plane and kinematic regions for $m'_S > 125$ GeV and b shows this for $m'_S < 125$ GeV 109

7.16 Distribution of $|\Delta\eta_{SH}|$ in (a) and X_{SH} in (b) for the BDT-reconstructed S and H candidates in the signal region for $m'_S = 300$ GeV. The background estimate from the normalizing flow method (brown) is compared with two simulated signal mass points, each normalized to 1 fb. 109

7.17 Distributions of m_H versus m_S for $Z \to b\bar{b}$ production, shown in the four BDT pairing configurations: (a) compressed/minH, (b) compressed/minS, (c) non-compressed/minH, and (d) non-compressed/minS. The signal band (encompassing all sub-signal regions) and control region are indicated. These distributions demonstrate that the $Z \to b\bar{b}$ background is smoothly distributed across the mass plane and does not introduce any problematic resonance within the signal region. 112

7.18 Distributions of kinematic variables of the S and Higgs candidate bosons for the predictions from the normalizing flows in the Control Region for the compressed and non-compressed BDT pairings when $m_S > m_H$. 116

7.19 Distributions of kinematic variables of the SH candidates and event level variables X_{SH} and X_{Wt} for the predictions from the normalizing flows in the Control Region for the compressed and non-compressed BDT pairings when $m_S > m_H$. 117

7.20 Gaussian process regression for the Higgs candidate mass (m_H) distribution in the (a) non-compressed minH and (b) non-compressed minS selection for the nominal 4b data in 2018. Points show data in the CR and SR, while the shaded band show the GP prediction and its uncertainty. 118

7.21 Twenty five predictions of the categorized m_{SH} distributions from the normalizing flows, each conditioned on a different GP sample for the non-compressed signal region (m_X = 400 GeV, m_S = 200 GeV). The average prediction is shown in red. See the SH4b Internal Note [145]. 120

7.22 Eigenvalues from the principal component analysis of the ensemble background prediction for the non-compressed signal region (m_X = 400 GeV, m_S = 200 GeV, 2016–2018 data). The largest eigenvalues correspond to the retained shape systematics. See the SH4b Internal Note [145]. 122

7.23 Leading principal component variations in the m_{SH} distribution from the deep ensemble for the non-compressed region ($m_X = 400$ GeV, $m_S = 200$ GeV). The shaded envelope indicates the standard deviation across the 25 ensemble predictions in each bin. See the SH4b Internal Note [145]. 123

7.24 The 3b non-closure uncertainty as a function of m_{SH}, derived for the non-compressed BDT pairing when $m_X = 400$ GeV, $m_S = 200$ GeV. The up and down variations from the non-closure is shown in red and blue respectively. See the SH4b Internal Note [145]. 124

7.25 Definition of the validation regions used to test the data-driven background estimate. These regions are obtained by inverting kinematic cuts or altering the b-tag multiplicity from the nominal selection. 126

7.26 GP fits to the m_H distribution in the blinded pseudo-signal region for: (a) 3b, $|\Delta\eta_{SH}| < 1.5$ and (b) 4b, $|\Delta\eta_{SH}| > 1.5$ events, both in the non-compressed region ($m_H < m_S$) for 2017 data. Non-closure observed in the 3b region is used to define the associated systematic uncertainty. 127

7.27 The categorized m_{SH} distribution in the 3b signal region for $m_S = 250$ GeV. The six subplots correspond to: (a) Category 0: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.0 < X_{SH} < 0.95$, (b) Category 1: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.0 < X_{SH} < 0.95$, (c) Category 2: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.0 < X_{SH} < 0.95$, (d) Category 3: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.95 < X_{SH} < 1.6$, (e) Category 4: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.95 < X_{SH} < 1.6$, and (f) Category 5: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.95 < X_{SH} < 1.6$. 128

7.28 Pre-fit and post-fit of a background-only profile likelihood fit for the sub-SR of m_X = 400 GeV, m_S = 200 GeV for the 4b nominal predictions and toy data. See the SH4b Internal Note [145]. 132

7.29 The categorized m_{SH} distribution in the unblinded $4b$ signal region for $m_S = 250$ GeV. The six subplots correspond to: (a) Category 0: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.0 < X_{SH} < 0.95$, (b) Category 1: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.0 < X_{SH} < 0.95$, (c) Category 2: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.0 < X_{SH} < 0.95$, (d) Category 3: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.95 < X_{SH} < 1.6$, (e) Category 4: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.95 < X_{SH} < 1.6$, and (f) Category 5: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.95 < X_{SH} < 1.6$. . . . 133

7.30 The major background systematic nuisance parameter pulls for (a) $(m_X, m_S) = (400\text{ GeV}, 250\text{ GeV})$ and (b) $(m_X, m_S) = (750\text{ GeV}, 500\text{ GeV})$. These systematics nuisance parameters include the dominant shape systematics, the $3b$ non-closure uncertainty, and the normalization uncertainty from the GP regression. . . . 135

7.31 The dominant pulls and impacts of the background systematic uncertainties, including the nominal shape systematic uncertainties with the $3b$ non-closure uncertainties, ranked by magnitude of post-fit impact and shown for (a) $(m_X, m_S) = (400\text{ GeV}, 250\text{ GeV})$ and (b) $(m_X, m_S) = (750\text{ GeV}, 500\text{ GeV})$. See the SH4b Internal Note [145]. . . . 136

7.32 Expected (a) and observed (b) 95% CL upper limits on the cross section times the branching ratio for $X \to SH \to b\bar{b}b\bar{b}$ for the resolved signal mass pairs including all background systematic uncertainties. . . . 137

7.33 Observed (solid black line) and expected (dashed black line) 95% CL upper limits on the cross section times the branching ratio. For each unique m_X value, where a scaling factor is applied to the limits so that all mass pairs can be visualized together. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties, respectively. The mass grid used in this analysis allows for interpolation across m_S, except in regions between $m_S = 100$ GeV and $m_S = 170$ GeV. . . . 138

8.1 The (m_{H1}, m_{H2}) mass planes for (a) $2b$ and (b) $4b$ events from 2016-2018 data, showing the SR (red oval) and CR (surrounding bands). The full ggF selection is applied including a kinematic cut for $X_{Wt} > 1.5$ which greatly reduces the $t\bar{t}$ background. . . . 144

8.2 Illustration of the background shape systematic derivation. The top row shows the nominal estimate from CR1, while the bottom row shows the alternative estimate from CR2. The right plot displays the symmetrized shape variation (pink histogram), used to define the "CR12 shape systematic". Variations are defined per $|\Delta\eta_{HH}|$ and X_{HH} category as appropriate. Figure from [155]. . . . 146

8.3 Distributions of m_{HH} in $4b$ data (black points) within Control Region 1 (CR1) of the ggF selection. (a) The yellow histogram shows the $2b$ CR1 m_{HH} distribution normalized to the $4b$ yield; (b) the yellow histogram shows the $2b$ CR1 after NN reweighting. Hatched bands indicate the statistical uncertainty on the 2b sample: Poisson only in (a), and including the bootstrap uncertainty in (b). 147

8.4 Distributions of the reweighted 2b events and 4b data for m_{HH} in Control Region 2 for 2017 data, shown for configurations 1 (with Poisson sampling) and 2 (without Poisson sampling). No significant difference is observed in the average prediction of the resulting distributions. 149

8.5 The bootstrap band procedure, shown as a ratio to the nominal (average) estimate in the CR1. Each lightly colored line is from the m_{HH} prediction for a single bootstrap training while the darker line shows the standard deviation of histograms for its corresponding dataset. All three years are included separately for ggF from configurations 1 (black) and 2 (red). 150

8.6 The bootstrap band procedure, shown as a ratio to the nominal (average) estimate in the CR1. Each lightly colored line is from the m_{HH} prediction for a single bootstrap training while the darker line shows the standard deviation of histograms for its corresponding dataset. Data for 2017 for the ggF analysis (a) and for all three years from the VBF analysis (b) are included for configurations 1 (black) and 3 (red). 151

8.7 The m_{HH} vs X_{Wt} distributions for (a) signal and (b) average estimated background in the SR illustrating the correlations between m_{HH} and X_{Wt}. The transformation into a 1D histogram for fitting purposes (c), shown for these distributions, where the 2D histogram is "rolled out" starting from the bottom-left hand corner to the top-right corner. 153

8.8 Normalized distributions of (a) m_{HH} and (b) X_{Wt} for several representative κ_λ values, and their transformation into a (c) 2D discriminant employed in the statistical fit. The use of a 2D discriminant consistently improves the statistical-only 95% CL limits on κ_λ. The effect of varying κ_λ on the distribution shapes is clearly observed. 155

8.9 Observed and expected 95% confidence level exclusion limits on the non-resonant HH signal strength and constraints on the Higgs boson trilinear self-coupling κ_λ in the resolved $b\bar{b}b\bar{b}$ channel, as published for the Run-2 analysis. The Standard Model value is indicated by the vertical line [102]. . 157

8.10 Performance of the b-jet trigger HLT selections in simulated $HH \to b\bar{b}b\bar{b}$ events. The trigger efficiency, evaluated relative to all inclusive $HH \to b\bar{b}b\bar{b}$ events, is shown as a function of the true di-Higgs boson invariant mass, m_{HH}. The efficiency is presented for the $2b2j^{asym}$ selection and for the logical OR of the $2b2j^{asym}$, $3b1j^{asym}$, $2b1j$, and $2b2j^{cent}$ triggers and is compared with the trigger strategy employed in Run 2 [156]. 159

9.1 Individual trigger efficiencies for non-resonant $HH \to b\bar{b}b\bar{b}$ production in the (a) mc23a and (b) mc23d simulated samples, shown after applying a $4b$ preselection. The preselection requires four jets with $p_T > 25$ GeV, $|\eta_{\text{jet}}| < 2.5$, and GN2 b-tagging at the 77% working point [158]. 163

9.2 Pairing efficiency as a function of the reconstructed m_{HH} for the min ΔR and min $\chi^2 = \sum(m_{jj} - m_X)^2$ pairing algorithms. The min ΔR pairing method, which is the nominal pairing method, is poor at low masses, but very accurate for $m_{HH} > 500$ GeV. 165

9.3 Distributions of simulated $4b$ QCD events for mc23a+d in the $m_{H1} - m_{H2}$ mass plane for the (a) min ΔR pairing and the (b) mass-based pairing. The min ΔR approach leads to much less background in the signal region (red contour). 166

9.4 The (m_{H1}, m_{H2}) mass planes for (a) simulated SM signal (mc23a) and (b) blinded $4b$ data from 2022, showing the signal region (red ellipse) and control region (dashed black box) defined for the analysis. Higgs candidate masses are reconstructed using the min ΔR pairing. 168

9.5 Distributions of the simulated SM mc23a signal and blinded four-b data from 2022, where events in the signal region (SR) are masked, for (a) $|\Delta\eta_{HH}|$ and (b) X_{Wt}. The red vertical lines indicate the kinematic selection cuts: $|\Delta\eta_{HH}| < 1.5$ and $X_{Wt} < 1.5$. The X_{Wt} distribution also shows the simulated $t\bar{t}$ background, demonstrating the effectiveness of the veto in suppressing this background while preserving signal events. 169

9.6 Distributions of 4b data and normalizing flow predictions for all data-taking years (2016–2018, 2022–2023) in the control region: (a) η_{H1}, (b) η_{H2}, (c) $p_{T,H1}$, and (d) $p_{T,H2}$. There is overall great closure between data and prediction. 171

9.7 Distributions of 4b data and normalizing flow predictions for all data-taking years (2016–2018, 2022–2023) in the control region: (a) $|\Delta\phi_{HH}|$, (b) $|\Delta\eta_{HH}|$, (c) m_{HH}, and (d) $p_{T,HH}$. There is overall great closure between data and prediction outside of the boundary of the $|\Delta\eta_{HH}|$ cut. 172

9.8 The distribution of the (a) data and the (b) average GP prediction over the (m_{H1}, m_{H2}) mass plane for the 4b data in 2022 for events where $|\Delta\eta_{HH}| < 1.5$. The 4b data is blinded in the SR (red) of (a). 174

9.9 The additional validation regions used to test the data-driven background estimate. These regions are obtained by inverting the $|\Delta\eta_{HH}| < 1.5$ kinematic cut, altering the b-tag multiplicity (N_{tag}), or both. 175

9.10 Binned distributions in the (m_{H1}, m_{H2}) mass plane for the $3b$ validation region with $|\Delta\eta_{HH}| < 1.5$ in 2022: (a) observed data, (b) GP-based background prediction, (c) ratio of data to prediction, and (d) the difference between data and prediction normalized to the statistical uncertainty in the data. The agreement between data and prediction is excellent across the phase space. The signal region ($X_{HH} < 1.6$) is indicated by the red contour. 176

9.11 The distribution of the data and the background prediction for (a) m_{HH} and (b) $|\Delta\eta_{HH}|$ for the $3b$ validation region with $|\Delta\eta_{HH}| < 1.5$ in 2022 in the SR. The modeling is generally well within the uncertainty of the standard deviation of the 25 predictions outside of the kinematic boundary. 177

9.12 Ratio of the total data yield to the predicted background yield in the signal region for each $2b$ and $3b$ validation region, shown separately for all data-taking years. Panel (a) shows the result when the normalizing flows and Gaussian processes are trained on the full statistics, while panel (b) presents the average ratio from models trained on 10 downsampled subsets (each averaged over 25 seeds). The comparison illustrates the degree of non-closure in each region, with a pronounced under-prediction of the total yield for Run 2 years, reaching up to 10% in the most extreme case, while the agreement is notably improved for the Run 3 years though still not perfect. 178

9.13 (a) Expected 95% CL upper limits on the signal strength for non-resonant $HH \to b\bar{b}b\bar{b}$ production via ggF in the resolved channel, shown for three discriminants: categorized m_{HH}, MVA, and MVA categorized in m_{HH}. The median expected limit is presented along with the $\pm1\sigma$ and $\pm2\sigma$ uncertainty bands, assuming no signal contribution. (b) Expected 68% and 95% confidence level upper limits on non-resonant $HH \to b\bar{b}b\bar{b}$ production as a function of κ_λ, for each discriminant strategy with all other couplings fixed to their Standard Model values. 183

10.1 The central role of the Higgs boson in modern particle physics. The Higgs field not only gives mass to elementary particles via electroweak symmetry breaking but also connects to a broad array of theoretical and experimental frontiers, including Higgs self-coupling, dark matter, baryon asymmetry, supersymmetric particles, the stability of the vacuum, composite Higgs models, and extensions to the scalar sector. Each branch highlights a key area where the Higgs boson could provide insight into physics beyond the Standard Model. 186

10.2 Expected values of the negative log-likelihood ratio as a function of κ_λ for an integrated luminosity of 3000 fb-1 at $\sqrt{s} = 14$ TeV for the ATLAS (ATLAS+CMS) combination of different channels and their statistical combination in the baseline scenario in Figure 10.2a (10.2b) [160, 161]. 187

A.1 Truth distribution of the angular distance between the b-quarks for the (a) leading and (b) subleading Higgs bosons in the mc23a signal, showing both correct and incorrect pairings for low-mass ($m_{HH} < 450$ GeV) and high-mass ($m_{HH} > 450$ GeV) events. 211

A.2 Truth distribution of the angular distance between the b-quarks for the (a) leading and (b) subleading Higgs bosons in the mc23a signal, shown for three m_{HH} intervals: [0, 250], [350, 400], and [400, 450] GeV. 212

A.3 Distributions of $3b$ events with $m_{HH} < 400$ GeV from 2017 data in the m_{H1}-m_{H2} plane for each of the three possible jet pairings. With jets ordered by decreasing p_T and indices [1,2,3,4]: pairing 1 is [(1,2), (3,4)], pairing 2 is [(1,3), (2,4)], and pairing 3 is [(1,4), (2,3)]. 212

A.4 (a) Stat-only 95% CL limits and (b) their percentage improvement when reducing the low-mass background by a factor of up to three, for several boundaries of m_{HH}. 213

A.5 Stat-only 95% CL limits when both increasing the signal and reducing the background by factors between 1 and 3 for low-mass events, with boundaries at (a) $m_{HH} < 396$ GeV and (b) $m_{HH} < 433$ GeV. 214

A.6 Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 2b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the SR. Good closure is observed across the kinematic phase space. 215

A.7 Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 3b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the SR. While closure improves, significant modeling challenges remain. 215

A.8 Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 4b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the CR. While the modeling shows some improvement, significant discrepancies remain, particularly in the control region used for training. 216

B.1 Distributions of the BDT output for signal (correct pairing, score near 1) and background (incorrect pairing, score near -1) for (a) odd and (b) even events in the compressed region, and (c) odd and (d) even events in the non-compressed region. The agreement between training and test samples indicates no overtraining. See SH4b Internal Note [145]. 220

B.2 (a) Number of simulated events passing selection in the training dataset. (b) BDT pairing efficiencies for different signal mass points, defined as the weighted fraction of selected events with the correct pairing identified. See SH4b Internal Note [145]. 221

B.3 Reconstructed mass distributions for H ((a), (b)), S ((c), (d)), and X ((e), (f)) for a compressed signal ($m_X = 1000$ GeV, $m_S = 750$ GeV; left column) and a non-compressed signal ($m_X = 750$ GeV, $m_S = 400$ GeV; right column). Results are shown for BDT pairing (green), min ΔR pairing (yellow), and truth-matching (pink). See SH4b Internal Note [145]. 222

C.1 Expected (dashed black lines) and observed (solid black lines) 95% CL upper limits on the cross section times the branching ratio of resonant HH production in the spin-0 model. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties. Expected limits using each of the resolved and boosted channels individually (dashed colored lines) are shown [105]. 224

C.2 Observed (solid black line) and expected (dashed black line) 95% CL upper limits on the cross section times the branching ratio. For each unique m_X value, a scaling factor is applied to the limits so that all mass pairs can be visualized together. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties, respectively. The observed and expected limits on the cross section times the branching ratio for the resonant $HH \to b\bar{b}b\bar{b}$ results are depicted (approximately) in gray. 226

ACKNOWLEDGMENTS

What a journey this has been! I started my Physics PhD program at the University of Washington in 2020 at the height of the COVID-19 pandemic. With remote courses and homework, no one was quite sure how things would unfold, but we eventually found our footing. Of course, I am grateful to my professors and friends in the PhD program for helping me navigate that first year remotely. Learning graduate physics online is a formidable task, and it was made immeasurably easier by the professors' successful use of technology and my friends' incredible collaboration. We worked together, helping one another grasp the difficult concepts.

The unique challenge the pandemic posed did give me a rare opportunity to spend my first year in various places I call "home" with friends and family. I will forever be grateful to my brother, Doug, his wife Annie, and their kids for hosting me for part of the fall quarter. Simon may be too old for this now, but I will always cherish the memory of us listening to "Baby Shark" and "The Final Countdown" on repeat. I am also deeply thankful to the Baldwin family for taking me in when I needed to focus on my coursework and exams. I have always been fortunate to have neighbors I consider family. Staying with them also afforded me the chance to secretly train for the "Beat 'em Stephen" 5k Thanksgiving Run, a race that came down to the wire and had some surprising results. In the end, I did beat Stephen, and I am so appreciative of Lauren and Stephen for putting the event together and being such good sports.

I'll also always treasure the time I lived with Kelvin and Fonzy at "Barb's house". While it could sometimes be a distraction from my studies, I enjoyed their charisma, humor, and the constant fun they brought. There was never a dull moment, and I could (and did) spend hours listening to them tell stories. My time in Utah also

allowed me to continue past traditions of Chinese hot pot with Kenan, Hodge, and Yance and I am not sure that any other Chinese fondue has lived up to those memories.

I am so happy that I had the chance to move to Seattle and the University of Washington after my first year. I love the Wallingford area, and I will miss the time I had living with Jack, Nick, John, Brandon, and Jesse. Our household was a hub of delicious meals, great TV shows, and spontaneous discussions on niche physics topics. I'll miss the incredible view of Lake Union and the many great food spots in the area, not to mention the amazing sunsets that can be a rare treat given the weather.

To the new friends I made in Seattle, thank you. Our hikes, climbing trips, and game nights with Jen, Nikki, Lindsay, and others were full of laughter. Playing intramural basketball with Eli, Murali, Obinna, and the rest of the physics team was a highlight of my time. We may not have had many wins, but we had a lot of heart! I also enjoyed sharing an office with Will, where our discussions on theoretical physics and other random subjects were always appreciated. I'm thankful to my sister, Jennie, for helping me move my belongings from Seattle, and to her husband for lending me a large trekking backpack that has since reached the highest points in several different countries.

After completing my coursework in Seattle, I had the incredible opportunity to conduct my research at CERN. Not only did I get to work on-site at the LHC, but I also attended lectures on everything from the physics of music theory to how quantum physics relates to Santa Claus. I made friends from all over the world, and I am especially grateful to Miguel, Naseem, Daniel, Sergio, and my other friends from Valencia. You all are an incredibly fun group and some of you may even be able to beat me in squash one day. The Geneva ward was also full of wonderful people. Chad and the Turner family were always so kind, even inviting me over for dinner on my birthday. Seeing Dune 2 in IMAX with Tanner was certainly an experience I'll

never forget. I'm also incredibly fortunate to have met my French-speaking friends. Coline, I never would have been able to navigate the French bureaucracy to get auto insurance (among so many other things) without your help. You also taught me so much about the local food and culture and I'm grateful for that exposure beyond my work.

To my friends and family who visited me abroad, I cherish those memories. I had many hikes and trips with Diana, and sometimes Ben, and then another separate trip seeing Ben, Paige, and the kids during their study abroad. Will and Maddie were the absolute best tour guides a person could hope for! I also got to go on flights and road trips with Kelvin and Amos, where Amos's generosity and patience knew no bounds. It was amazing to travel through Switzerland with Matt and Daniel and see the picturesque landscapes. I am also so glad that my parents had the chance to see CERN near the end of the program. Both of my parents are great examples of people who attained a high education, but more importantly, they are great examples of good character. There were countless others in my extended family and friends who helped me throughout this program, and I am thankful for all of them.

I am particularly grateful to Anna, my advisor, for guiding me throughout these years. Your expertise, positivity, and support have been crucial to my success. While we didn't often see each other in person, I always knew you were there for me. I am also thankful to Jana. There was never a question you didn't have an answer to, and the things you have taught me have made me a much better physicist.

Countless others from the ATLAS Experiment helped me along the way. I am grateful for the opportunity to work with the "Resolvedbbbbesties" in my analyses. Going through the analysis workflow with Varsha, Malin, and Thandi made the work so much easier, especially with the support and expertise from Nicole. It was great to see how much progress we made together and I am proud of the work we accomplished. It was also a great opportunity to be a part of the JetEtMiss subgroup for jet definitions and calibrations, and I am thankful to "G", Chris, Ana, Margherita,

Hannsjorg, and others who helped me learn how Athena works and how to use the various code repositories for the calibrations. I am also grateful for the help and leadership of others from JetEtMiss like Reina, Fabrice, and Jeff.

I want to thank everyone on ATLAS who helped make this the best experiment in the world. Michael always had such a depth of knowledge on the physics in the di-Higgs analyses, as well as creative ideas on how to use the newest machine learning methods. It was a privilege to see how an expert conducts research and I learned a great deal. Lidija was amazing in supporting us students in the SH4b analysis. Your knowledge on the triggers always helped me understand more since this was one of the areas I wasn't as familiar with. To the UW EPE group, including Haoran, Yue, Yulei, Yuan-Tang, Bayley, Ali, Miles, Abdel, and others, thank you. It was always great spending time together in the office and outside of it. I hope you all find great success in your future endeavors.

Last of all, I would like to thank my Grandpa Smoot. He was the reason I applied to the University of Washington in the first place, as he earned his PhD in chemical engineering here. He was the best engineer and scientist I knew and a great inspiration. It was unfortunate that he passed before I accepted and started the graduate program. Nevertheless, I will always remember his example, and I dedicate this thesis to him.

Chapter 1

INTRODUCTION

"Sometimes the smallest things take up the most room in your heart."

– Winnie the Pooh

Particle physics studies the fundamental building blocks of nature and the interactions between them that shape our universe. The idea that matter consists of indivisible units dates back to Democritus, who proposed the concept of "atomos" nearly 2,500 years ago. However, it was only in the late 19th and early 20th centuries that experiments began to uncover the subatomic structure of matter. The discovery of the electron by J.J. Thomson in 1897, followed by Rutherford's atomic model and Chadwick's identification of the neutron in 1932, challenged the notion that atoms were indivisible. Further research revealed that protons and neutrons are themselves composed of even smaller particles called quarks. These discoveries laid the foundation for the Standard Model, a theoretical framework discussed in Section 2.1 that describes the fundamental particles and their interactions, excluding gravity [1, 2].

Despite its success, the Standard Model is incomplete. It does not account for dark matter, the dominance of matter over antimatter, or the integration of gravity at the quantum level [3]. As a result, physicists are not only testing its predictions but also searching for signs of new physics beyond it. The Large Hadron Collider (LHC) at CERN has been instrumental in this effort, most notably with the 2012 discovery of the Higgs boson, which confirmed the mechanism by which particles acquire mass [4, 5]. However, no new fundamental particles, beyond the Higgs boson, have yet been conclusively detected. Beyond the LHC, international experiments are probing the limitations of the Standard Model. The Deep Underground Neutrino Experiment (DUNE) in the U.S. investigates neutrino properties, while Japan's Super-Kamiokande and future Hyper-Kamiokande detect neutrinos from cosmic and terrestrial sources [6, 7]. The Belle II and LHCb examine CP violation, which

may help explain the matter-antimatter asymmetry of the universe [8, 9]. These global collaborations drive the search for physics beyond the Standard Model.

Several experimental anomalies hint at new physics. One of the most significant comes from the Muon g-2 experiment at Fermilab, where measurements of the muon's magnetic moment deviate from Standard Model predictions, suggesting the presence of unknown forces or particles [10]. Meanwhile, astrophysical evidence for dark matter and dark energy points to phenomena beyond the Standard Model's scope [11, 12]. While these results are not yet conclusive, they provide compelling motivation for continued experimental and theoretical exploration.

Over the past century, particle physics has not only deepened our understanding of the universe but also driven technological advancements with broad societal impact. The field has led to the development of transformative technologies such as PET scans, radiation therapy, and semiconductor advancements. One of its most influential contributions is the creation of the World Wide Web at CERN, originally developed to facilitate data sharing among physicists. These innovations underscore the broader significance of particle physics beyond fundamental research.

This thesis presents my contributions to particle physics research as a member of the ATLAS Collaboration. My work has focused on the search for Higgs pair production in the $b\bar{b}b\bar{b}$ (4 b-quark) decay channel, where two Higgs bosons are produced in the same event and each decays into a pair of bottom quarks. Additionally, I have contributed to jet calibration in the fast simulation software, benefiting the collaboration as a whole, and explored the future use of simulated quantum chromodynamics (QCD) samples. I have also investigated the integration of machine learning (ML) techniques into physics analyses. This thesis is structured as follows: Chapter 2 outlines the theoretical motivations behind the analyses, Chapter 3 introduces the LHC and the ATLAS Experiment, Chapter 4 details my technical work for jet calibrations and systematics, Chapter 5 provides an overview of Higgs pair production with its theoretical motivations and experimental results, Chapter 6 discusses common aspects to HH analyses in a $b\bar{b}b\bar{b}$ final-state, Chapter 7 presents the $X \to SH \to b\bar{b}b\bar{b}$ analysis using the ATLAS Run 2 data, Chapter 9 discusses the non-resonant $HH \to b\bar{b}b\bar{b}$ analysis using the ATLAS Run 2 data, while Chapter 9 covers the

same non-resonant $HH \to b\bar{b}b\bar{b}$ analysis using the ATLAS Run 2 and early Run 3 data, and Chapter 10 summarizes the findings and future directions.

Chapter 2

THEORETICAL FOUNDATIONS

2.1 The Standard Model

Our universe is made up of different types of matter which interact through the following fundamental forces: gravity, electromagnetism, the weak nuclear force, and the strong nuclear force. These four fundamental forces that govern particle interactions exhibit varying strengths and have unique properties, as summarized in Table 2.1. The electromagnetic force governs interactions between charged particles, the weak nuclear force is responsible for phenomena such as beta decay, and the strong nuclear force binds quarks within composite particles like protons and neutrons. Given how weak the gravitational force is on a particle level, it is generally excluded from consideration when discussing interactions in particle physics though it dominates on the cosmological scale in forming planets, solar systems, and galaxies.

Force	**Force Carrier**	**Interaction Type**	**Relative Strength**	**Interaction Length**
Strong Nuclear	Gluon (g)	Color Charge	1	10^{-15} m
Electromagnetic	Photon (γ)	Electric Charge	$\approx 10^{-2}$	Infinite
Weak Nuclear	$W^{\pm}, Z^0$ bosons	Weak Isospin	$\approx 10^{-13}$	10^{-18} m
Gravity	Graviton†	Mass/Energy	$\approx 10^{-38}$	Infinite

Table 2.1: The four fundamental forces, their corresponding force carriers, the interaction types they govern, their relative strengths, and their average interaction lengths. † Note that the graviton is hypothetical.

As previously described, the Standard Model of particle physics is a highly successful

theoretical framework that describes the fundamental particles and their interactions, encompassing three of these four known fundamental forces, namely electromagnetism, the weak nuclear force, and the strong nuclear force [13, 14]. It emerged in the 1970's and remains the dominant model for making and testing predictions of phenomena in particle physics. It classifies all known elementary particles into two broad categories: fermions, which constitute matter, and bosons, which mediate fundamental interactions (see Figure 2.1). These classifications are determined by the quantum number for the spin of the particle, the intrinsic angular momentum. The fermions have half-integer spin numbers and obey the Pauli exclusion principle, which asserts that no two identical fermions can occupy the same quantum state simultaneously. Bosons, on the other hand, have integer spin numbers and can occupy the same quantum state.

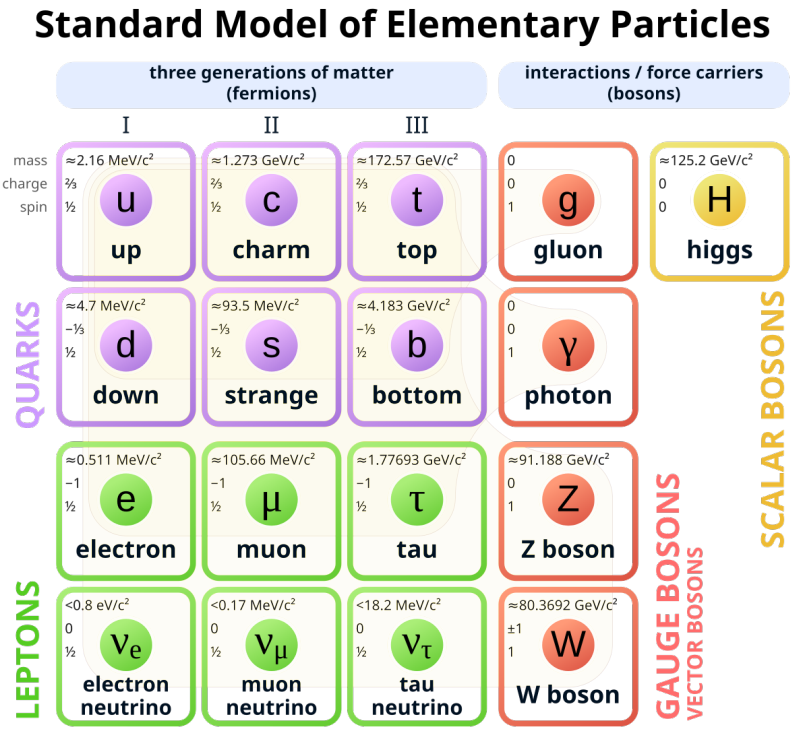


Figure 2.1: The fundamental particles of the Standard Model illustrating the different classifications and generations of matter and the force carriers that mediate their interactions [15].

Fermions, the fundamental building blocks of matter, are divided into two families with different properties: **quarks** and **leptons**. Both families are organized into three generations with each successive generation containing heavier and generally less stable particles.

Quarks are the only fermions that participate in all three fundamental forces of the Standard Model: the electromagnetic force, the weak nuclear force, and the strong nuclear force. There are six quark flavors: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). Each flavor carries electric charge, either $+\frac{2}{3}e$ (up, charm, top) or $-\frac{1}{3}e$ (down, strange, bottom), where e is the elementary charge of an electron. Having electric charge allows them to interact through electromagnetic forces where same-signed charges repel and opposite-signed charges attract. In addition, quarks possess a property called **color charge**, which can be red, green, or blue. This is the source of their interaction via the strong nuclear force, mediated by gluons. Due to *color confinement*, quarks are never observed in isolation [16]; instead, they form color-neutral composite particles called hadrons, such as protons and neutrons which each consist of the three valence quarks uud and udd respectively.

Leptons, on the other hand, do not carry color charge and thus do not interact via the strong force. The six known leptons consist of three charged particles, the electron (e), muon (μ), and tau (τ), and their corresponding neutral neutrinos, the electron neutrino (ν_e), muon neutrino (ν_μ), and tau neutrino (ν_τ). The charged leptons carry an electric charge of $-e$, while neutrinos are electrically neutral. All leptons participate in the weak interaction, which is the only force capable of changing a particle's flavor and is a key feature in certain decay processes [17, 18, 19]. Like quarks, leptons are grouped into three generations, with particles in higher generations being more massive and unstable. Although neutrinos were once thought to be massless, experiments have confirmed they have tiny but nonzero masses, as evidenced by the phenomenon of neutrino oscillation where the neutrino can spontaneously oscillate between its generations [20].

As stated previously, bosons act as the force carrier for particle interactions and include the photon, gluon, $W^{\pm}$, Z^0, and Higgs bosons. These first four particles are all examples of **vector bosons**, meaning they have a spin of 1 and are associated with vector fields in quantum field theory. In contrast, the **Higgs boson** is a **scalar boson** with spin 0. It

is responsible for giving mass to the W and Z bosons (and to fermions indirectly) via the Brout-Englert-Higgs (BEH) mechanism, or simply the Higgs mechanism. This mechanism operates through spontaneous symmetry breaking in the Higgs field and is discussed in more detail in Section 2.2 [21, 22].

Electromagnetism governs the interactions between electrically charged particles and is mediated by the photon. Here, **electric charge** refers to a particle's positive or negative charge, such as the $+e$ of the proton or the $-e$ of the electron. The photon is a massless, electrically neutral boson and represents the quantum of the electromagnetic field. Owing to its zero mass, the photon always travels at the speed of light in a vacuum, which defines the ultimate speed limit for information and energy transfer in the universe. Because photons do not carry electric charge, they do not directly self-interact at the fundamental level. Nevertheless, rare quantum effects allow photons to interact indirectly through processes such as light-by-light scattering, which have been observed experimentally and serve as a precise test of quantum electrodynamics [23].

The strong nuclear force, which binds quarks together to form protons, neutrons, and other hadrons, is mediated by the gluon. Like the photon, the gluon is massless and electrically neutral. However, gluons carry **color charge**, and can interact not only with quarks but also with other gluons. Quarks come in three possible color charges (**red**, **green**, and **blue**) and the strong force acts to ensure that all observable particles are color-neutral combinations of these. This direct self-interaction among gluons is a key feature of the strong force and gives rise to phenomena such as **color confinement**, where quarks are never observed in isolation. As a result, the strong force is extremely strong at short distances (on the order of a femtometer) but does not extend beyond the atomic nucleus.

The weak nuclear force is mediated by the W^+, W^-, and Z^0 bosons. The $W^\pm$ bosons are electrically charged, while the Z^0 is neutral. All three are **massive**, with the W and Z bosons being among the heaviest known elementary particles. Because of their large masses, these bosons are highly unstable and have very short lifetimes, which is why the weak force operates only over extremely short distances.

Unlike the electromagnetic and strong forces, which act only on electric or color charge, the weak interaction couples to a property called **weak charge**, which is related to a

quantum number known as **weak isospin**. Only particles with nonzero weak isospin, such as left-handed fermions, interact via the charged weak force mediated by the $W^{\pm}$ bosons. The neutral Z^0 boson also couples to particles based on their weak charge, enabling weak neutral current interactions. This charge structure gives the weak force its distinctive behavior, including the ability to change a particle's flavor and violate parity symmetry in certain processes.

Gravity, though not incorporated into the Standard Model, is hypothesized to be mediated by the graviton, a theoretical massless boson of spin 2 [24].

Antimatter consists of particles that share the same quantum numbers of their counterparts in the Standard Model, but with opposite electric charge. For example, the **positron** is the antiparticle of the electron, so it has the same mass and spin as the electron but carries a positive charge instead of a negative one. Similarly, for each quark, there is an antiparticle with the opposite charge, such that an **antiproton** is composed of *antiquarks* (antidown and antiup quarks) and carries a negative charge. An antiparticle is denoted either by reversing the electric charge of the corresponding lepton (e.g., e^+ for the positron) or by placing a bar over the particle symbol (e.g., $\bar{b}$ for the anti-bottom quark). The existence of antimatter was predicted by Paul Dirac in the 1920s and experimentally confirmed with the discovery of the positron in 1932. In fact, the reason the universe predominantly consists of matter and not antimatter remains one of the unsolved mysteries in physics.

In addition to fermions, **bosons** can also have corresponding antiparticles. However, an important property of bosons is that neutral ones are their own antiparticles. For example, the photon (the mediator of the electromagnetic force) and the gluon (the mediator of the strong force) are their own antiparticles.

2.2 Gauge Symmetry and the Higgs Mechanism

The Standard Model is built upon the principles of gauge symmetry, where certain transformations can be applied to a physical system without changing the observable quantities. The model is specifically described by the symmetry group:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{2.1}$$

This structure defines the interactions among elementary particles. The $SU(3)_C$ group corresponds to quantum chromodynamics (QCD), which governs the strong nuclear force and the interactions of quarks via gluon exchange. The electroweak interaction is described by the $SU(2)_L \times U(1)_Y$ symmetry, which unifies electromagnetism and the weak nuclear force at high energies. A more thorough treatment and explanation for these group theories can be found in [25, 26].

Gauge symmetries dictate the permissible interactions and enforce conservation laws, such as conservation of color charge in QCD and electric charge in electroweak theory. Although the weak and electromagnetic forces appear distinct at low energies, they are fundamentally unified in the Standard Model. However, in nature, the weak force is short-ranged due to the non-zero mass of the W and Z bosons acquired via the Higgs mechanism. The following discussion of this mechanism draws on the principal sources in [27, 28, 29].

In the Standard Model, the Higgs field, a complex scalar doublet Φ, is designed to have a nonzero vacuum expectation value (VEV). The Higgs potential is given by:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2, \tag{2.2}$$

where μ^2 and λ are parameters. For $\mu^2 < 0$, the Higgs field develops a nonzero VEV:

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.3}$$

where $v \approx 246$ GeV is the electroweak scale. This process spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry down to $U(1)_{EM}$, which corresponds to electromagnetism. As a result, the $W^\pm$ and Z^0 bosons acquire mass, while the photon remains massless.

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

$$\langle 0|\phi|0\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}, \tag{2.4}$$

where we have explicitly chosen a real, nonzero vacuum expectation value for the neutral component of the Higgs doublet to maintain a massless photon. The vacuum itself is not electrically charged.

To describe quantum fluctuations around this vacuum, we expand the Higgs field as:

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + \eta(x) + i\phi_3 \end{pmatrix}. \tag{2.5}$$

Substituting this expansion into the Higgs potential:

$$V(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{2.6}$$

$$= \mu^2 \left(\sum_i \phi_i^2 + (v + \eta(x))^2 \right) + \lambda \left(\sum_i \phi_i^2 + (v + \eta(x))^2 \right)^2. \tag{2.7}$$

Expanding the potential in terms of η and ϕ_i:

$$V(\Phi) = -\frac{1}{4}\lambda v^4 + \lambda v^2 \eta^2 + \lambda v \eta^3 + \frac{1}{4}\lambda \eta^4 \tag{2.8}$$

$$+ \frac{1}{2}\lambda \sum_i \phi_i^2 \sum_j \phi_j^2 + \lambda v \eta \sum_i \phi_i^2 + \frac{1}{2}\lambda \eta^2 \sum_i \phi_i^2 + \frac{1}{4}\sum_i \phi_i^4, \tag{2.9}$$

where we have used the relation $\mu^2 = -\lambda v^2$.

From this expression, we identify the mass of $\eta(x)$ as $m_\eta = \sqrt{2\lambda v^2}$. The ϕ_i fields remain massless and correspond to the Goldstone bosons, representing quantum fluctuations along the minimum of the potential. The terms $\lambda v \eta^3$ and $\frac{1}{4}\lambda \eta^4$ describe trilinear and quartic self-interactions of the Higgs field.

Expanding the kinetic term:

$$(\partial_\mu \Phi)^\dagger (\partial^\mu \Phi) = \frac{1}{2}\sum_i (\partial_\mu \phi_i)(\partial^\mu \phi_i) + \frac{1}{2}(\partial_\mu \eta)(\partial^\mu \eta), \tag{2.10}$$

we confirm that three massless degrees of freedom correspond to the Goldstone bosons, while one massive degree of freedom remains as the Higgs boson. The Goldstone bosons are absorbed by the weak gauge bosons via the Higgs mechanism, giving mass to the $W^\pm$ and Z^0 bosons, while the photon remains massless due to the unbroken $U(1)_{EM}$ symmetry. This mass generation mechanism is critical for ensuring the consistency of the Standard Model with experimental observations. Furthermore, the Higgs field couples to fermions via Yukawa interactions, generating their masses in proportion to their coupling strengths.

2.3 Experimental Verification of the Standard Model

While the theoretical structure of the Standard Model was established by the early 1970s, experimental verification unfolded over several decades. The charm quark was first observed in 1974 through the independent discovery of the J/ψ meson at Brookhaven National Laboratory and SLAC [30, 31]. Its unexpectedly high mass and narrow width matched predictions for a bound state of a charm and anti-charm quark, providing the first strong evidence for quark substructure and sparking what became known as the November Revolution. Evidence for the bottom quark followed in 1977 at Fermilab through the observation of bottomonium ($b\bar{b}$) states decaying into muon pairs, later confirmed at DESY [32, 33].

The W and Z bosons were discovered in 1983 by the UA1 and UA2 experiments at CERN's Super Proton Synchrotron [34, 35, 36, 37]. Their measured masses and decay properties confirmed the electroweak unification, with the W boson mediating charged current interactions and the Z boson mediating neutral current interactions, as predicted by the Standard Model. The gluon, the carrier of the strong force in QCD, was indirectly observed in 1979 at the PETRA collider at DESY through characteristic three-jet events in electron-positron collisions, consistent with gluon bremsstrahlung from quarks [38].

The top quark, the last missing piece of the quark sector, was discovered in 1995 by the CDF and DØ experiments at Fermilab's Tevatron collider [39]. Distinguished by its exceptionally large mass of approximately 173 GeV, the top quark decays before hadronization, offering a unique opportunity to study a bare quark. Its large mass also plays a key role in electroweak radiative corrections, making its precise measurement important for global fits of Standard Model parameters. Its discovery confirmed the third generation of quarks and further solidified the framework of the Standard Model.

Despite the remarkable success in discovering all other Standard Model particles, the Higgs boson remained elusive. Extensive searches were carried out at CERN's LEP collider and at Fermilab's Tevatron, setting progressively stronger exclusion limits on the Higgs boson's mass. While some hints of Higgs-like signals were seen, no statistically significant discovery could be achieved. The challenge was that the Higgs boson, if it existed within the expected mass range, would have a relatively small production cross-section and branching

ratios into clean final states were rare, requiring higher energy and luminosity to observe. It became clear that a new, more powerful collider was needed to probe the Higgs sector decisively.

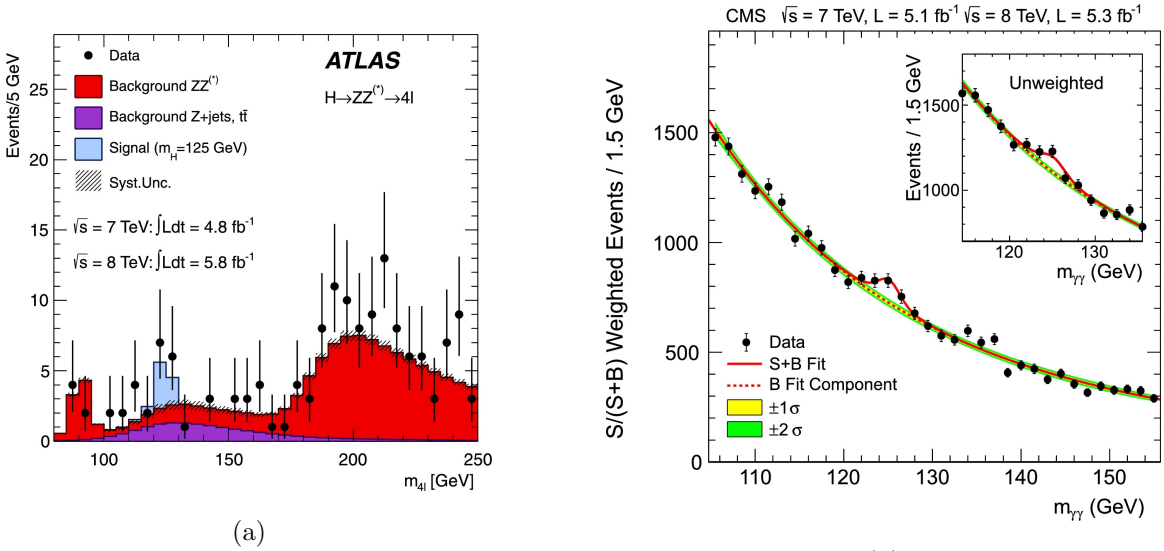
2.3.1 The Higgs Boson

A major objective of the LHC and its general-purpose experiments, ATLAS and CMS, was the discovery of the Higgs boson and the measurement of its properties. With proton-proton collisions at a center-of-mass energy of 7 and 8 TeV during Run 1, both collaborations accumulated sufficient integrated luminosity to search for the Higgs across a wide mass range. In 2012, both ATLAS and CMS announced the discovery of a new boson with a mass around 125 GeV, consistent with the long-sought Higgs boson.

The discovery was driven primarily by two key decay channels: $H \to \gamma\gamma$ and $H \to ZZ^* \to 4\ell$. The diphoton channel, though having a small branching ratio ($\sim$0.2%), offered excellent mass resolution, while the four-lepton channel provided an exceptionally low-background signature. These channels, illustrated in Figure 2.2, were pivotal for a robust and unambiguous discovery.

Understanding the decay modes and branching ratios of the Higgs boson has been critical for both its discovery and subsequent measurements. Figure 2.3 shows the SM predictions for branching ratios across a range of possible Higgs masses, guiding the experimental strategy for which decay channels to prioritize [42].

Recent precision measurements by ATLAS and CMS have further refined our understanding. The Higgs mass is now measured to be 125.11 ± 0.11 GeV (ATLAS) and 125.04 ± 0.12 GeV (CMS), with indirect constraints placing the Higgs total width below 10.5 MeV (95% CL), consistent with the SM prediction of about 4 MeV [43, 44]. The measured cross-section for gluon-gluon fusion (ggF), where two gluons interact via a top-quark loop to produce a Higgs boson, is $55.5^{+4.0}_{-3.8}$ pb at $\sqrt{s} = 13$ TeV which aligns with theory. Several decay modes, such as $\tau\tau$, $b\bar{b}$, and $t\bar{t}H$ production have been observed at rates compatible with the SM, confirming Higgs couplings to both fermions and bosons [45, 46, 47]. Studies of the Higgs boson's CP properties show no significant deviations from a CP-even scalar



(a) (b)

Figure 2.2: Distributions used in the Higgs boson discovery: (a) the four-lepton invariant mass $m_{4\ell}$ from the ATLAS analysis of $H \to ZZ^* \to 4\ell$, showing data compared to the background expectation and the signal prediction for a 125 GeV Higgs, combining $\sqrt{s} = 7$ TeV and 8 TeV data; (b) the diphoton invariant mass distribution from the CMS analysis of $H \to \gamma\gamma$, with events weighted by $S/(S+B)$ to enhance signal sensitivity [40, 41].

nature, while efforts to probe the Higgs self-coupling via Higgs pair production are ongoing (see Chapter 5). These comprehensive results cement the Higgs boson's role within the Standard Model, while imposing stringent constraints on new physics beyond it.

2.4 Unanswered Questions

Following the landmark discovery of the Higgs boson as direct evidence for the Higgs mechanism, the Standard Model of particle physics was considered complete. This represented a major triumph for modern physics, as the model successfully described three of the four fundamental forces while making precise, testable predictions that have been repeatedly confirmed in experiments. Yet, despite its extraordinary success, the Standard Model leaves many foundational questions unanswered.

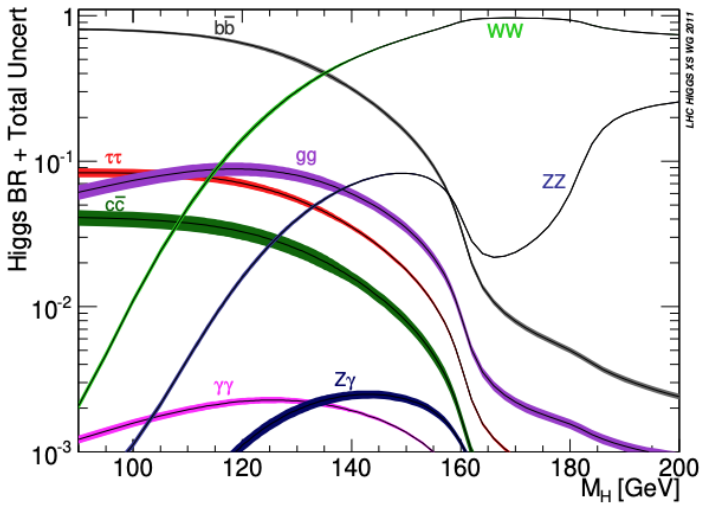


Figure 2.3: SM Higgs branching ratios and their uncertainties for possible Higgs boson masses in the range from 90 - 200 GeV [42].

One of its well-known shortcomings lies in its treatment of neutrinos. In the original formulation, the Standard Model assumes neutrinos to be massless. However, the observation of neutrino oscillations, a quantum phenomenon in which neutrinos change flavor as they propagate, provides compelling evidence that they possess small, yet nonzero masses. This finding necessitates physics beyond the Standard Model, as no mechanism within it accounts for neutrino mass.

The model also explains all forms of ordinary matter such as protons, neutrons, and electrons that compose the visible universe. Nevertheless, astronomical and cosmological observations reveal the existence of another form of matter: dark matter. This elusive substance interacts gravitationally but not electromagnetically, making it invisible to direct detection. It is estimated to comprise roughly 26% of the universe's total energy content [48]. Despite decades of searching, no dark matter particles have been directly observed, and their fundamental nature remains a mystery.

Another unresolved issue concerns the observed imbalance between matter and antimatter. The Standard Model predicts that the Big Bang should have produced matter and

antimatter in nearly equal quantities. Yet today, the observable universe is overwhelmingly composed of matter, with antimatter present only in trace amounts. The known sources of the violation of CP-symmetry, where the laws of physics should be the same if a particle is interchanged with its antiparticle (C-symmetry) while its spatial coordinates are inverted (P-symmetry), within the Standard Model are insufficient to explain this asymmetry, pointing to the need for additional mechanisms beyond the current theory.

There also remains an open question how to similarly incorporate the strong nuclear force into the Standard Model, similar to how electromagnetism and the weak force were unified in electroweak theory. To this end, physicists have developed Grand Unified Theories (GUTs), which propose that the electromagnetic, weak, and strong interactions are manifestations of a single force at extremely high energies. Although GUTs are mathematically compelling and have motivated extensive theoretical exploration, they currently lack direct experimental support.

Another compelling theoretical framework is supersymmetry (SUSY), which extends the Standard Model by proposing that every known particle has a corresponding supersymmetric partner [49]. In this framework, each fermion (such as the electron) is paired with a bosonic counterpart (like the selectron), and each boson with a fermionic partner, differing in spin but sharing other quantum numbers. Supersymmetry offers elegant solutions to several outstanding problems in particle physics, such as the hierarchy problem, but despite extensive searches, no supersymmetric particles have been experimentally detected to date.

While these theories provide promising avenues for extending and unifying our understanding of the fundamental particles and forces, none have yet been confirmed by experiment. The search for a more complete theory of nature remains one of the most profound and enduring challenges in modern physics. These and other unresolved questions will be revisited in the context of my work in Chapter 5.

Chapter 3

THE ATLAS EXPERIMENT

Modern research in high energy physics relies on large-scale experiments, international collaborations, and highly advanced facilities to probe the fundamental nature of matter and the universe. At the forefront of this effort is the European Organization for Nuclear Research (CERN), located on the Franco-Swiss border near Geneva. CERN is a global center for particle physics and supports a wide range of research, from studying antimatter to testing predictions of the Standard Model and beyond. Its most prominent facility is the LHC, the world's most powerful particle accelerator, designed to recreate conditions similar to those just after the Big Bang and to explore the frontiers of fundamental physics.

Among the major experiments housed at the LHC is the ATLAS experiment, one of four primary detectors alongside CMS, ALICE, and LHCb. ATLAS, along with CMS, is a general-purpose detector designed to explore a wide range of physics including searches for new particles and high-precision testing of the Standard Model. The experiment is a truly global collaboration, comprising over 5,000 scientists, engineers, and students from more than 180 institutions across 40 countries. The name ATLAS stands for the **A** **T**oroidal **L**HC **A**pparatu**s**, while also alluding to the Greek titan condemned to bear the weight of the heavens which is a fitting metaphor for the scale and ambition of the project. Physically, ATLAS is the largest particle detector ever built, measuring 46 meters in length, 26 meters in both height and width, and weighing approximately 7,000 tons, roughly the same as the Eiffel Tower.

3.1 The Large Hadron Collider

The LHC is the centerpiece of CERN's experimental program and the most powerful particle accelerator ever built. Spanning a 27-kilometer ring, the LHC was designed to explore fundamental questions in physics by recreating conditions similar to those of the early

universe. It accelerates two beams of protons in opposite directions to near light speed, bringing them into head-on collisions at energies up to 6.8 TeV per beam. These extreme energies are necessary to produce the heaviest known particles, such as the Higgs boson and top quark, and to probe rare and previously inaccessible processes that may point toward physics beyond the Standard Model.

The energies achieved by protons in the LHC are truly extraordinary by particle physics standards. Using Einstein's relativistic energy formula, $E = \gamma mc^2$, and the rest mass of the proton (938 MeV/c^2), one finds a Lorentz factor of approximately 7000. This corresponds to protons traveling at ultra-relativistic speeds (over 99.999999% of the speed of light).

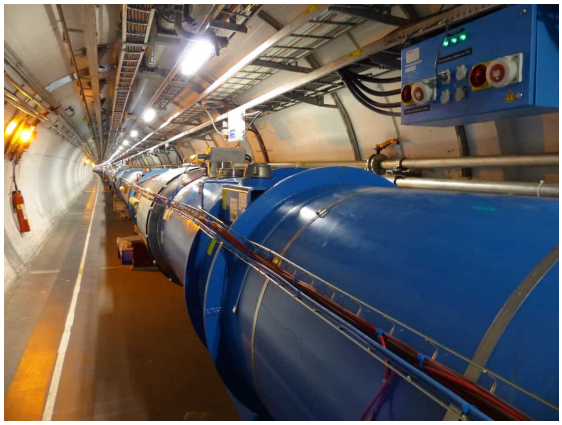
To achieve and maintain such conditions, the LHC accelerates counter-rotating beams of protons in the ring under an ultrahigh vacuum of approximately 1.013×10^{-10} mbar, nearly as rarefied as the surface of the Moon. The beams are guided by a powerful magnetic field generated by thousands of superconducting electromagnets. Dipole magnets bend the proton paths to keep them on their circular trajectory, while quadrupole magnets focus the beams into tight bunches to maximize collision rates. To sustain superconductivity, the magnets are cooled to $-271.3°$C using a cryogenic system based on liquid helium, making the LHC one of the coldest places in the universe. These precisely controlled magnets steer the beams toward designated interaction points (IPs), where detectors such as ATLAS and CMS are positioned to record the resulting high-energy collisions. A schematic view of the LHC tunnel and a cross section of a dipole magnet are shown in Figure 3.1.

The operation of the LHC relies fundamentally on the manipulation of charged particles using electric and magnetic fields. This is governed by the Lorentz force law, which describes how a charged particle experiences a force in the presence of electromagnetic fields:

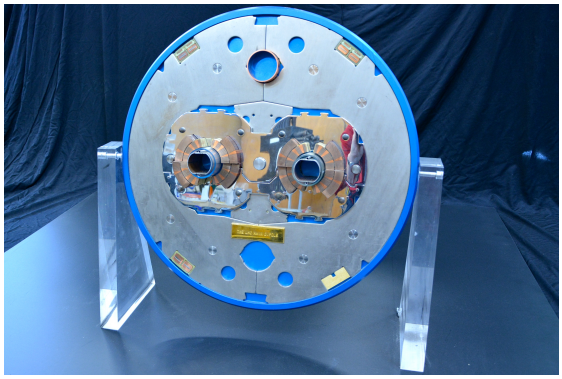
$$\vec{F} = q\left(\vec{E} + \vec{v} \times \vec{B}\right), \tag{3.1}$$

where q is the particle's charge, $\vec{E}$ is the electric field, $\vec{B}$ is the magnetic field, and $\vec{v}$ is the particle's velocity.

Magnetic fields, while incapable of performing work since the magnetic force is always perpendicular to the particle's velocity, are used to steer the proton beams through the



(a)



(b)

Figure 3.1: A view of the LHC in its underground tunnel, alongside a cross sectional illustration of the beam pipe. The diagram shows the two separate beam pipes, which carry protons in opposite directions, and highlights the surrounding dipole magnets used to steer the beams.

circular accelerator. This is achieved through the magnetic component of Eq. 3.1, $\vec{F} = q\,\vec{v} \times \vec{B}$, which causes the protons to follow curved trajectories within the beam pipe.

The actual speed increase or tangential acceleration, on the other hand, is provided by electric fields within radiofrequency (RF) cavities placed at key points along the ring. These cavities generate oscillating electric fields that act tangentially to the beam's motion, increasing the kinetic energy of the protons. The RF field is precisely synchronized such that protons arriving slightly early are decelerated, while those arriving late are accelerated, helping to compress the beam into tightly grouped bunches. Over roughly 20 minutes or approximately 10 million revolutions of the ring, these cavities boost the proton energy from an injection energy of 450 GeV up to the nominal collision energy of 6.5 TeV per beam.

As Hydrogen gas molecules are made up of two atoms with a proton and electron each, a tank of Hydrogen gas in a strong electric field serves as a source of protons for the proton-proton collisions. Following the diagram in Figure 3.2, the protons then enter a linear accelerator which then fires them into the PS Booster accelerator. These accelerators

use the aforementioned RF cavities to accelerate the protons to higher energies while the magnets maintain the curved trajectories. Once at a high enough energy level, the protons enter the Super Proton Synchrotron (SPS) which accelerates them to an even higher energy level. From there, the protons are injected into the two opposing beams of the LHC where they circle the LHC ring over 10,000 times per second. The magnet systems described above guide the proton beams to collision points at specific interaction points (IP), the center of the different detector systems for analyses. During routine operation, the LHC circulates up to nearly 2800 bunches of protons in each beam. These bunches are brought into collision every 25 ns at each interaction point, resulting in a nominal bunch crossing rate of 40 MHz. This high collision frequency is essential for accumulating the large datasets needed to study rare processes such as Higgs boson pair production.

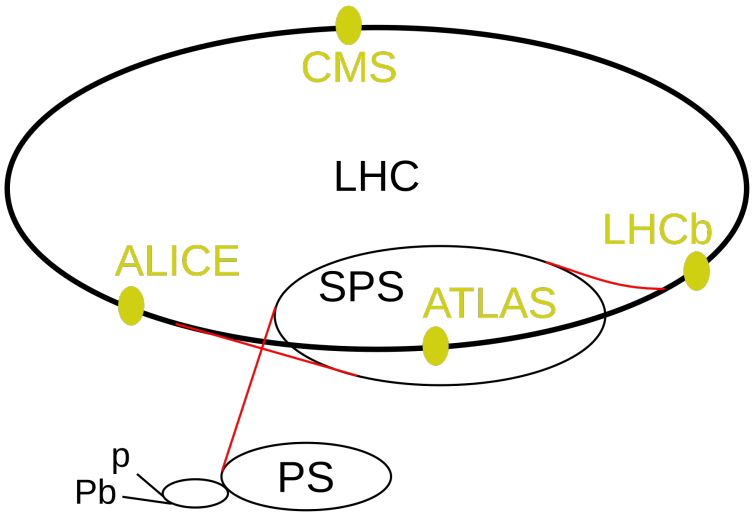


Figure 3.2: A general diagram illustrating the path that protons (or lead in the case of the ion collisions) take to reach the ultra-relativistic velocities for colliding in the CERN accelerator complex.

When high energy particles collide, they are able to create new, massive particles, the likes of which existed in the early universe. Most of these rare particles aren't found in

nature, but have to be produced in collision experiments like these. The production is related to the principle of energy-mass equivalence which describes mass as one of the portrayals of energy and so we are able to produce particles with higher masses as we collide particles with higher energies. These particles decay instantaneously from our perspective and cannot be observed directly. However, we are able to observe them indirectly by looking for their decay products in all of the events that are produced from the collisions. This isn't an easy thing to do as there are other processes which can have the same decay products and so our task lies in correctly identifying a maximum number of events that originate from the rare process of interest from the other processes that are more common that we don't care about. We call the former our signal and the latter our background.

3.1.1 LHC Operational Schedule

Given the immense complexity and scale of the LHC and its experiments, continuous operation requires periodic shutdowns for essential maintenance and substantial upgrades. Both the accelerator and its detectors, including ATLAS, undergo these scheduled long shutdowns to implement technical improvements and ensure sustained high performance. As a result, the LHC's physics program is organized into alternating periods of data-taking, called *Runs*, and extended *Long Shutdowns* (LS) for upgrades. This structured schedule allows the LHC to remain at the cutting edge of experimental high-energy physics. Analyses in this thesis utilize data from both Run 2 and (partial) Run 3 of the ATLAS experiment.

- **Run 1 (2010–2013)**: The initial data-taking period, with collisions at 7 TeV (2010–2011) and 8 TeV (2012–2013), culminating in the discovery of the Higgs boson in 2012.

- **Long Shutdown 1 (LS1, 2013–2015)**: Major upgrades to the accelerator complex and detectors to prepare for higher energy collisions.

- **Run 2 (2015–2018)**: Data-taking resumed at 13 TeV, expanding searches for new physics and enabling precision measurements of the Higgs boson and Standard Model processes.

- **Long Shutdown 2 (LS2, 2019–2021)**: Further upgrades to increase luminosity, enhance detector performance, and accommodate higher data rates.

- **Run 3 (2022–2026)**: Ongoing data-taking at 13.6 TeV, with improved detector capabilities and higher instantaneous luminosity.

- **Long Shutdown 3 (LS3, expected 2026–2029)**: Preparations for the High-Luminosity LHC (HL-LHC) including major enhancements to the accelerator and detectors.

- **High-Luminosity LHC (HL-LHC, expected 2030–2041)**: Operation at 14 TeV, with an expected tenfold increase in integrated luminosity, enabling precision studies of the Higgs boson, rare processes, and potential new physics.

A central quantity in collider physics is the *luminosity*, which determines the rate of collisions and thus the potential to observe rare processes. The *instantaneous luminosity* $\mathcal{L}$ is given by:

$$\mathcal{L} = f\, N_b\, \frac{n_1 n_2}{A}, \tag{3.2}$$

where f is the revolution frequency, N_b is the number of bunches per beam, n_1 and n_2 are the number of protons per bunch in each beam, and A is the effective transverse area of the beam overlap. The units of instantaneous luminosity are $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$.

The total *integrated luminosity* $\mathcal{L}_{\mathrm{int}}$ is the time integral of the instantaneous luminosity and quantifies the total dataset size collected:

$$\mathcal{L}_{\mathrm{int}} = \int \mathcal{L}\, dt. \tag{3.3}$$

The expected number of events for a given process with cross section σ is then $N = \sigma \times \mathcal{L}_{\mathrm{int}}$.

Luminosity is especially critical in the search for rare processes, such as Higgs boson pair production, which have extremely small cross sections compared to common background processes. Increasing both the instantaneous and integrated luminosity directly enhances the potential to observe these rare events and to perform precision measurements. Over the course of its operation, the LHC has steadily increased its luminosity, often exceeding

its original design parameters as shown in Table 3.3. For example, during Run 2, the instantaneous luminosity regularly surpassed 2×10^{34} cm^{-2} s^{-1}, nearly double the original design value. The ongoing Run 3 continues this trend with further increases in collision energy and luminosity, see Figure 3.4.

Run	**Period**	**Integrated Luminosity (fb^{-1})**
Run 1	2010	0.04
Run 1	2011	5.46
Run 1	2012	22.8
Run 2	2015	4.2
Run 2	2016	38.5
Run 2	2017	50.2
Run 2	2018	63.3
Run 3	2022	38.5
Run 3	2023	31.7
Run 3	2024	125.0

Figure 3.3: Total integrated luminosity delivered by the LHC to ATLAS in Runs 1, 2, and 3 (to date) for the *pp* collisions. Each row corresponds to the total luminosity collected during the respective data-taking years: 2010–2012 (Run 1), 2015–2018 (Run 2), and 2022–2024 (Run 3 to date) [50].

Looking ahead, the High-Luminosity LHC (HL-LHC) will provide a transformative increase in dataset size, aiming for an integrated luminosity roughly ten times greater than all previous runs combined. This will enable unprecedented sensitivity to rare processes and subtle effects in the Higgs sector and beyond, making luminosity the key to future discoveries in particle physics.

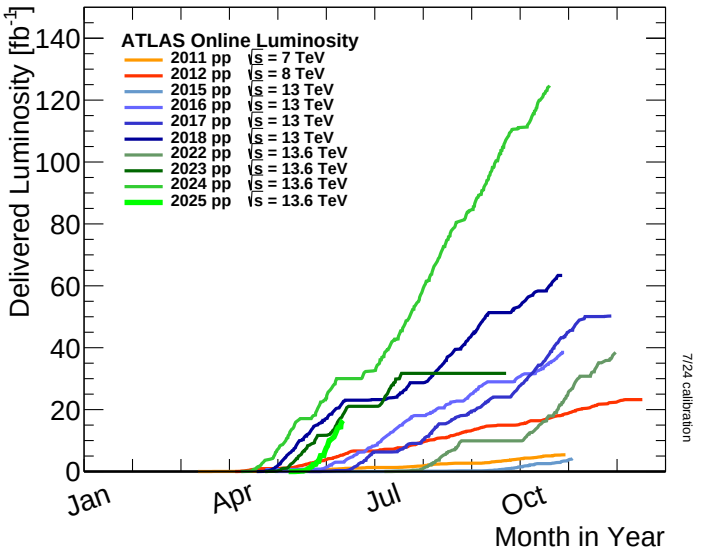


Figure 3.4: Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy pp collisions for each data-taking year. Note that data-taking for 2025 is ongoing [51].

3.1.2 Coordinate System

In the context of the LHC, a right-handed Cartesian coordinate system is employed where the z-axis is aligned along the beam direction, representing the longitudinal axis of the accelerator. The y-axis points vertically, typically referred to as "up", while the x-axis lies in the horizontal plane, pointing inward toward the center of the ring and outward. This standard (x, y, z) coordinate system is adapted to describe particle trajectories in the collider, but for tracking, detection, and analysis purposes, a cylindrical coordinate system is often more convenient due to the geometry of the LHC and its detectors.

In this cylindrical system, the coordinates are described in terms of transverse momentum, pseudorapidity, and the azimuthal angle. The transverse plane, perpendicular to the beam axis, is particularly important in collider physics, where transverse momentum (p_T) is the primary measurable component of a particle's momentum. Instead of directly using polar coordinates (r, θ, ϕ), the pseudorapidity η is often employed as a convenient variable

for the longitudinal component of particle motion. The pseudorapidity is related to the polar angle θ by the following expression:

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{3.4}$$

This formula is useful in high-energy physics because it simplifies the kinematic analysis of particle motion, especially when particles have high energies, where η can be a good approximation to rapidity. The overall coordinate system, including η, is crucial for the accurate tracking of particles as they move through the LHC detectors, as depicted along with the overall coordinate system in Figure 3.5.

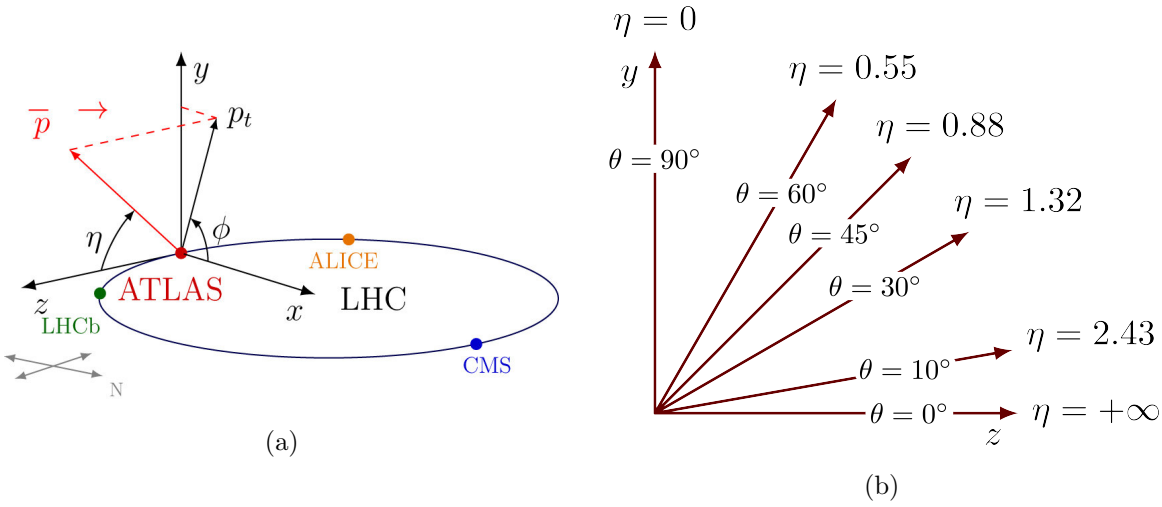


Figure 3.5: The typical coordinate system used in analyzing collisions at the LHC in 3.5a based on the four-vector of (p_T, η, ϕ, m). The pseudorapidity, η, is related to the polar angle in 3.5b.

Additionally, particles are characterized by a four-vector in the space defined by transverse momentum p_T, pseudorapidity η, azimuthal angle ϕ, and mass m. Here, p_T represents the magnitude of the momentum projected onto the transverse plane (the x-y plane) and it serves as a key observable in collider experiments. The importance of p_T stems from the fact that it is easily measurable using the tracking and calorimeter detectors which are

designed to capture the transverse components of particle motion. Notably, p_T is the most relevant momentum component for detecting hard collisions and high-energy interactions.

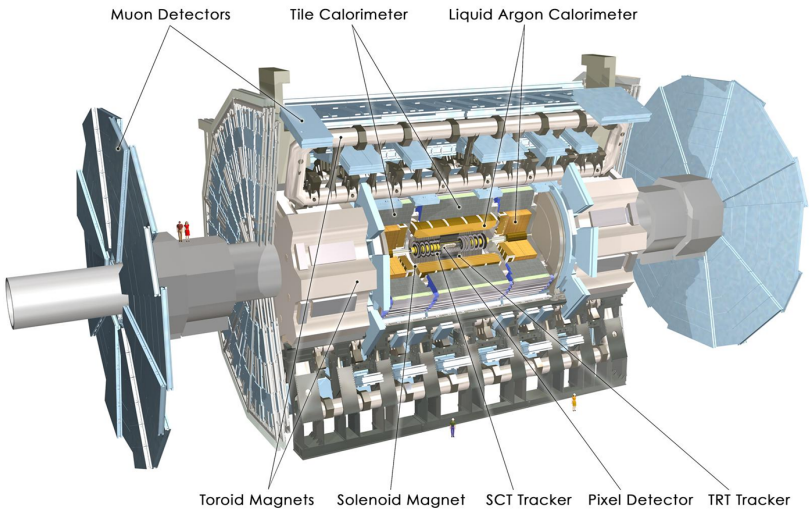
Furthermore, the concept of "missing transverse energy" (E_T^{miss}) arises when there is an imbalance in the transverse momentum of detected particles. This imbalance is indicative of particles that are undetectable by the experiment, such as neutrinos, or even new, hypothetical particles that could point to new physics beyond the Standard Model. This system of using transverse momentum and pseudorapidity is not only convenient but also essential for analyzing events in high-energy particle colliders like the LHC, where many phenomena are best understood in the transverse plane.

3.2 The ATLAS Detector

To fully exploit the large number of high energy particle collisions generated by the LHC, a highly sophisticated and precise detection system is required in order to identify, measure, and reconstruct the resulting particles and their interactions. This essential role is fulfilled by the ATLAS detector. Designed to capture and analyze collision events with exceptional spatial and temporal resolution, ATLAS enables researchers to probe a broad spectrum of physics phenomena. The detector is cylindrical in shape with large endcap components on either side to ensure near-complete solid-angle coverage around the interaction point as illustrated in Figure 3.6.

Precise tracking of charged particles is a fundamental component of event reconstruction in high-energy physics experiments. It enables accurate measurements of particle momenta, the identification of decay vertices, and plays a critical role in distinguishing particle types. In the ATLAS detector, this function is performed by the Inner Detector (ID), which operates within the pseudorapidity range $|\eta| < 2.5$. The ID is composed of three nested subsystems: the Pixel Detector, the Semiconductor Tracker (SCT), and the Transition Radiation Tracker (TRT), each contributing complementary information to the reconstruction of particle trajectories.

The innermost component, the Pixel Detector, consists of four concentric layers of high-resolution silicon pixels. Charged particles traversing these layers deposit small amounts of energy, producing signals that are measured with a spatial resolution of approximately



(a)

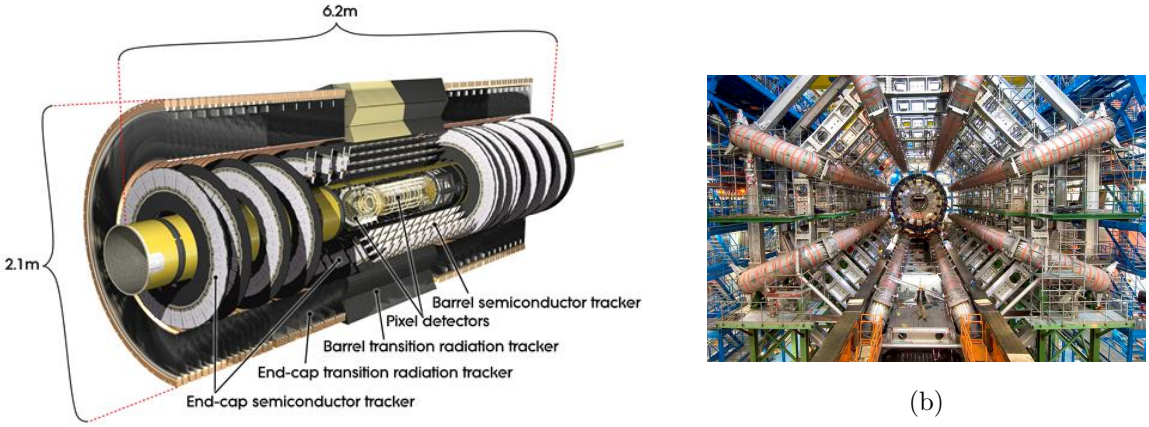


(b)

Figure 3.6: A layout of the ATLAS subdetector systems in 3.6a with two people shown left for size. In 3.6b is myself standing in front of the ATLAS Detector during one of the winter shutdowns when cavern visits are allowed.

10 μm in the transverse plane and about 115 μm along the beam direction [52]. This exceptional granularity is made possible by over 92 million pixels and allows for precise reconstruction of particle origins, impact parameters, and short-lived decay vertices. Surrounding the Pixel Detector is the SCT, which includes 4,088 modules of silicon microstrips totaling about 6 million sensing elements. The SCT provides a spatial resolution of about 17 μm in the r–ϕ plane and 580 μm in the z direction [53]. The layers are arranged such that most particles pass through at least four distinct measurement planes. The curvature of these particle trajectories in the presence of a magnetic field reveals both their charge (based on direction) and momentum (inversely related to curvature radius), enabling detailed kinematic reconstruction. The outermost tracking layer is the TRT, composed of approximately 300,000 gas-filled drift tubes, or "straws." The TRT provides a hit resolution of about 130 μm per straw and enhances the overall tracking capability as well as provides electron identification via detection of transition radiation [54, 55].

Surrounding the Inner Detector is the Central Solenoid Magnet, shown in Figure 3.7, which provides a uniform axial magnetic field of 2 T. This magnetic field enables the measurement of charged particle momenta with a resolution of approximately 1% for tracks with $p_T \sim 10$ GeV in the central region, rising to about 10% at $p_T \sim 1$ TeV [56]. Constructed using niobium-titanium superconducting wire embedded in aluminum stabilizers, the solenoid ensures the curvature of charged particle tracks necessary for momentum measurements. Beyond the solenoid lie three large toroidal magnet systems: two End-Cap Toroids and the Barrel Toroid, which is the largest toroidal magnet ever constructed. This barrel magnet is composed of eight massive superconducting coils arranged symmetrically around the detector and plays a crucial role in bending muon trajectories in the outermost regions of the ATLAS detector.



(a)

(b)

Figure 3.7: Overview of ATLAS tracking and magnet systems. (a) The Inner Detector (ID), composed of the Pixel Detector, Semiconductor Tracker (SCT), and Transition Radiation Tracker (TRT), provides high-precision tracking of charged particles. (b) The eight superconducting coils of the Barrel Toroid magnet, part of the ATLAS toroidal system, generate a magnetic field used primarily to bend muon trajectories in the outer detector region. Together with the Central Solenoid, these systems are essential for tracking and identifying particles across a wide momentum range.

After passing through the Inner Detector, particles enter the calorimeters, which are crucial components of the ATLAS detector. Calorimeters are specifically designed to measure the energy of particles by detecting the cascades of secondary particles, known as particle showers, that result when high-energy particles interact with dense materials. These showers are categorized into two main types: electromagnetic showers and hadronic showers, each of which follows distinct physical processes and requires different detection strategies.

Electromagnetic showers occur through the combination of electron-positron pair production and braking radiation. For example, when high-energy photons interact with the calorimeter material through pair production, they create an electron-positron pair. These electrons and positrons then undergo bremsstrahlung (braking radiation), emitting additional photons and further contributing to the cascade of particles. This process results in a shower of particles that deposit their energy in a relatively compact region, making electromagnetic showers well-suited to precise energy measurements.

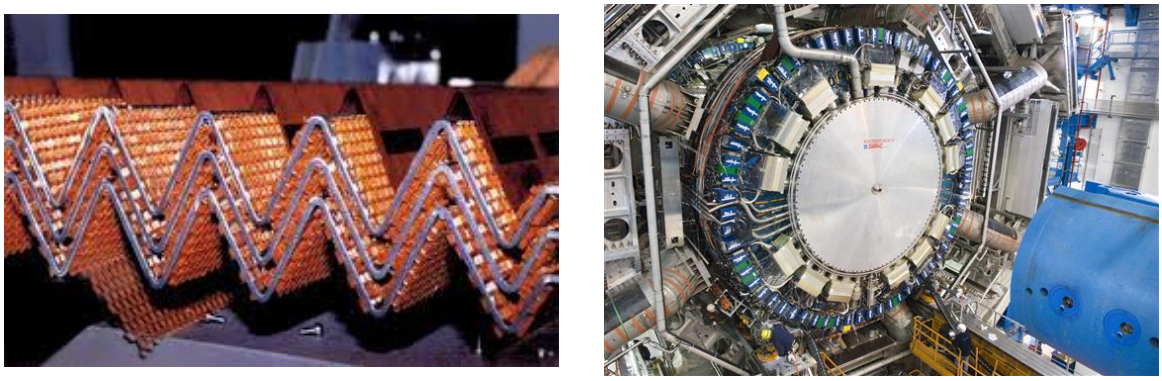
As such, the first calorimeter in the ATLAS detector is the Electromagnetic Calorimeter (EM Calorimeter), which is optimized to measure the energy of electrons and photons with high accuracy. The EM Calorimeter consists of alternating layers of metal, as shown in Figure 3.8a, which absorb the incoming particles and produce secondary particles that ionize the liquid argon between the layers. These ionizations generate an electric current that is measured, allowing for precise energy measurements from the interaction of particles in different cells. The EM Calorimeter achieves an energy resolution of

$$\frac{\sigma_E}{E} \approx \frac{10\%}{\sqrt{E\ [\text{GeV}]}} \oplus 0.7\%$$

for isolated electrons and photons, where E is the particle energy [57].

In contrast, hadronic showers involve particles that interact strongly with atomic nuclei, making these showers more complex and involving a broader range of processes. The heavier hadrons produced in these interactions do not undergo significant bremsstrahlung and instead primarily ionize the material as they pass through it. These interactions, with varying radiation lengths depending on the process, generate a shower of secondary particles that deposit energy over a larger volume compared to electromagnetic showers.

The second calorimeter in the ATLAS detector, the Tile Hadronic Calorimeter, is de-



(a) (b)

Figure 3.8: Illustrations of the two main calorimeter systems in the ATLAS detector. (a) View of the accordion-shaped structure of alternating absorber and readout layers of the Electromagnetic Calorimeter. (b) View of the Tile Hadronic Calorimeter, which consists of alternating layers of steel and plastic scintillator tiles. This system is optimized for measuring the energy of hadrons and forms the largest and most massive component of the calorimeter system.

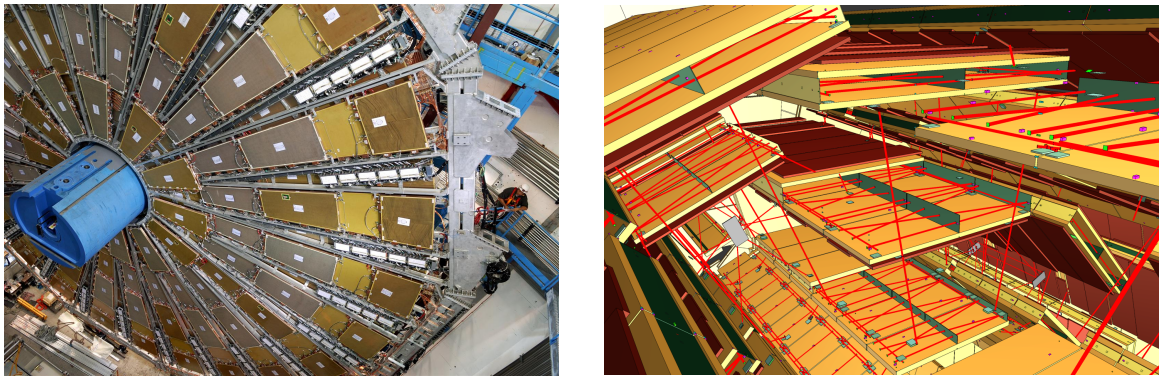
signed to measure the energy of hadrons. This calorimeter consists of alternating layers of steel and plastic scintillator tiles, as shown in Figure 3.8b, with over 420,000 plastic scintillator tiles forming the largest and heaviest part of the detector. When hadronic particles strike the steel layers, they create a secondary shower of particles, which then interact with the scintillator tiles. The resulting interactions produce photons that are converted into an electric current, which is detected by photomultiplier tubes. The intensity of this current is proportional to the energy of the incoming hadron, allowing for precise energy measurements. The hadronic calorimeter achieves an energy resolution of approximately

$$\frac{\sigma_E}{E} \approx \frac{50\%}{\sqrt{E\ [\text{GeV}]}} \oplus 3\%$$

for isolated hadrons [58].

The electromagnetic and hadronic calorimeters together provide comprehensive coverage of particle energy measurement across the full range of interactions.

At this stage, all stable particles of the Standard Model would have been detected by the calorimeters and Inner Detector, with the notable exceptions of muons and neutrinos. To address this, the ATLAS detector includes a specialized Muon Spectrometer, located just before the endcaps, which is designed to detect muons, particles that are highly penetrating and can pass through the calorimeter and Inner Detector without significant energy loss. The Muon Spectrometer consists of a toroidal magnet and three layers of precise drift chambers, as shown in Figure 3.9. The system achieves a momentum resolution of about 3% for muons with $p_T = 10$ GeV in the central region, increasing to 10% at $p_T = 1$ TeV, depending on the track angle and position [59, 60]. This system allows for accurate measurement of the muons' momentum and trajectory, crucial for studying their behavior and interactions in particle collisions.



(a) (b)

Figure 3.9: ATLAS Muon Spectrometer overview. (a) Schematic of the spectrometer, showing the toroidal magnet system and layered drift chambers surrounding the calorimeters. (b) Diagram of the alignment system that ensures precise chamber positioning for accurate momentum measurements.

Neutrinos, on the other hand, present a unique challenge. As neutral, nearly massless particles that interact only via the weak nuclear force, they cannot be directly detected by the ATLAS detector. However, their presence can be inferred through the concept of missing

transverse energy. When neutrinos are produced in particle collisions, they carry away energy and momentum that cannot be detected directly by any of the subdetector systems as shown in Figure 3.10. By applying the principles of energy and momentum conservation, the missing energy in the transverse plane can be calculated, providing indirect evidence of the neutrinos' existence and enabling researchers to reconstruct their properties. This method is crucial for many physics analyses, including those involving processes that are sensitive to neutrinos, such as certain decay modes and the search for new physics.

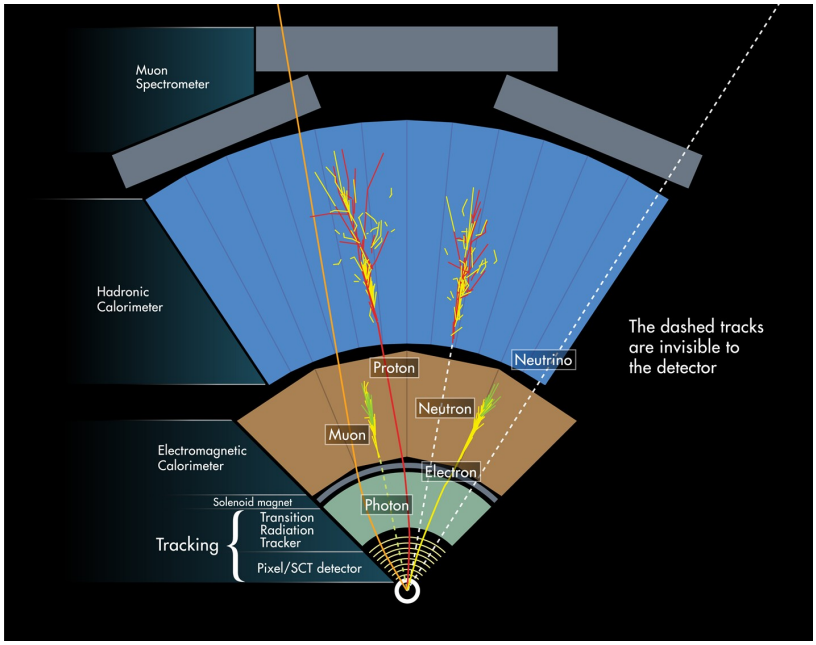


Figure 3.10: A 2D slice of the ATLAS Detector showing the different subdetector systems and how they function for object reconstruction in the events. The dashed lines indicate the neutrinos that pass through undetected.

3.2.1 Phase-1 Upgrades

The transition from Run 2 to Run 3 of the LHC brought a substantial increase in both luminosity and collision energy, necessitating a suite of targeted upgrades to the ATLAS detector during Long Shutdown 2 (LS2, 2019–2021). Key developments during this Phase-1

upgrade period included the enhancement of pixel and tracking electronics in the Insertable B-Layer which strengthened tracking performance in the inner detector. The endcap regions of the muon spectrometer were equipped with the New Small Wheel (NSW) system, which significantly boosted muon triggering efficiency and momentum resolution, particularly under high-background conditions. In addition, the calorimeter and forward detector readout systems were modernized with advanced electronics to reliably handle the elevated data throughput expected with higher collision rates [61, 62].

3.2.2 Phase-2 Upgrades

The advent of the High-Luminosity LHC (HL-LHC) will usher in an unprecedented leap in instantaneous luminosity, as well as a modest increase in collision energy. To meet the demanding conditions expected during this era, a comprehensive set of upgrades, collectively referred to as Phase-2, will be implemented during Long Shutdown 3 (LS3, projected for 2026–2029) [63]. The most transformative change will be the complete replacement of the current inner detector with the all-silicon Inner Tracker (ITk), featuring extended geometric coverage, increased granularity, and enhanced resilience to radiation. The muon spectrometer will undergo further modernization of its detectors and trigger systems, while the calorimeter electronics will be replaced with faster, more robust components to support substantially higher data rates [64, 65]. A major innovation will be the introduction of the High-Granularity Timing Detector (HGTD) in the forward region, designed to provide precision timing measurements for each track with a target time resolution better than 30 ps [66]. This capability will significantly improve the ability to distinguish between interactions occurring within the same bunch crossing (pile-up). Together, these upgrades will ensure that ATLAS retains its sensitivity to rare processes and new physics, even under the challenging conditions of the HL-LHC.

3.3 Trigger and Data Storage

The ATLAS detector is capable of observing up to 1.7 billion proton-proton collisions per second, generating approximately 60 million megabytes of data each second. Given the sheer volume of these events, most of which are not relevant for the research goals, a sophisticated

trigger and data acquisition system is essential to efficiently select the most promising events that could lead to new discoveries [67, 68]. The selection process begins with the first-level hardware trigger, which operates in real-time within just 2.5 microseconds. This initial trigger filters events using a subset of information from the calorimeters and Muon Spectrometer to identify potentially interesting collisions. The data from these events is then transferred to a storage buffer, capable of handling up to 100,000 events per second, for a second round of screening.

In the second stage, the software trigger comes into play. This trigger is executed by a large farm of about 40,000 CPU cores, where a more detailed analysis of each event is performed in approximately 200 microseconds. Using more complex algorithms, the second-level software trigger identifies the top 1000 events per second, which are then recorded in the data storage system. These selected events are preserved for offline analysis by ATLAS physicists, who conduct deeper investigations into the fundamental physics behind the collisions. Overall, the data-taking process achieves an impressive efficiency of 95%, ensuring that only the most relevant data is retained for further analysis and discovery.

3.4 Processing of Simulated and Real Data

Simulation is a critical component of ATLAS analyses, providing the foundation for understanding both signal and background processes. By generating synthetic data based on theoretical models and simulating the detector's response, it becomes possible to interpret experimental observations, optimize event selection strategies, and estimate background contributions with high precision. Accurate simulations are essential for quantifying systematic uncertainties and ensuring reliable measurements in the complex environment of the LHC.

The simulation workflow consists of four main stages: event generation, detector simulation, digitization, and reconstruction [69].

- **Event Generation:** The process begins with generating the underlying physics events based on theoretical models. This involves simulating the interactions of quarks and gluons within the colliding protons, governed by the Standard Model or beyond-

the-Standard Model (BSM) theories. Event generators such as PYTHIA, HERWIG, and MADGRAPH are commonly used to model the hard scattering process, parton showers, hadronization, and decays of unstable particles [70]. The output is a detailed kinematic description of all final-state particles produced in the collision.

- **Detector Simulation:** The next step involves simulating the passage of these particles through the ATLAS detector. Tools such as GEANT4 are used to model the interactions between particles and detector materials, including processes like ionization, bremsstrahlung, multiple scattering, Compton scattering, and nuclear interactions [71]. A realistic and detailed geometry of the detector is used to ensure that the response is accurately represented. This stage produces simulated detector signals such as energy deposits, hit times, and spatial locations.

- **Digitization:** In the next stage, the simulated detector signals are converted into a format that mimics the raw data collected during real data-taking. This process involves modeling the electronics response of the detector, including digitized outputs like voltages, currents, and time stamps. The result is a simulated dataset that can be processed by the same reconstruction and analysis algorithms used for real collision data.

- **Reconstruction:** Finally, both the digitized simulated data and the actual raw collision data are passed through the same reconstruction algorithms. These algorithms translate detector signals into high-level physics objects including tracks, jets, electrons, muons, and missing transverse energy which can then be used for analysis. The reconstruction step ensures that simulated events can be directly compared to experimental data and are subject to the same object identification criteria, calibration, and selection procedures [72].

Together, these simulation stages provide a consistent and controlled environment for evaluating the performance of the detector and analysis algorithms. They also enable the

construction of control and signal regions, validation of background models, and interpretation of observed data within a theoretical framework.

Chapter 4

JET CALIBRATION IN FAST SIMULATION

4.1 Jet Definitions

The ATLAS detector is capable of observing all Standard Model particles either directly or indirectly through decay products or missing energy signatures. Each particle type interacts differently with the detector subsystems allowing for specialized identification strategies as discussed in the preceding chapter. Among the most commonly produced in proton-proton collisions are quarks and gluons which are collectively known as partons. Due to the property of color confinement in quantum chromodynamics (QCD), partons cannot exist in isolation. Instead, they hadronize, forming bound color-neutral states.

When high-energy quarks or gluons are produced, they rapidly fragment into collimated sprays of hadrons, a process known as hadronization. These sprays, referred to as **jets**, are observed in the detector as clusters of energy deposits and tracks aligned along the original parton's trajectory. An important exception is the top quark, which decays via the weak interaction before it can hadronize.

In a typical LHC event, multiple partons may be produced simultaneously, complicating the association between observed energy deposits and their partonic origin. Consequently, sophisticated reconstruction algorithms are employed to define and identify jets in a way that is both efficient and robust to experimental effects like pile-up, additional pp collisions occurring in the same or nearby bunch crossings.

4.1.1 An Aside on Pile-Up

An essential consideration in jet reconstruction is the effect of pile-up, since these additional interactions can introduce significant background activity that overlaps with the primary hard-scatter event under study [74]. Pile-up arises in two forms: *in-time pile-up*, which refers to multiple collisions occurring within the same bunch crossing, and *out-of-time pile-up*,

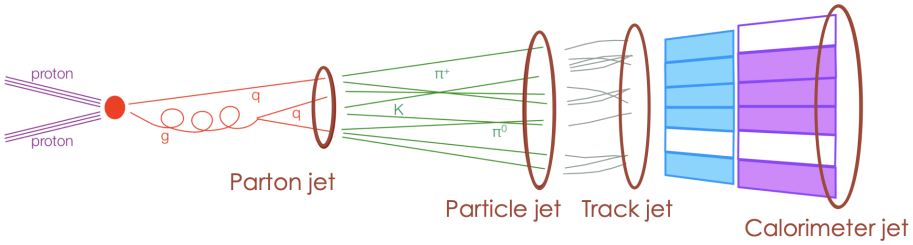


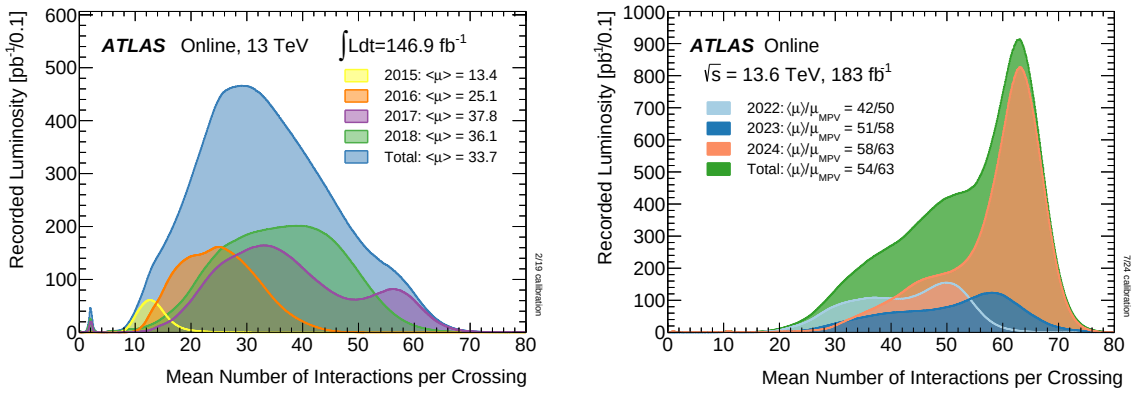
Figure 4.1: The different stages of jet development going from the pp collision to the produced partons to stable particle jets to track jets and then the final calorimeter jets. Figure taken from [73].

resulting from interactions in adjacent bunch crossings that affect the detector response due to finite readout times. This is particularly relevant for subdetectors like the calorimeters, which have relatively long integration windows.

The impact of pile-up is widespread. It contributes spurious tracks and calorimeter energy deposits, distorts the measurement of missing transverse momentum, and complicates the identification of physics objects [75]. Most critically for jet reconstruction, pile-up can lead to incorrect clustering of energy deposits, degradation of jet energy and mass resolution, and misidentification of jet substructure. To mitigate these effects, modern jet algorithms incorporate pile-up suppression techniques at both the object reconstruction and jet clustering stages. As such, pile-up mitigation is an integral part of defining and calibrating jets in the high-luminosity environment of the LHC.

The level of pile-up is typically characterized by the average number of interactions per bunch crossing, $\langle\mu\rangle$. During Run 2 (2015–2018), ATLAS operated with $\langle\mu\rangle$ values typically between 20 and 40, with peak values exceeding 60 in the later years of the run, see Figure 4.2. In Run 3, the increased instantaneous luminosity has led to $\langle\mu\rangle$ values in the range of 40 to 60 and the trend is expected to continue upwards. Looking ahead to Run 4 and the HL-LHC, pile-up conditions will become significantly more challenging with $\langle\mu\rangle$ projected to reach 140–200 [76]. These extreme conditions will place even greater demands on pile-up mitigation strategies and detector technology, making effective pile-up suppression a central

challenge for precision measurements and new physics searches in the HL-LHC era.



(a) (b)

Figure 4.2: The average number of interactions per bunch crossing, $\langle\mu\rangle$, illustrating the pile-up density profiles for the data-taking years of Run 2 (4.2a) and Run 3 (4.2b) [50, 51].

The energy resolution for traditional calorimeter-based (EMTopo) jets is typically around 5–15% for jets with p_T of a few hundred GeV, and mass resolution is even worse, especially in the presence of pile-up. The need for better resolution and pile-up resilience motivates the development of more advanced jet algorithms.

4.1.2 Jet Clustering

Jet reconstruction in ATLAS begins with clustering algorithms that group together individual particles or energy deposits into composite objects interpreted as jets. The most widely used algorithm is the anti-k_t algorithm, a sequential clustering method that operates on a set of input constituents using a distance metric defined by their transverse momentum and angular separation [77]. In ATLAS, jets are typically categorized by the radius parameter R used in the anti-k_t algorithm, where $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. **Small-$R$ jets** refer to jets clustered with $R = 0.4$, optimized for the reconstruction of individual quark and gluon jets from the hard scatter. **Large-R jets** use a larger radius parameter of $R = 1.0$ and are particularly suited for capturing the decay products of highly boosted heavy particles, such as top

quarks or W, Z, and Higgs bosons. The choice of radius parameter is analysis-dependent and balances capturing all relevant energy while minimizing contamination from unrelated activity.

The anti-k_t algorithm defines a distance metric based on the transverse momentum and angular separation between particles and then iteratively combines the closest pairs until all are assigned to a jet. Its tendency to cluster around high-p_T particles makes it especially effective in complex hadronic environments. While it is robust and computationally efficient, it also may obscure internal jet substructure and can be challenged by nearby or overlapping jets.

4.1.3 Electromagnetic Topological Jets (EMTopo)

The earliest jet definitions in ATLAS were based solely on calorimeter information. Electromagnetic Topological (EMTopo) jets are reconstructed from topological clusters built at the electromagnetic (EM) energy scale in the calorimeter [78]. This approach represents a direct measurement of the energy deposited in the calorimeter, before full jet calibration is applied. EMTopo jets served as a baseline for many physics analyses during Run 1 of the LHC. However, as the LHC's luminosity increased, the limitations of calorimeter-only jets, such as their susceptibility to pile-up and relatively coarse spatial resolution, became a significant factor in high-precision measurements. This spurred the development of more sophisticated jet definitions that combine information from both the calorimeter and the tracking system.

4.1.4 Electromagnetic Particle Flow Jets (EMPFlow)

A major advancement in jet reconstruction came with the introduction of Electromagnetic Particle Flow (EMPFlow) jets, which are built from Particle Flow Objects (PFOs) [79]. This method leverages the strengths of both the inner tracking system and the calorimeter: the high-resolution tracking detector provides precise momentum measurements for charged particles, while the calorimeter measures the energy of neutral particles. The core of the PFlow algorithm lies in a sophisticated process of combining these measurements

and mitigating double-counting. Specifically, the energy deposited by charged tracks in the calorimeter is subtracted from the corresponding topological clusters, and the remaining energy is attributed to neutral particles. These PFOs are then clustered into jets using an algorithm such as anti-k_t.

The EMPFlow approach offers substantial improvements over EMTopo jets, particularly for low- and moderate-momentum jets. EMPFlow jets achieve an improved jet energy resolution of approximately 8–10% and a mass resolution of 15–20% at $p_T \sim 200$ GeV, compared to 10–15% and 20–25% respectively for EMTopo jets [79]. Furthermore, they are more resilient to pile-up due to techniques like Charged Hadron Subtraction (CHS), which removes tracks not originating from the primary vertex. These improvements make EMPFlow jets better suited for analyses requiring high-precision measurements of jet properties and substructure.

4.1.5 Track-CaloClusters (TCCs)

Another method to combine information from the inner detector and calorimeter is the Track-CaloClusters (TCCs) approach [80]. TCCs are composite objects created by matching charged particle tracks to nearby calorimeter topological clusters. This association allows for the precise momentum of the track to be combined with the energy and shower shape information from the calorimeter. This is particularly effective in dense environments, where it helps resolve ambiguities from overlapping particle showers. Unmatched tracks and calorimeter clusters are also retained as TCCs, ensuring a comprehensive representation of the event.

While the TCC procedure is a type of particle flow, it is distinct from the EMPFlow algorithm. The primary difference lies in how energy is shared and corrected. The TCC approach uses the spatial coordinates from the tracker to guide the energy clustering but relies on the calorimeter's energy scale. In contrast, the EMPFlow algorithm uses an energy subtraction method that allows the tracking detector measurements to fully correct the 4-vector of charged particles [81]. TCCs were specifically optimized for improving jet substructure in high-momentum jets, making them complementary to the EMPFlow ap-

proach which excels at lower momenta. Both methods, however, share the common goal of combining tracker and calorimeter information to enhance the understanding of hadronic showers and improve jet performance.

4.1.6 Unified Flow Object Jets (UFOs)

The latest evolution in jet reconstruction in ATLAS is the Unified Flow Object (UFO) approach, which combines the strengths of both Particle Flow Objects and TCCs into a single, cohesive framework [80]. This is a sophisticated refinement of the PFlow technique, where neutral PFOs are further subdivided and repositioned based on their associated tracks. The UFO algorithm correlates low-level objects, including tracks and individual calorimeter energy deposits, before any jet finding algorithms are run. This hybrid strategy takes advantage of the inner detector's superior angular resolution and the calorimeter's precise energy measurements, providing optimal performance across a wide kinematic range.

UFO jets significantly outperform both EMTopo and EMPFlow jets, particularly in terms of jet mass, angular, and substructure resolution. For instance, UFO jets achieve a jet mass resolution as low as 10% at $p_T \sim 200$ GeV, a substantial improvement over EMPFlow jets, and about 9% at $p_T = 500$ GeV, compared to 13% for EMTopo jets [80]. Their robust pile-up mitigation strategy, which combines Constituent Subtraction (CS) with the SoftKiller (SK) algorithm, ensures stable performance under the challenging conditions of high-luminosity runs [82]. This two-step process estimates the pile-up energy density and dynamically subtracts it from jet constituents, minimizing contamination and stabilizing jet energy measurements. As a result, UFO jets are becoming a new standard for large-radius jet reconstruction in ATLAS, providing an ideal foundation for high-precision measurements in Run 3 and the High-Luminosity LHC era.

Although multiple jet definitions are maintained within ATLAS for different use cases, EMPFlow and UFO jets represent modern advancements that integrate tracking and calorimeter data for improved performance. The optimal choice of jet definition depends on the specific requirements of a physics analysis including balancing resolution, pile-up resilience, and sensitivity to jet substructure features. Calibration procedures, which further refine jet

energy and mass measurements, are addressed in the following section.

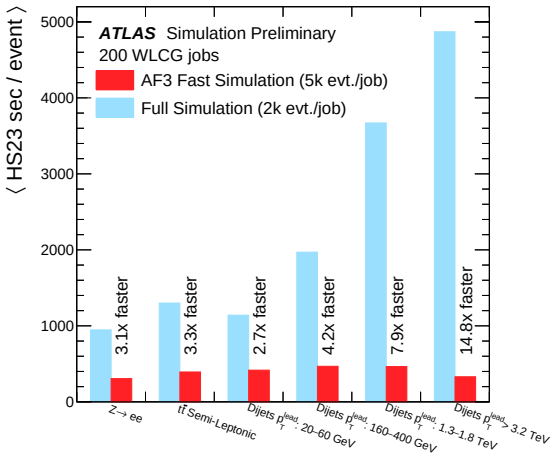
4.2 AtlFast3 Simulation

AtlFast3 (AF3) is the newest fast simulation framework used in the ATLAS experiment to model particle interactions with the detector while significantly reducing computational requirements [83, 84]. It follows a top-down approach focusing on improving the slowest parts of ATLAS simulation, the calorimeters. Unlike full Geant4 simulations (FullSim) [85, 86], which track every particle interaction in detail, AF3 uses parameterized models for certain detector responses, accelerating simulation while maintaining high accuracy. This allows researchers to simulate large datasets efficiently, which is crucial for physics analyses requiring extensive Monte Carlo statistics.

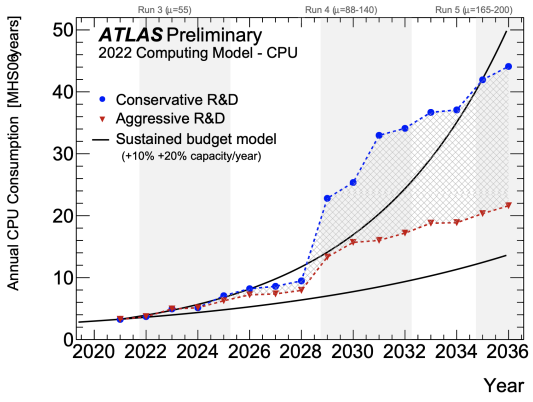
A key advancement in AF3 is its incorporation of machine learning, particularly in calorimeter simulation, which is the most computationally intensive part of full simulation. The framework combines two approaches: FastCaloSim V2, a parametric model that simulates electromagnetic and hadronic showers using predefined energy distributions, and FastCaloGAN, a Generative Adversarial Network (GAN)-based model that improves the simulation of complex hadronic showers [87]. The GANs learn from detailed Geant4 simulations to reproduce correlated fluctuations in the energy distribution, especially for medium-energy hadrons. This combination allows AF3 to achieve a balance between speed and accuracy with machine learning models ensuring that the simulated calorimeter responses closely match those of Geant4 while drastically reducing CPU consumption [88].

Fast simulation is particularly critical due to the enormous computational demands at the LHC. A full Geant4 simulation can take minutes per event, making it impractical for datasets containing billions of collisions. AF3 reduces this significantly which enables large-scale studies within manageable computing time. With the upcoming High-Luminosity LHC (HL-LHC) upgrade, which will increase collision rates and event complexity, the need for optimized fast simulation tools like AF3 is more pressing than ever [89].

CERN's computing infrastructure must handle petabytes of data annually, requiring a distributed Grid Computing system with thousands of processing nodes [92]. Even with this vast resource, balancing precision and efficiency is essential, as computational costs continue



(a)

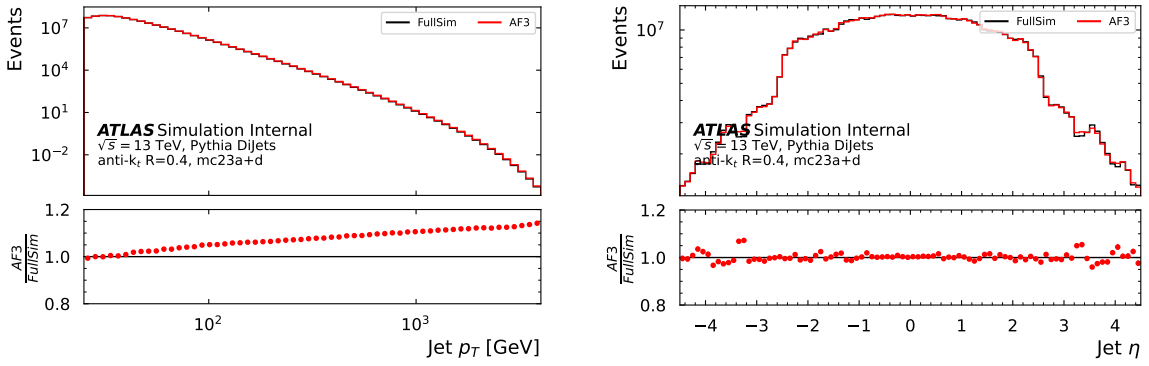


(b)

Figure 4.3: The mean CPU time per event simulated in Fast Simulation (AF3) and Full Simulation measured in standardized HS23 seconds for six different physics processes in 4.3a. Figure 4.3b show projected CPU consumption for the ATLAS experiment from 2020 to 2036 varies based on R&D investment [90, 91].

to grow. The integration of fast simulation methods helps ensure that large-scale physics analyses remain feasible while maintaining the necessary accuracy for discoveries in particle physics. Figure 4.3 illustrates both the faster simulation time when using AF3 instead of FullSim for reconstructing photons and the future CPU needs in the ATLAS Experiment.

While AF3 generally provides modeling performance comparable to FullSim, notable discrepancies remain in specific observables. In particular, reconstructed jets in AF3 tend to exhibit higher transverse momentum (p_T) than those in FullSim, with the deviation increasing at higher p_T, as illustrated in Figure 4.4. Additionally, the reconstructed pseudorapidity (η) distribution shows localized mismodeling in certain regions of the detector. These discrepancies are most pronounced near the calorimeter transition boundaries, where the parametric reconstruction struggles to accurately capture the complex detector geometry and material distribution. For example, non-closure effects in η are correlated with the interface between the hadronic and forward calorimeters, as shown in Figure 4.5. These lim-



(a) Reconstructed jet p_T distribution.

(b) Reconstructed jet η distribution.

Figure 4.4: Comparison of reconstructed small-R PFlow jet distributions in AF3 and FullSim for simulated Pythia dijet events. (a) The jet transverse momentum (p_T) distribution, showing that AF3 tends to overestimate p_T, particularly at high energies. (b) The jet pseudorapidity (η) distribution, highlighting discrepancies near calorimeter transition regions due to limitations in the parametric modeling. These examples illustrate key non-closure effects in AF3, motivating further improvements in jet reconstruction performance.

itations highlight areas where further refinement of the AF3 parametrizations is necessary to improve accuracy in jet reconstruction.

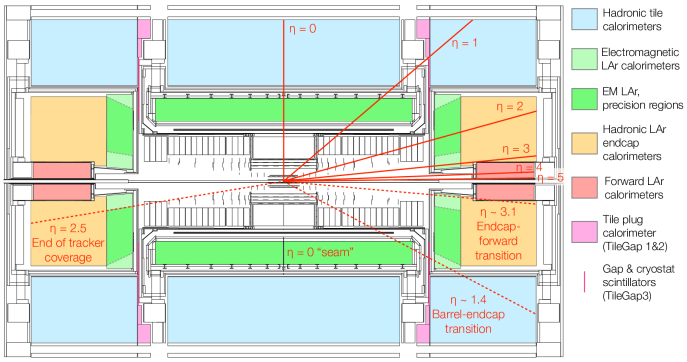


Figure 4.5: A two-dimensional slice of the ATLAS detector showing the calorimeter geometries at specific pseudorapidities along the beam axis. The poor modeling of jet η in AF3 tends to occur near the calorimeter boundaries [84].

4.2.1 AF3 Derived small-RJet Calibrations

Jet calibration in ATLAS is a multi-step process that corrects reconstructed jet energies to match particle-level truth jets, ensuring accurate measurements for physics analyses. This procedure involves an initial set of simulation-based corrections, followed by in situ adjustments using real collision data to refine the jet energy scale [93]. The overall process for small-R jets is defined in Figure 4.6.

My work involved taking the reconstructed small-R jets provided in the AF3 samples and deriving the pile-up corrections, absolute MC-based calibration, and the global calibration for the small-R UFO jets as will be shown hereafter. All calibrations and systematic uncertainty studies in this work are based on simulated dijet ($pp \rightarrow$ jets) samples generated with Pythia8 and the A14 tune, produced in both the mc20 and mc23 Monte Carlo campaigns. The mc20 campaign corresponds to Run 2 data-taking conditions (2015–2018) at a center-of-mass energy of 13 TeV, while mc23 reflects Run 3 conditions (2022 onward) with an updated detector geometry and a center-of-mass energy of 13.6 TeV. For each campaign, the full range of JZ slices (`JZ0`–`JZ12`) was used, covering the full p_T spectrum relevant for jet analyses. Each JZ slice represents a different jet p_T range, ensuring good statistical precision across all kinematic regions. The analysis uses the derived DAOD_JETM1 format

885 with Small- and Large-R UFO and PFlow jets for all years.

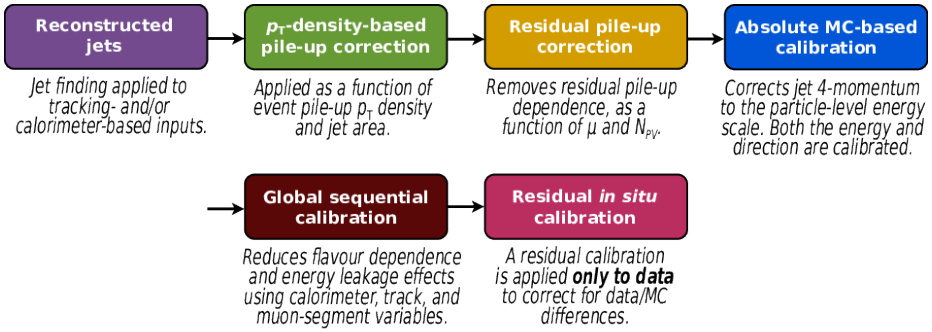


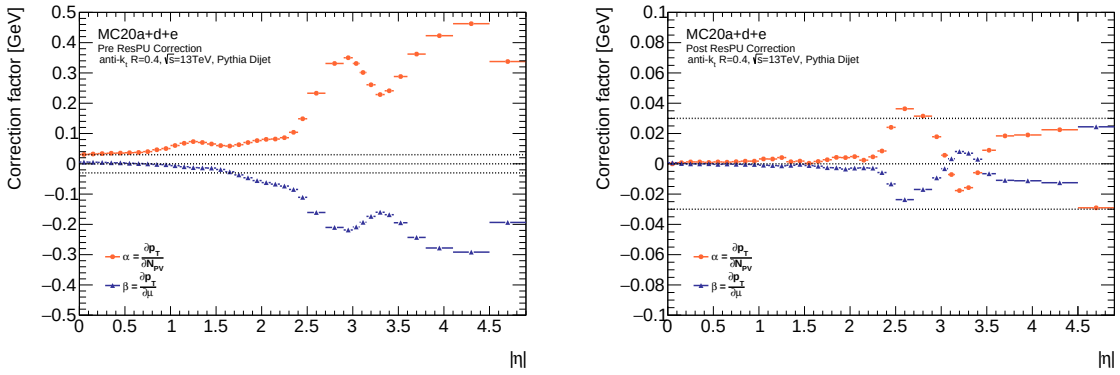
Figure 4.6: Jet calibration sequence for small-R jets in the ATLAS experiment. Starting from reconstructed jets, successive corrections are applied: pile-up subtraction, a residual pile-up correction, an absolute Monte Carlo–based calibration (also known as η-JES corrections), a Global Sequential Calibration to improve resolution and reduce flavor dependence, and a final in situ correction to align data with simulation [93].

886 The calibration begins with pile-up corrections to remove contamination from multiple
887 proton-proton interactions occurring within the same or nearby bunch crossings. A jet-area-
888 based correction is first applied, estimating the additional energy from pile-up by measuring
889 the transverse momentum density across the event and subtracting the excess energy based
890 on each jet's area. A residual pile-up correction, Equation 4.1, further refines the jet energy
891 by accounting for dependencies on the number of reconstructed primary vertices and the
892 average number of interactions per event.

$$p_T^{corr} = p_T^{reco} - \underbrace{\rho \times A}_{\substack{p_T\text{-density-based} \\ \text{pile-up correction}}} \underbrace{- \alpha \times (N_{PV} - 1) - \beta \times \mu}_{\substack{\text{residual} \\ \text{pile-up correction}}} \tag{4.1}$$

893 The correction factors are defined as $\alpha = \frac{\partial p_T}{\partial N_{PV}}$ and $\beta = \frac{\partial p_T}{\partial \mu}$ and are initially calculated
894 in bins of $|\eta|$ and applied to their corresponding jets, they are then recalculated. If the
895 second calculation of correction factors yields values less than 0.03, then this is deemed
896 to be acceptable as can be seen in the pre and post-application of the correction factors

in Figure 4.7. These steps help mitigate pile-up effects, which can artificially inflate jet energies and degrade resolution.



(a) (b)

Figure 4.7: Residual pile-up correction factors for AF3-simulated jets as a function of $|\eta|$ using Pythia dijet samples from the MC20 campaign. (a) Correction factors α and β obtained from the first iteration, before applying any residual pile-up correction. (b) Correction factors recalculated after the initial correction is applied. Values below 0.03 across all $|\eta|$ regions indicate that pile-up effects have been effectively mitigated.

Next, an absolute jet energy scale (JES) calibration is applied to align the reconstructed jet energy and direction with those of truth jets from MC simulations. This correction accounts for non-compensating calorimeter response, energy losses in passive material, and out-of-cone effects. It works by evaluating the mean jet response in a 2D-binning of η and E_{true} and fitting the jet response as a function of E_{true} with a polynomial fit and a p-spline. Then a numerical inversion, $E_{NI} = E_{reco} \times R(\eta, E_{true})$, is evaluated and fitted to determine the calibration function $C(E_{reco})$ which should provided closure where $C(E_{reco}) = 1 = \frac{E_{calib}}{E_{true}}$. When applied to the small-R UFO jets, there is great closure in the average energy response in the central detector, $|\eta| < 2.5$, as demonstrated in Figure 4.11a.

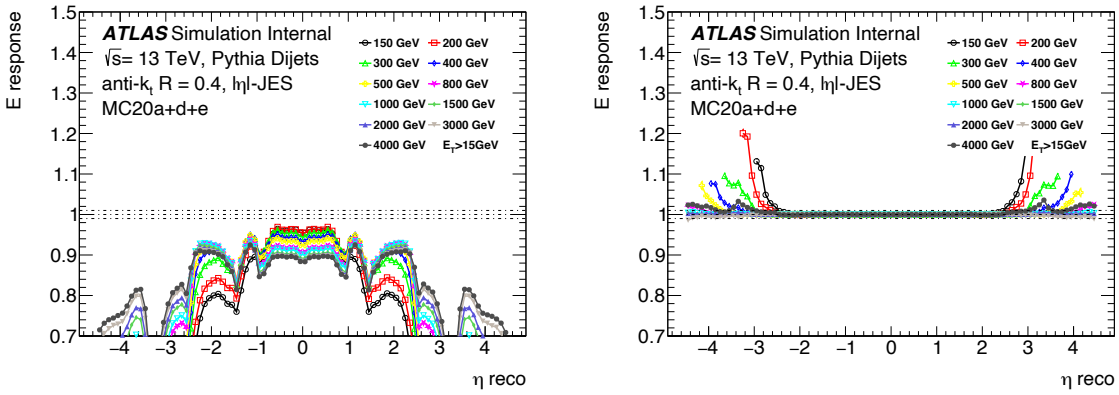


Figure 4.8: Jet energy response before and after applying the absolute JES calibration for small-R UFO jets in AF3 simulation. (a) Mean jet energy response, defined as $E_{\text{reco}}/E_{\text{truth}}$, as a function of jet energy and $|\eta|$ prior to calibration, showing significant deviations from unity across the detector. (b) Response after applying the JES correction function $C(E_{\text{reco}})$ obtained from numerical inversion and polynomial fitting. Closure is achieved in the central region ($|\eta| < 2.5$), indicating that the calibrated jet energies now accurately match the corresponding truth-level energies.

Additionally, either a global sequential calibration (GSC) or a newer global neural network calibration (GNNC) is performed to improve jet energy resolution and reduce systematic dependencies. The GSC applies multiplicative corrections based on jet observables such as the fraction of energy deposited in different calorimeter layers, the number of associated charged-particle tracks, and information from muon chambers. On the other hand, the GNNC uses a deep neural network with these same observables as inputs to leverage all the information. In either case, these observables help address variations due to jet flavor, energy leakage, and shower development within the calorimeter [94].

For the small-R UFO jets in AF3, the newer GNNC is used and greatly improves both the p_T response and p_T resolution in Figure 4.9. As an example, the Gaussian distributions of p_T response are shown in Figure 4.10 to further illustrate the improvement.

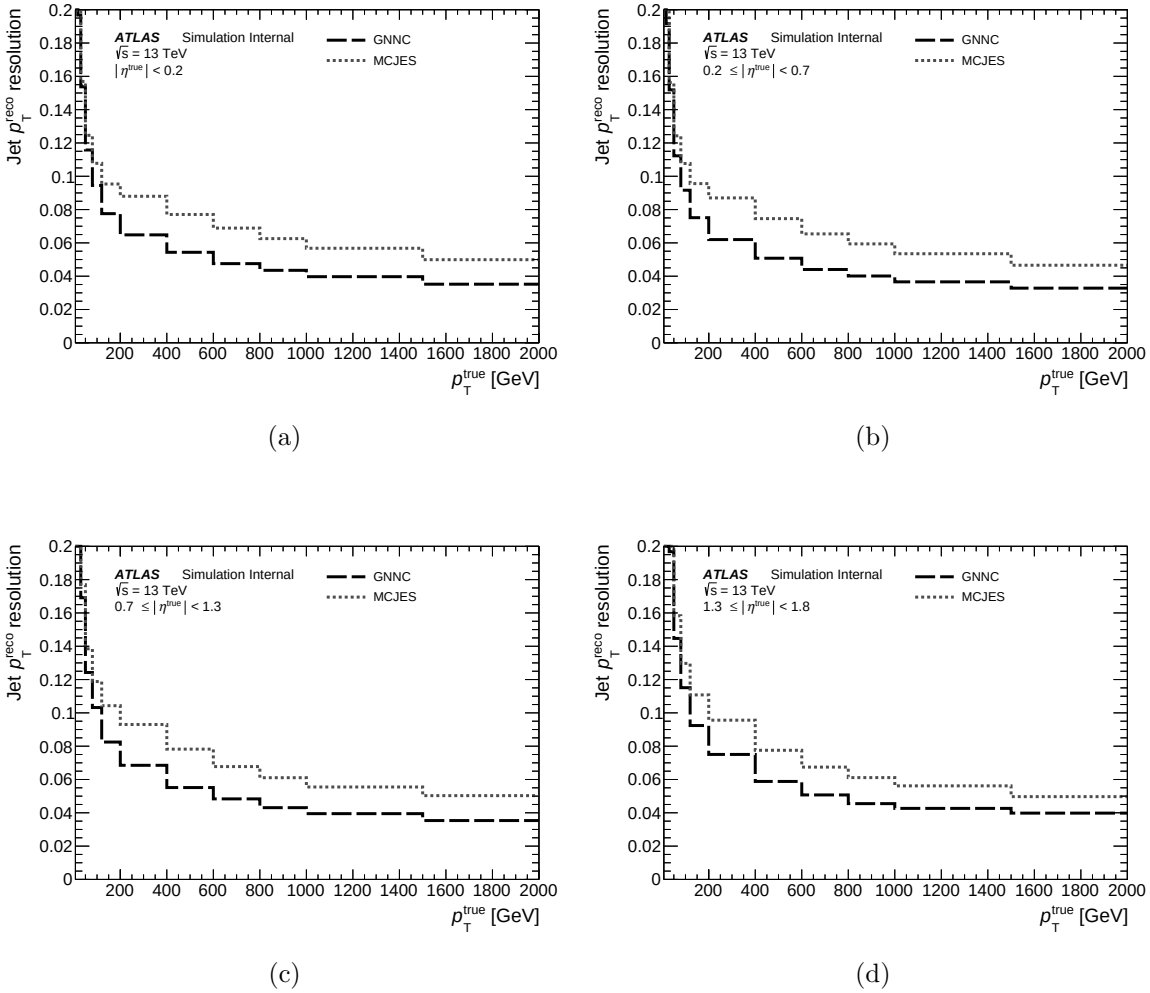


Figure 4.9: The jet p_T resolution for AF3 small-R UFO jets with the application of the global neural network calibration (GNNC) after the previous MC-JES calibration for the Pythia dijets in MC20 as shown for the η ranges of 0.0-0.2 in 4.9a, 0.2-0.7 in 4.9b, 0.7-1.3 in 4.9c, and 1.3-1.8 in 4.9d. All areas of the phase space see great improvement in p_T response and resolution.

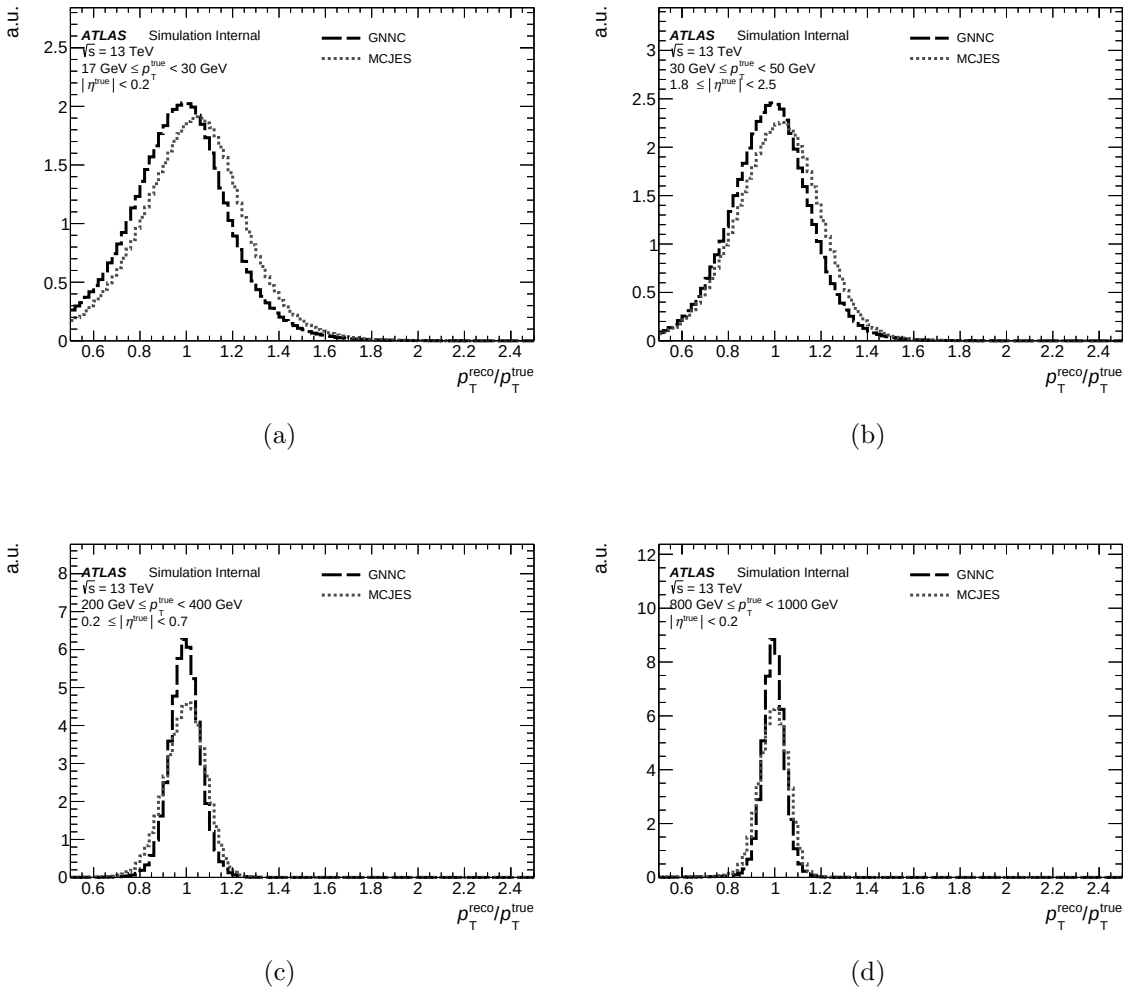


Figure 4.10: The jet p_T response distribution for AF3 small-R UFO jets with the application of the global neural network calibration (GNNC) after the previous MC-JES calibration for the Pythia dijets in MC20 as shown for the jets with 17 GeV $< p_T <$ 30 GeV and $|\eta| < 0.2$ in 4.10a, 30 GeV $< p_T <$ 50 GeV and $1.8 < |\eta| < 2.5$ in 4.10b, 300 GeV $< p_T <$ 400 GeV and $0.2 < |\eta| < 0.7$ in 4.10c, and 800 GeV $< p_T <$ 1000 GeV and $|\eta| < 0.2$ in 4.10d. All areas of the phase space see great improvement in p_T response and resolution.

Following simulation-based calibrations, in situ corrections are applied to account for residual differences between data and MC simulations. These corrections are derived us-

ing well-measured reference objects, ensuring that jets in real collision data have the same calibrated energy as those in simulations. Several balancing techniques are used, including dijet balance, where the transverse momentum asymmetry of back-to-back jets is measured to correct response variations across detector regions [95, 96]. The Z/γ+jet method calibrates jets by comparing their momentum balance against well-reconstructed photons or Z-bosons, while the multijet balance (MJB) technique uses a system of low-pt jets to calibrate a high-pt jet. Additionally, an eta intercalibration aligns the jet energy scale across different detector regions, ensuring uniform response across the calorimeter. As these calibrations are complicated, there is a dedicated subgroup for the in-situ calibrations and these are still in progress for the small-R UFO jets.

This comprehensive calibration strategy is essential for maintaining high precision in jet energy measurements, reducing systematic uncertainties, and ensuring consistency between simulation and experimental data. By refining the jet energy scale at multiple levels, ATLAS ensures that reconstructed jets accurately reflect the underlying physics, which is crucial for both Standard Model precision measurements and searches for new physics.

4.2.2 large-R Jet Calibrations

While my work focused on dedicated calibrations for small-R UFO jets in AF3 samples, the general calibration strategy for large-R jets in ATLAS follows a similar sequence of simulation-based and in situ corrections. The main differences are that large-R jets undergo an initial grooming step (such as trimming) to reduce pile-up and soft radiation and require additional calibration for the jet mass as well as the energy. In situ corrections for large-R jets also target both the energy and mass scales, often using boosted W, Z, or top quarks as reference objects. In some analyses, calibrations of jet substructure variables are performed as well. Despite these differences, the overarching goal remains the same: to ensure that reconstructed jets accurately reflect the underlying particle-level jets in both data and simulation.

4.2.3 AF3 Systematic Uncertainties for FullSim Calibrations

Dedicated calibrations for AF3 samples always provide the best results, but unfortunately, there isn't always time or person-power to derive such calibrations. When AF3 calibrations are needed quickly with less precision, the next easiest option is to apply the FullSim calibrations and add additional systematics with jet energy scale (JES) and jet energy response (JER) uncertainties for the small-R jets or even additional jet mass scale (JMS) and jet mass response (JMS) uncertainties for the large-R jets.

The JES and JMS uncertainties are calculated as the difference between the average p_T or mass response (R) or resolution (σ_R where $\sigma_R = \frac{\sigma}{R}$) between AF3 and FullSim,

$$\frac{R_{\text{FullSim}} - R_{\text{AF3}}}{R_{\text{FullSim}}}, \tag{4.2}$$

while the JMR uncertainty is

$$\frac{\sigma_{R,\text{FullSim}} - \sigma_{R,\text{AF3}}}{\sigma_{R,\text{FullSim}}}, \tag{4.3}$$

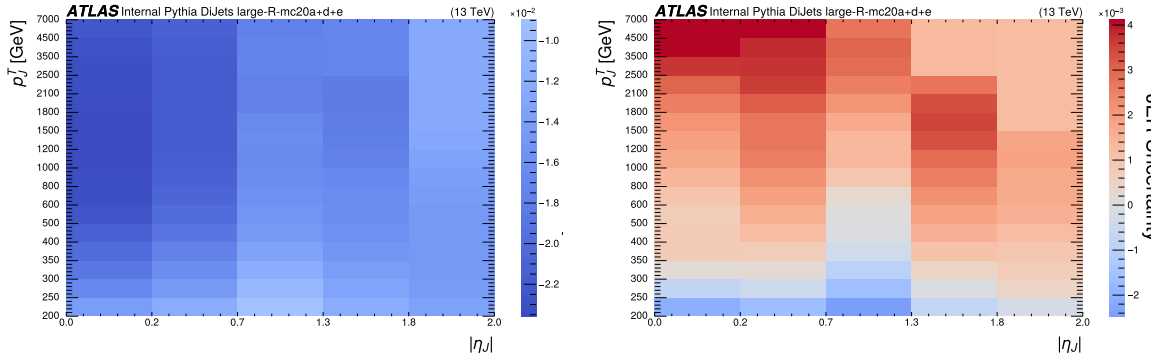
and the JER uncertainty is

$$\sigma_{R,\text{FullSim}} - \sigma_{R,\text{AF3}}. \tag{4.4}$$

The Jet Energy Resolution (JER) and Jet Mass Resolution (JMR) uncertainties are derived using different propagation tools, resulting in distinct treatment for each. The JER uncertainty is modeled by applying a Gaussian smearing function to the jet energy, while the JMR uncertainty is calculated using a dedicated jet mass smearing tool that employs the Forward-Folding technique [97]. Due to differences in implementation and formatting between these tools, they require separate input formats and procedures for uncertainty propagation.

I performed the derivation of jet energy scale (JES) and jet energy resolution (JER) uncertainties for small-R particle-flow (PFlow) jets using AF3-simulated dijet samples from the mc20 and mc23 Monte Carlo campaigns. For large-R PFlow jets in AF3 (mc20), I also evaluated the JES, JER, jet mass scale (JMS), and jet mass resolution (JMR) uncertainties.

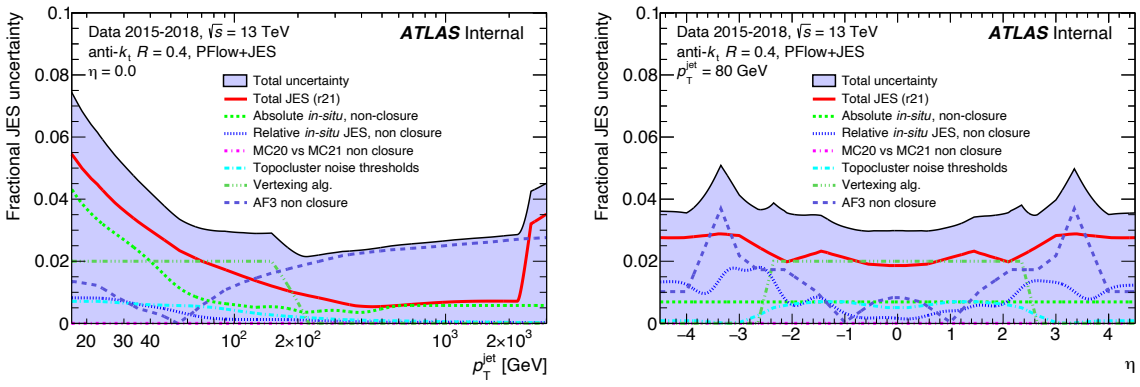
These systematic uncertainty evaluations are essential for accurately modeling jet-related observables under the corresponding data-taking conditions. The AF3 non-closure uncertainties for the aforementioned large-R jets are shown in Figures 4.11 and 4.13, while com-



(a) (b)

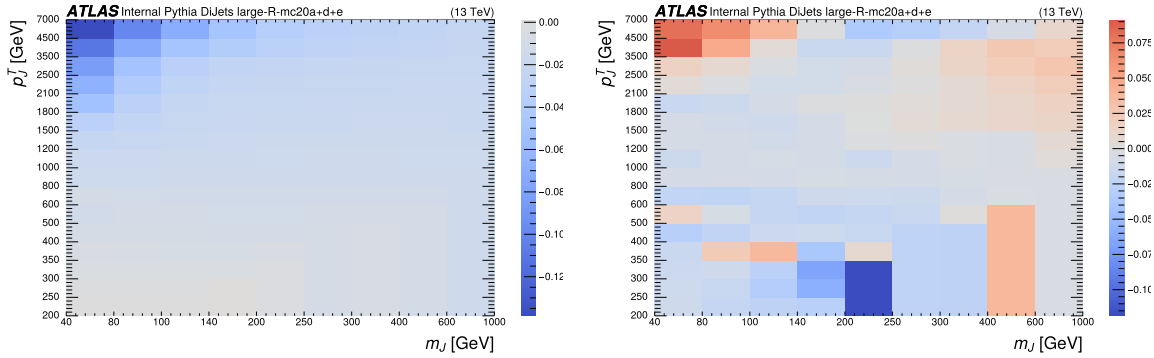
Figure 4.11: Jet energy uncertainties as a function of jet η and p_T, derived by comparing FullSim and AF3 Pythia DiJet samples for large-R PFlow jets in the mc20 campaign. (a) Jet Energy Scale (JES) uncertainty, calculated according to Equation 4.2, indicates a general overestimation of the response in AF3 relative to FullSim. (b) Jet Energy Resolution (JER) uncertainty, defined in Equation 4.4, tends to increase with higher p_T at fixed η. Irregular features at high p_T and in forward regions are attributed to limited MC statistics.

parisons of JES uncertainties with other methods are presented in Figure 4.12. In several kinematic regions, these AF3-specific uncertainties dominate, reinforcing the need for dedicated calibrations such as those introduced in Section 4.2.1.



(a) (b)

Figure 4.12: AF3 JES non-closure uncertainties for PFlow jets in the central region ($0.0 < |\eta| < 0.2$). (a) JES uncertainties dominate for high-p_T jets above 200 GeV. (b) For jets with $80 < p_T < 100$ GeV, the uncertainties are particularly large in the forward region, highlighting AF3's degraded performance near subdetector boundaries, as discussed in Section 4.2.



(a) (b)

Figure 4.13: Jet mass uncertainties as a function of jet mass and p_T for large-R PFlow jets in AF3 (mc20). (a) Jet Mass Scale (JMS) uncertainty, from Equation 4.2, remains generally small except at low jet mass and high p_T. (b) Jet Mass Resolution (JMR) uncertainty, from Equation 4.3, is largest at low p_T and high jet mass. Irregular features in the distributions reflect limited MC statistics in sparsely populated regions of phase space.

Chapter 5

HIGGS BOSON PAIR PRODUCTION

5.1 Theoretical Motivations

Higgs boson pair production is a crucial process for probing the scalar sector and testing the Standard Model's predictions. This process investigates events where two Higgs bosons are produced simultaneously in the proton-proton collisions, providing a direct probe of the Higgs self-coupling which appears in the cubic term of the Higgs potential from Equation 2.8. This self-coupling plays a central role in shaping the Higgs potential, as shown in Figure 5.1, and it directly relates to the stability of our universe and the mechanism for electroweak symmetry breaking.

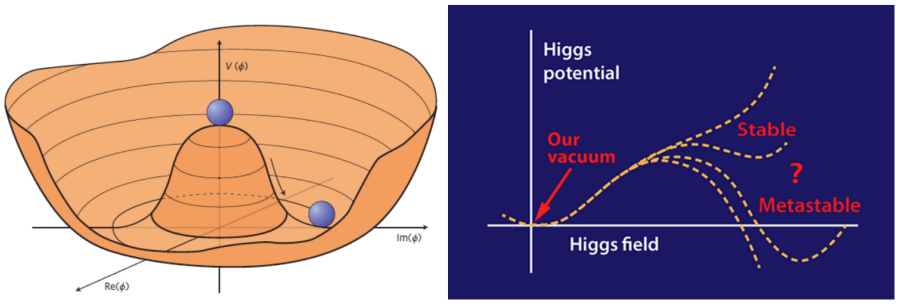


Figure 5.1: The Higgs potential, featured on the left, for $\mu^2 < 0$. Choosing any of the points at the bottom of the potential spontaneously breaks the rotational U(1) symmetry. Quantum corrections, featured on the right, can change the shape of the Higgs potential indicating whether the universe is stable or metastable. Here, the minimum of our vacuum is taken at $|\phi| = \frac{v}{\sqrt{2}}$. [98]

A central concept in this framework is the vacuum expectation value (VEV) of the Higgs field, denoted by v. The VEV represents the average value that the Higgs field acquires in

its lowest energy (vacuum) state. Unlike most fields in the Standard Model, the Higgs field has a nonzero VEV, filling all of space even in the absence of particles. This nonzero VEV is what spontaneously breaks the electroweak symmetry through the Brout–Englert–Higgs mechanism, giving rise to the masses of elementary particles such as the W and Z bosons, while the photon remains massless. Thus, the VEV is not merely a parameter, but the cornerstone of the entire mass-generation mechanism in the Standard Model.

The value of the Higgs VEV is not a theoretical prediction, but rather is determined experimentally from the precisely measured Fermi constant (G_F), which governs the strength of weak interactions in processes such as muon decay. This relationship is given by

$$v = (\sqrt{2}G_F)^{-1/2} \approx 246 \text{ GeV}.$$

Unlike the Higgs VEV, the Standard Model predicts specific values for the Higgs self-couplings, which arise from the Higgs potential

$$V(H) = \mu^2 H^\dagger H + \lambda (H^\dagger H)^2,$$

where λ is the dimensionless quartic coupling. After electroweak symmetry breaking, this leads to effective Higgs boson interaction terms, most notably the trilinear and quartic self-couplings.

The trilinear Higgs self-coupling, denoted λ_{HHH}, governs the three-Higgs vertex and has units of mass:

$$\lambda_{HHH} = 3\frac{m_H^2}{v} \approx 190 \text{ GeV}, \tag{5.1}$$

where $m_H \approx 125$ GeV is the Higgs boson mass. This coupling is directly accessible via processes such as Higgs boson pair production, and any measured deviation from the Standard Model prediction would provide vital insight into physics beyond the Standard Model, potentially pointing to extended scalar sectors, composite Higgs scenarios, or modifications to the Higgs potential relevant for electroweak baryogenesis.

Similarly, the quartic Higgs self-coupling, λ_{HHHH}, governs the four-Higgs vertex and is related to the dimensionless quartic coupling as

$$\lambda_{HHHH} = 3\lambda = \frac{3m_H^2}{2v^2} \approx 0.78, \tag{5.2}$$

and is therefore dimensionless.

The precise measurement of these couplings is crucial for understanding the shape of the Higgs potential, the mechanism of electroweak symmetry breaking, and the overall stability of the vacuum of our universe.

5.2 Production Modes

5.2.1 Non-resonant Searches

Higgs boson pair production can occur through both resonant and non-resonant processes. Non-resonant Higgs boson pair production refers to the direct production of two Higgs bosons from contributions from an exchange of virtual or off-shell particles [99]. This process is highly sensitive to the Higgs boson's self-coupling and provides a fundamental test of the Standard Model's predictions where any significant deviation could suggest new physics.

There are two main processes for non-resonant Higgs boson pair production. The first and dominant process is through gluon-gluon fusion (ggF), where two gluons combine with enough center-of-mass energy to create a Higgs boson pair final state via a top-loop-mediated process [100]. This occurs through two competing Feynman diagrams shown in Figure 5.2: the triangle diagram, which involves the Higgs self-coupling, and the box diagram, which does not. Notably, these two diagrams interfere destructively, reducing the overall ggF production cross section, as shown in Figure 5.3. Since the Higgs self-coupling is exactly predicted by the Standard Model, we often refer to κ_λ, the measured self-coupling modifier normalized to the SM expectation, where:

$$\kappa_\lambda = \frac{\lambda}{\lambda_{HHH}}. \tag{5.3}$$

As such, the SM predicts that $\kappa_\lambda = 1$.

The second and subdominant production process occurs through vector boson fusion (VBF), where two weak bosons scatter off each other to produce a di-Higgs final state. Although its cross section is significantly smaller than ggF, VBF provides a complementary probe of Higgs interactions due to its unique kinematic features and sensitivity to new physics scenarios that could modify Higgs couplings.

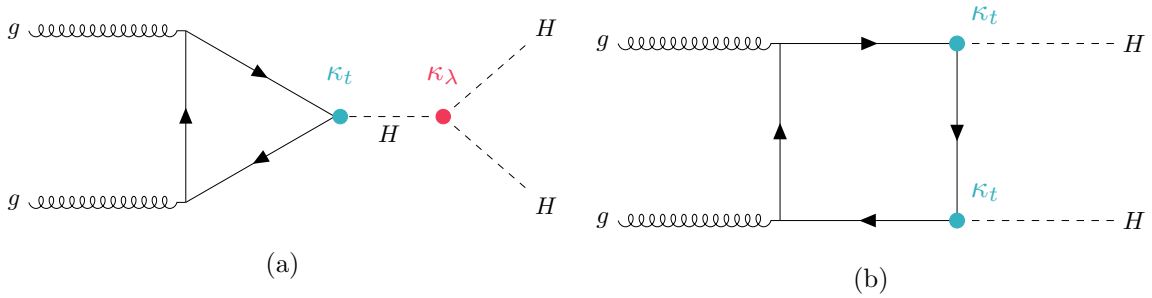


Figure 5.2: The Feynman diagrams for the non-resonant Higgs boson pair production through gluon-gluon fusion (ggF) where the triangle diagram in 5.2a works through a top loop process including the Higgs self-coupling while the box diagram in 5.2b does not include the self-coupling. These diagrams destructively interfere [101, 102].

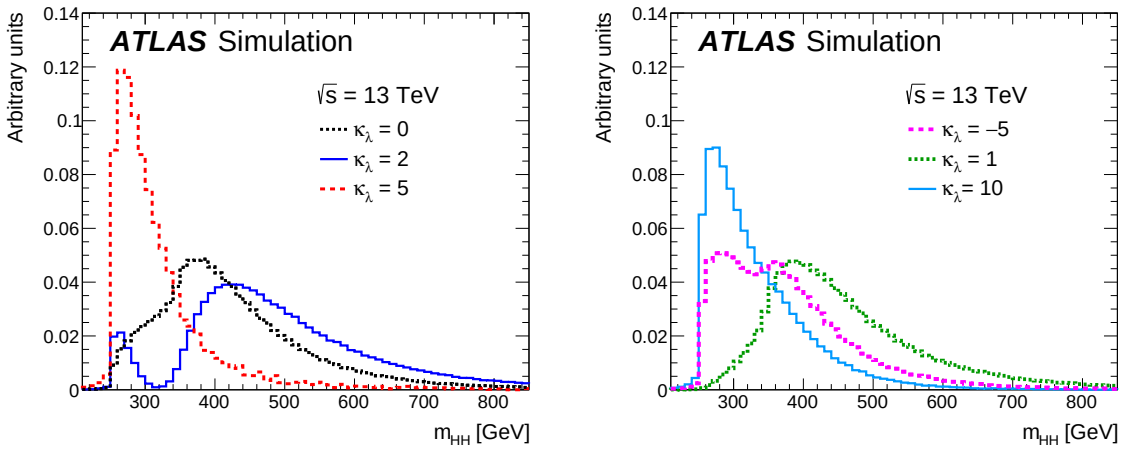


Figure 5.3: The destructive interference between the triangle and box diagrams of the ggF Higgs boson pair production showing how the distribution for m_{HH}, the invariant mass of the di-Higgs system, changes depending on the Higgs self-coupling value. The green line in 5.3b where $\kappa_\lambda = 1$ indicates the distribution for the Standard Model prediction [103].

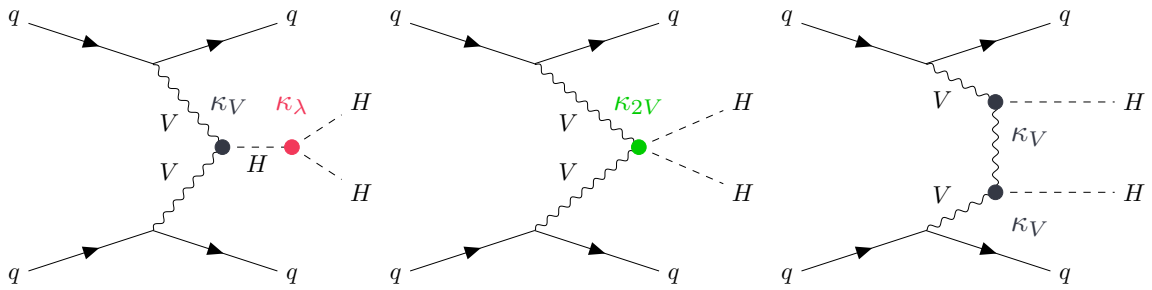


Figure 5.4: Feynman diagrams illustrating the three main contributions to non-resonant Higgs pair production via vector boson fusion (VBF). The process proceeds through: 5.4a a diagram involving the Higgs self-coupling, proportional to κ_λ; 5.4b a diagram featuring a direct coupling of two Higgs bosons to vector bosons, governed by κ_{2V}; 5.4c a diagram with a single-Higgs coupling to vector bosons, proportional to κ_V [102].

Both ggF and VBF di-Higgs production are sensitive to modifications in the Higgs boson's couplings. In addition to the Higgs self-coupling modifier κ_λ, which quantifies deviations from the Standard Model prediction, other important coupling modifiers include κ_t (top quark Yukawa coupling), κ_V (Higgs–vector boson coupling), and κ_{2V} (quartic Higgs–vector boson interaction). These parameters are defined as

$$\kappa_\lambda = \frac{\lambda}{\lambda_{HHH}}, \quad \kappa_t = \frac{y_t}{y_t^{\mathrm{SM}}}, \quad \kappa_V = \frac{g_{HVV}}{g_{HVV}^{\mathrm{SM}}}, \quad \kappa_{2V} = \frac{g_{HHVV}}{g_{HHVV}^{\mathrm{SM}}}, \tag{5.4}$$

where y_t is the top Yukawa coupling, g_{HVV} is the Higgs boson–vector boson coupling, and g_{HHVV} is the quartic coupling between two Higgs bosons and two vector bosons. Di-Higgs events directly probe these couplings, while κ_t and κ_V can also be constrained through measurements of single Higgs boson production and decay. By combining information from both production modes, one can perform stringent tests of the Standard Model and enhance sensitivity to new physics.

The Standard Model predicts cross sections for Higgs boson pair production via ggF and VBF of approximately $30.8^{+2.0}_{-7.1}$ fb ($34.1^{+2.2}_{-7.9}$ fb) and $1.69^{+0.05}_{-0.05}$ fb ($1.87^{+0.05}_{-0.05}$ fb), respectively, for pp collisions at $\sqrt{s} = 13$ TeV ($\sqrt{s} = 13.6$ TeV) [104]. By comparison, single Higgs boson

production cross sections through ggF and VBF are 48.57 pb and 3.78 pb at $\sqrt{s} = 13$ TeV, more than a thousand times larger than their di-Higgs counterparts. This stark contrast, summarized in Table 5.1, highlights the exceptional rarity of Higgs boson pair production. With the full ATLAS Run 2 dataset (139 fb^{-1}), one would expect only about 4,550 di-Higgs events in total. Consequently, precise measurements of the di-Higgs production cross section and the Higgs self-coupling will require the increased statistics of the High-Luminosity LHC, unless we discover that new physics significantly enhances the production rate.

Production Mode	$\sigma(pp \to H + X)$ [pb]	$\sigma(pp \to HH + X)$ [fb]
ggF	48.57	31.05
VBF	3.782	1.73

Table 5.1: Production mode and cross section of SM single Higgs and Higgs pair production for pp collisions at $\sqrt{s} = 13$ TeV.

Following production, each Higgs boson decays according to its Standard Model branching fractions as illustrated in Figure 5.5. The $b\bar{b}b\bar{b}$ final state, arising from the dominant $H \to b\bar{b}$ decay ($\sim 58\%$ branching ratio), constitutes nearly 34% of di-Higgs events, making it a statistically significant channel for studying Higgs boson pair production. This thesis primarily focuses on this channel despite the significant challenge of reconstructing and distinguishing Higgs bosons from quantum chromodynamics (QCD) multijet backgrounds. To enhance selection efficiency and improve signal-to-background discrimination, advanced machine learning techniques, such as boosted decision trees, neural networks, and transformers, are frequently employed as a multi-variate analysis (MVA) discriminant.

5.2.2 Resonant Searches

In contrast to non-resonant di-Higgs production, resonant di-Higgs production occurs when two Higgs bosons are produced via an intermediate particle which could be a heavy scalar or other new particle predicted by Beyond-the-Standard Model (BSM) theories. Such a process

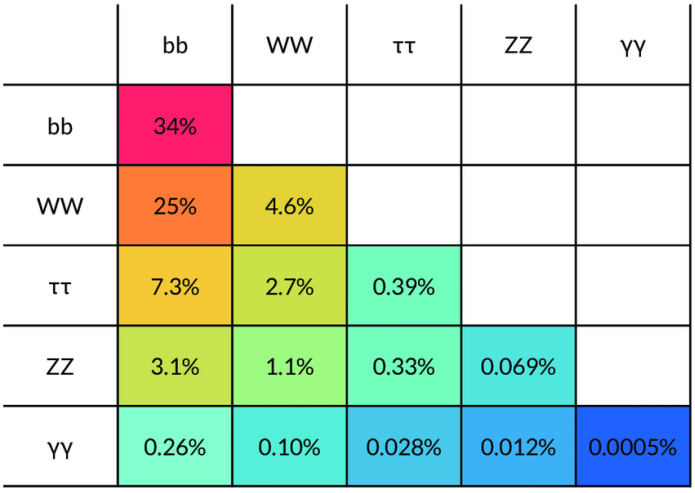


Figure 5.5: The branching ratios of the possible decay channels for Higgs pair production for pp collisions at $\sqrt{s} = 13$ TeV. The $b\bar{b}b\bar{b}$ decay channel has the largest branching ratio at 34% [105].

would lead to a characteristic resonant peak in the di-Higgs production rate with the peak mass corresponding to the mass of the intermediate particle. Detecting such a resonance could offer crucial insights into the existence of new particles or interactions beyond the Standard Model.

Symmetric Di-Higgs Production via Resonance

Symmetric di-Higgs production refers to events in which both produced Higgs bosons are Standard Model-like, each with a mass of approximately $m_H = 125$ GeV. Several Beyond Standard Model (BSM) theories predict the existence of a new heavy resonance that can decay into two such Higgs bosons. One common framework is the Two-Higgs-Doublet Model (2HDM), which extends the Standard Model Higgs sector by introducing a second scalar doublet [106, 107]. This leads to a richer Higgs spectrum containing five physical Higgs bosons: two CP-even neutral scalars (h, H), one CP-odd pseudoscalar (A), and a pair of charged Higgs bosons ($H^{\pm}$). The additional scalar degrees of freedom allow for the possibility of resonant di-Higgs production through the decay of a heavy CP-even Higgs

boson ($H \to hh$), where h is identified as the observed 125 GeV Higgs.

A particularly well-motivated and constrained realization of a 2HDM arises in the Minimal Supersymmetric Standard Model (MSSM), which imposes supersymmetry on the particle spectrum and structure of interactions. In the MSSM, the Higgs sector must include two Higgs doublets to provide masses to both up- and down-type fermions and to cancel anomalies in the supersymmetric theory [108]. The resulting Higgs bosons in the MSSM form a Type-II 2HDM with specific relationships among the parameters, tightly constraining the phenomenology. In this context, the heavy CP-even scalar (H) can decay into two SM-like Higgs bosons (h), producing a resonant di-Higgs signature.

Other theories, such as models with extra spatial dimensions, also predict resonant di-Higgs production. For example, spin-2 Kaluza-Klein gravitons may couple to Standard Model particles and decay into Higgs boson pairs. However, no significant excess in di-Higgs production has been observed in either scenario in the latest ATLAS Run 2 analyses, as discussed further in Section 5.3.2 [105, 109].

Resonance Production for Asymmetric Di-Higgs

In extended scalar sectors, there are also possibilities for asymmetric di-Higgs production where a heavy resonant scalar boson can decay into a SM-like Higgs boson and another Higgs-like scalar boson having its own unique mass. Several theories posit extensions to the scalar sector of the Standard Model that would help explain these phenomena. These extensions often include additional scalar bosons but with different masses or symmetries [110]. Such BSM theories could further explain electro-weak phase transition and baryogenesis [111]. These BSM theories provide strong theoretical motivation for analyses searching for large scalar resonances decaying into a Standard Model Higgs boson and a lighter or heavier scalar boson. In this case, the scalar resonance X would decay into a lighter scalar S and an SM Higgs H, so this asymmetric di-Higgs process occurs in the decay $X \to SH$.

The Higgs field is the sole member of the SM scalar sector and is formed from a complex doublet under the weak isospin SU(2) symmetry [112]. To extend this scalar sector, singlet and doublet extensions are common, both of which can be explored at the LHC through

precise measurements of the SM Higgs boson signal rates and scalar mixing structures. These extensions could also open the possibility of discovering new decay modes of the SM Higgs boson, or they could provide direct searches for new scalar particles that would reveal the existence of other scalar bosons.

While the Two-Higgs-Doublet Models (2HDMs) introduce an additional SU(2) doublet, further extending them by a scalar singlet leads to 2HDM+S models. These models allow for a rich scalar spectrum, including multiple CP-even and CP-odd neutral scalars. If the singlet does not acquire a vacuum expectation value (VEV), it can also act as a dark matter candidate [113]. In scenarios where the singlet mixes with the doublets, asymmetric di-Higgs decays such as $X \to SH$ can occur, providing characteristic signals at the LHC.

Another important model is the Next-to-Minimal Supersymmetric Standard Model (NMSSM), which introduces an additional complex scalar singlet along with the additional SU(2) complex doublet from the MSSM, thereby adding five neutral and two charged Higgs bosons to the SM [111]. In the NMSSM, both neutral and charged Higgs bosons can decay into one another or into other Higgs states, provided these processes are kinematically allowed. This leads to a rich phenomenology with various possible cascade decays among the extended scalar sector. Though the parameter space of the MSSM has largely been excluded, the parameter space of the NMSSM is less constrained [110].

Another model of interest is the Two-Real-Scalar-Singlet Model (TRSM), which adds two real scalar singlets to the SM [114]. This extension reduces the number of scalar potential parameters and predicts three CP-even neutral scalar bosons, one of which is identified as the SM Higgs boson. The singlets interact with the SM Higgs boson through the scalar potential and can mix with the SM Higgs boson if they acquire a non-zero VEV. This interaction gives rise to new scalar fields that could potentially be discovered at the LHC, especially through their asymmetric decays, where one of the singlets decays into the SM Higgs boson and another scalar field. Such decays could provide valuable insights into the structure of the scalar sector and open up new avenues for searching for BSM physics.

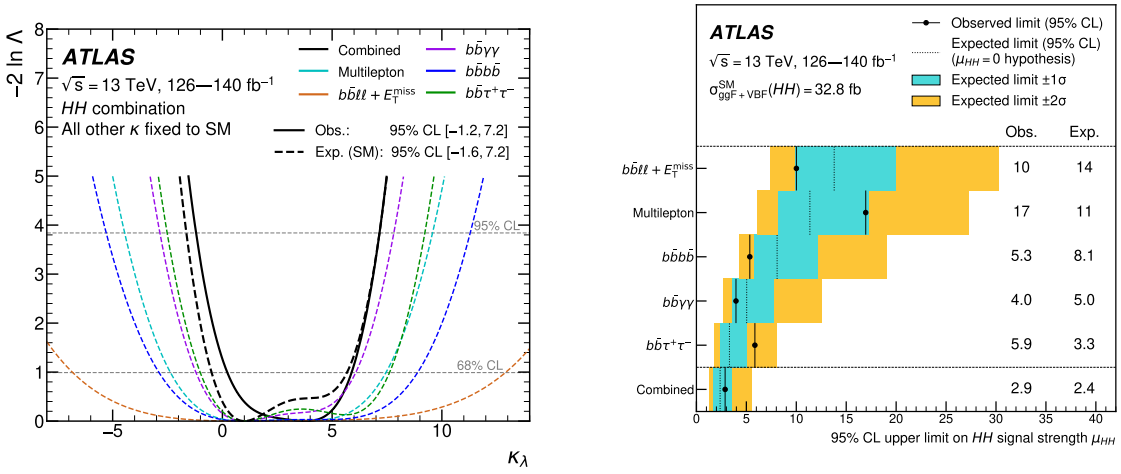
Table 5.2: Summary of selected BSM theories extending the scalar sector of the Standard Model.

Model	**Extension to SM**	**Dark Matter Candidate**	**Scalar Sector**
Singlet Extension	Adds one (real or complex) scalar singlet	Possible	One additional neutral scalar boson
2HDM (Two Higgs Doublet Model)	Adds a second SU(2) Higgs doublet	Possible	2 CP-even, 1 CP-odd, and 2 charged Higgs bosons
2HDM + Singlet (2HDM+s)	2HDM extended with a scalar singlet	Possible	3 CP-even, 2 CP-odd, and 1 charged Higgs bosons
NMSSM (Next-to-Minimal Supersymmetric SM)	MSSM plus an additional complex scalar singlet	Possible	3 CP-even, 2 CP-odd, and 2 charged Higgs bosons
TRSM (Two Real Scalar Singlet Model)	Adds two real scalar singlets to the SM	Possible	3 CP-even neutral Higgs bosons

5.3 Experimental Results

5.3.1 Non-resonant Results

Both ATLAS and CMS have performed extensive searches in a variety of final states while combining the results to maximize their sensitivity. Each decay channel presents a distinct trade-off between signal branching fraction, background complexity, and mass resolution. For example, high-branching channels such as $b\bar{b}b\bar{b}$ (34%) have large statistical power but face enormous QCD backgrounds, while rare channels such as $b\bar{b}\gamma\gamma$ (0.26%) feature excellent mass resolution and background rejection, albeit with limited event yields. Intermediate channels, such as $b\bar{b}\tau^+\tau^-$ (7.3%), $b\bar{b}W^+W^-$ (25%), $b\bar{b}\ell^+\ell^- + E_T^{\text{miss}}$ (2.9%), and various multilepton final states (6.5%), further complement the overall sensitivity since each channel is subject to a different mixture of backgrounds and systematics [115, 116, 117, 118, 119].



(a) (b)

Figure 5.6: (a) Expected values (dashed lines) of the test statistic $-2\ln\Lambda$ as a function of κ_λ, shown for the $b\bar{b}\gamma\gamma$ (purple), $b\bar{b}\tau^+\tau^-$ (green), multilepton (cyan), $b\bar{b}b\bar{b}$ (blue), and $b\bar{b}\ell\ell + E_T^{\text{miss}}$ (brown) channels, as well as their combination (black). Observed values from the combined data are indicated by the solid black line [120].
(b) Observed and expected 95% CL upper limits on the signal strength for inclusive ggF HH and VBF HH production from the $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}b\bar{b}$, multilepton, and $b\bar{b}\ell\ell + E_T^{\text{miss}}$ decay channels, and their combination. The expected limit and its $\pm1\sigma$ and $\pm2\sigma$ bands assume no HH production and profile all nuisance parameters to the observed data [120].

ATLAS has combined the following channels for the Run 2 results: $HH \to b\bar{b}b\bar{b}$ (discussed in Chapter 8), $HH \to b\bar{b}\tau^+\tau^-$, $HH \to b\bar{b}\gamma\gamma$, $HH \to b\bar{b}\,\ell^+\ell^- + E_T^{\text{miss}}$, and $HH \to$ multi-lepton [120]. The observed (expected) 95% CL upper limit on the non-resonant HH signal strength is

$$\mu_{HH}^{\text{obs}} = 2.9\ (2.4) \times \text{SM}\,,$$

with an observed (expected) significance of $0.4\,\sigma$ $(1.0\,\sigma)$. Compared to the previous publication [101], this represents a 17% overall improvement in sensitivity, with 13% arising from refinements in the $b\bar{b}\,b\bar{b}$, $b\bar{b}\,\tau^+\tau^-$ and $b\bar{b}\,\gamma\gamma$ analyses, and a further 4% from the inclusion

of the multi-lepton and $b\bar{b}\,\ell^+\ell^- + E_T^{\text{miss}}$ channels. The dominant systematic uncertainties remain the theoretical cross section predictions and background modeling.

CMS has also published results from Run 2 with the non-resonant HH production in the following channels: $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}W^+W^-$, multilepton, $\gamma\gamma\tau^+\tau^-$, $b\bar{b}ZZ, 4l$, and $WW\gamma\gamma$ with their combination as well [121, 122, 123, 124, 125, 126, 127, 128, 129]. The combined CMS result gives an observed (expected) 95% CL upper limit on the HH signal strength that is similar to ATLAS, with small differences in sensitivity reflecting the different detector designs and analysis strategies. The order of channel sensitivities also varies between the experiments.

These combined results can be interpreted as constraints on the Higgs trilinear self-coupling modifier, κ_λ. For the combination of Run 2 analyses, ATLAS obtains

$$\kappa_\lambda \in [-1.2,\ 7.2] \quad @\ 95\%\ \text{CL} \quad (\text{expected } [-1.6,\ 7.2])\,,$$

while CMS reports

$$\kappa_\lambda \in [-1.39,\ 7.02] \quad @\ 95\%\ \text{CL} \quad (\text{expected }) [-1.02,\ 7.19]\,.$$

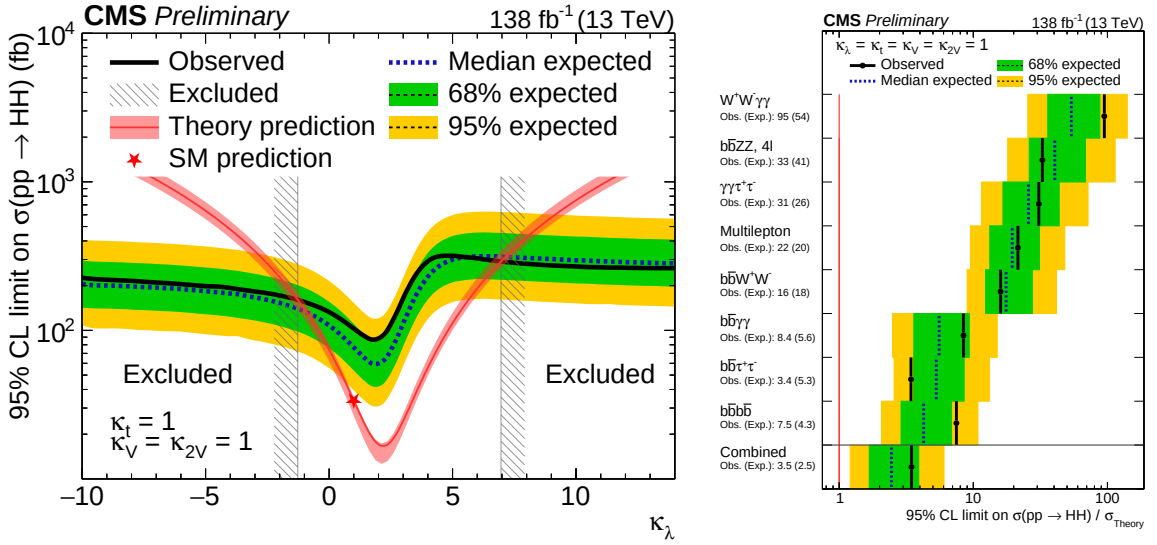
The results from ATLAS and CMS are thus highly consistent, both excluding large deviations from the Standard Model prediction of $\kappa_\lambda = 1$. The remaining allowed regions reflect the limited sensitivity of Run 2 and the impact of systematic uncertainties, particularly those associated with background modeling in the $b\bar{b}b\bar{b}$ final state.

5.3.2 Resonant Results

Symmetric Resonant Decays

ATLAS and CMS have also both conducted extensive searches for resonant Higgs boson pair production in various final states. These analyses target new heavy resonances decaying into pairs of Standard Model Higgs bosons, with particular focus on scenarios motivated by spin-0 heavy scalar theories discussed previously.

A recent ATLAS combination of resonant HH analyses was performed in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, and $bb\gamma\gamma$ final states under the heavy spin-0 hypothesis [109]. This combination

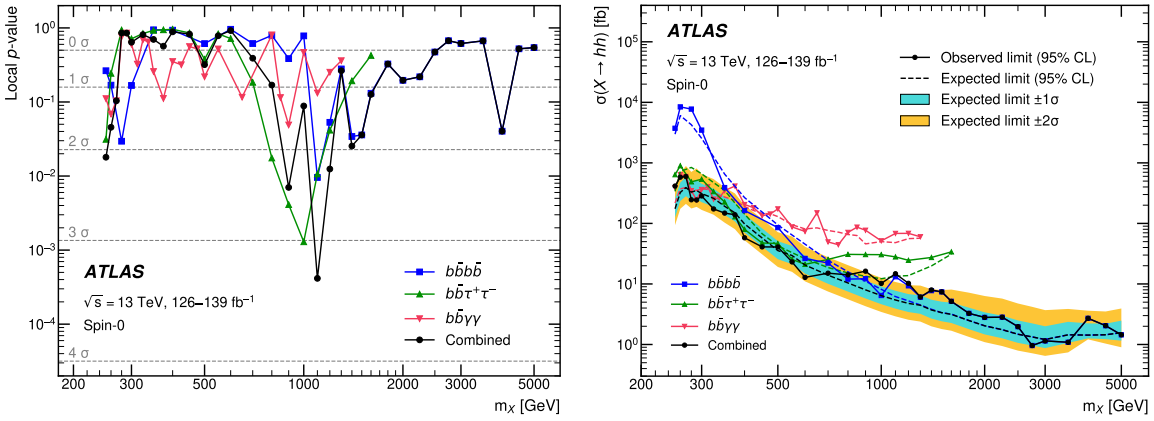


(a) (b)

Figure 5.7: (a) The 95% CL upper limits on the inclusive HH cross section as a function of κ_λ when all other couplings are set to values predicted by the SM. The inner (green) band and the outer (yellow) band indicate the 68% and 95% CL intervals [129].
(b) Observed and expected 95% CL upper limits on the signal strength for inclusive ggF HH and VBF HH production from the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}W^+W^-$, multilepton, $\gamma\gamma\tau^+\tau^-$, $b\bar{b}ZZ, 4l$, and $W^+W^-\gamma\gamma$ decay channels, and their combination. The expected limit and its $\pm1\sigma$ and $\pm2\sigma$ bands assume no HH production and profile all nuisance parameters to the observed data [129].

provides among the highest sensitivity to gluon–gluon fusion (ggF) production modes. Each decay channel contributes in different regions of resonance mass m_X: the $b\bar{b}\gamma\gamma$ channel offers strong sensitivity at low masses due to excellent diphoton mass resolution and efficient triggering and reconstruction, while the $b\bar{b}b\bar{b}$ channel dominates at high mass, particularly in the boosted regime, where the Higgs bosons are highly Lorentz-boosted and reconstructed as merged jets. The results of the combination are shown in Figure 5.8, which includes (a) the local p-value scan across m_X, where the p-value quantifies the probability of observing a

deviation at least as extreme as the one seen in data under the background-only hypothesis, and (b) the observed and expected upper limits at 95% confidence level on the cross section of resonant Higgs boson pair production.



(a) (b)

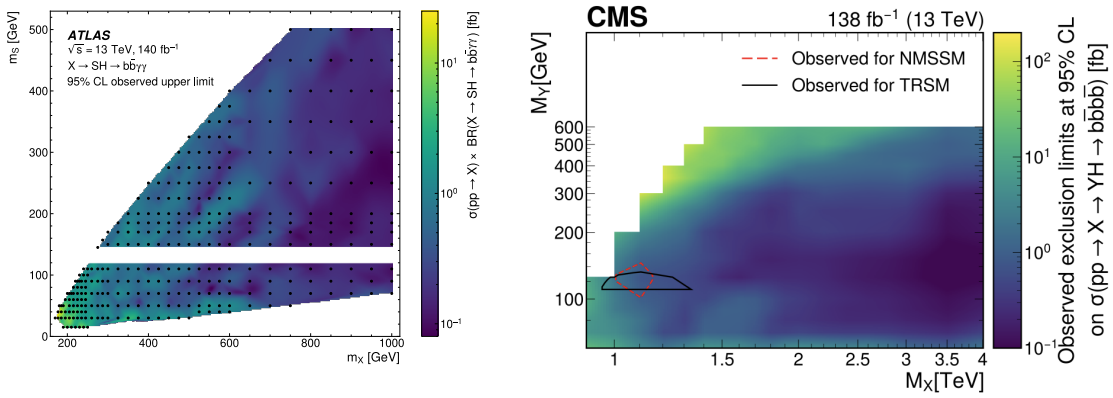
Figure 5.8: (a) Local p-value and (b) observed and expected upper limits at the 95% CL on the resonant Higgs boson pair production cross section as a function of the resonance mass m_X [109].

Asymmetric Resonant Decays

ATLAS has published searches for asymmetric resonant di-Higgs decays in the $X \to SH \to VV\tau^+\tau^-$, $X \to SH \to VV\gamma\gamma$, and $X \to SH \to b\bar{b}\gamma\gamma$ channels, though no combination has yet been performed [130, 131, 132]. In all channels, no significant excess above Standard Model expectations is observed. The analyses set upper limits at 95% CL on the production cross section times branching ratio, while excluding values down to a few tenths of a femtobarn over a broad resonance mass range.

CMS has also published similar searches for heavy and lighter scalar resonances in the $b\bar{b}b\bar{b}$, $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$, and $\gamma\gamma\tau^+\tau^-$ through ggF production [133, 134, 135]. They typically refer to the S scalar boson as Y in their literature. Results for the 95% CL observed upper

limits are shown in Figure 5.9 for the ATLAS $X \to SH \to b\bar{b}\gamma\gamma$ Run 2 analysis and the CMS boosted $X \to YH \to b\bar{b}b\bar{b}$ Run 2 analysis.



(a) (b)

Figure 5.9: (a) Observed 95% CL upper limits on the signal cross section times branching fraction for the $X \to SH \to b\bar{b}\gamma\gamma$ signal in the ATLAS Run 2 analysis, in the (mX, mS) plane with the points show where the limits were evaluated [132]. (b) Observed 95% CL upper limits on the signal cross section times branching fraction for the boosted $X \to YH \to b\bar{b}b\bar{b}$ signal in the CMS Run 2 analysis, in the (mX, mS) plane with the boundaries of the excluded regions of the NMSSM and TRSM included [133].

While most asymmetric resonance searches have not observed significant deviations from Standard Model expectations, one notable exception is the CMS resolved $X \to YH \to b\bar{b}b\bar{b}$ analysis using the Run 2 data, which reported a local (global) significance of 4.1σ (2.8σ) at $(m_X, m_Y) = (700, 400)$ GeV [136]. The broader program of resonant Higgs boson pair production searches, including asymmetric decays of the form $X \to SH$, continues to expand across multiple final states and kinematic regimes, combining model independence with sensitivity to a range of well-motivated BSM scenarios. The ATLAS resolved-channel search for $X \to SH \to b\bar{b}b\bar{b}$, detailed in Chapter 7, is a key component of this effort, targeting a region of phase space where future discoveries may yet arise.

Chapter 6

COMMON ASPECTS TO $b\bar{b}b\bar{b}$ ANALYSES

This chapter outlines the common methodological components underlying the analyses presented in the following chapters. These analyses focus on searches for both resonant ($X \to SH \to b\bar{b}b\bar{b}$) and non-resonant ($HH \to b\bar{b}b\bar{b}$) Higgs boson pair production through ggF using the ATLAS Run 2 dataset, as well as a combined Run 2 and partial Run 3 dataset for the non-resonant search. To avoid redundancy, the discussion here centers on elements shared by these analyses, including flavor tagging, $b\bar{b}b\bar{b}$ analysis topologies, background composition, and data-driven background estimation strategies. While specific implementations of these strategies may differ, they are all based on the same underlying principles. This chapter also provides a general overview of systematic uncertainties, statistical modeling, and limit-setting procedures. The detailed results are presented in subsequent chapters. Here, the focus is on establishing the methodological foundation for Higgs pair analyses in the $b\bar{b}b\bar{b}$ final state.

6.1 b-Tagging

Since both analyses target final states with four bottom quarks, the ability to identify jets originating from b-quarks is fundamental to the event selection. This is achieved through a process known as flavor tagging, which refers to the identification of the flavor of the quark that initiated a jet. Among the different types of flavor tagging, b-tagging plays a central role in distinguishing b-jets from jets initiated by lighter quarks or gluons.

Flavor tagging relies on a variety of observables derived from the substructure of jets and the properties of tracks and secondary vertices reconstructed in the ATLAS detector [137]. Jets that are likely to have originated from a b-quark are referred to as *b-tagged* jets as shown in Figure 6.1. Since the origin of a jet cannot be directly observed, dedicated b-tagging algorithms are trained using simulated Monte Carlo samples containing

jets of known flavors. These algorithms utilize machine learning techniques such as boosted decision trees or deep neural networks to learn distinguishing features of b-jets. Common inputs include the presence of displaced secondary vertices, impact parameter significances of tracks, and the number of associated tracks with high transverse momentum [138, 139].

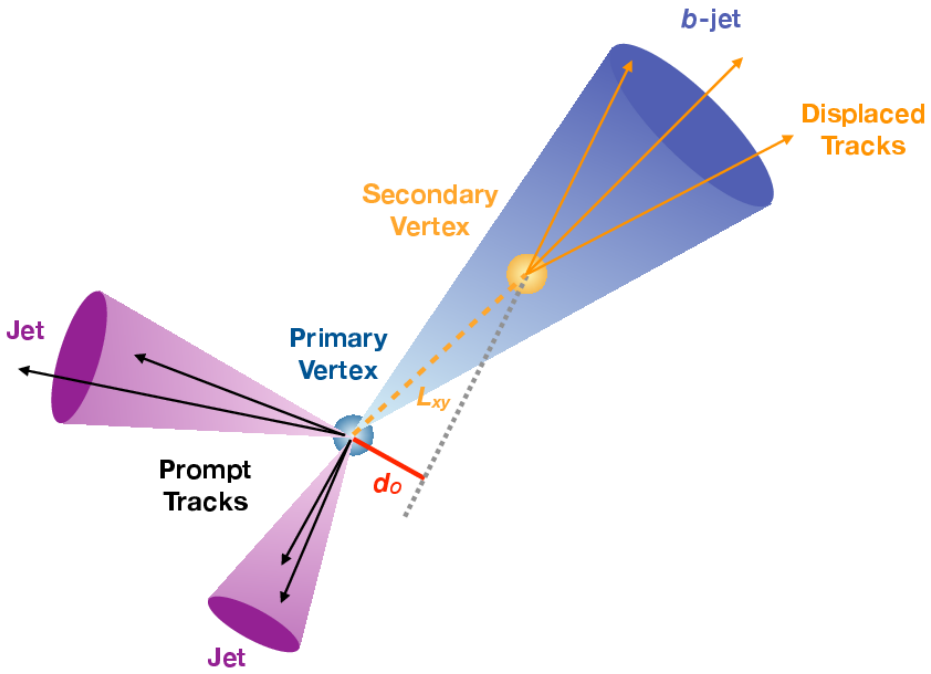


Figure 6.1: A b-jet can be identified by tracks that originate from a displaced secondary vertex due to the relatively long lifetime of b-hadrons. These tracks have a large transverse impact parameter (d0), unlike those in light-flavour jets, which typically come directly from the primary vertex [140].

Each b-tagging algorithm is associated with one or more working points (WPs), which correspond to fixed operating thresholds chosen to achieve specific efficiencies for correctly identifying b-jets. These working points are typically defined by the fraction of true b-

jets that are correctly tagged (efficiency), and are tuned to balance signal retention with background rejection. For example, a 70% WP implies that 70% of true b-jets are correctly identified, while maintaining a low misidentification rate for light- or charm-quark jets.

B-tagging is particularly important both for the analyses described here and for high-energy physics more broadly, because many processes of interest feature b-jets in the final state. In these types of analyses, requiring multiple b-tagged jets greatly enhances sensitivity to the signal by suppressing backgrounds that predominantly contain jets from light quarks or gluons.

Some of the main indicators of a b-tagged jet include:

- **Long-lived B-hadrons**: B-hadrons produced in b-quark fragmentation have relatively long lifetimes (~ 1.5 ps), resulting in displaced decay vertices several millimeters from the primary vertex.

- **High-mass secondary vertices**: The decay products of B-hadrons often reconstruct to a significant invariant mass, which helps distinguish them from lighter quark jets.

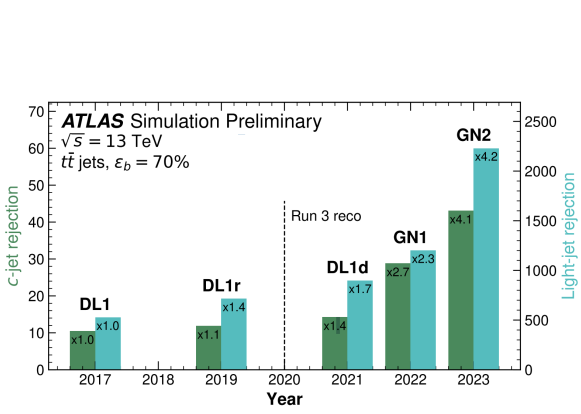
- **Multiple associated tracks**: Jets from b-quarks typically contain several charged particle tracks with large impact parameters.

- **Soft leptons**: Semileptonic decays of B-hadrons can produce low-energy (soft) leptons, which are another discriminating feature.

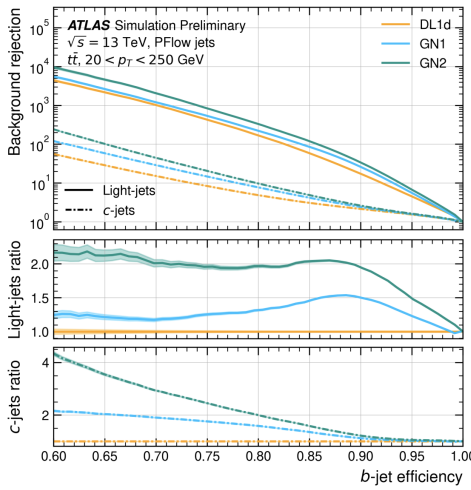
Altogether, b-tagging provides a powerful tool for enhancing signal-to-background discrimination in searches involving b-jets and is a crucial component of the event selection strategy in the following analyses.

The most advanced b-tagging algorithm currently used in ATLAS is the GN2 tagger [141]. GN2 stands for the second-generation Graph Neural Network tagger, an evolution from the earlier GN1 model. Unlike previous taggers based on dense neural networks, such as DL1, DL1r (a retrained version optimized for Run 2), and DL1d (a DL1r variant tuned for improved data–MC agreement and stability), GN2 leverages a graph-based architecture

enhanced with transformer layers. This architecture allows GN2 to process information from a variable number of tracks and associated objects within a jet. It performs multiple tasks simultaneously: jet flavor classification, partitioning of tracks into vertices, and prediction of the physical origin of each track. GN2 demonstrates significant performance improvements over both GN1 and earlier DL1 variants, achieving higher b-jet identification efficiency, while substantially reducing misidentification rates for charm- and light-flavor jets at fixed working points, as shown in Figure 6.2.



(a)



(b)

Figure 6.2: The c-jet and light-jet rejections of the different ATLAS flavor tagging algorithms over time in the $t\bar{t}$ MC simulation in Figure 6.2a provided at the 70% efficiency b-tagging working point where the improvement is normalized to the DL1 tagger. The background rejection is also plotted against b-jet efficiency for the same samples in Figure 6.2b where the ratio for the different taggers is taken according to the DL1d tagger [141].

The analyses in Chapters 9 and 7 both employ the GN2v01 tagger to identify small-R jets originating from b-quarks. Two working points (WPs) are used: the 77% and 85% b-tagging efficiency WPs, corresponding to the fraction of true b-jets correctly identified by the tagger. At these WPs, the GN2v01 tagger achieves substantial suppression of background

jets. Specifically, the light-flavor jet rejection factors are 92.33 and 29.36, and the charm-jet rejection factors are 5.88 and 3.08 for the 77% and 85% WPs, respectively.

These rejection rates quantify the algorithm's ability to distinguish b-jets from non-b jets. Formally, the rejection rate R is defined as the inverse of the misidentification efficiency ϵ for a given jet flavor:

$$R = \frac{1}{\epsilon_{\text{mistag}}}.$$

For example, a light-jet rejection rate of 92.33 at the 77% WP means that only about 1 in 92 light-flavor jets is misidentified as a b-jet. Higher rejection values indicate greater purity in the selected b-jet sample, which is critical for suppressing backgrounds in analyses targeting final states with multiple b-jets. The GN2v01 tagger thus offers excellent performance across a wide range of jet kinematics, enabling robust flavor tagging essential for the sensitivity of the analyses that follow.

6.2 Analysis Topologies

In searches for di-Higgs production in the $b\bar{b}b\bar{b}$ final state, a critical challenge is the correct reconstruction of the event topology. Since the signal involves two scalar bosons decaying into four b-quarks, the analysis strategy must account for the different kinematic configurations in which these quarks may appear as they depend on the momentum of the Higgs bosons.

The first reconstruction approach is the **resolved topology**, which targets events where the Higgs bosons are produced with relatively low transverse momentum. In this scenario, the four b-quarks hadronize into well-separated, narrow jets that can be individually reconstructed as small-radius ($R = 0.4$) jets. Events are selected by requiring the presence of at least four b-tagged small-R jets where each jet corresponds to a distinct b-quark from the decay.

However, not all di-Higgs events present this resolved topology. When one or both Higgs bosons are produced with substantial transverse momentum (**boosted**), their decay products can become collimated such that the two resulting b-quarks form a single, large-radius ($R = 1.0$) jet with substructure consistent with a two-prong decay. These large-R jets

are often *double b-tagged* to reflect their Higgs origin and are reconstructed using advanced substructure techniques.

In the **intermediate topology**, one Higgs boson is reconstructed in a resolved configuration with two small-R b-jets, while the other is sufficiently boosted to form a single large-R jet that is double b-tagged. In the fully **boosted topology**, both Higgs bosons are reconstructed as large-radius jets each being double b-tagged. These complementary categories extend the analysis sensitivity to a wider range of Higgs boson kinematics.

By classifying events into these different topologies, namely **resolved**, **intermediate**, and **boosted**, analyses are able to maximize sensitivity to the wide range of possible di-Higgs kinematics. This approach enables more comprehensive coverage of signal phase space and improves the ability to distinguish signal from background in each regime. An example of these topologies is shown in Figure 6.3 for the resonant $X \to SH \to b\bar{b}b\bar{b}$ analysis where each topology is used. The following chapters will discuss my work in the resolved topologies of the non-resonant $HH \to b\bar{b}$ and resonant $X \to SH \to b\bar{b}$ analyses.

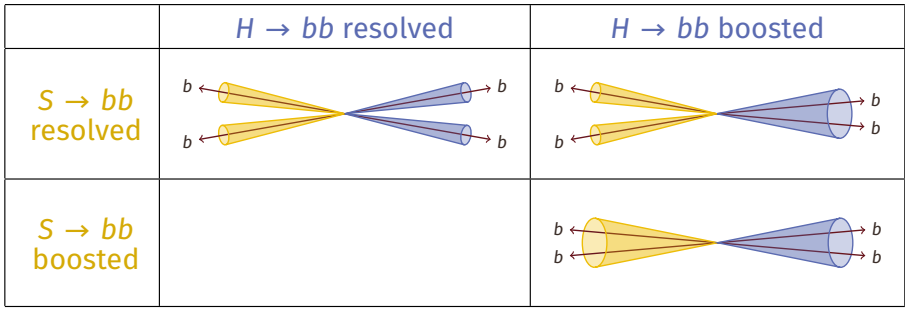


Figure 6.3: Illustration of possible topological configurations for the $b\bar{b}b\bar{b}$ final state in the SH $\to 4b$ analysis. In the resolved topology, all four b-quarks hadronize into separate, individually reconstructed small-R jets. In the intermediate topology, one $b\bar{b}$ pair is sufficiently boosted that both quarks are reconstructed within a single large-R jet. In the boosted topology, both $b\bar{b}$ pairs are sufficiently boosted that they are reconstructed within two separate large-R jets.

As my analyses deal with the resolved topology, pairing algorithms are needed to pair the

small-R jets to correctly reconstruct the intermediate scalar bosons. For a final state of four small-R jets, there are six possible permutations for the different pairings for the asymmetric case where the two intermediate scalars are unique and three possible permutations for the symmetric case where the two scalars are identical as shown in Figure 6.4. This is a vital part of resolved analyses in correctly reconstructing the signal and each analysis for the resolved topology has a dedicated pairing algorithm.

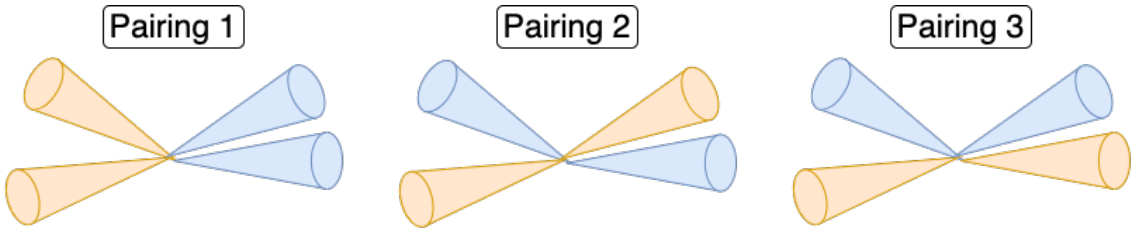
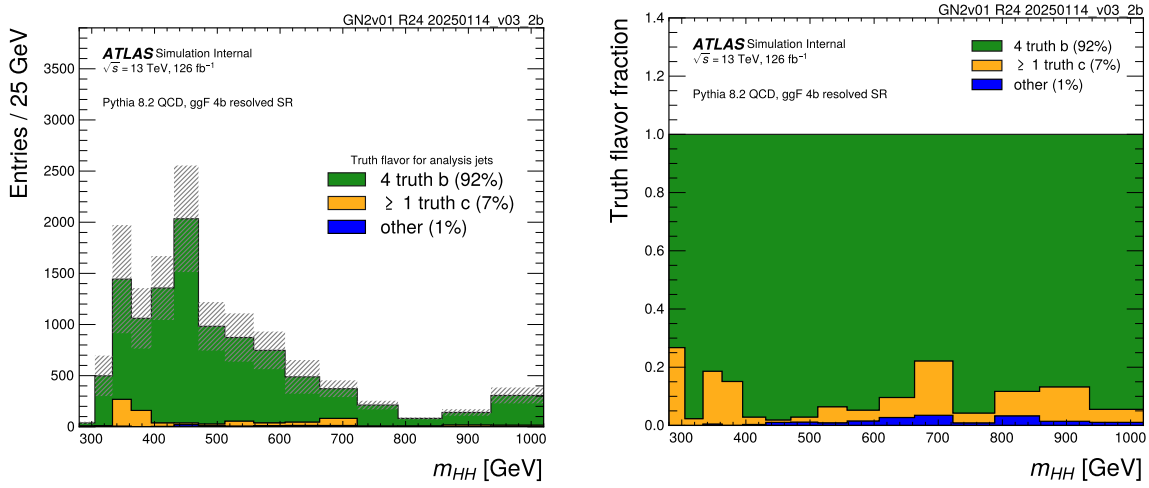


Figure 6.4: There are six possible ways to reconstruct four small-R jets into two unique scalars and three possible ways to reconstruct them into two identical scalars.

6.3 Background Composition

In both the resonant $X \to SH \to b\bar{b}b\bar{b}$ and non-resonant $HH \to b\bar{b}b\bar{b}$ analyses, the dominant background arises from QCD multi-jet production, which accounts for approximately 90–95% of the total background across much of the relevant di-Higgs phase space. The precise fraction depends on the kinematic selections applied, particularly those designed to suppress other backgrounds.

The composition of the QCD background in terms of jet flavor is a key consideration for analysis optimization. As shown in Figure 6.5, the majority of QCD events in the signal region consist of genuine b-jets, although contributions from charm and light-flavor jets are non-negligible. The bin-by-bin flavor breakdown highlights the effectiveness of b-tagging requirements in shaping the background composition. Notably, the transition from the DL1r b-tagger in Run 2 to the GN2 tagger in Run 3 has significantly improved the rejection of misidentified jets, enhancing the purity of the selected b-jet sample and improving overall signal sensitivity.



(a) (b)

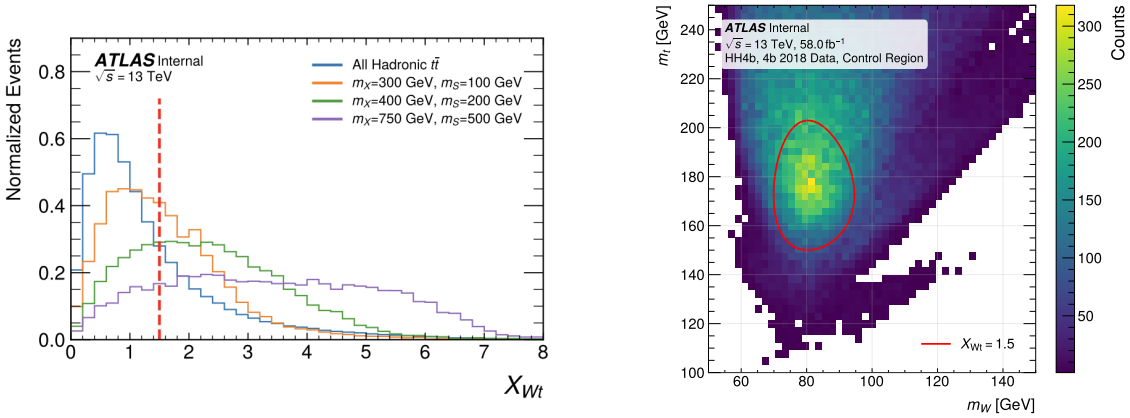
Figure 6.5: (a) Flavor composition of the QCD background in the $4b$ signal region of the $HH \to b\bar{b}b\bar{b}$ analysis. (b) Bin-by-bin normalized version of (a), showing the fractional contribution of different jet flavors.

The second most significant background arises from top quark pair production ($t\bar{t}$). To suppress this background, both analyses have at times a top veto discriminant known as X_{Wt}, defined as:

$$X_{Wt} = \sqrt{\left(\frac{10(m_W - 80.4\text{ GeV})}{m_W}\right)^2 + \left(\frac{10(m_t - 172.5\text{ GeV})}{m_t}\right)^2} \tag{6.1}$$

This variable is computed for all valid jet triplets in an event, where two jets form a W boson candidate and a third, b-tagged jet forms the top quark candidate. Among all combinations, the triplet yielding the smallest X_{Wt} is taken as the event-level discriminant. This technique is particularly effective in identifying and rejecting top-like jet configurations in multi-jet final states.

Figure 6.6 highlights the effectiveness and (some) limitations of the X_{Wt} discriminant. Subfigure (a) presents its distribution for $t\bar{t}$ background events and several $X \to SH \to b\bar{b}b\bar{b}$ signal hypotheses. While applying a cut of $X_{Wt} > 1.5$ significantly suppresses top quark



(a) (b)

Figure 6.6: (a) Distribution of the top veto discriminant X_{Wt} for simulated $t\bar{t}$ background and $X \to SH \to b\bar{b}b\bar{b}$ signal hypotheses, showing reduced separation when m_S is near the top quark or W boson mass. (b) Reconstructed W boson and top quark candidate masses in the non-resonant $HH \to b\bar{b}b\bar{b}$ Control Region (2018 data), using the jet combination with minimum X_{Wt} (see Eq. 6.1). The ellipse marks the region excluded by the X_{Wt} cut.

contamination, it also leads to substantial signal loss for mass points with $m_S = 70$ or 170 GeV which are regions where the reconstructed masses for the signal are close to the true W and top quark masses. As a result, this cut was ultimately not used in the resonant $X \to SH \to b\bar{b}b\bar{b}$ analysis to preserve signal efficiency. In contrast, subfigure (b) shows the two-dimensional mass plane of reconstructed W and top candidates with at least four reconstructed b-jets of a defined control region where the red ellipse indicates the kinematic regime where the $t\bar{t}$ decays are predominantly populated. The non-resonant $HH \to b\bar{b}b\bar{b}$ analysis retains the X_{Wt} cut, as its signal hypotheses remain largely unaffected by this selection.

Other Standard Model processes, including $Z \to b\bar{b}$ and single Higgs production via gluon-gluon fusion (ggF) or vector boson fusion (VBF), contribute only marginally to the total background in these analyses. Their yields have been evaluated using Monte Carlo simulation and are found to be negligible after final selections.

6.4 Background Estimation Method

Given the overwhelming contribution of QCD multi-jet events to the background in $4b$ final state analyses, a purely simulation-based modeling approach is not feasible due to the inherent difficulty in accurately describing such processes. Instead, both the $X \to SH \to b\bar{b}b\bar{b}$ and $HH \to b\bar{b}b\bar{b}$ analyses employ data-driven techniques to estimate the background. Historically, these types of analyses have relied on reweighting strategies—including both traditional and neural network-based methods—to extrapolate the background shape from a control region (CR) to a signal region (SR), under the assumption that the two regions are kinematically similar. While this approach has been widely used, it is limited by the assumption of transferability between regions and often suffers from large statistical uncertainties in the reweighting functions.

In contrast, the analyses presented in this thesis utilize a more flexible and expressive generative modeling framework based on the combination of **normalizing flows** and **Gaussian processes**. This method avoids the assumption of shared kinematic structure between CR and SR by directly learning the multivariate event distributions in the CR and using this to construct a predictive model of the SR. Specifically, normalizing flows are first trained

on events in the control region to model the full multivariate distribution of higher-level event kinematics, conditioned on variables that define the phase space of interest. Then, Gaussian processes are used to interpolate the distribution of the conditioning variables into the blinded signal region. Finally, the trained normalizing flow models are applied using the interpolated signal region conditions to generate synthetic event kinematics, which together define the predicted background in the SR.

This generative modeling strategy is conceptually distinct from traditional reweighting. Rather than learning a transfer function between CR and SR, it directly synthesizes events in the SR by sampling from a learned probability distribution. This provides greater modeling flexibility and statistical precision, and aligns closely with modern developments in machine learning, where generative models are increasingly used to capture complex high-dimensional data distributions.

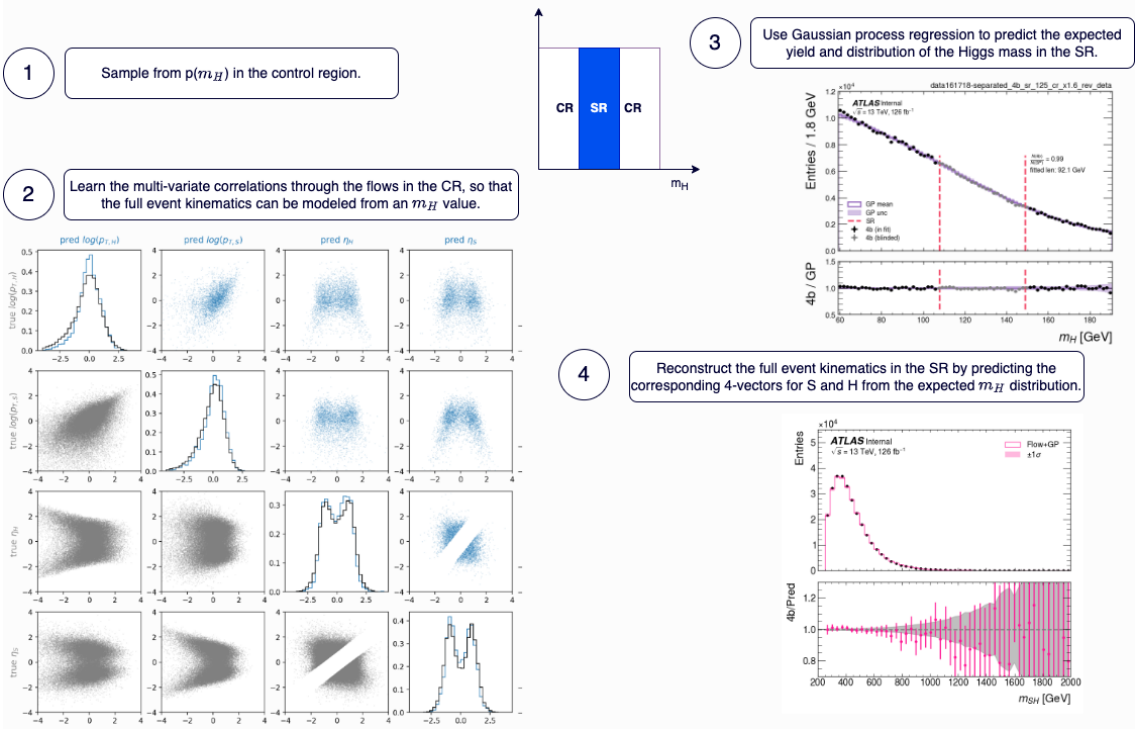


Figure 6.7: A general illustration of the data-driven background estimation method employed in the $X \to SH \to b\bar{b}b\bar{b}$ analysis. Further technical details of the method and the results will follow in the subsequent chapters.

Figure 6.7 illustrates the basic idea of this method employed in the $X \to SH \to b\bar{b}b\bar{b}$ analysis as a general introduction. The technical implementation of this method, including the training configurations of the normalizing flows and Gaussian processes, as well as validation and uncertainty estimation strategies, is discussed in detail in the chapters dedicated to each analysis.

6.4.1 Normalizing Flows

Normalizing flows are a class of generative models used to learn complex, high-dimensional probability distributions from data. They operate by transforming a simple base distribution (typically a multivariate Gaussian) into the target data distribution through a series of invertible, differentiable mappings. Formally, one defines a mapping T such that a data

point $\vec{x}$ is expressed as $\vec{x} = T(\vec{z})$, where $\vec{z} \sim p_z(\vec{z})$ is drawn from the base distribution.

The key property of normalizing flows is that the transformation is invertible, allowing the probability density of the target distribution to be computed exactly using the change of variables formula:

$$p(\vec{x}) = p_z(\vec{z}) \left|\det J_T(\vec{z})\right|^{-1}, \tag{6.2}$$

where J_T is the Jacobian matrix of the transformation. The model is trained by minimizing the Kullback–Leibler (KL) divergence between the empirical data distribution and the model prediction. This is equivalent to maximizing the likelihood of the data under the model:

$$\mathcal{L}(\vec{\theta}) = -\mathbb{E}_{p_{\text{data}}(\vec{x})} \left[\log p_{\text{model}}(\vec{x}; \vec{\theta})\right], \tag{6.3}$$

where $\vec{\theta}$ denotes the parameters of the flow model.

In the analyses presented here, conditional normalizing flows are employed. These extend the above framework to learn the conditional distribution $p(\vec{x} \mid \vec{y})$, where $\vec{x}$ represents higher-level event kinematic variables and $\vec{y}$ is a set of conditioning variables that define regions of phase space relevant to the analysis. The model is trained exclusively on data from the control region, where the kinematic distributions are known, and later applied to signal region configurations inferred via Gaussian processes.

A statistical ensemble of independently trained flows is used to account for model uncertainty that include variations due to random initialization and finite training statistics. The final background prediction is taken as the mean prediction across the ensemble, with the standard deviation providing an associated uncertainty.

6.4.2 Gaussian Processes

To predict the distribution of the conditioning variable(s) in the signal region, both analyses employ Gaussian processes (GPs). GPs are non-parametric Bayesian models that define a prior over functions and, once trained on data, provide a posterior distribution over possible fits. This probabilistic framework naturally yields both a central prediction (the posterior

mean function) and an associated uncertainty (the posterior standard deviation), making GPs well-suited for extrapolating distributions with quantified confidence.

In general, a GP models a function $f(x) \sim \mathcal{GP}(m(x), k(x, x'))$, where $m(x)$ is the mean function (typically taken as zero) and $k(x, x')$ is the kernel function, which encodes the correlation between inputs x and x'. In this analysis, the kernel of choice is the widely-used radial basis function (RBF), also known as the squared exponential kernel:

$$k(x_i, x_j) = \exp\left(-\frac{d(x_i, x_j)^2}{l^2}\right), \tag{6.4}$$

where $d(x_i, x_j)$ is the Euclidean distance between inputs x_i and x_j, and l is the characteristic length scale that controls the smoothness of the interpolation. A shorter length scale yields a model more sensitive to local fluctuations in the training data, while a longer length scale enforces smoother behavior.

The GP is trained using histogrammed data from the control region, where the x-values correspond to the centers of the bins in the conditioning variable(s), and the y-values are standardized bin counts. This standardization ensures numerical stability and centers the data for improved convergence:

$$N' = \frac{N - \mu}{\sigma}, \tag{6.5}$$

where N is the raw bin count, μ is the mean of all bin counts in the control region, and σ is the corresponding standard deviation.

Once trained, the GP is used to predict the shape and normalization of the conditioning variable(s) in the signal region. The posterior mean of the GP provides the expected value of the distribution, while the uncertainty band gives a measure of confidence in the extrapolation. To obtain continuous samples from the distribution within each bin, a smearing procedure is applied by uniformly sampling from the width of the bin corresponding to the GP prediction. This ensures that downstream applications of the conditioning variable receive a smooth and continuous input distribution.

Crucially, the total predicted normalization in the signal region, as inferred from the GP, is used to rescale the output of the normalizing flow to yield the final background estimate.

In this way, the GP plays a dual role: it informs the shape of the conditioning variable distribution and governs the total event yield in the region of interest.

6.5 Systematic Uncertainties

Systematic uncertainties in collider analyses generally fall into two main categories: **normalization uncertainties**, which affect the expected event yields, and **shape uncertainties**, which affect the kinematic distributions of events. Both types of uncertainties are critical, as they significantly influence the sensitivity and interpretation of analyses.

6.5.1 Normalization Uncertainties

- **Integrated Luminosity:** The uncertainty on the total integrated luminosity affects all simulated processes uniformly. Typical values range from 1–2%, varying slightly by data-taking year during the LHC Run 2 period (2015–2018).

- **Theoretical cross sections:** cross section uncertainties affect the predicted yields for signal and background processes. These uncertainties arise primarily from higher-order QCD corrections, the choice of PDF sets, and scale uncertainties in theoretical calculations. Depending on the process, they typically range from a few percent up to around 20%. When multiple signal production modes contribute within the same signal region, variations in their cross sections alter the signal composition and can affect both the normalization and shape of kinematic distributions.

6.5.2 Shape Uncertainties

- **Jet Energy Scale (JES) and Jet Energy Resolution (JER):** As detailed in Chapter 4 concerning the AF3 simulated samples, uncertainties in jet energy calibration and measurement significantly affect reconstructed jet-related variables. JES uncertainties typically range from 2–4% per jet, while JER uncertainties are larger, around 5–15%.

- **Flavor Tagging (b-tagging):** The identification efficiency for jets originating from

b-quarks introduces notable shape uncertainties. Mis-tagging rates for charm or light-flavor jets also contribute. In $b\bar{b}b\bar{b}$ analyses, these uncertainties are typically among the largest shape-related contributions, often on the order of 2–10% per jet.

- **Trigger Efficiency:** Trigger uncertainties arise from imperfect knowledge of the ratio between trigger efficiencies in data and simulation. These scale factors are applied to simulated events, and systematic variations are derived by shifting the scale factors up and down by one standard deviation. The variations are informed by measurements of per-jet online reconstruction and b-tagging efficiencies, which are then propagated to the event level. The resulting uncertainty is applied as an overall event weight variation in the simulation.

- **Background Estimation Uncertainties:** The primary background in the resolved $b\bar{b}b\bar{b}$ decay channel arises from multijet QCD processes, typically estimated from data-driven methods. The uncertainties associated with this estimation are generally substantial and can affect both normalization and shape.

Careful treatment of systematic uncertainties, especially those related to jet and flavor tagging calibrations and precise background modeling, is essential for achieving optimal sensitivity in searches for Higgs boson pair production. More detailed discussions of these uncertainties are provided in the corresponding chapters for each analysis.

6.6 Limit Setting

This section outlines the statistical framework employed in both the $X \to SH \to b\bar{b}b\bar{b}$ and $HH \to b\bar{b}b\bar{b}$ analyses to set exclusion limits on potential signals of new physics. The statistical interpretation begins with a binned profile likelihood fit to the reconstructed invariant mass spectrum of the heavy resonance candidate in the signal region. While the likelihood formalism provides an exact approach, practical calculations often rely on asymptotic approximations for computational efficiency, as detailed below.

To improve the stability of the statistical fits, a bin-merging algorithm is applied to the mass distribution using the `pyhf` [142] framework. The binning is determined by maximizing

a bin-merging score defined as:

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s}, \tag{6.6}$$

where n_b and n_s are the background and signal yields in a given bin, and N_b and N_s are the total background and signal yields, respectively. The coefficients z_b and z_s are user-defined parameters that control the relative importance of background and signal when determining bin boundaries, with the constraint $z_s + z_b = N_{\text{bins}}$.

A likelihood-based approach is adopted for the statistical inference. The likelihood function for a given signal strength parameter μ and nuisance parameters $\boldsymbol{\theta}$ is defined as:

$$\mathcal{L}(\mu, \boldsymbol{\theta}) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k}, \tag{6.7}$$

where n_j is the observed count in bin j, and s_j and b_j are the expected signal and background counts, respectively. The second product accounts for auxiliary measurements that constrain nuisance parameters through observed counts m_k and expected values $u_k(\boldsymbol{\theta})$.

To evaluate the compatibility of the observed data with different hypotheses, a test statistic is constructed from the profile likelihood ratio:

$$\lambda(\mu) = \frac{\mathcal{L}\left(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu)\right)}{\mathcal{L}\left(\hat{\mu}, \hat{\boldsymbol{\theta}}\right)}, \tag{6.8}$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ are the maximum likelihood estimates (MLEs) from the unconditional fit, and $\hat{\hat{\boldsymbol{\theta}}}(\mu)$ are the conditional MLEs given μ. This ratio ranges between 0 and 1, with values close to 1 indicating better agreement between the model and data.

A commonly used form of the test statistic is:

$$t_\mu = -2 \log \lambda(\mu), \tag{6.9}$$

where larger values of t_μ indicate decreasing compatibility of the hypothesis μ with the observed data.

The p-value for a given test statistic t_μ is then computed as:

$$p_\mu = \int_{t_{\text{obs}}}^{\infty} f(t_\mu | \mu), \tag{6.10}$$

where $f(t_\mu|\mu)$ is the probability density function of t_μ under the assumption of signal strength μ, and t_{obs} is the observed value of the test statistic.

For setting exclusion limits, a modified test statistic q_μ is defined as:

$$q_\mu = \begin{cases} -2\log\lambda(\mu) & \text{if } \hat{\mu} \leq \mu, \\ 0 & \text{if } \hat{\mu} > \mu, \end{cases} \tag{6.11}$$

which ensures that only upward fluctuations in the signal-like direction contribute to the exclusion limit.

Limits are set using the CL_s method [143, 144], defined as:

$$CL_s = \frac{p_{s+b}}{p_b}, \tag{6.12}$$

where p_{s+b} is the p-value under the signal-plus-background hypothesis, and p_b is the p-value under the background-only hypothesis:

$$p_{s+b} = P(q_\mu > q_{\text{obs}}|s+b), \tag{6.13}$$

$$p_b = P(q_\mu > q_{\text{obs}}|b). \tag{6.14}$$

A signal hypothesis is excluded at the 95% confidence level if $CL_s < 0.05$.

While the exact distributions of these test statistics can be determined through repeated pseudo-experiments, in practice, analytic asymptotic formulae are often used to approximate the distribution of t_μ in the large-sample limit. This approach provides computationally efficient p-values and confidence intervals, and is widely adopted in LHC analyses.

In the analyses described in this thesis, model-independent limits are typically placed assuming a reference cross section of $\sigma(pp \to X \to SH \to b\bar{b}b\bar{b}) = 1$ fb. This allows for reinterpretation across a range of theoretical models, since specific branching ratios and production cross sections can vary with the details of the extended Higgs sector or resonance structure under consideration.

Chapter 7

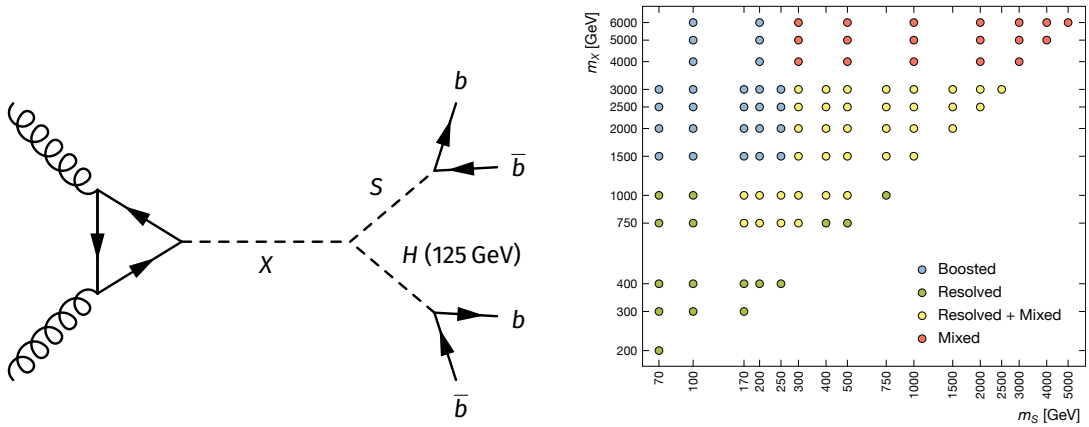
RESONANT $X \to SH \to b\bar{b}b\bar{b}$ WITH RUN 2 DATA

Having established the theoretical motivation for resonant asymmetric Higgs boson decays in Chapter 5, this chapter presents the search for such processes in the $X \to SH \to b\bar{b}b\bar{b}$ final state with the ATLAS Run 2 dataset. In this model-independent analysis, a heavy Higgs-like resonance X decays into a Standard Model Higgs boson (H) and a second scalar boson (S), where both X and S can have variable masses. The asymmetric decay arises from the mass difference between H and S. Both scalars subsequently decay to $b\bar{b}$, yielding four b-quarks in the final state. The $H \to b\bar{b}$ decay follows the Standard Model branching ratio of approximately 58%, while no specific branching fraction is assumed for $S \to b\bar{b}$, preserving model independence.

7.1 Data and Monte Carlo Datasets

The signal Monte Carlo (MC) samples used in this analysis simulate the production of a heavy resonance X via gluon-gluon fusion, with subsequent decays $X \to SH \to b\bar{b}b\bar{b}$, as illustrated in Figure 7.1. The samples span a wide range of mass hypotheses, with m_X varying from 200 GeV to 6 TeV and m_S from 70 GeV to 5 TeV. All signal samples were generated under the narrow-width approximation, setting the widths of both the X and S scalar bosons to a fixed value of 10 MeV. Event generation and parton showering are performed using Pythia 8 interfaced with EvtGen and the NNPDF23LO PDF set. Detector simulation is carried out with AF3, with additional JES and JER uncertainties included to account for differences relative to Geant4, as discussed in Chapter 4. This thesis focuses on the resolved channel, which has its own specific set of mass pairs and a unique recombination scheme in the mixed channel. The relevant signal mass pairs for the resolved channel are indicated in Figure 7.1b.

The data used in this analysis was collected by the ATLAS detector during Run 2 of



(a) (b)

Figure 7.1: The Feynman diagram in (a) for the production of the heavy scalar resonance X from gluon-gluon fusion as it decays into a lighter scalar S and a Standard Model Higgs boson, where the S and Higgs bosons each decay into a pair of b-quarks. All signal mass pairs in (b) for the SH4b analysis are shown for the different channels in which they are reconstructed. The resolved channel has its own unique mass pairs analyzed and is also included in a combination scheme with the mixed channel for other mass pairs. See the SH4b Internal Note [145].

the LHC in proton-proton collisions at a center-of-mass energy of $\sqrt{s} = 13$ TeV. The total integrated luminosity utilized for the analysis employing small-R b-jet triggers is $126.9\,\text{fb}^{-1}$. Data collected in 2015 were excluded due to differences in b-tagging algorithms and minimal impact relative to resources required. Additionally, a portion of 2016 data was excluded due to b-jet trigger inefficiencies caused by a bug in the online vertex finding algorithm, which was fixed only from data-taking period J onward. A small loss (1.5%) in total luminosity occurred across Run 2 due to invalid beamspot positions at the start of some runs [140].

Table 7.1: Integrated luminosities of each data-taking year for the resolved analysis which uses b-jet triggers.

Year	Luminosity (fb^{-1})
2016	24.9
2017	44.0
2018	58.0
Total	126.9

7.1.1 Signal MC Validation

Before formal approval to begin the $X \to SH \to bbbb$ analysis, a dedicated signal Monte Carlo (MC) validation was required as part of the ATLAS analysis pre-approval process. This validation step serves to confirm that the requested signal samples are both technically correct and physically meaningful. Specifically, the procedure involves checking that the event generator configuration accurately reflects the desired physics process, inspecting truth-level kinematic distributions (such as masses, transverse momenta, and angular separations), and verifying that the sample contains sufficient statistics for all signal mass points of interest. I produced truth-level signal MC samples across the grid of (m_X, m_S) mass pairs and inspected key kinematic variables to ensure that the mass peaks appeared in the expected regions and that the shapes of transverse momentum distributions matched theoretical expectations. This process allowed us to catch potential configuration errors or

filtering issues at an early stage, and provided assurance that the samples could support a robust analysis.

An example of this validation is shown in Figure 7.2, which displays the mass and transverse momentum distributions of the scalar boson S for signal events with $m_X = 3$ TeV. Such checks were essential for demonstrating the viability of the analysis and played a critical role for beginning the analysis.

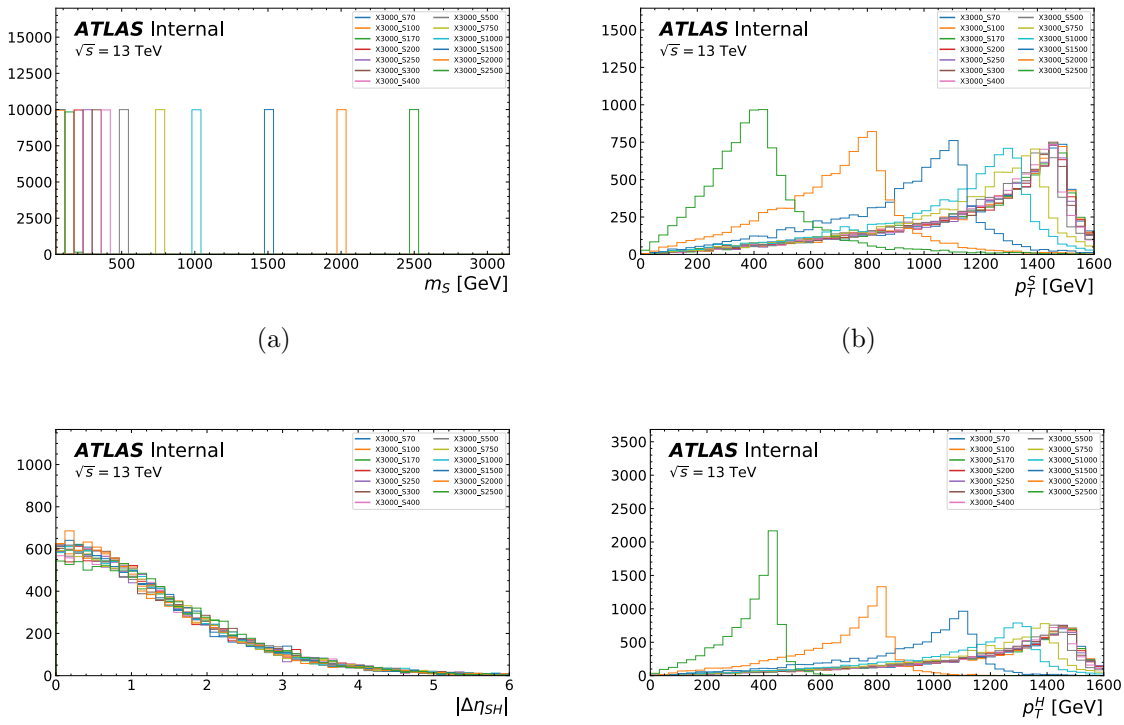


Figure 7.2: Truth-level distributions from the signal MC validation: (a) mass and (b) transverse momentum of the scalar boson S in addition to (c) $|\Delta\eta_{SH}|$ and (d) the transverse momentum of the Higgs boson for all signal mass pairs at $m_X = 3$ TeV. These checks were used to verify correct event generation and kinematic coverage for the $X \to SH \to bbbb$ analysis.

7.2 Triggers

Trigger studies for the resolved analysis were performed using events containing at least four jets that satisfy the following offline selection criteria:

- $p_{\mathrm{T}} > 40$ GeV
- $|\eta| < 2.5$
- b-tagged using the `GN2v00NewAliasWP` algorithm at the 77% working point
- Passing the Neural Network Jet Vertex Tagger (NN JVT)

The selected jets are ordered in decreasing transverse momentum. Both individual trigger efficiencies and trigger bucketing strategies were evaluated using simulated signal events across a range of mass hypotheses which span $m_X = 200$–3000 GeV. Two primary trigger bucket schemes were investigated. In this context, a trigger bucket strategy refers to the grouping of events based on which trigger(s) fired and allows for a consistent treatment of overlapping triggers with varying thresholds and efficiencies.

Bucket Strategy 1, the nominal trigger bucket scheme used in this analysis as illustrated in Figure 7.5, employs two triggers: the **2b+1j** and **2b+2j** triggers. The **2b+1j** trigger, which exhibits higher individual efficiency for most mass points, is applied in the first bucket defined by offline selections requiring the first jet to have $p_{\mathrm{T}} > 170$ GeV and the third jet to have $p_{\mathrm{T}} > 70$ GeV, implicitly ensuring the second jet also meets this threshold. This region targets the efficiency turn-on of the **2b+1j** trigger. Events failing the first bucket selection but passing the **2b+2j** trigger are collected in the second bucket. This trigger strategy maximizes efficiency, as the **2b+2j** trigger uniquely selects events not captured by the **2b+1j** trigger.

Bucket Strategy 1, used previously in the non-resonant $HH \to 4b$ Run 2 analysis, was ultimately chosen as the nominal trigger strategy for this analysis due to its combination of simplicity and high efficiency. Other bucket strategies based on the resonant $HH \to 4b$ Run 2 analysis in addition to individual triggers were checked, but these all had worse performance.

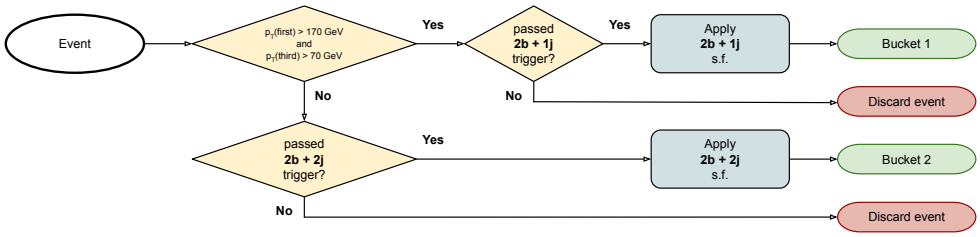
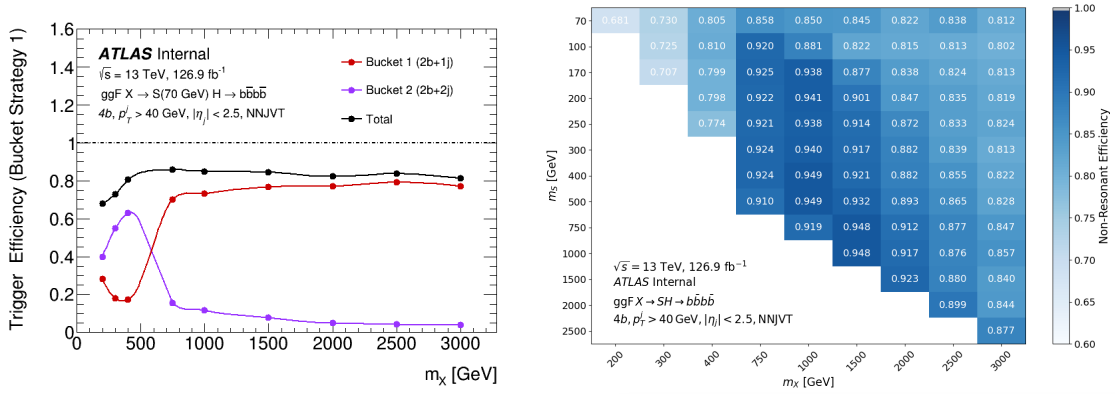


Figure 7.3: The nominal trigger bucket strategy (Bucket Strategy 1) employed that uses both **2b+1j** and **2b+2j** triggers and results in the highest efficiency across the phase space of the resolved channel. See the SH4b Internal Note [145].



(a) (b)

Figure 7.4: (a) Total trigger efficiency of Bucket Strategy 1 as a function of m_X for $m_S = 70$ GeV, showing the contribution from each individual bucket. Bucket 1 provides higher efficiency at large m_X values, while Bucket 2 is more efficient at lower m_X values. (b) Total efficiency of Bucket Strategy 1 for all signal mass pairs considered up to $m_X = 3000$ GeV. See the SH4b Internal Note [145].

7.3 Event Selection and Reconstruction

The event selection is designed to efficiently reconstruct the SH topology while suppressing contributions from the background processes discussed in Section 6.3. In the resolved channel of the $4b$ final state, events are required to contain at least four small-R jets ($R = 0.4$, anti-k_t) with transverse momentum $p_T > 40$ GeV and $|\eta| < 2.5$. The p_T threshold ensures compatibility with the trigger requirements, while the $|\eta|$ cut reflects the acceptance of the ATLAS inner tracking detector, which is essential for b-tagging performance (see Chapter 3). Jets must also pass b-tagging requirements using the GN2v01 tagger at the 85% efficiency working point. Additionally, each jet is required to pass the NN JVT, a machine learning-based algorithm that associates jets with the correct primary vertex and effectively suppresses jets originating from pileup interactions.

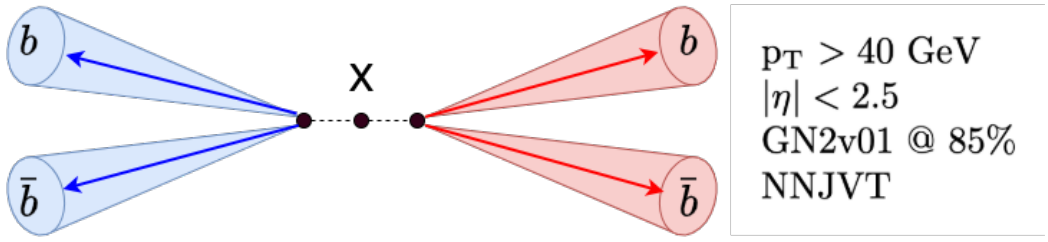


Figure 7.5: Schematic representation of the event selection criteria in the resolved channel of the SH4b analysis. Events are required to contain at least four small-R ($R = 0.4$) jets satisfying the specified kinematic and tagging requirements.

If more than four jets satisfy all selection criteria, only the four jets with the highest p_T are retained for analysis. A $3b$ validation region is constructed using the same selection criteria as the nominal $4b$ region, with the exception of the b-tagging requirement: exactly three of the four jets must satisfy the b-tag working point, while the remaining jet must fail it. In this case, the four jets are chosen by first prioritizing b-tagging status and then by descending p_T.

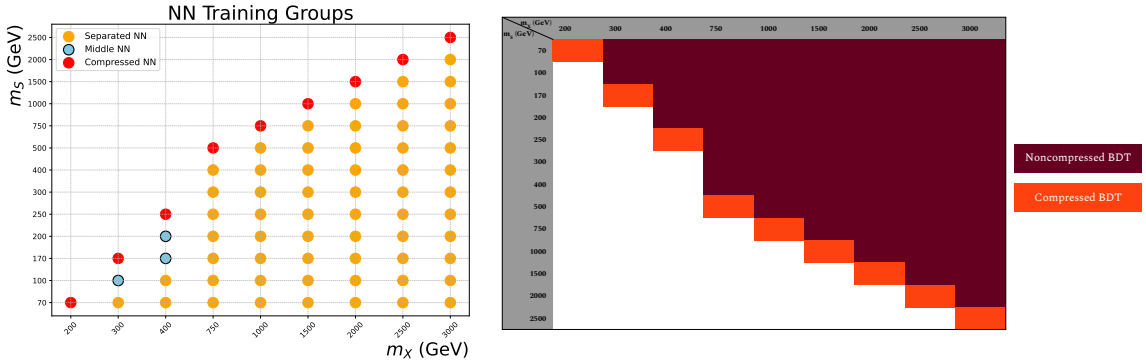
7.3.1 Pairing

After the event selection, the primary challenge is to correctly pair the four leading jets in order to reconstruct the $SH \to b\bar{b}b\bar{b}$ signal, as outlined in Section 6.2. In this analysis, the nominal pairing method is based on a Boosted Decision Tree (BDT), which is used for the background estimation, systematic uncertainty evaluation, and the final results. However, much of the development and optimization focused on a neural network (NN)–based pairing method. Due to the central role of the pairing algorithm in the analysis sensitivity, additional details on the BDT setup and training are provided in Appendix B, while this section details the motivations and strategies behind the grouping and comparison of the two methods.

The kinematic diversity of the signal, driven by the wide range of resonance masses considered, necessitates a tailored approach to jet pairing. Because the mass of the SM Higgs boson is fixed, the event topology is predominantly determined by the mass difference between the heavy resonance X and the lighter scalar S. When m_X is only slightly greater than $m_S + m_H$, the limited available phase space results in both the S and H bosons being produced nearly at rest. This "compressed" regime, characterized by $\frac{m_X}{m_S+m_H} < 1.33$, leads to final-state b-jets that are close in angle, making correct jet pairing particularly challenging.

Conversely, when the mass splitting is large, the S and H bosons are highly boosted. In these cases, the two b-jets from each boson are close in angular distance to each other but are well separated from the other pair, simplifying the pairing task. Not all signal mass points fall neatly into these groups; some are intermediate and display kinematics that blend features of both regimes. These observations motivate the definition of kinematic groups for the ML-based pairing algorithms. For the NN, three groups are defined: "compressed," "separated," and "middle," corresponding to small, large, and intermediate mass splittings, respectively. The BDT method uses nearly identical training groups, defined as "compressed" and "non-compressed." The "non-compressed" group is functionally equivalent to the NN's "separated" group, with the addition of the signal points grouped as "middle" by the NN. Figure 7.6 illustrates the assignment of mass points to these groups for both algo-

rithms. The convergence of independent studies toward similar groupings across different kinematic regimes demonstrates the robustness and physical motivation of this approach. Dividing the signal samples in this way allows each ML pairing model to be optimized for the unique challenges of its respective kinematic regime, ultimately improving the pairing performance across the full signal phase space.



(a) (b)

Figure 7.6: Assignment of signal mass points to the training groups of the NN (a) and the BDT (b). The compressed group consists of signal points where the mass difference between X and $S + H$ is small, resulting in S and H being produced nearly at rest. The non-compressed group in the BDT and the separated group in the NN contain signal points where both scalars are significantly boosted. The NN includes an additional group, referred to as the middle group, representing mass points whose kinematics are intermediate between the compressed and separated groups.

NN Training

To construct the NN (and BDT) jet pairing training set in the resolved $SH \to b\bar{b}b\bar{b}$ channel, additional criteria are imposed beyond the nominal event selection:

- $\Delta R_{bb} > 1.0$ between the two b-quarks originating from the H boson at truth level,

- $\Delta R_{bb} > 1.0$ between the two b-quarks from the S boson at truth level,
- **Truth matching:** Each truth-level b-quark must be uniquely matched to a reconstructed small-R jet within $\Delta R(\text{jet}, b\text{-quark}) < 0.4$.

These requirements ensure that the training set contains events where each b-quark can be reconstructed as a single jet and reliably matched to the truth information.

All signal MC samples with m_X between 200 and 3000 GeV are used for NN training and validation. For each selected event, the three possible pairings of the four b-tagged jets are constructed and included among the training variables. Crucially, both the NN and the BDT treat the two jet pairs symmetrically, without explicitly assigning jets to S or H, in order to avoid sculpting background distributions and introducing analysis biases, a problem observed in earlier NN pairing prototypes. Correct jet pairings are assigned by matching reconstructed b-jets to generator-level b-quarks using the minimum angular distance criterion, ensuring unbiased training.

To maximize the use of available data and prevent overfitting, a five-fold cross-validation scheme is employed for NN training [146, 147]. Events are partitioned into five folds based on their event number modulo five. For each fold, the NN is trained on the remaining four folds and validated on the held-out fold. In this way, every event is ultimately evaluated by a network that was not trained on it, and a consistent approach can be applied to both MC and data.

The NN is trained on a set of 27 kinematic variables, including $\Delta\eta_{jj}$, $\Delta\phi_{jj}$, and ΔR_{jj} for both pairs and all three possible pairings (18 variables), along with $\Delta\eta_{SH}$, $\Delta\phi_{SH}$, and ΔR_{SH} for the S and H candidates of each pairing (9 variables). Table 7.2 summarizes these variables. Early stopping and a grid search over network hyperparameters (number of layers, neurons, normalization, batch size, activation functions, learning rate) were used to prevent overfitting and optimize model performance.

The final neural network architecture comprises an input layer with 64 neurons, followed by dense layers of 32, 16, and 3 neurons, respectively. The last three output neurons represent the probability that each respective pairing is correct. The network is trained with a batch size of 2000, input normalization, a learning rate of 0.003, and uses a softmax

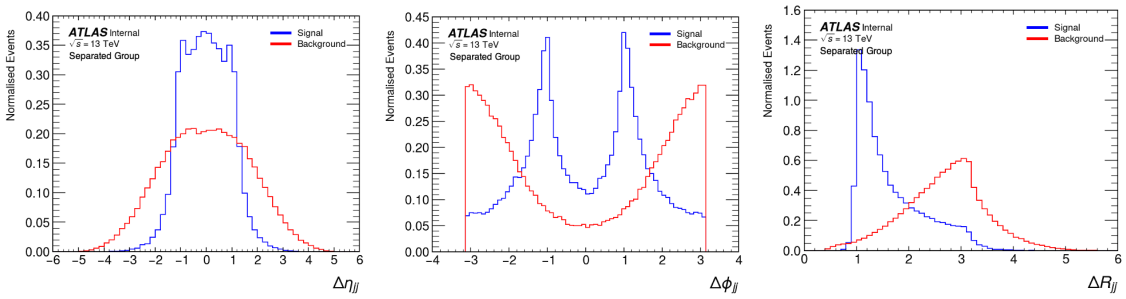
Table 7.2: List of neural network training variables and their descriptions. Variables are computed from four jets labeled $\{i, j, k, l\}$ for all three pairings [{1,2,3,4}, {1,3,2,4}, {1,4,2,3}]. The S and H candidates are treated symmetrically; the S/H assignment is independent of the training.

NN Training Variable	Description
$\Delta\eta_{ij}$	Difference in η between i-th and j-th jet
$\Delta\phi_{ij}$	Difference in ϕ between i-th and j-th jet
ΔR_{ij}	Angular distance between i-th and j-th jet
$\Delta\eta_{kl}$	Difference in η between k-th and l-th jet
$\Delta\phi_{kl}$	Difference in ϕ between k-th and l-th jet
ΔR_{kl}	Angular distance between k-th and l-th jet
$\Delta\eta_{SH}$	Difference in η between S_{ij} and H_{kl} candidates
$\Delta\phi_{SH}$	Difference in ϕ between S_{ij} and H_{kl} candidates
ΔR_{SH}	Angular distance between S_{ij} and H_{kl} candidates

output with a categorical cross-entropy loss. While this is a relatively simple architecture, it is well suited for the pairing task, balancing efficiency and background modeling, which is essential for robust data-driven background estimation. The performance of the NN pairing is demonstrated in Figure 7.8, which shows the reconstructed mass distributions for S and H in a representative mass scenario.

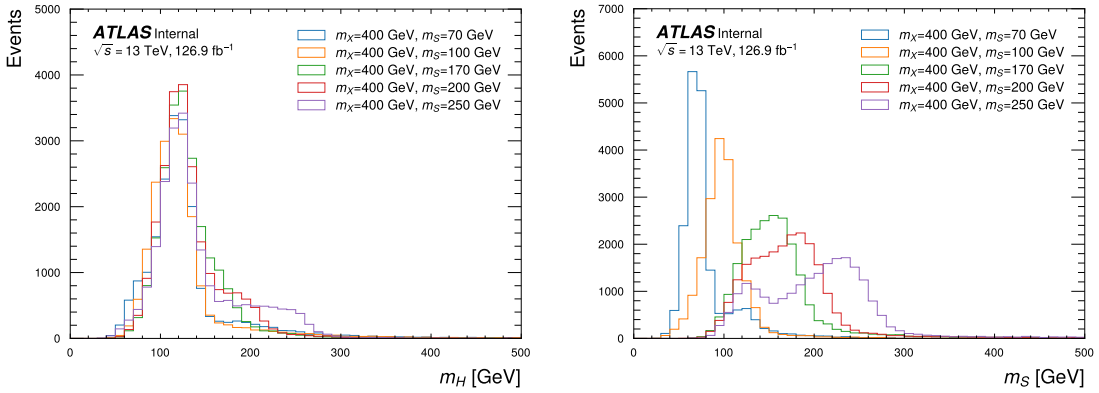
Comparison of NN and BDT Performance

Both the NN and BDT algorithms provide robust jet pairing performance across the considered mass range. The BDT pairing employs a variable set similar to the NN, with the addition of mass and p_T differences between the reconstructed bosons for the different pairings, as described in Appendix B. This additional information allows the BDT to achieve superior mass reconstruction, particularly at high m_S. By contrast, the NN pairing produces a slightly more accurate background model with reduced modeling uncertainties.



(a) (b) (c)

Figure 7.7: Distributions of key NN training inputs for the separated group: (a) $\Delta\eta_{jj}$, (b) $\Delta\phi_{jj}$, and (c) ΔR_{jj}, for S and H candidates. The signal corresponds to correctly paired jets, while the background consists of incorrect pairings. The histograms are normalized to account for the fact that there are twice as many incorrect pairings as correct pairings per event.



(a) (b)

Figure 7.8: Reconstructed mass distributions using the NN pairing for $m_X = 400$ GeV: (a) m_H and (b) m_S. Clear mass peaks are observed, with a small excess near 125 GeV in (b) due to events where jets are correctly paired but S and H assignments are swapped.

The final choice of pairing method for the nominal analysis was determined by a comparison of the expected statistical-only 95% confidence level limits for the $X \to SH \to b\bar{b}b\bar{b}$ decay, as shown in Figure 7.9. While the NN-based pairing yields marginally improved sensitivity for some lower-mass signal points, the BDT-based pairing offers a significant sensitivity improvement at higher masses. Consequently, the BDT pairing was selected as the default for the resolved channel analysis and all the results in the following sections use the BDT pairing.

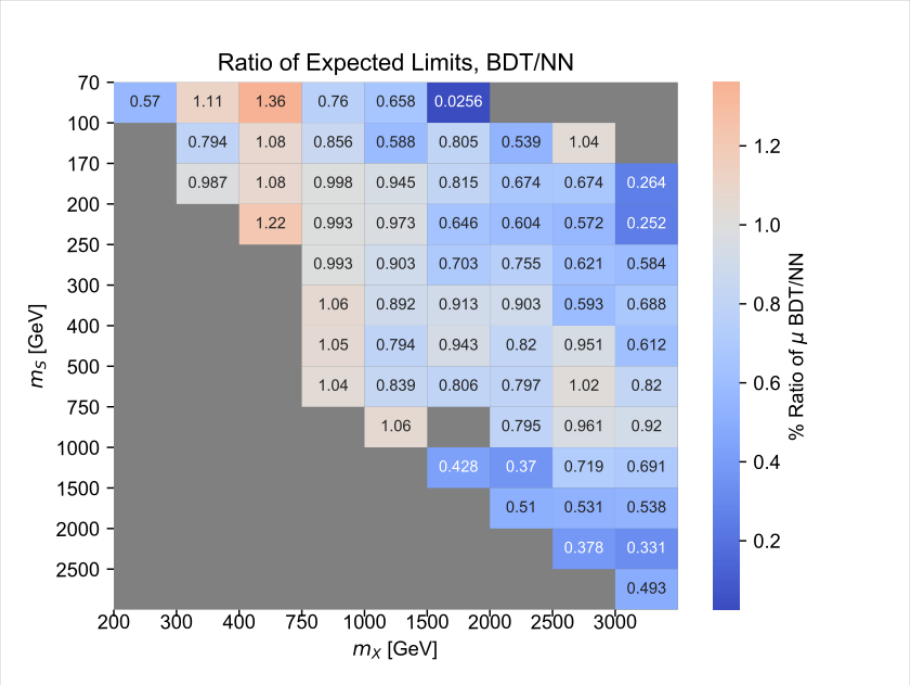


Figure 7.9: Ratio of statistical-only 95% confidence level upper limits for $X \to SH \to b\bar{b}b\bar{b}$ between the NN and BDT pairing methods, evaluated using resolved signal samples. The NN pairing shows slightly better sensitivity for some low-mass points, while the BDT provides a substantial improvement at higher masses.

7.3.2 *Assignment of S and H Candidates*

As this is an asymmetric decay, the intermediate scalar bosons S and H are not identical, making it essential to correctly assign each reconstructed b-jet pair to its parent boson after the BDT determines the best pairing. This assignment leverages a key kinematic correlation observed in MC studies: the pair of b-jets with the larger angular separation ΔR_{bb} often originates from the heavier scalar boson, since the more massive particle is typically less boosted and its decay products are more widely separated. This trend is illustrated for various mass hypotheses in Figure 7.10, where the ΔR distribution is shown for b-quark pairs from both S and H. The efficiency of the S and H assignment is nearly 100% when m_S and m_H are well separated, but drops to as low as 60% for signal samples with $m_S = 100$ GeV, where the masses are closer together.

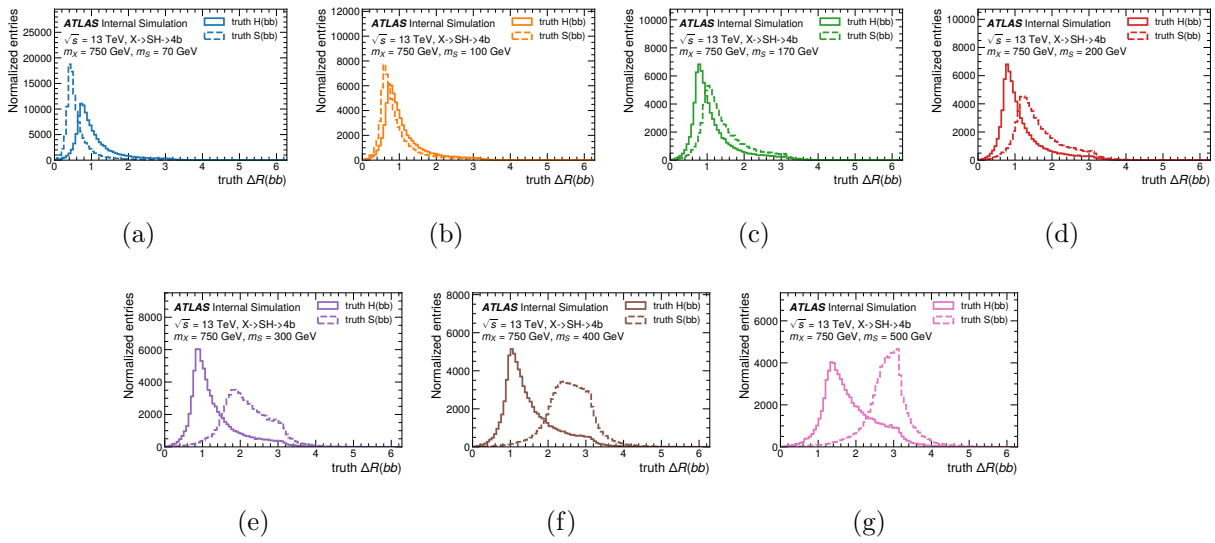


Figure 7.10: Distribution of ΔR between the two b-quarks from the decays of the scalar S (dashed) and the SM Higgs boson (solid) for signal samples with $m_X = 750$ GeV and m_S as indicated in each panel. See the SH4b Internal Note [145].

The assignment procedure is therefore straightforward: for each event, after selecting the optimal b-jet pairing, the pair with the larger ΔR_{bb} is assigned to the heavier boson, and

the pair with the smaller ΔR_{bb} is assigned to the lighter boson. This method is robust across different mass combinations and efficiently exploits the differences in kinematics between the S and H decays. Figure 7.11 provides a schematic of this assignment strategy.

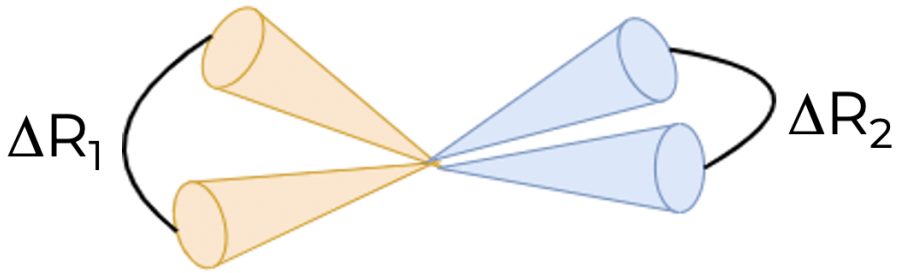


Figure 7.11: Illustration of the assignment of S and H candidates to the two b-jet pairs based on their angular separation. The pair with smaller ΔR is assigned to the lighter scalar boson for a given signal mass pair.

Depending on the signal mass hypothesis, events are grouped both by the BDT training group (compressed or non-compressed) and by whether S or H is lighter (denoted as "minS" or "minH"). Each resulting configuration—compressed minS, compressed minH, non-compressed minS, and non-compressed minH—defines a unique background group for the analysis, ensuring a tailored treatment for all relevant kinematic regions.

7.3.3 Background Reduction

Background processes present a major challenge in analyses with $4b$ final-states, particularly due to the overwhelming presence of multijet QCD and $t\bar{t}$ production, as discussed in Section 6.3. To enhance signal purity, additional selection criteria are applied based on the reconstructed kinematics of the S and H candidates.

A key variable used to suppress background is the pseudorapidity separation between the reconstructed S and H bosons, denoted as $\Delta\eta_{SH}$. Signal events, particularly those from the asymmetric decay topology considered here, tend to produce S and H bosons with relatively small separation in η, whereas background processes, such as QCD multijet production, often produce more widely separated jets. For this reason, a requirement of $|\Delta\eta_{SH}| < 1.5$ is imposed. This cut was determined to provide an optimal balance between signal efficiency and background rejection across the full range of signal mass points.

In addition to $\Delta\eta_{SH}$, other discriminating variables were considered for further background reduction. One example is the top veto discriminant, X_{Wt}, which is designed to reject events consistent with hadronic $t\bar{t}$ decays as defined in Equation 6.1. However, in the present analysis, imposing a cut on X_{Wt} was found to reduce signal acceptance, especially for signal models with masses near the W boson or top quark mass (approximately 80.4 GeV and 173 GeV, respectively). For lower m_X and m_S values, this selection would remove a significant fraction of signal events as shown in Figure 7.12, so it is not included in the nominal resolved analysis.

The overall strategy is thus to apply a $\Delta\eta_{SH}$ requirement to increase the signal-to-background ratio while preserving high acceptance for a broad range of signal hypotheses. The effects of this selection on signal acceptance and efficiency are illustrated in Figure 7.13 for representative mass points.

7.3.4 Kinematic Region Definition

Accurate definition of the signal and control regions is crucial for maximizing sensitivity and ensuring a robust background estimation. In the resolved channel, the regions are defined based primarily on the reconstructed mass of the Higgs boson candidate, m_H, in each

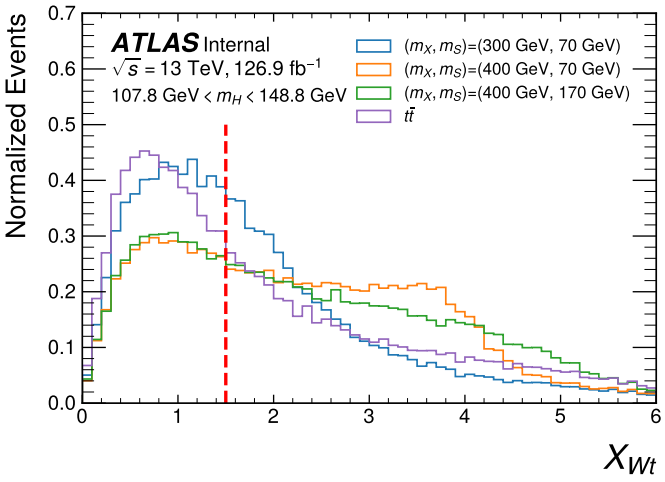
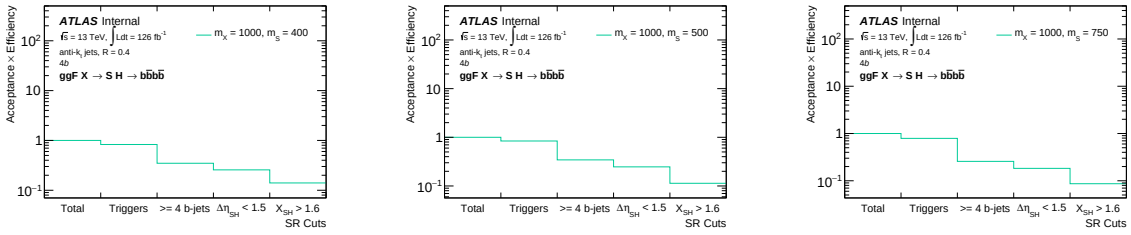


Figure 7.12: The distribution of the top veto discriminant, X_{Wt} with signal samples of (m_X GeV,m_S GeV): (300,70), (400,70), (400,170) and the simulated $t\bar{t}$. Requiring that $X_{Wt} > 1.5$ as an additional kinematic cut would result in a significant loss in signal events for these samples despite the background reduction.



(a) (b) (c)

Figure 7.13: Acceptance × efficiency plots obtained for the trigger, 4b selections, QCD background rejection ($\Delta\eta_{SH}$) selections in SRs defined by $X_{SH} < 1.6$, when using the BDT-paired jets for $m_X = 1000$ GeV and (a) $m_S = 400$ GeV, (b) $m_S = 500$ GeV, (c) $m_S = 750$ GeV. See the SH4b Internal Note [145].

event. A one-dimensional blinding strategy is implemented along the m_H axis. The signal region (SR) is centered at the known Higgs boson mass, $m'_H = 125$ GeV, with boundaries

determined by the parameter $\beta = 0.16$, a value obtained from optimization studies that maximizes the expected sensitivity based on statistics-only limits. The SR is defined as follows:

$$\frac{m'_H}{1+\beta} < m_H < \frac{m'_H}{1-\beta} \tag{7.1}$$

This choice yields a symmetric interval around the Higgs mass and is consistent with the kinematic properties of the signal. Control regions (CRs) are defined as sidebands to the SR and are used for data-driven background modeling and validation. The width of the CRs is controlled by the parameter α_{CR}, set to 1.6, resulting in the following boundaries:

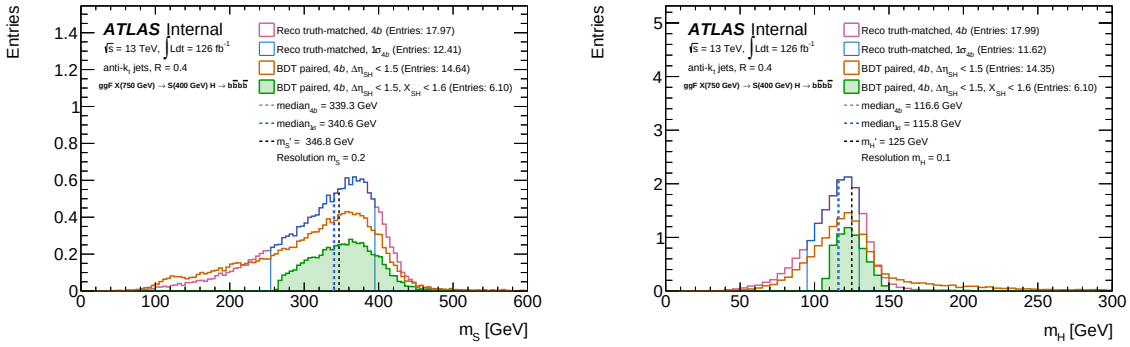
$$\begin{aligned} \text{Left sideband (CR):} \quad & 59.3 \text{ GeV} < m_H < 107.8 \text{ GeV} \\ \text{Signal region (SR):} \quad & 107.8 \text{ GeV} < m_H < 148.8 \text{ GeV} \\ \text{Right sideband (CR):} \quad & 148.8 \text{ GeV} < m_H < 190.7 \text{ GeV} \end{aligned} \tag{7.2}$$

For each signal hypothesis, sub-signal regions (sub-SRs) are constructed in the two-dimensional m_H–m_S plane. These sub-SRs are defined as elliptical regions, or "ovals," around each S mass point to account for the variable resolution of the reconstructed S mass. The boundaries of the sub-SRs are set according to the variable X_{SH}:

$$X_{SH} = \sqrt{\left(\frac{m_H - m'_H}{r_{f,H} \cdot m_H}\right)^2 + \left(\frac{m_S - m'_S}{r_{f,S} \cdot m_S}\right)^2} < 1.6 \tag{7.3}$$

Here, m'_H is fixed at 125 GeV, and $r_{f,H} = 0.1$ is the Higgs mass resolution factor. The center m'_S and resolution $r_{f,S}$ for the S candidate are obtained from the distributions of the reconstructed m_S of the signal samples, as summarized in Table 7.3, to optimize signal coverage for each mass point.

This approach ensures that the signal region is tightly matched to the kinematic properties of each signal hypothesis, maximizing the fraction of signal events retained while minimizing contamination from background. The resulting SRs and CRs are illustrated in the m_H–m_S plane in Figure 7.15. This kinematic region definition forms the basis for subsequent background estimation and statistical analysis.



(a) (b)

Figure 7.14: Example reconstructed mass distributions for selected signal points. The pink solid line shows the median after the $4b$ selection, with the pink dashed line indicating the same for the initial selection. The blue solid line marks the innermost 68% quantile, while the blue dashed line shows its median. The black dashed line represents the average median across all signals with the same m_S. Orange indicates BDT-paired events with the $\Delta\eta_{SH} < 1.5$ selection, while the green filled area shows the 1D signal region after the X_{SH} selection. a displays m_S for $m_X = 750$ GeV, $m_S = 400$ GeV, and b displays m_H for the same mass point. See the SH4b Internal Note [145].

7.3.5 Analysis Categorization

To enhance the sensitivity of the search for scalar resonances, the signal region is subdivided into analysis categories, where the main discriminant is separated into different subdivisions of other observables. This approach follows the strategy used in the Run 2 non-resonant $HH \to 4b$ analysis and leverages additional kinematic information to improve discrimination between signal and background.

Two key variables are used to define the categories: $|\Delta\eta_{SH}|$, the pseudorapidity separation between the reconstructed S and H bosons, and X_{SH}, the elliptical variable described in Section 7.3.4 that defines the sub-signal regions in the m_H–m_S mass plane. Both variables already serve critical roles in background rejection and signal region definition. By catego-

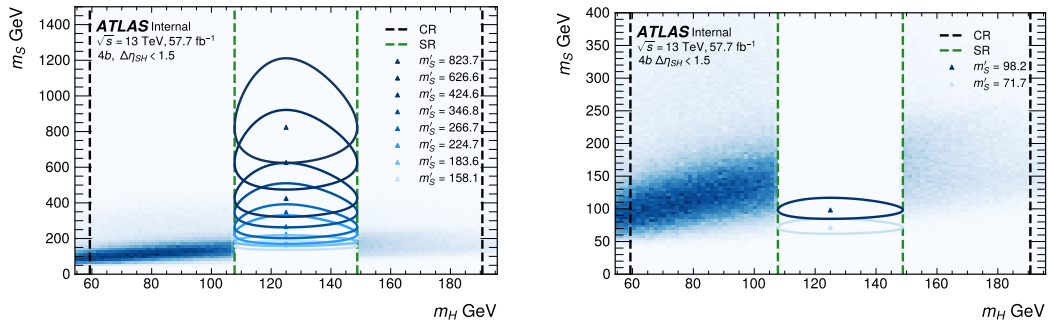
Table 7.3: The optimal values for the center of the oval of each sub-SR and the width of the ellipse along m_S for all the 13 signal regions that are used in the search for the 56 signal mass points.

True m'_S	Reconstructed m'_S	$r_{f,S}$
70	71.7	0.1
100	98.2	0.1
170	158.1	0.1
200	183.6	0.1
250	224.7	0.2
300	266.7	0.2
400	346.8	0.2
500	424.6	0.2
750	626.6	0.2
1000	823.7	0.2
1500	1200.9	0.2
2000	1571.6	0.3
2500	1952.0	0.3

rizing events along these axes, it is possible to target regions of phase space with distinct signal-to-background characteristics and thereby improve overall analysis sensitivity.

Figure 7.16 displays the distributions of $|\Delta\eta_{SH}|$ and X_{SH} in the signal region for events reconstructed with the BDT pairing method, shown for $m'_S = 300$ GeV. In both panels, the estimated background distribution from the normalizing flow method is compared to the expected distributions from two representative signal mass points, each normalized to a reference cross section of 1 fb. This visualization illustrates how event categorization in these two variables enhances the ability to distinguish signal from background.

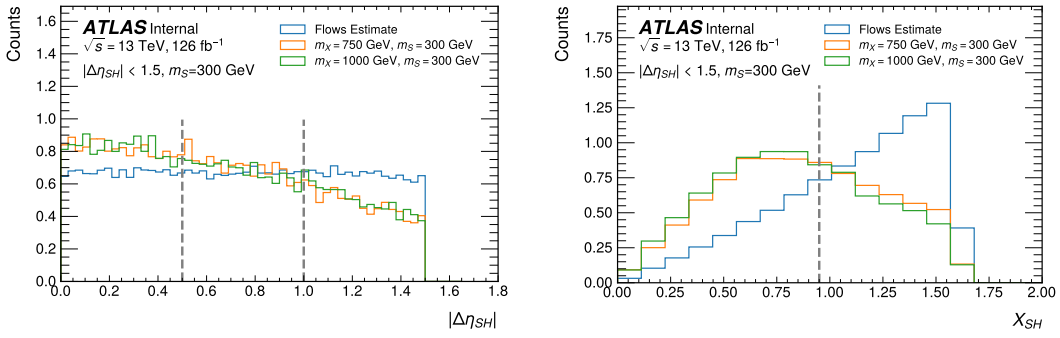
By dividing the analysis into categories based on $|\Delta\eta_{SH}|$ and X_{SH}, the search maximizes



(a) $m'_S > 125$ Gev (b) $m'_S < 125$ GeV

Figure 7.15: Mass plane and kinematic regions for 2018 data, using the BDT pairing in the non-compressed region for a selected number of m'_S mass points considered in the resolved analysis. a shows the m_H-m_S mass plane and kinematic regions for $m'_S > 125$ GeV and b shows this for $m'_S < 125$ GeV

.



(a) (b)

Figure 7.16: Distribution of $|\Delta\eta_{SH}|$ in (a) and X_{SH} in (b) for the BDT-reconstructed S and H candidates in the signal region for $m'_S = 300$ GeV. The background estimate from the normalizing flow method (brown) is compared with two simulated signal mass points, each normalized to 1 fb.

sensitivity across a broad range of signal kinematics and mass hypotheses, ensuring that regions with varying signal-to-background ratios contribute optimally to the final statistical analysis.

In total, we have 6 analysis categories defined as follows:

- Category 0: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.0 < X_{SH} < 0.95$
- Category 1: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.0 < X_{SH} < 0.95$
- Category 2: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.0 < X_{SH} < 0.95$
- Category 3: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.95 < X_{SH} < 1.6$
- Category 4: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.95 < X_{SH} < 1.6$
- Category 5: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.95 < X_{SH} < 1.6$

7.3.6 Small Standard Model Backgrounds

While the dominant backgrounds in the $b\bar{b}b\bar{b}$ final state arise from QCD multijet and $t\bar{t}$ production, several smaller Standard Model (SM) processes can also contribute to the event yield in the signal region. These subdominant backgrounds include $H \to b\bar{b}$ decays produced via gluon-gluon fusion (ggF) and vector boson fusion (VBF), as well as $Z \to b\bar{b}$ production. The cross sections for these processes are provided in Table 7.4.

Although these backgrounds make up only a few percent of the total, it is important to assess both their event yields and kinematic distributions to ensure that they do not introduce significant biases or localized resonances within the signal regions defined for the analysis. In particular, care must be taken to confirm that these backgrounds are properly modeled by the data-driven background estimation techniques used in this analysis.

The treatment of these backgrounds follows the same reconstruction and categorization steps applied to the signal and primary backgrounds. Specifically, events are paired using the BDT method for both compressed and non-compressed regions, and the assignment between the S and H bosons is performed according to the angular separation of the

Table 7.4: Cross sections for $H \to b\bar{b}$ and $Z \to b\bar{b}$ SM processes.

Sample	$\boldsymbol{\sigma}$ **(pb)**
ggF $H \to b\bar{b}$	28.30
VBF $H \to b\bar{b}$	2.18
$Z \to b\bar{b}$	137.74

paired b-jets, resulting in four configurations: compressed/minH, compressed/minS, non-compressed/minH, and non-compressed/minS. Table 7.5 summarizes the expected yields of these SM backgrounds in the blinded signal band for each configuration and data-taking year. The event yields are found to be negligible compared to the QCD and $t\bar{t}$ backgrounds, with even the largest contributions from $Z \to b\bar{b}$ amounting to only a few percent of the total background in specific regions.

The m_H versus m_S distributions for $Z \to b\bar{b}$, shown in Figure 7.17, further confirm that no significant resonant structure is present in the signal region. The smooth kinematic behavior of these backgrounds means they are readily absorbed into the overall background estimate as it interpolates from the control to the signal region.

Although $H \to b\bar{b}$ and $Z \to b\bar{b}$ processes are present, their contribution to the total background is negligible both in rate and in their impact on the discriminating variables. The data-driven background estimation method used in this analysis is well-equipped to model their residual effect without introducing bias or spurious structure in the search for $X \to SH \to b\bar{b}b\bar{b}$.

7.4 Background Estimation

The background estimation in the resolved analysis is performed using a fully data-driven generative modeling framework through the method of normalizing flows and Gaussian processes discussed in Section 6.4, leveraging the statistical power of data in the control regions

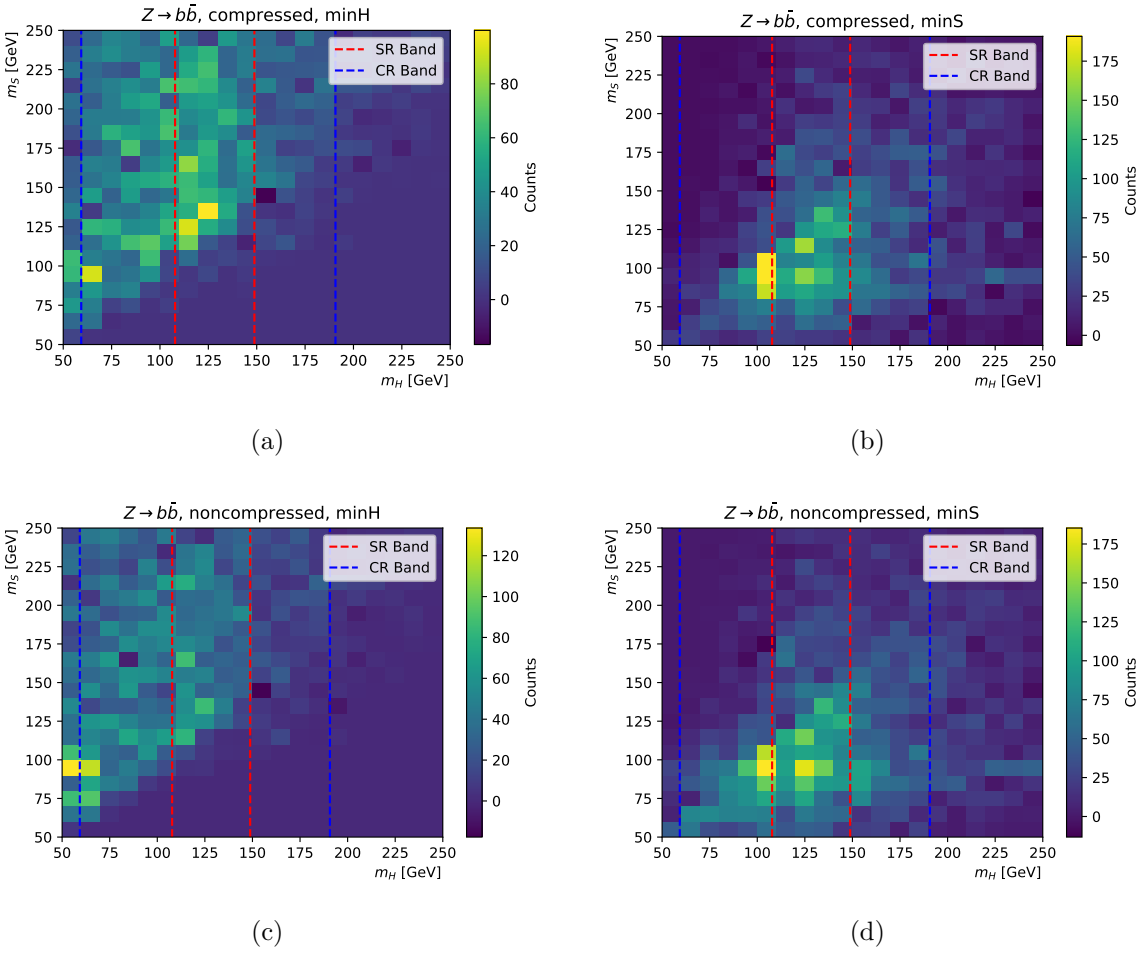


Figure 7.17: Distributions of m_H versus m_S for $Z \to b\bar{b}$ production, shown in the four BDT pairing configurations: (a) compressed/minH, (b) compressed/minS, (c) non-compressed/minH, and (d) non-compressed/minS. The signal band (encompassing all sub-signal regions) and control region are indicated. These distributions demonstrate that the $Z \to b\bar{b}$ background is smoothly distributed across the mass plane and does not introduce any problematic resonance within the signal region.

Table 7.5: Signal band yields for the ggF $H \to b\bar{b}$, VBF $H \to b\bar{b}$, and $Z \to b\bar{b}$ processes, shown for different event categories and years. Uncertainties are statistical only.

Process	**Pairing**	**Selection**	**2016**	**2017**	**2018**	**Total**
ggF $H \to b\bar{b}$	Compressed	minH	23.0 ± 3.69	33.32 ± 8.93	50.96 ± 13.83	107.28 ± 16.86
	Compressed	minS	21.14 ± 3.30	33.70 ± 9.08	48.93 ± 13.27	103.77 ± 16.41
	Non-compressed	minH	12.50 ± 2.01	17.68 ± 4.56	33.51 ± 8.88	63.69 ± 10.18
	Non-compressed	minS	17.69 ± 2.78	31.05 ± 8.25	43.23 ± 11.78	91.97 ± 14.65
VBF $H \to b\bar{b}$	Compressed	minH	2.14 ± 0.06	3.57 ± 0.15	5.39 ± 0.30	11.10 ± 0.34
	Compressed	minS	1.92 ± 0.05	3.09 ± 0.12	3.95 ± 0.22	8.96 ± 0.26
	Non-compressed	minH	1.47 ± 0.04	2.55 ± 0.11	4.08 ± 0.23	8.10 ± 0.26
	Non-compressed	minS	1.71 ± 0.05	2.94 ± 0.12	4.64 ± 0.26	9.29 ± 0.29
$Z \to b\bar{b}$	Compressed	minH	886.9 ± 58.2	1460.9 ± 84.5	2029.1 ± 108.5	4376.9 ± 147.9
	Compressed	minS	903.9 ± 64.3	1652.3 ± 98.5	2240.2 ± 117.4	4796.4 ± 166.2
	Non-compressed	minH	742.2 ± 55.4	1151.7 ± 77.1	1911.4 ± 107.4	3805.3 ± 143.3
	Non-compressed	minS	905.2 ± 62.7	1505.2 ± 95.8	2196.1 ± 116.4	4606.5 ± 166.3

(CRs) and minimizing dependence on Monte Carlo simulation. This approach is essential given the complexity of QCD multi-jet backgrounds and the limitations of simulation in the $4b$ phase space.

The central objective is to predict the multivariate distribution of background events in the blinded signal region (SR), using only the observed data in the CRs. This is accomplished by training the normalizing flows to learn the joint distribution of relevant kinematic variables for the S and H candidates, conditioned on the Higgs candidate mass (m_H) and the data-taking year. The generative model thus captures both the event-level correlations and the year-to-year variations in detector performance.

The variables used in flow training, conditioning, and reconstruction are summarized in

Table 7.6. Specifically, the flows are trained to model

$$p_t(\vec{x} \mid \vec{y})$$

where $\vec{x}$ includes $p_{T,H}$, $p_{T,S}$, η_H, η_S, $\Delta\phi_{SH}$, m_S, and X_{Wt}, while $\vec{y}$ comprises m_H and the data-taking year. Event-level observables such as m_{SH} and $\Delta\eta_{SH}$ are reconstructed from the sampled flow outputs via four-vector addition.

Table 7.6: Kinematic variables used in normalizing flow training, conditioning, and downstream reconstruction for the background estimate.

Flow Training Variables	**Description**
$p_{T,H}$	Transverse momentum of Higgs candidate
$p_{T,S}$	Transverse momentum of scalar S candidate
η_H	Pseudorapidity of Higgs candidate
η_S	Pseudorapidity of scalar S candidate
$\Delta\phi_{SH}$	Difference in ϕ between S and H candidate
m_S	Scalar S candidate mass
X_{Wt}	Top veto variable
Flow Conditioning Variables	**Description**
m_H	Higgs candidate mass
Year	Data collection year
Reconstructed Variables	**Description**
m_{SH}	SH system mass
$p_{T,SH}$	SH system transverse momentum
$\Delta\eta_{SH}$	Difference in η between S and H candidate

The normalizing flows utilize rational quadratic spline coupling layers [148]. The hyperparameters shown in Table 7.7 define the flexibility and training configuration of the model. Specifically, the number of transformation layers (L) and residual blocks per layer (n_b) control the depth and expressivity of the network, while the number of hidden nodes

per residual block (H) determines the network's capacity for learning non-linear features. The number of spline bins (K) sets the granularity of the spline used for each coupling transformation. The learning rate (ϵ) is chosen to ensure stable convergence and the dropout rate (p) helps prevent overfitting. These hyperparameters were optimized to provide robust modeling and reliable background predictions across both signal and control regions.

Table 7.7: Hyperparameters and configuration settings used to train the normalizing flows.

Parameter	**Description**	**Value**
L	Number of transformation layers	10
n_b	Number of residual blocks per layer	1
H	Number of hidden nodes per residual block	32
K	Number of spline bins	4
ϵ	Learning rate	0.001
p	Dropout rate	0.1

To quantify the modeling uncertainty due to neural network initialization and stochastic training, an ensemble of 25 normalizing flows is independently trained. The final background prediction is taken as the ensemble mean, with the spread (standard deviation) representing the associated uncertainty in relevant plots, such as the Figures 7.18 and 7.19 below which demonstrate the modeling of the normalizing flows in the CR of the compressed and non-compressed BDT pairing with the minH assignment.

7.4.1 Signal Region Prediction Using Gaussian Processes

As the normalizing flows are conditioned on m_H, an accurate prediction of the m_H distribution in the SR is required. This is provided by a Gaussian process (GP) regression, which is trained on the binned m_H distribution observed in the CR for each category and pairing (compressed/non-compressed, minH/minS, and year). The GP, defined as $f(x) \sim \mathcal{GP}(0, k(x, x'))$ with an RBF kernel (see Eq. 6.4 in Chapter 6), yields a central value and associated uncertainty band for the m_H spectrum in the blinded region.

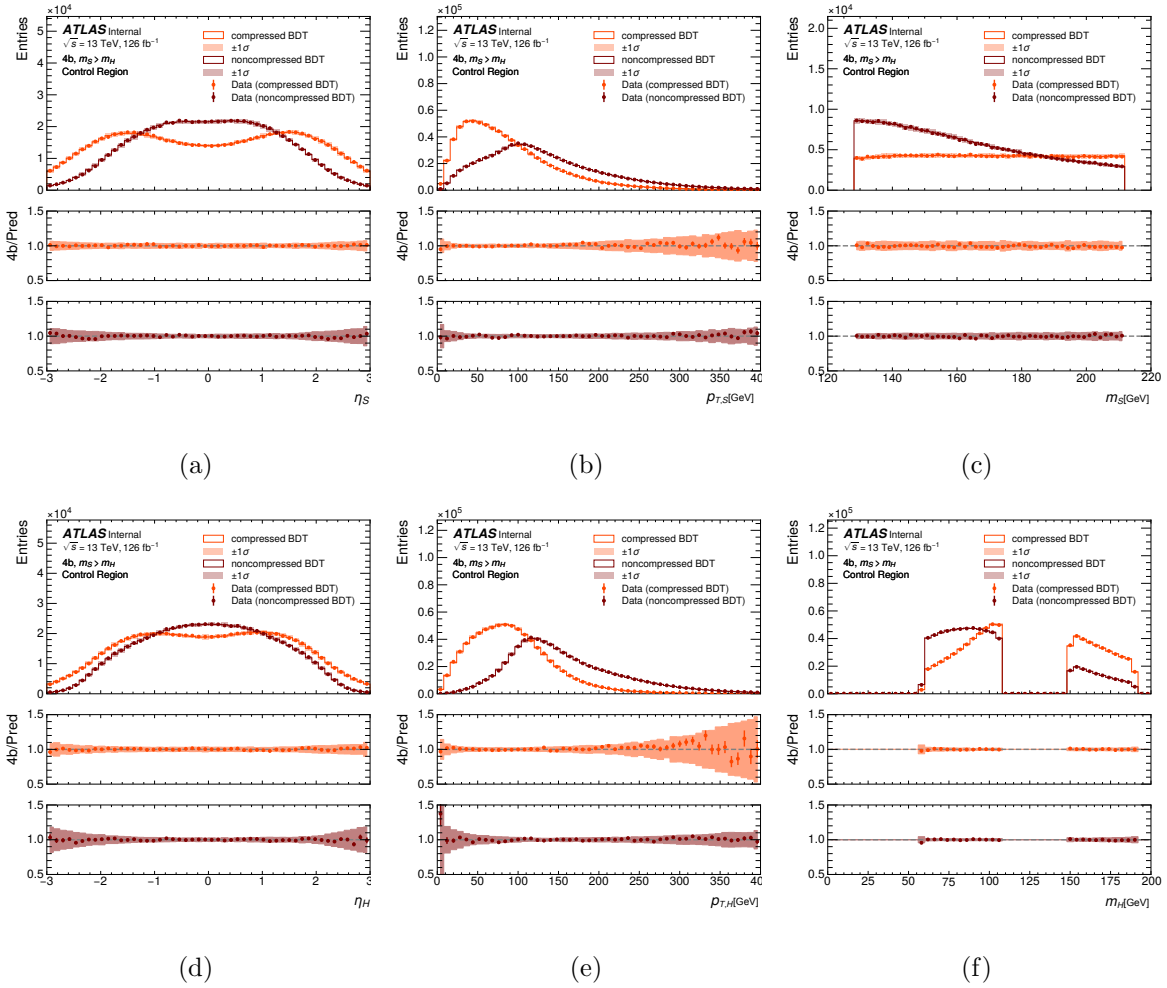


Figure 7.18: Distributions of kinematic variables of the S and Higgs candidate bosons for the predictions from the normalizing flows in the Control Region for the compressed and non-compressed BDT pairings when $m_S > m_H$.

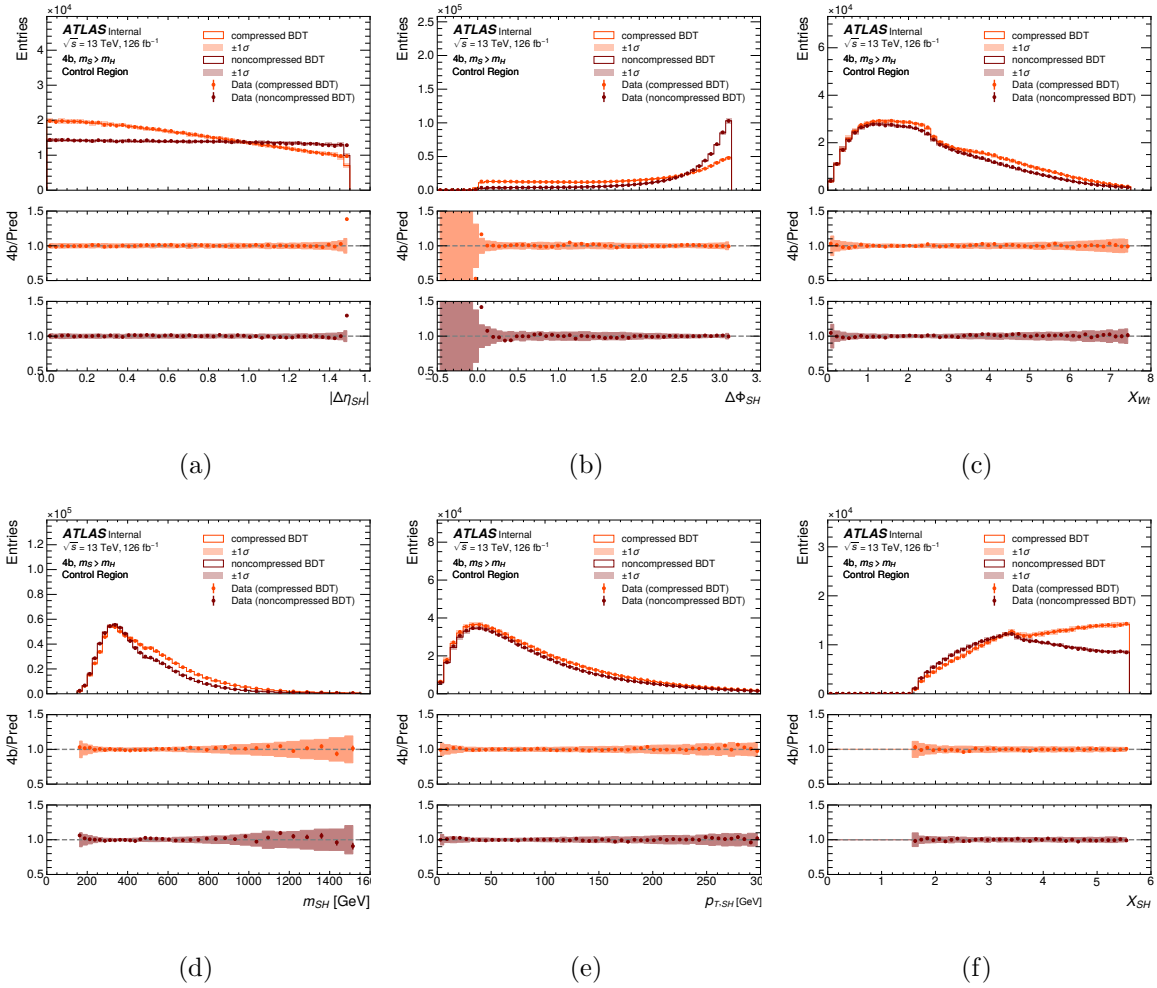
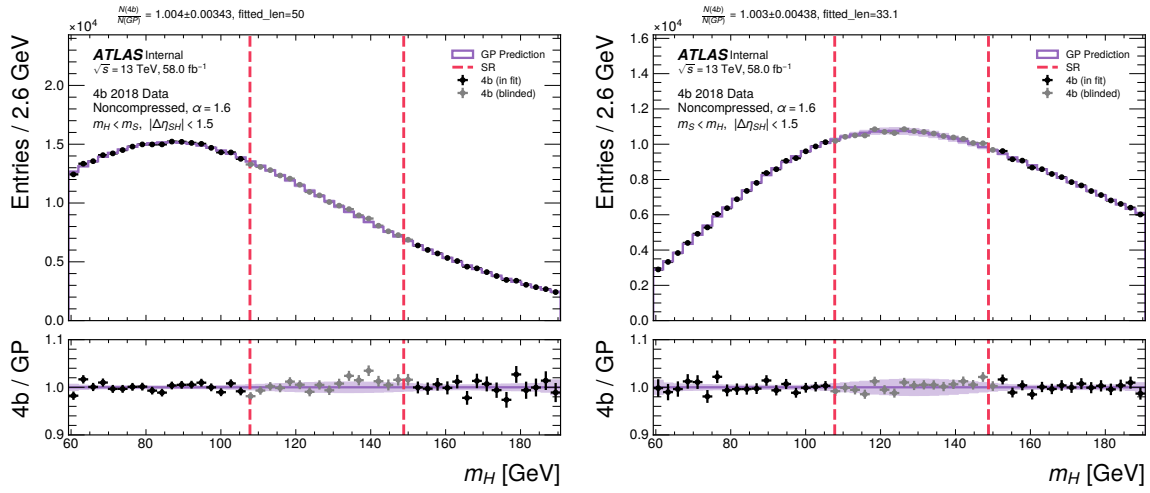


Figure 7.19: Distributions of kinematic variables of the SH candidates and event level variables X_{SH} and X_{Wt} for the predictions from the normalizing flows in the Control Region for the compressed and non-compressed BDT pairings when $m_S > m_H$.

The fitted GP is evaluated within the boundaries of the SR, providing the predicted m_H normalization and shape. This m_H spectrum is then sampled as input to the normalizing flow, enabling generation of background events in the SR with both correct normalization and associated uncertainty. Figure 7.20 shows example GP fits for the non-compressed and compressed minH pairings in 2018 4b data. The band reflects the uncertainty from the spread in the distribution of the 25 predictions. The normalization of the background in the SR is determined directly from the integral of the GP prediction in the blinded region.



(a) (b)

Figure 7.20: Gaussian process regression for the Higgs candidate mass (m_H) distribution in the (a) non-compressed minH and (b) non-compressed minS selection for the nominal $4b$ data in 2018. Points show data in the CR and SR, while the shaded band show the GP prediction and its uncertainty.

In summary, the background estimate in the $X \to SH \to b\bar{b}b\bar{b}$ resolved analysis proceeds as follows:

1. Train conditional normalizing flows on CR data from the sidebands of the m_H distribution to model the joint kinematics of S and H, conditioned on m_H and year.

2. Train a Gaussian process on the m_H distribution in the CR, and use it to predict the m_H distribution in the blinded SR.

3. Generate high-statistics synthetic background events in the SR by sampling from the trained flows while using the GP prediction for m_H.

4. Assign systematic and statistical uncertainties from the ensemble, the GP regression, and normalization, and propagate them into the statistical analysis.

This approach provides a robust, flexible, and high-precision background estimate anchored in data, well-suited to the high-dimensional, limited phase space. Further details on the systematics, performance, and closure of this method are presented in the following sections.

7.5 Background Systematics

Accurately quantifying uncertainties in the background estimate is essential to the integrity of this search. The main sources of systematic uncertainty in the data-driven background prediction arise from: (1) the modeling flexibility and initialization of the neural network architectures (normalizing flows and Gaussian processes) used for background estimation, and (2) residual mismodeling, or non-closure, observed in data control samples.

7.5.1 Shape Uncertainties from Deep Ensembles

As the background estimation in this analysis leverages two neural network models (normalizing flows and GPs), systematic uncertainties arising from both models must be rigorously propagated to the final background prediction. To achieve this, a deep ensemble approach is employed [149, 150]. Twenty five independent normalizing flows are trained, each conditioned on a random realization of the GP m_H spectrum for the signal region. This ensures that both the stochastic fluctuations in the neural network training (initialization, optimization) and the statistical uncertainty from the GP background shape are consistently propagated to the flow predictions. The resulting ensemble of m_{SH} predictions for a rep-

resentative non-compressed signal point, $m_X = 400$ GeV, $m_S = 200$ GeV, is illustrated in Fig. 7.21.

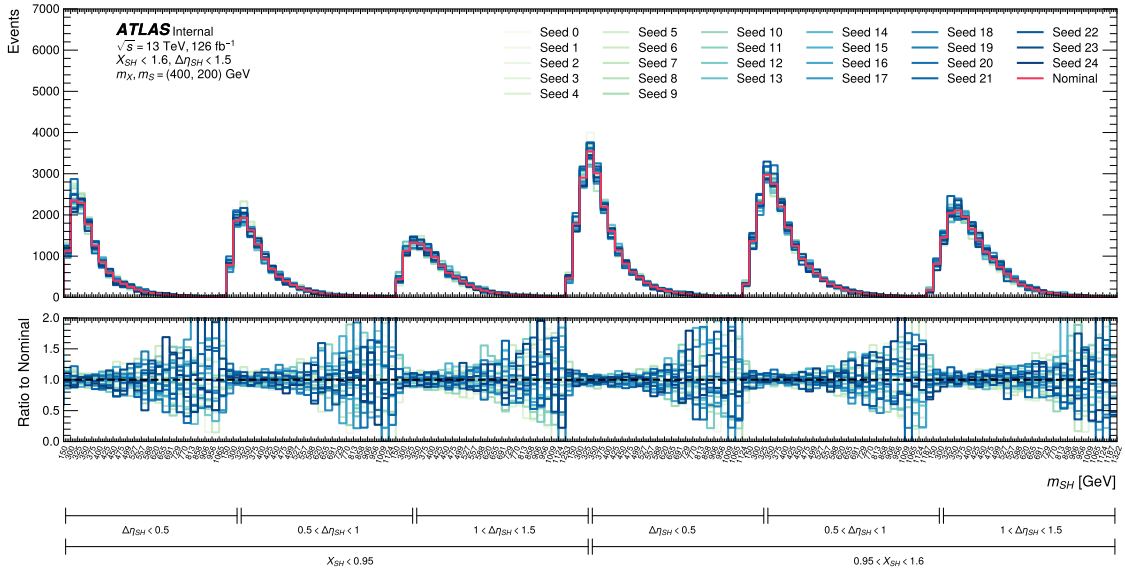


Figure 7.21: Twenty five predictions of the categorized m_{SH} distributions from the normalizing flows, each conditioned on a different GP sample for the non-compressed signal region ($m_X = 400$ GeV, $m_S = 200$ GeV). The average prediction is shown in red. See the SH4b Internal Note [145].

To quantify the correlated bin-by-bin uncertainties for use in the statistical analysis, a principal component analysis (PCA) [151] is performed on the ensemble predictions. The first step is to construct the sample covariance matrix $\mathbf{Q}$ from the normalized binned predictions of the ensemble:

$$Q_{jk} = \frac{1}{n+1} \sum_{i=1}^{n} \left(b_{ij} - \bar{b}_j\right) \left(b_{ik} - \bar{b}_k\right), \tag{7.4}$$

where b_{ij} is the value in the j-th bin of the i-th normalized ensemble prediction, and $\bar{b}_j$ is the mean over all n predictions in bin j.

The next step is to diagonalize $\mathbf{Q}$ by eigen-decomposition, yielding eigenvalues λ and corresponding eigenvectors v_λ, which represent the principal components (directions of max-

imum variance in the prediction space):

$$\mathbf{Q} = \mathbf{PDP}^T, \tag{7.5}$$

$$\mathbf{D} = \mathrm{diag}(\lambda_1, ..., \lambda_m), \tag{7.6}$$

where the columns of $\mathbf{P}$ are the principal component eigenvectors v_λ.

The principal components with the largest eigenvalues represent the dominant shape variations, while those with small eigenvalues are treated as statistical noise. As only a limited number of principal components have corresponding eigenvalues significantly larger than zero, we retain only a subset of the eigenvariations to capture the most significant uncertainty structure, with the remaining components treated as noise [151]. To determine the optimal number of principal components to use as systematic shape uncertainties, an autocorrelation-based approach is employed. Autocorrelation quantifies the similarity between a series and a lagged version of itself. For each eigenvector, the autocorrelation function is computed, and the resulting correlation coefficients are combined in quadrature to yield a single scalar metric. This metric serves as an indicator of whether the eigenvector exhibits systematic patterns or behaves more like uncorrelated noise. A threshold criterion is applied to classify the eigenvectors: those with autocorrelation sums exceeding a certain threshold are designated as systematic shape uncertainties, while those below the threshold are treated as uncorrelated uncertainties. This method ensures that the number of shape uncertainties retained is data-driven and avoids overfitting the background model to random fluctuations.

The systematic uncertainty variation for each principal component is constructed as

$$s_\lambda = \mathcal{N}\sqrt{\lambda}\, v_\lambda, \tag{7.7}$$

where $\mathcal{N}$ is the ensemble normalization factor, λ is the eigenvalue, and v_λ is the corresponding eigenvector.

Figure 7.22 shows the eigenvalue spectrum, while Figure 7.23 displays the principal component shape variations for the m_{SH} distribution for the SR of $m_X = 400$ GeV, $m_S = 200$ GeV. These principal component variations are implemented as correlated shape

systematics in the statistical model, ensuring that the fit is sensitive to the true uncertainty structure of the flow and GP-based background estimate.

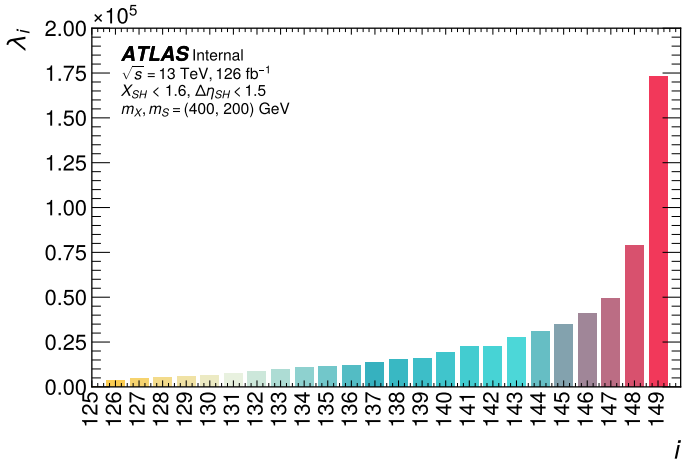


Figure 7.22: Eigenvalues from the principal component analysis of the ensemble background prediction for the non-compressed signal region (m_X = 400 GeV, m_S = 200 GeV, 2016–2018 data). The largest eigenvalues correspond to the retained shape systematics. See the SH4b Internal Note [145].

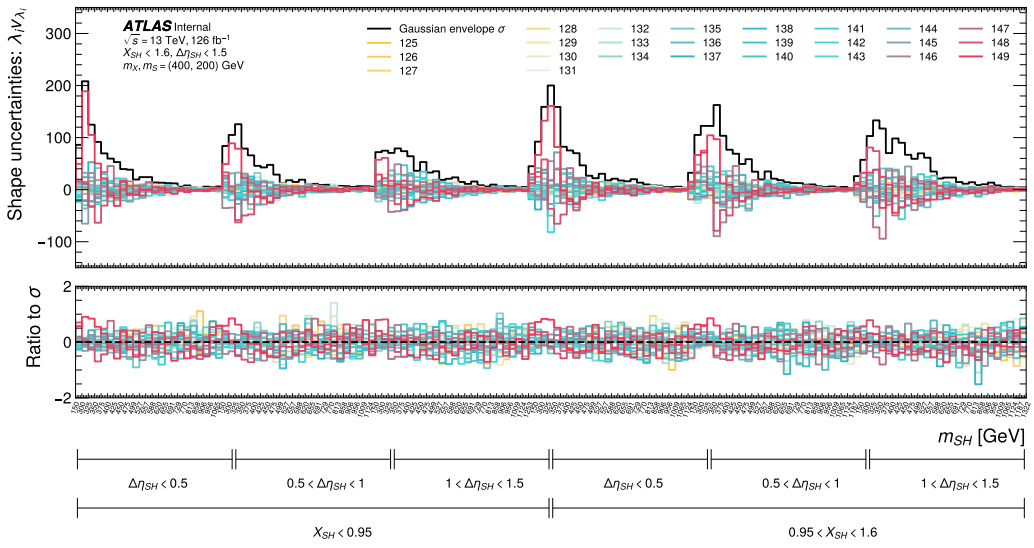


Figure 7.23: Leading principal component variations in the m_{SH} distribution from the deep ensemble for the non-compressed region ($m_X = 400$ GeV, $m_S = 200$ GeV). The shaded envelope indicates the standard deviation across the 25 ensemble predictions in each bin. See the SH4b Internal Note [145].

7.5.2 *Background Non-Closure Uncertainty*

To further guard against possible mismodeling of the background, an additional non-closure uncertainty is included. This is derived by comparing the background prediction to observed data in the $3b$ kinematic region, which serves as an independent validation sample not used in the fit. For each m_{SH} bin, the non-closure uncertainty is calculated as

$$3b \text{ non-closure uncertainty}_i = \frac{\text{obs}_{3b,i} - \text{pred}_{3b,i}}{n_{3b,\,\text{tot}}} \times n_{4b,\,\text{tot}}, \tag{7.8}$$

where $\text{obs}_{3b,i}$ and $\text{pred}_{3b,i}$ are the observed and predicted counts in bin i of the $3b$ signal region, and $n_{3b,\,\text{tot}}$, $n_{4b,\,\text{tot}}$ are the total observed event counts in the $3b$ and $4b$ signal regions, respectively. This scaling ensures that the uncertainty is normalized to the $4b$ yield.

This bin-by-bin uncertainty accounts for shape differences between observed and predicted data, while the non-closure in total yield between the data and prediction is taken as a normalization uncertainty. Figure 7.24 shows the resulting non-closure shape uncertainty for a representative non-compressed BDT pairing of $m_X = 400$ GeV, $m_S = 200$ GeV.

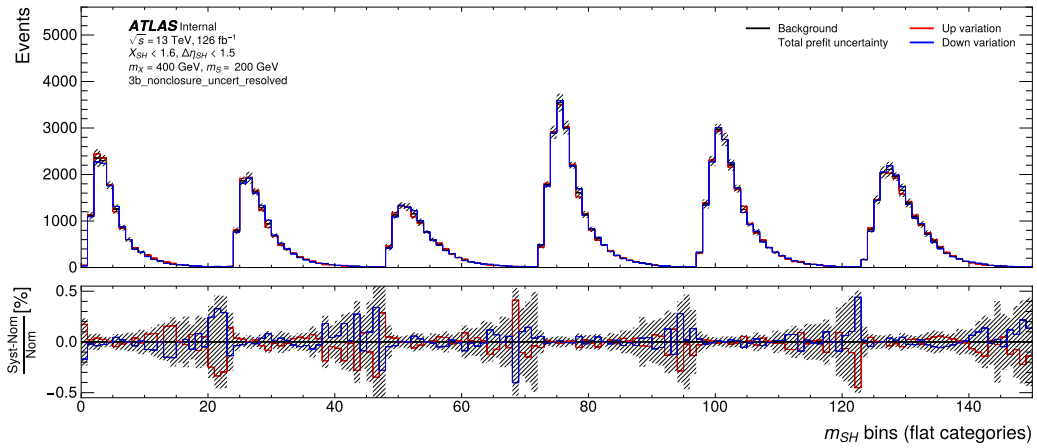


Figure 7.24: The $3b$ non-closure uncertainty as a function of m_{SH}, derived for the non-compressed BDT pairing when $m_X = 400$ GeV, $m_S = 200$ GeV. The up and down variations from the non-closure is shown in red and blue respectively. See the SH4b Internal Note [145].

Thus, the complete background systematics incorporated in the analysis include:

- Shape uncertainties derived from principal component variations in the deep ensemble of neural network background predictions, propagated for both compressed and non-compressed regions.

- An additional non-closure uncertainty, derived from direct comparison of prediction and data in the $3b$ validation region, included bin-by-bin in the m_{SH} distribution.

Together, these uncertainties provide a comprehensive and data-driven treatment of potential background mismodeling, ensuring the robustness of the statistical interpretation of the results.

7.5.3 Background Validation Studies

The background estimation strategy in this analysis is fully data-driven, relying on generative models trained on control region data. As a result, direct validation of the method within the blinded signal region is not possible prior to unblinding. Instead, the accuracy and robustness of the background prediction, including the assigned systematic uncertainties, are tested using dedicated validation regions constructed by modifying the nominal event selection.

Two principal validation regions are defined, as shown in Figure 7.25:

- Events with $|\Delta\eta_{SH}| > 1.5$ (inverted kinematic selection).

- Events with exactly 3 b-tagged jets, where the fourth jet is the highest p_T jet that fails the b-tagging requirement.

Each validation region contains its own pseudo-control and pseudo-signal regions, defined analogously to those in the nominal analysis, to facilitate direct testing of the modeling approach. The validation region found to be most representative of the nominal signal region is the $3b$ region with $|\Delta\eta_{SH}| < 1.5$, as it most closely matches the kinematic properties of the $4b$ selection used in the analysis. In the $3b$ validation region, events are required to have exactly three jets passing the full selection and b-tagging requirements, while the fourth jet is selected as the highest p_T jet that fails the same b-tagging criteria. This selection

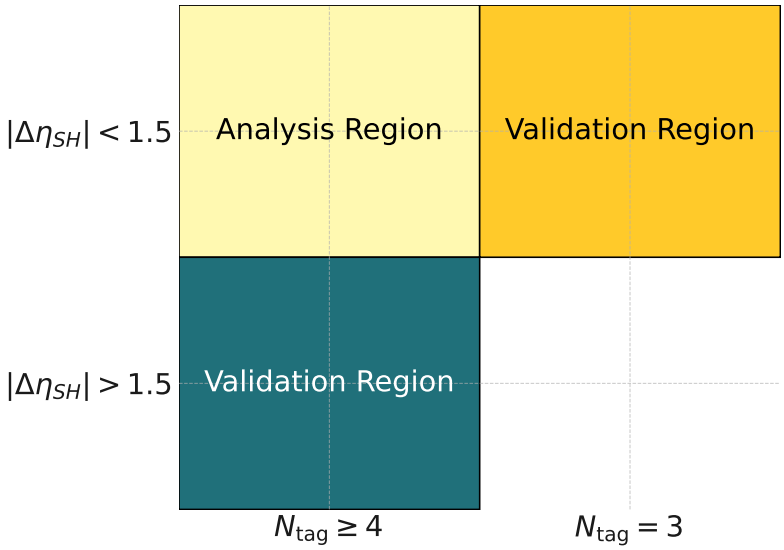


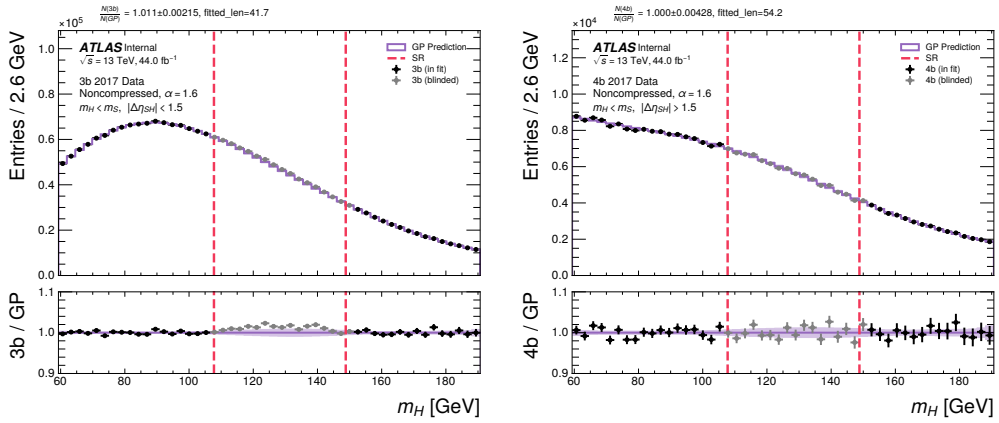
Figure 7.25: Definition of the validation regions used to test the data-driven background estimate. These regions are obtained by inverting kinematic cuts or altering the b-tag multiplicity from the nominal selection.

provides a statistically independent sample for testing the performance of the background estimation framework.

The predictive power of the background model is validated by comparing the predicted and observed distributions for the discriminant variable, m_{SH}, in the pseudo-signal regions of these validation samples. Where discrepancies between the prediction and the observed data are identified, a non-closure systematic uncertainty is derived and included in the final fit, as discussed in Section 7.5.2. This ensures that any limitations in the modeling are propagated to the main result. The total systematic uncertainty assigned to the background prediction in the signal region therefore includes both the statistical uncertainty from the modeling ensemble and the non-closure uncertainty derived from these validation studies, providing a conservative and data-driven estimate of the background modeling uncertainty.

GP Fit Performance and Validation

The performance of the Gaussian process (GP) regression used to predict the m_H spectrum in the blinded signal region is validated in both $3b$ and $4b$ samples. Figure 7.26 shows side-by-side comparisons of the GP fit for the m_H distribution in the $3b$ and $4b$ validation regions for the non-compressed pairing with $m_H < m_S$ in the 2017 data. This comparison demonstrates that the GP regression adequately models the shape and normalization of the m_H distribution, with any non-closure manifest in these plots and subsequently covered by the assigned systematic uncertainties.



(a) (b)

Figure 7.26: GP fits to the m_H distribution in the blinded pseudo-signal region for: (a) $3b, |\Delta\eta_{SH}| < 1.5$ and (b) $4b, |\Delta\eta_{SH}| > 1.5$ events, both in the non-compressed region ($m_H < m_S$) for 2017 data. Non-closure observed in the $3b$ region is used to define the associated systematic uncertainty.

Background Validation in 3b Pseudo-signal Regions

Figure 7.27 displays the m_{SH} distribution in the $3b$ signal region for representative signal mass points where $m_S = 250$ GeV and all six analysis categories for both non-compressed and compressed pairings. Across all regions, the data-driven prediction generally describes

the observed data within the assigned uncertainties. Where non-closure occurs, the derived non-closure systematic uncertainty ensures coverage in the statistical interpretation.

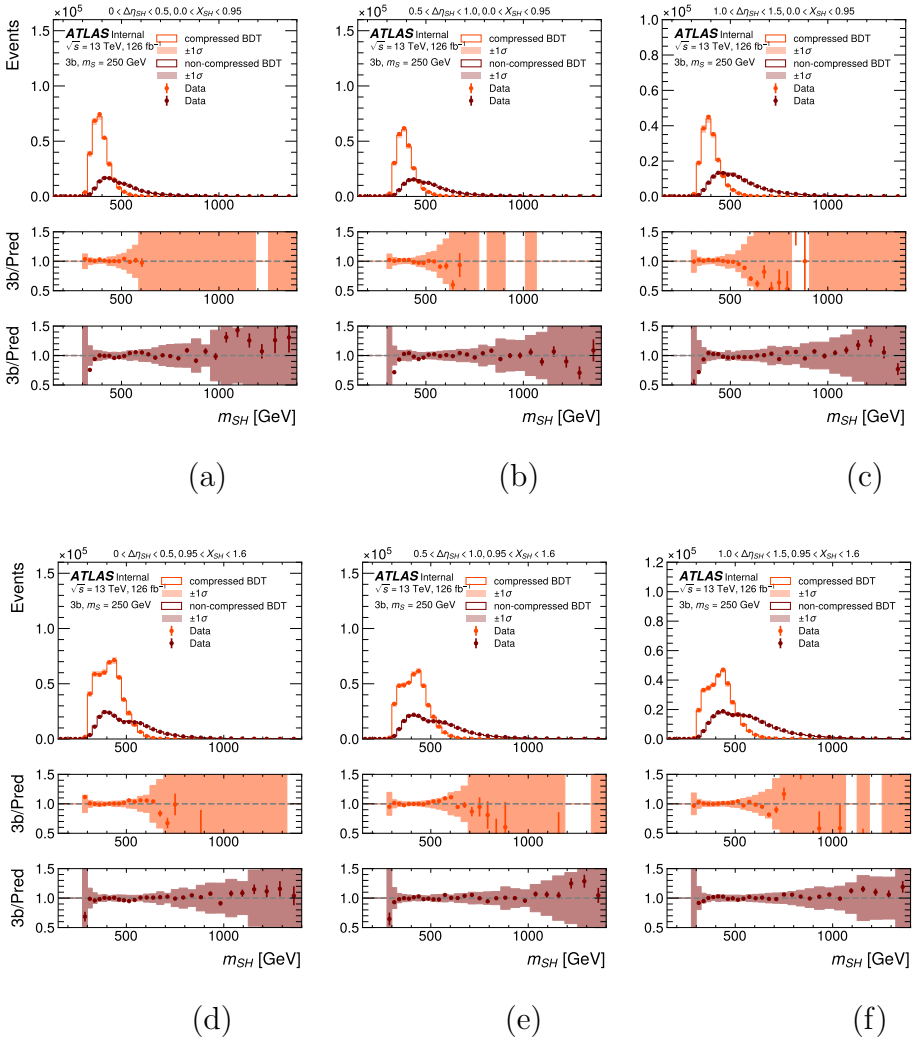


Figure 7.27: The categorized m_{SH} distribution in the $3b$ signal region for $m_S = 250$ GeV. The six subplots correspond to: (a) Category 0: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.0 < X_{SH} < 0.95$, (b) Category 1: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.0 < X_{SH} < 0.95$, (c) Category 2: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.0 < X_{SH} < 0.95$, (d) Category 3: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.95 < X_{SH} < 1.6$, (e) Category 4: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.95 < X_{SH} < 1.6$, and (f) Category 5: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.95 < X_{SH} < 1.6$.

7.6 Results

7.6.1 m_{SH} Distributions

The m_{SH} variable, categorized in $\Delta\eta_{SH}$ and X_{SH}, serves as the primary observable for statistical interpretation. Initial fixed-width and logarithmic binning resulted in many bins with low or zero expected background, especially at high masses, which led to instability in the statistical fit.

To mitigate this, an adaptive rebinning algorithm was applied to each signal region and category: adjacent bins are merged until each contains at least 10 predicted background events. This approach maintains statistical robustness and preserves resolution where statistics are sufficient. In regions with very low background, a lower threshold (e.g., 3 or 4 events) was tested but a minimum of 10 events per bin was chosen as the default for uniformity and fit stability. This strategy optimizes sensitivity to resonant signals while ensuring reliable signal and background extraction across all kinematic regimes.

A background-only profile likelihood fit is run for the m_{SH} distribution in the categorized signal region and results are shown in Figure 7.28. 4b data yields, estimated background, and signal event yields are extracted in the sub-SR for each signal mass hypothesis. The results of the yield comparisons are shown in Table 7.8. Note that the plots and tables show the sum across all years, but the signal extraction fit and background estimate are run for the years separately.

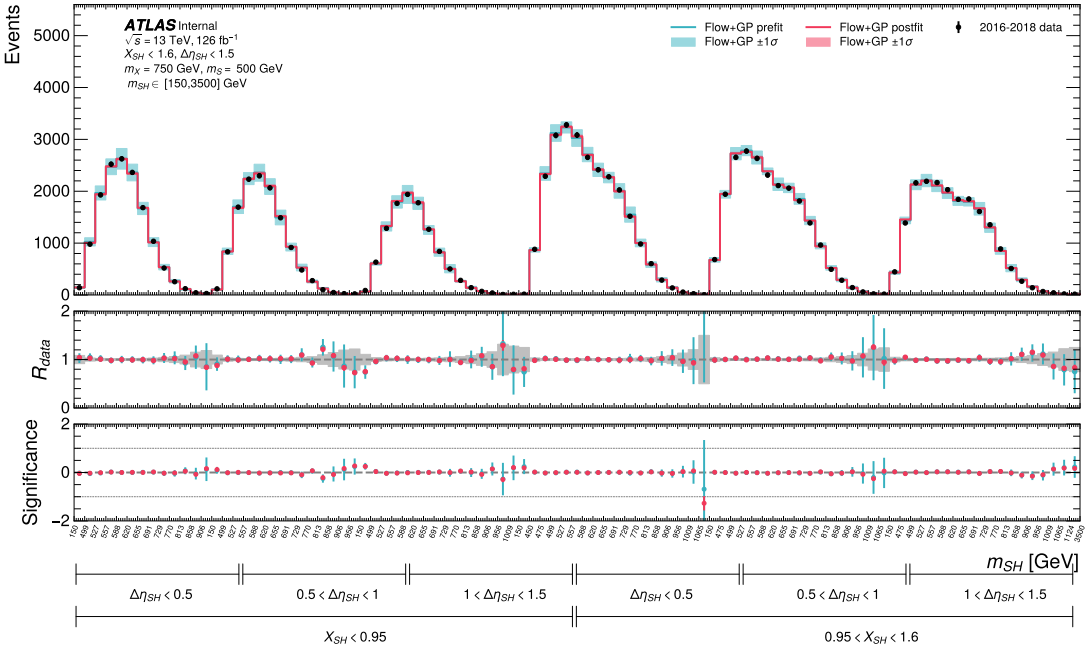
Table 7.8: Observed, background model, and signal model yields in the signal region, summed over 2016–2018, for each simulated mass point assuming a production cross section of 1 fb before the statistical fitting. The background model uncertainties correspond to the standard deviation of the 25 predictions, while the signal uncertainties are statistical only.

m_X [GeV]	m_S [GeV]	Data	Background Model	Signal Model
300	70	65314	64500 ± 1100	0.88 ± 0.94
300	100	88027	88100 ± 1300	0.95 ± 0.98
300	170	79281	79100 ± 1900	2.0 ± 1.4

m_X [GeV]	m_S [GeV]	Data	Background Model	Signal Model
400	70	65314	64500 ± 1100	2.2 ± 1.5
400	100	88027	88100 ± 1300	2.4 ± 1.6
400	170	92555	90330 ± 850	2.5 ± 1.6
400	200	90042	88660 ± 940	2.2 ± 1.5
400	250	233655	231100 ± 2200	4.0 ± 2.0
750	70	65314	64500 ± 1100	5.4 ± 2.3
750	100	88027	88100 ± 1300	9.5 ± 3.1
750	170	92555	90330 ± 850	9.1 ± 3.0
750	200	90042	88660 ± 940	9.6 ± 3.1
750	250	109846	109300 ± 1300	12.4 ± 3.5
750	300	66803	67200 ± 1100	12.0 ± 3.5
750	400	34080	34360 ± 870	11.2 ± 3.3
750	500	106464	104800 ± 2000	9.6 ± 3.1
1000	70	65314	64500 ± 1100	0.88 ± 0.94
1000	100	88027	88100 ± 1300	7.4 ± 3.0
1000	170	92555	90330 ± 850	12.9 ± 3.5
1000	200	90042	88660 ± 940	12.8 ± 3.6
1000	250	109846	109300 ± 1300	16.9 ± 4.1
1000	300	66803	67200 ± 1100	16.2 ± 4.2
1000	400	34080	34360 ± 870	16.0 ± 4.1
1000	500	20450	20110 ± 530	13.0 ± 3.9
1000	750	37975	37700 ± 1000	9.6 ± 3.0
1500	300	66803	67200 ± 1100	9.7 ± 3.2
1500	400	34080	34360 ± 870	10.8 ± 3.5
1500	500	20450	20110 ± 530	12.3 ± 3.6
1500	750	4094	4050 ± 130	14.0 ± 3.8
1500	1000	14459	14225 ± 590	12.7 ± 3.6
2000	300	66803	67200 ± 1100	2.74 ± 1.7

m_X [GeV]	m_S [GeV]	Data	Background Model	Signal Model
2000	400	34080	34360 ± 870	2.93 ± 1.9
2000	500	20450	20110 ± 530	3.19 ± 2.0
2000	750	4094	3950 ± 200	4.51 ± 2.1
2000	1000	1785	1619 ± 95	6.49 ± 2.7
2000	1500	2838	2877 ± 180	14.3 ± 3.9
2500	300	66803	67200 ± 1100	0.63 ± 0.79
2500	400	34080	34360 ± 870	0.89 ± 0.94
2500	500	20450	20110 ± 530	0.94 ± 0.97
2500	750	4094	3950 ± 200	1.2 ± 1.1
2500	1000	1785	1619 ± 95	1.6 ± 1.3
2500	1500	830	727 ± 59	4.2 ± 2.0
2500	2000	1296	1460 ± 130	14.3 ± 3.8
3000	300	66803	67200 ± 1100	0.11 ± 0.33
3000	400	34080	34360 ± 870	0.18 ± 0.43
3000	500	20450	20110 ± 530	0.20 ± 0.45
3000	750	4094	3950 ± 200	0.32 ± 0.57
3000	1000	1785	1619 ± 95	0.46 ± 0.67
3000	1500	830	727 ± 59	1.0 ± 1.0
3000	2000	476	488 ± 43	3.3 ± 1.8
3000	2500	414	528 ± 59	12.8 ± 3.6

Figure 7.29 shows the unblinded m_{SH} distribution in the 4b signal region for representative signal mass points with $m_S = 250$ GeV across all six analysis categories for both non-compressed and compressed pairings. The data-driven prediction is in good agreement with the observed data within the assigned uncertainties. Although not shown, an additional $3b$ non-closure systematic is included to ensure adequate coverage in the statistical interpretation.



(a)

Figure 7.28: Pre-fit and post-fit of a background-only profile likelihood fit for the sub-SR of $m_X = 400$ GeV, $m_S = 200$ GeV for the 4b nominal predictions and toy data. See the SH4b Internal Note [145].

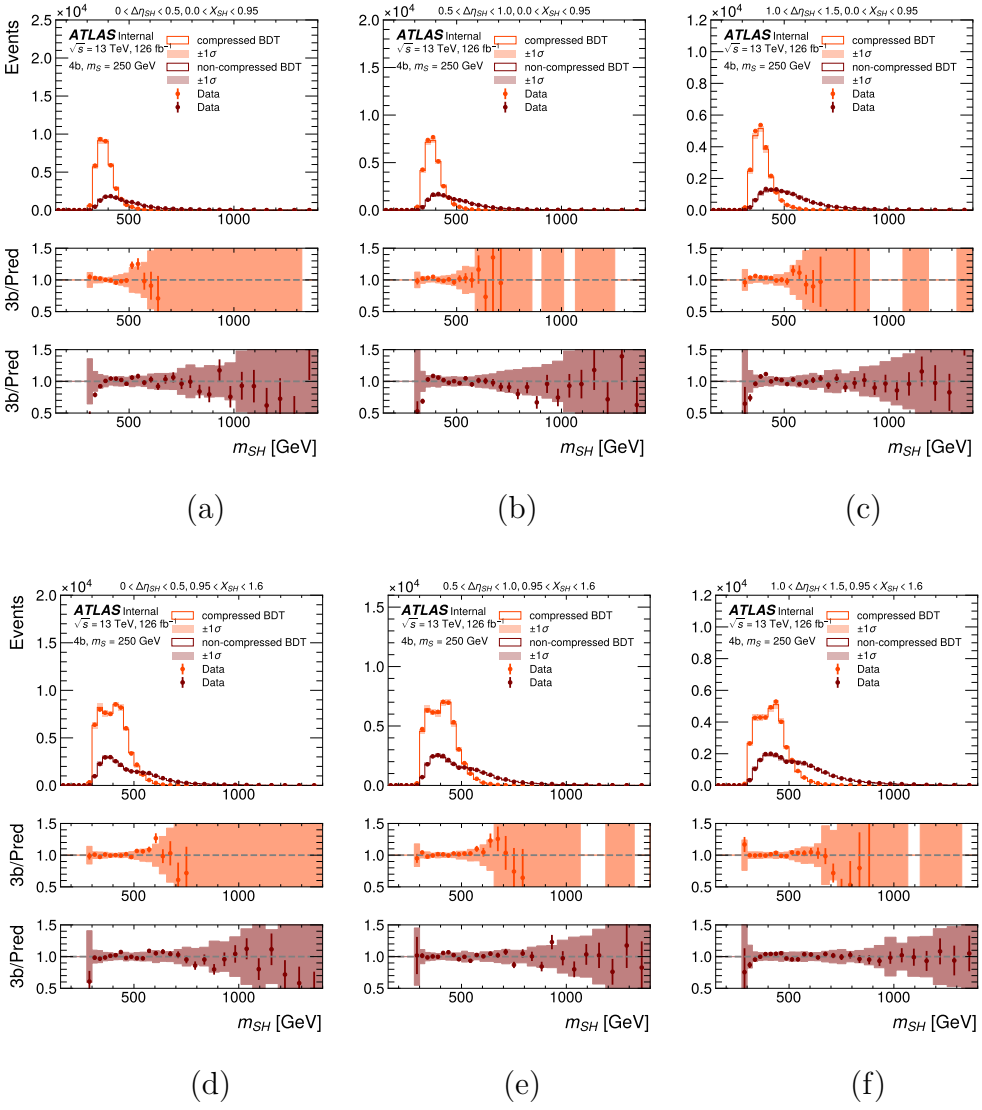


Figure 7.29: The categorized m_{SH} distribution in the unblinded $4b$ signal region for $m_S = 250$ GeV. The six subplots correspond to: (a) Category 0: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.0 < X_{SH} < 0.95$, (b) Category 1: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.0 < X_{SH} < 0.95$, (c) Category 2: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.0 < X_{SH} < 0.95$, (d) Category 3: $0.0 < |\Delta\eta_{SH}| < 0.5$, $0.95 < X_{SH} < 1.6$, (e) Category 4: $0.5 < |\Delta\eta_{SH}| < 1.0$, $0.95 < X_{SH} < 1.6$, and (f) Category 5: $1.0 < |\Delta\eta_{SH}| < 1.5$, $0.95 < X_{SH} < 1.6$.

7.6.2 *Confidence Level Limits, Nuisance Parameter Impacts, and Correlations*

The statistical interpretation of the analysis is performed through a combined signal-plus-background fit across all analysis categories. This fit fully incorporates the common and background-specific systematic uncertainties described in Sections 6.5.2 and 7.5, respectively. The primary result of this procedure is the set of expected 95% confidence level (CL) upper limits on the production cross section for the resonant $X \to SH \to b\bar{b}b\bar{b}$ process, derived for all tested signal mass hypotheses.

The fit utilizes a data-driven background estimate from the generative modeling framework and predicted signal distributions from simulation. Systematic uncertainties are included as nuisance parameters in the likelihood function, and their correlations and impacts are evaluated directly within the fit.

The key results are summarized as follows:

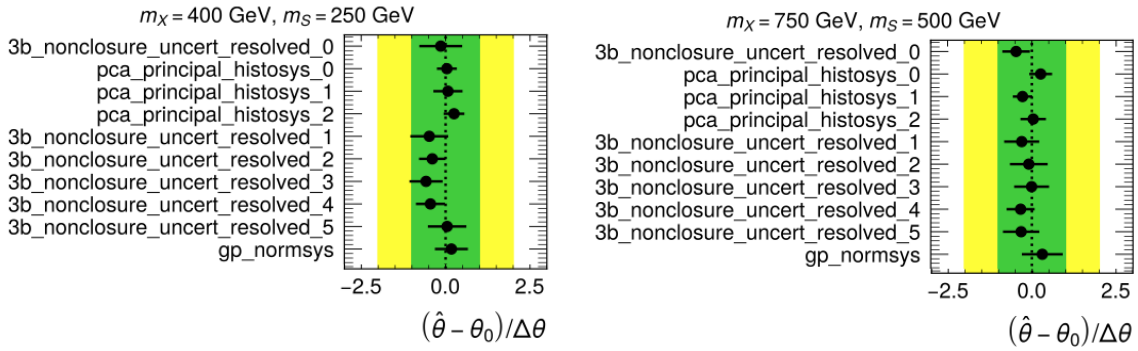
- **Expected and Observed Limits:** The expected 95% CL upper limits on the cross section times the branching ratio are derived for each signal mass point using the CL_s method. Figure 7.32 displays a heatmap of both the expected and observed limits for all scanned (m_X, m_S) mass pairs.

- **Uncertainty Bands:** Figure 7.33 illustrates the expected and observed limits for the signal hypotheses of the resolved channel. The inner (green) and outer (yellow) bands represent the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainties on the expected limit, respectively.

- **Nuisance Parameter Pulls:** The pulls for the dominant background nuisance parameters in the fit for the $(m_X, m_S) = (400\ \mathrm{GeV}, 250\ \mathrm{GeV})$ and $(m_X, m_S) = (750\ \mathrm{GeV}, 500\ \mathrm{GeV})$ hypotheses is shown in Figure 7.30. The pulls are a standardized measure of the difference between a nuisance parameter's fitted value and its a priori value, where large values indicate a significant shift in this parameter to achieve a good fit.

- **Nuisance Parameter Impacts:** The ranking plots of the impacts of nuisance parameters for the same mass points are given in Figure 7.31. This ranking highlights

the dominant uncertainties that most significantly influence the fit result.

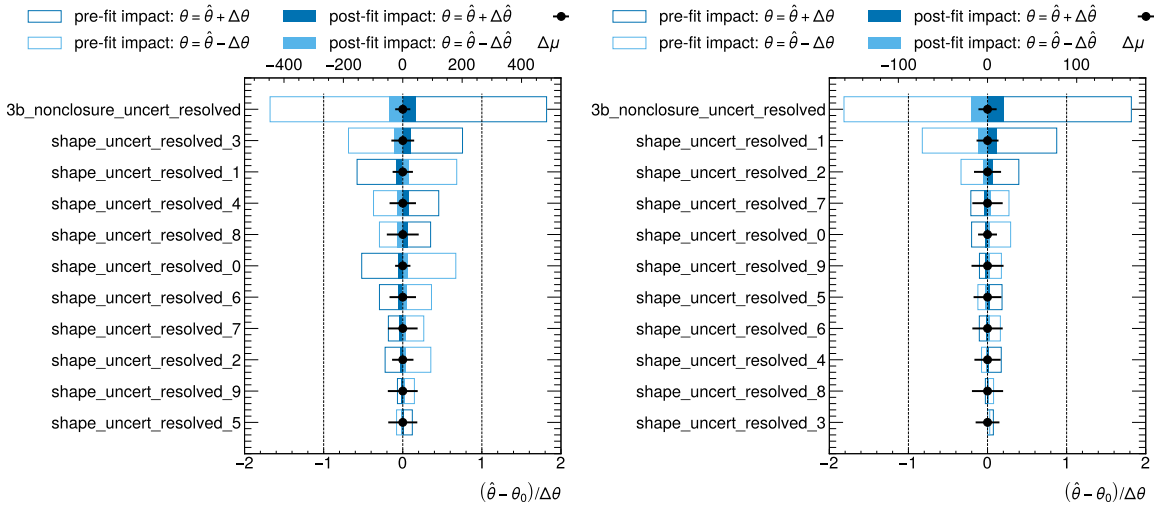


(a) (b)

Figure 7.30: The major background systematic nuisance parameter pulls for (a) $(m_X, m_S) = (400\text{ GeV}, 250\text{ GeV})$ and (b) $(m_X, m_S) = (750\text{ GeV}, 500\text{ GeV})$. These systematics nuisance parameters include the dominant shape systematics, the $3b$ non-closure uncertainty, and the normalization uncertainty from the GP regression.

The expected and observed 95% confidence level upper limits on the cross section times branching ratio for $X \to SH \to b\bar{b}b\bar{b}$ are displayed in Figure 7.32. In the resolved channel, the analysis uses a mass grid designed for interpolation across m_S, except for the range $100\text{ GeV} < m_S < 170\text{ GeV}$. This specific region is excluded because ATLAS has separate analyses dedicated to symmetric Higgs boson decays from heavy scalar resonances. In contrast, the mass grid does not provide comparable granularity for interpolation across m_X. The derived limits are fully model-independent and do not assume a specific branching ratio for $S \to b\bar{b}$.

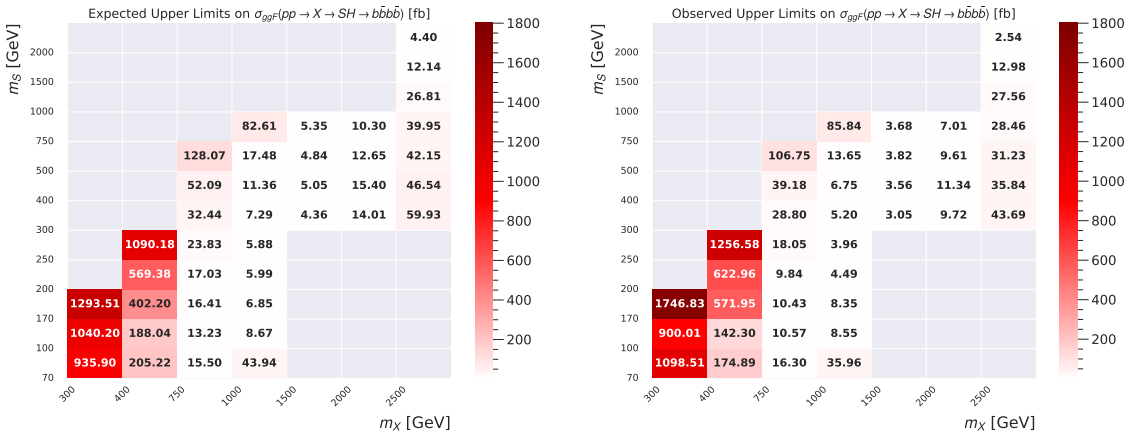
Ultimately, no significant excess was observed over the full phase space with the local significance for each signal hypothesis remaining below 2σ as shown in Figure 7.33. While the presented results do not include all uncertainties, the remaining signal and theoretical systematic uncertainties are anticipated to have a negligible impact, affecting the results by only a couple percent.



(a) (b)

Figure 7.31: The dominant pulls and impacts of the background systematic uncertainties, including the nominal shape systematic uncertainties with the $3b$ non-closure uncertainties, ranked by magnitude of post-fit impact and shown for (a) $(m_X, m_S) = (400\text{ GeV}, 250\text{ GeV})$ and (b) $(m_X, m_S) = (750\text{ GeV}, 500\text{ GeV})$. See the SH4b Internal Note [145].

In regards to published CMS results for the resolved channel, their largest excess was reported at $(m_X, m_S) = (700\text{ GeV}, 400\text{ GeV})$ [136]. While this analysis does not have an identical signal mass pair for a direct comparison, no significant excess is seen at the nearby mass point of $(m_X, m_S) = (750\text{ GeV}, 400\text{ GeV})$. Given the relative width of the m_X distribution for $m_X = 750$ GeV is around 53 GeV, the most conservative interpretation suggests that the excess seen by CMS may be a statistical fluctuation. To provide greater precision on this estimate, ATLAS is currently producing new derivations for the specific signal hypothesis at $(m_X, m_S) = (700\text{ GeV}, 400\text{ GeV})$, which will allow for a fair, one-to-one comparison with CMS.



(a) (b)

Figure 7.32: Expected (a) and observed (b) 95% CL upper limits on the cross section times the branching ratio for $X \to SH \to b\bar{b}b\bar{b}$ for the resolved signal mass pairs including all background systematic uncertainties.

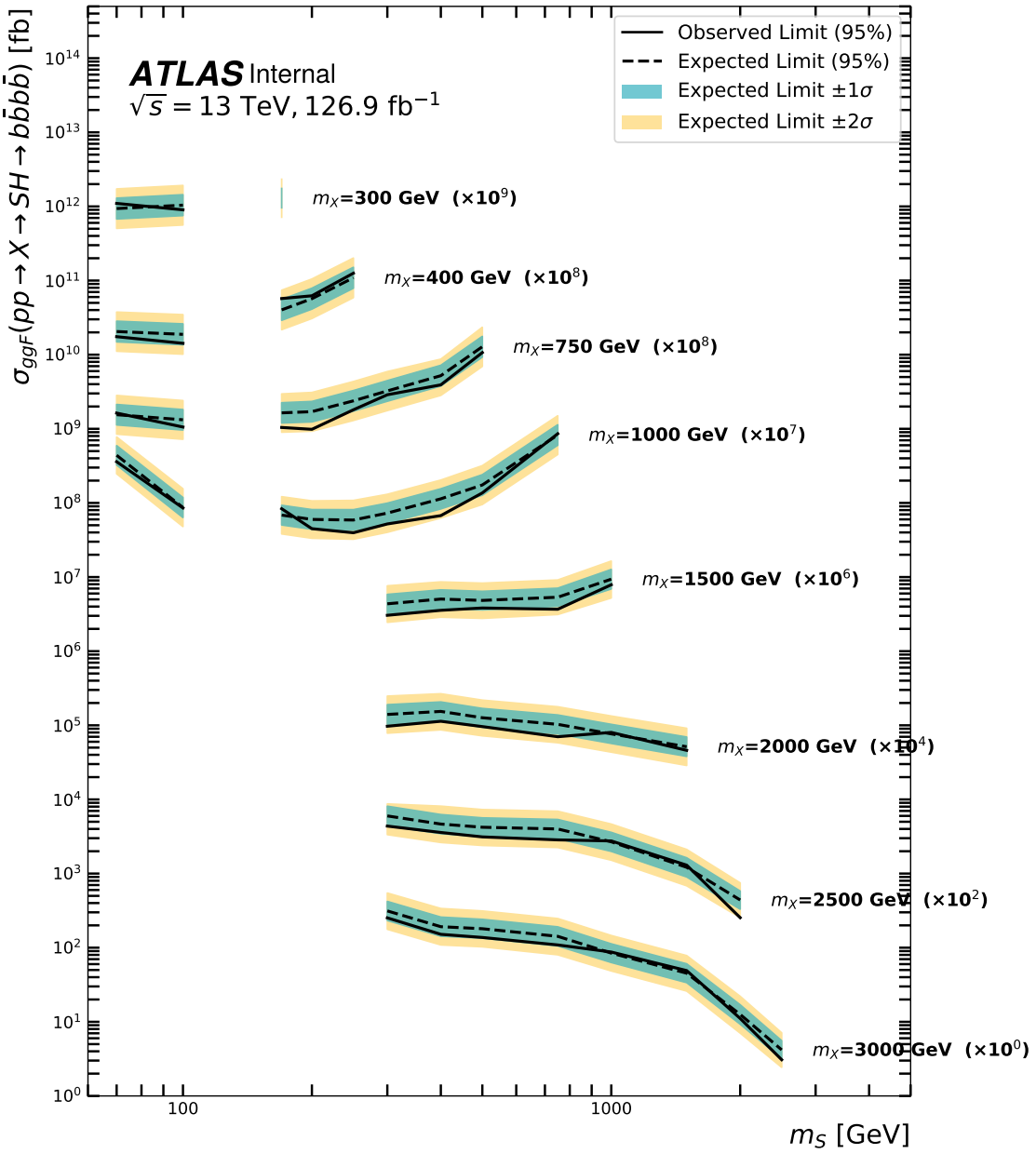


Figure 7.33: Observed (solid black line) and expected (dashed black line) 95% CL upper limits on the cross section times the branching ratio. For each unique m_X value, where a scaling factor is applied to the limits so that all mass pairs can be visualized together. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties, respectively. The mass grid used in this analysis allows for interpolation across m_S, except in regions between $m_S = 100$ GeV and $m_S = 170$ GeV.

7.6.3 Interpretation within the TRSM

The True Real Singlet Model (TRSM) is discussed in Section 5.2.2 as a straightforward extension of the Standard Model that introduces an additional scalar singlet. Using the package developed in [152], values for the cross section of the $X \to SH \to b\bar{b}b\bar{b}$ process were calculated and are shown in Table 7.9 for signal samples with $m_X \leq 2500$ GeV and relative widths determined from the mass distributions presented in this analysis.

Table 7.9: Cross sections (fb) for the $X \to SH \to b\bar{b}b\bar{b}$ process within the TRSM. The values are derived by using the relative widths of the scalar resonances from the mass distributions of this analysis and applying reasonable assumptions to the parameters for scalar mixing.

m_X [GeV] \ m_S [GeV]	70	100	170	200	250	300	400	500	750	1000	1500	2000
300	2.87E-01	2.52E-01	2.28E-03	–	–	–	–	–	–	–	–	–
400	7.79E-02	1.56E-01	2.06E-03	6.14E-04	1.85E-04	–	–	–	–	–	–	–
750	5.70E-03	3.94E-03	7.57E-05	2.86E-05	1.29E-05	6.34E-06	2.74E-06	1.29E-06	–	–	–	–
1000	5.79E-04	5.93E-04	6.72E-06	2.06E-06	1.12E-06	7.43E-07	3.12E-07	1.82E-07	7.00E-08	–	–	–
1500	9.11E-06	1.64E-05	1.78E-07	4.23E-08	1.62E-08	8.34E-09	3.54E-09	2.11E-09	2.12E-09	1.02E-9	–	–
2000	4.70E-07	1.00E-06	1.05E-08	2.55E-09	8.14E-10	2.10E-10	8.55E-11	5.98E-11	5.71E-11	6.71E-11	2.82E-11	–
2500	4.41E-08	1.10E-07	9.58E-10	2.19E-10	5.05E-11	2.08E-11	6.10E-12	4.21E-12	3.60E-12	2.40E-12	2.86E-12	1.34E-12

Based on the observed 95% CL upper limits, the signal hypothesis of the TRSM at $(m_X, m_S) = (400\text{ GeV}, 100\text{ GeV})$ can be excluded. However, the remaining phase space of the TRSM cannot be excluded, as the predicted cross section for the signal samples is smaller than the observed upper limits. While the sensitivity of this analysis is approaching the predicted values for some of the lower-mass signal hypotheses, the majority of the phase space remains at least two magnitudes of sensitivity away.

7.6.4 Future Run 3 Analysis

As the first ATLAS search for scalar resonances in the $X \to SH \to b\bar{b}b\bar{b}$ process, this analysis establishes an essential benchmark for future studies. The upcoming Run 3 dataset will offer a significant opportunity to extend this work, benefiting from a larger integrated luminosity, improved object reconstruction techniques, and more advanced analysis strategies.

One clear avenue for improvement lies in the treatment of background estimation. In the present analysis, the background systematic uncertainties reduce the sensitivity of the expected 95% CL upper limits on the signal cross section by approximately 5–15% for each signal hypothesis compared to a scenario where the relevant nuisance parameters are fixed to their best-fit values. The current method was designed for computational efficiency, producing a large-scale background dataset that spans the full kinematic phase space of the blinded signal region. This approach is effective for covering the wide range of signal hypotheses considered here, but it does not fully exploit the potential gains from more tailored modeling. A future analysis could instead generate background datasets separately for each (m_X, m_S) pair or for clusters of nearby mass points. Such a targeted approach would enable more precise modeling of the background in each region of phase space, potentially reducing systematic uncertainties and improving overall sensitivity.

Another limitation of the present strategy is the need to divide signal hypotheses into separate regimes, compressed and non-compressed, and further split them according to whether m_S is below or above m_H. This division simplifies the current analysis but adds complexity to the workflow. In Run 3, machine learning algorithms could be employed to learn the distinct kinematic features across all signal hypotheses simultaneously, optimize the pairing of b-jets, and maintain strong signal–background discrimination without requiring hard boundaries between categories. Such an approach would streamline the analysis, reduce manual categorization, and potentially improve statistical power.

In summary, while the present work delivers the first ATLAS constraints on ggF production of $X \to SH \to b\bar{b}b\bar{b}$ and sets a strong foundation for future studies, substantial gains in sensitivity and efficiency are expected in Run 3. Through refined background modeling, the adoption of more flexible and powerful machine learning techniques, and the benefits of

larger datasets, the next iteration of this search will be well positioned to probe deeper into the parameter space of scalar resonances and to test extended scalar sector models with unprecedented precision.

Chapter 8

NON-RESONANT $HH \to b\bar{b}b\bar{b}$ WITH RUN 2 DATA

At the onset of my PhD research, I contributed to the non-resonant $HH \to b\bar{b}b\bar{b}$ Run 2 analysis. While not a primary analyzer, I performed several key studies that informed critical decisions on background systematics and the choice of final discriminant. This chapter provides a general overview of the analysis and highlights my specific contributions, illustrating their impact on the final results. The discussion also serves to set the context for the subsequent chapter, which presents the updated analysis using the Run 2 and partial Run 3 data. It is instructive to first review the methods and findings of the Run 2 analysis, as they form the foundation for the work described in Chapter 9.

8.1 Background Estimation

As discussed in 6.3, the background estimation in the $4b$ final state is one of the key challenges of the non-resonant $HH \to 4b$ analysis as simulation of the multi-jet QCD background is unreliable in both normalization and shape for high-multiplicity b-jet final states. Unlike the $X \to SH \to b\bar{b}b\bar{b}$ analysis which uses generative methods for estimating the background, a neural network (NN)–based kinematic reweighting method is employed to reweight lower b-tag multiplicity events as higher b-tag multiplicity events.

The background prediction relies on two categories of events: the $2b$ events, consisting of events with exactly two b-tagged jets, and the $4b$ events, containing events with four or more b-tagged jets. The b-tagging is determined by the DL1r algorithm with a working point efficiency of 77%. This is the b-tagging algorithm which preceded the GN2 as discussed in Section 6.1. Both samples are required to satisfy the same analysis selection, with the only difference being the number of b-tagged jets. The two untagged jets in the $2b$ events are selected as the highest-p_{T} non-b-tagged jets, excluding any VBF jets in VBF-like categories.

8.1.1 Control and Signal Regions

The definition of the signal and control regions is the result of dedicated optimization studies aimed at maximizing the analysis sensitivity while ensuring a reliable background prediction. The SR is designed to provide the highest expected significance for the $HH \to 4b$ signal, based on the kinematic properties of the reconstructed Higgs boson candidates. In contrast, the CRs are chosen to be adjacent to the SR in kinematic space, such that they are enriched in background events and have minimal signal contamination. This proximity ensures that the background composition and shape in the CRs closely resemble those in the SR, enabling a data-driven extrapolation from the CRs to the SR with minimal bias. The SR is defined in the plane of reconstructed Higgs candidate masses (m_{H1}, m_{H2}) using the variable

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{0.1\,m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{0.1\,m_{H2}}\right)^2}, \tag{8.1}$$

with $X_{HH} < 1.6$ defining the SR. The CR is defined as the annular region outside the SR but within

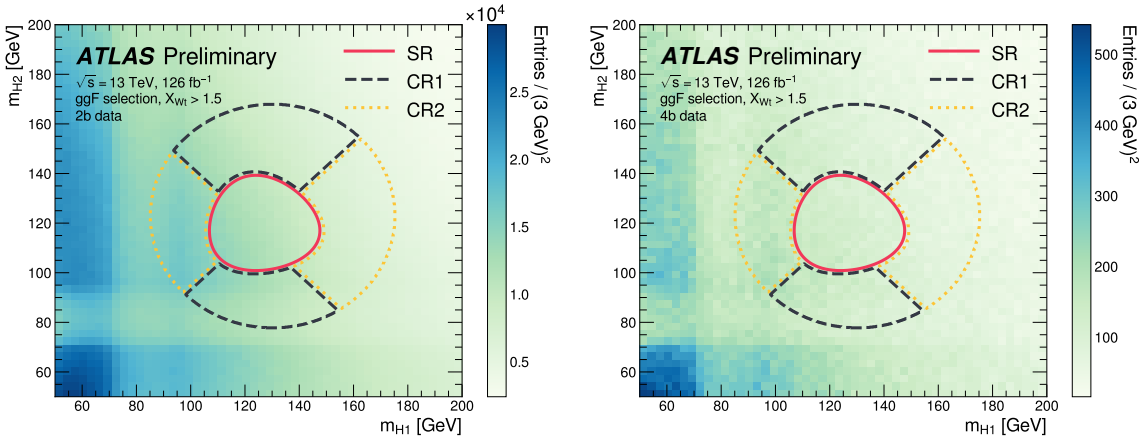
$$\sqrt{(m_{H1} - 1.05 \times 124\,\text{GeV})^2 + (m_{H2} - 1.05 \times 117\,\text{GeV})^2} = 45\,\text{GeV}. \tag{8.2}$$

The CR is split into two subregions (CR1 and CR2) based on their orientation in the (m_{H1}, m_{H2}) plane, as shown in Fig. 8.1. CR1 (the upper and lower crescents) is used to derive the nominal background estimate, while CR2 (the left and right crescents) is used to evaluate the shape systematic uncertainty.

In symmetric Higgs boson pair production, the two Higgs bosons are identical. However, it is conventional to designate the Higgs boson with the higher transverse momentum as $H1$ (the leading Higgs) and the one with the lower transverse momentum as $H2$ (the subleading Higgs). This convention enables more consistent characterization of kinematic trends across the phase space.

8.1.2 $2b \to 4b$ Reweighting

The principal aim of the background estimation is to learn a multivariate reweighting function $w(\vec{x})$ such that when applied to $2b$ events, their distributions match those of the $4b$



(a) (b)

Figure 8.1: The (m_{H1}, m_{H2}) mass planes for (a) $2b$ and (b) $4b$ events from 2016-2018 data, showing the SR (red oval) and CR (surrounding bands). The full ggF selection is applied including a kinematic cut for $X_{Wt} > 1.5$ which greatly reduces the $t\bar{t}$ background.

background in the SR. This is framed as a density ratio estimation problem,

$$w(\vec{x}) = \frac{p_{4b}(\vec{x})}{p_{2b}(\vec{x})} \tag{8.3}$$

where p_{4b} and p_{2b} are the probability densities of selected kinematic variables $\vec{x}$ in the $4b$ and $2b$ samples, respectively.

Instead of learning the reweighting function on a binned distribution, the density ratio is modeled by a neural network trained using an exponential loss to directly approximate the log-ratio:

$$\mathcal{L}(Q) = \mathbb{E}_{\vec{x}\sim p_{2b}}\left[e^{Q(\vec{x})/2}\right] + \mathbb{E}_{\vec{x}\sim p_{4b}}\left[e^{-Q(\vec{x})/2}\right] \tag{8.4}$$

where $Q(\vec{x}) = \log w(\vec{x})$. The optimal $Q^*(\vec{x})$ that minimizes $\mathcal{L}$ recovers $\log\left(p_{4b}(\vec{x})/p_{2b}(\vec{x})\right)$ [153].

A feed-forward neural network is trained on events from CR1 for each data-taking year separately (to capture year-specific trigger and pileup effects), using the Adam optimizer and early stopping [154]. For the ggF no-resonant $HH \to b\bar{b}b\bar{b}$ analysis, the network has three hidden layers with 50 nodes each and ReLU activation. An ensemble of 100 networks

is trained with different random seeds to reduce the impact of statistical fluctuations. The median weight prediction across the ensemble is taken as the nominal event weight.

The input variables for reweighting are chosen to be sensitive to differences between $2b$ and $4b$ samples, including jet kinematics, angular separations, event-level variables, and integer-encoded trigger categories. The list of features for the ggF selections is summarized in Table 8.1. After reweighting, the $2b$ sample is normalized such that the sum of event weights in the CR matches the observed $4b$ event yield. Since the median across the ensemble may not exactly preserve normalization, a final correction factor is applied. This normalization is performed independently for each year and selection category.

Table 8.1: Variables used as input to the neural network for reweighting in the ggF category.

Variable
$\log(p_T)$ of the 2nd leading Higgs boson candidate jet
$\log(p_T)$ of the 4th leading Higgs boson candidate jet
$\log(\Delta R)$ between the closest two Higgs boson candidate jets
$\log(\Delta R)$ between the other two Higgs boson candidate jets
Average $\lvert\eta\rvert$ of the Higgs boson candidate jets
$\log(p_T)$ of the di-Higgs system
ΔR between the two Higgs boson candidates
$\Delta\phi$ between jets in the leading Higgs boson candidate
$\Delta\phi$ between jets in the subleading Higgs boson candidate
$\log(X_{Wt})$
Number of jets in the event
Trigger category (integer-encoded)

To assess the shape systematic, an alternative NN ensemble is trained in CR2. The difference between the predictions using weights from CR1 and CR2 in the SR is symmetrized and assigned as the shape systematic uncertainty. This approach is illustrated in Fig. 8.2.

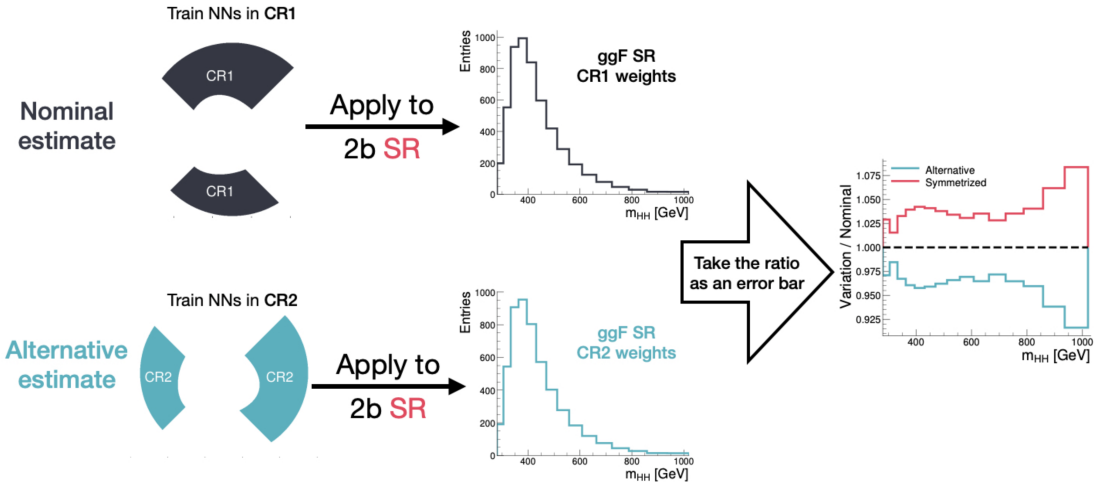


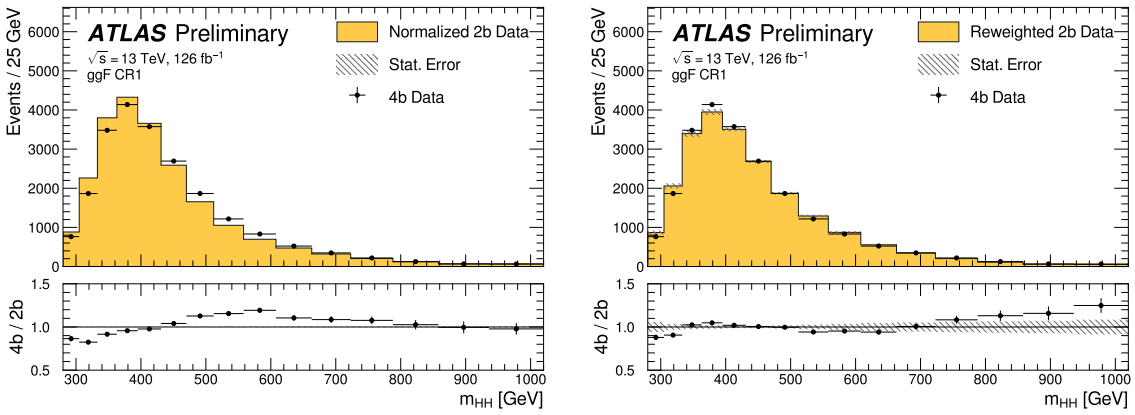
Figure 8.2: Illustration of the background shape systematic derivation. The top row shows the nominal estimate from CR1, while the bottom row shows the alternative estimate from CR2. The right plot displays the symmetrized shape variation (pink histogram), used to define the "CR12 shape systematic". Variations are defined per $|\Delta\eta_{HH}|$ and X_{HH} category as appropriate. Figure from [155].

The closure and modeling performance of the background estimation are validated in the CR and in independent validation regions using a wide range of kinematic observables. Typically, excellent agreement is found between the reweighted $2b$ and observed $4b$ distributions as shown in Figure 8.3, with residual discrepancies covered by assigned systematics. The method is robust to changes in variable selection, network architecture, and training statistics, as tested in dedicated cross-checks.

8.1.3 Neural Network Initialization on Reweighting Statistical Uncertainty

The statistical uncertainty in the kinematic reweighting of $2b$ events to $4b$ events is composed of two main components. The first is the standard Poisson error from the $2b$ data, which is computed for each bin as $\delta n_i = \sqrt{\sum_{j\in i} w_j^2}$, where w_j is the neural network-assigned weight for event j in bin i.

Additionally, the neural network training itself introduces a further source of statistical



(a) (b)

Figure 8.3: Distributions of m_{HH} in $4b$ data (black points) within Control Region 1 (CR1) of the ggF selection. (a) The yellow histogram shows the $2b$ CR1 m_{HH} distribution normalized to the $4b$ yield; (b) the yellow histogram shows the $2b$ CR1 after NN reweighting. Hatched bands indicate the statistical uncertainty on the 2b sample: Poisson only in (a), and including the bootstrap uncertainty in (b).

uncertainty. Because the network is trained on finite data, its predictions can vary due to both random initialization and the inherent stochasticity in the training procedure. To quantify this, we employ an ensemble of 100 neural networks, each initialized with different random weights and trained on bootstrapped versions of the training set. The bootstrapping is implemented by assigning event weights drawn from a Poisson distribution with mean one, ensuring that each network sees a statistically distinct dataset.

The final background prediction for each event is defined as the ensemble average. The per-bin uncertainty due to neural network initialization and bootstrapping is computed as the standard deviation across the 100 ensemble members:

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(h_{ij} - \bar{h}_j)^2}$$

where h_{ij} is the yield in bin j for network i, and $\bar{h}_j$ is the ensemble mean.

These uncertainties are combined in quadrature with the Poisson error from the $2b$ data to yield the total per-bin statistical uncertainty. Both are incorporated as per-bin Poisson nuisance parameters in the likelihood fit. The normalization of the reweighted $2b$ background to the $4b$ yield in the control region introduces a further normalization uncertainty, which is propagated as an independent nuisance parameter.

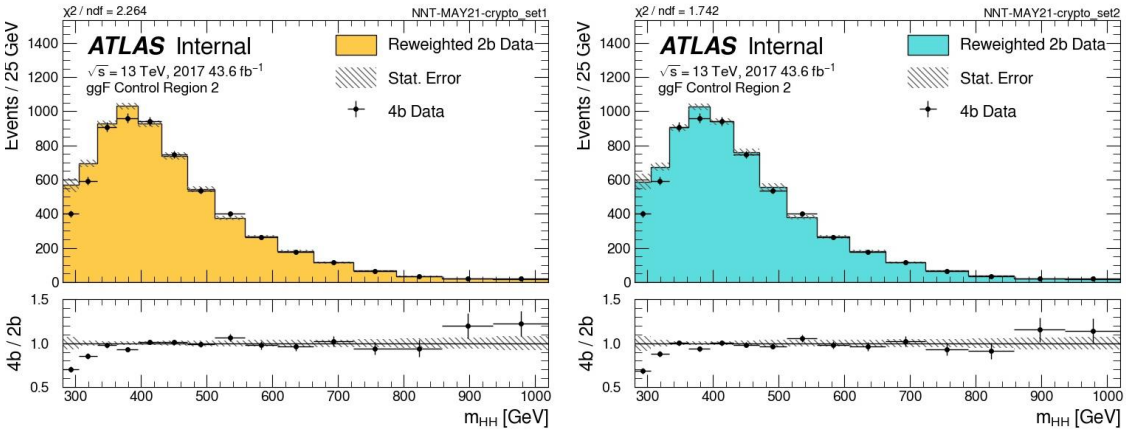
To validate the procedure, we tested several configurations for random seed initialization and Poisson sampling, summarized in Table 8.2. Comparisons of key distributions (e.g., m_{HH}, m_{H1}, and m_{H2}) for different configurations showed negligible differences in the average predictions (see Figure 8.4), and only minor differences in the uncertainty envelopes (see Figure 8.5 and 8.6). This demonstrates that the dominant contribution to the statistical uncertainty arises from the bootstrap procedure and finite training data, rather than random seed initialization.

Table 8.2: Configurations tested to evaluate the impact of neural network initialization and Poisson sampling on reweighting uncertainty for SM non-resonant $HH \to b\bar{b}b\bar{b}$ production. All seeds refer to Python's random number generators: numpy, random, TensorFlow, and Python hash.

Configuration #	Initialization and Sampling Conditions
1	Poisson sampling, fixed numpy seed (baseline configuration)
2	No Poisson sampling, variable numpy seed
3	Poisson sampling, variable numpy seed
4	Poisson sampling, all seeds fixed for full reproducibility (Python hash, numpy, tensorflow, random)

Comparisons of distributions from configurations 1 and 2 (with and without Poisson sampling, respectively) showed negligible differences for key kinematic variables (m_{HH}, m_{H1}, and m_{H2}) in Control Region 2, illustrated in Figure 8.4. Though no perceived differences are seen in the average predictions across various kinematic variables among the different

configurations, the uncertainty envelopes shown in



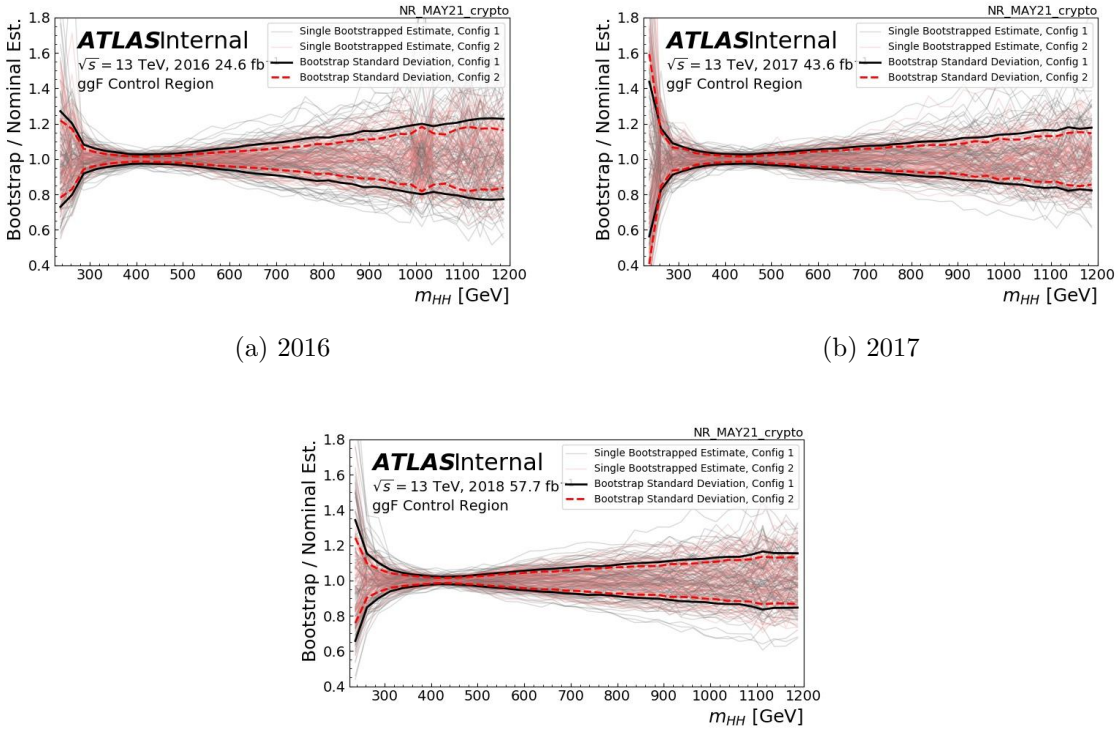
(a) (b)

Figure 8.4: Distributions of the reweighted 2b events and 4b data for m_{HH} in Control Region 2 for 2017 data, shown for configurations 1 (with Poisson sampling) and 2 (without Poisson sampling). No significant difference is observed in the average prediction of the resulting distributions.

Figure 8.5, calculated bin-by-bin from the standard deviation across bootstrap samples, indicate a very small entanglement between Poisson sampling and neural network training convergence. This was deemed to be a very small difference and suggested that the dominant source of uncertainty is due to finite statistics rather than initialization details.

Further comparisons between configurations 1 and 3, which differ only in the numpy random seed, reaffirmed that the initialization of the random seed has a negligible effect on the resulting reweighting uncertainty (see Figure 8.6). Comparison across all configurations showed that even when random seeds are held constant (or not), the bootstrap resampling step produces statistically independent replicas, resulting in variations that originate from the inherent randomness of the resampling itself rather than from the seed initialization. This behavior demonstrates that the primary source of variation across the ensemble arises from the bootstrap procedure, not from differences in the neural network initialization.

Overall, these studies demonstrate that the dominant contribution to the reweighting



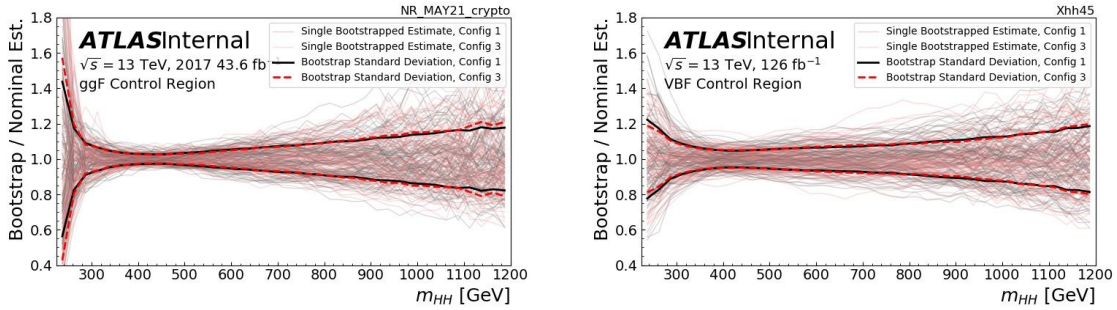
(a) 2016

(b) 2017

(c) 2018

Figure 8.5: The bootstrap band procedure, shown as a ratio to the nominal (average) estimate in the CR1. Each lightly colored line is from the m_{HH} prediction for a single bootstrap training while the darker line shows the standard deviation of histograms for its corresponding dataset. All three years are included separately for ggF from configurations 1 (black) and 2 (red).

statistical uncertainty stems from the finite size of the training data as captured by the bootstrap ensemble, while factors such as seed initialization and Poisson sampling play a minimal role. It is important to emphasize that the results here focus specifically on the statistical uncertainty in the kinematic reweighting. Additional systematic uncertainties, including those associated with the transfer of the background model from CR1 to the signal region as evaluated in the alternative CR2 control region and those due to potential non-closure as assessed in orthogonal 3b1f (three b-tagged jets plus one failing jet) samples, are



(a) (b)

Figure 8.6: The bootstrap band procedure, shown as a ratio to the nominal (average) estimate in the CR1. Each lightly colored line is from the m_{HH} prediction for a single bootstrap training while the darker line shows the standard deviation of histograms for its corresponding dataset. Data for 2017 for the ggF analysis (a) and for all three years from the VBF analysis (b) are included for configurations 1 (black) and 3 (red).

addressed through separate dedicated studies. For a detailed discussion of these systematic treatments, the reader is referred to the published analysis [102].

8.2 Two-Dimensional Discriminants

Selecting a final discriminant for statistical interpretation in a high-energy physics analysis requires balancing several factors: separation power between signal and background, reliable modeling, manageable systematic uncertainties, and robust statistical properties. While multivariate analysis (MVA) techniques through machine learning algorithms are commonly used for their superior separation, this analysis prioritized robust modeling and controlled systematics. As a result, the invariant mass of the Higgs boson pair, m_{HH}, was primarily used as the final discriminant in early studies.

However, a single variable cannot always capture all available information, especially when multiple observables exhibit nontrivial correlations. For example, as shown in Figure 8.7, there is a strong correlation between the reconstructed m_{HH} and the top-quark veto discriminant X_{Wt}. Notably, these correlations differ significantly between signal and background, suggesting that a two-dimensional (2D) fit using both variables could improve sensitivity.

To implement these 2D fits, the two-dimensional histograms were unrolled into one-dimensional distributions by sequentially ordering the bins. To maintain statistical reliability, bins with large uncertainties (specifically, those with relative uncertainties exceeding 50%) are merged with adjacent bins, and the same binning is applied to both signal and background. This approach ensures adequate bin populations, particularly in sparsely populated regions. The resulting statistical-only exclusion limits at 95% confidence level (CL) are summarized in Table 8.3. Both single-variable (1D) and two-variable (2D) fits were tested, using combinations of m_{HH}, X_{Wt}, $|\Delta\eta_{HH}|$, and the reconstructed masses of the leading and subleading Higgs candidates (m_{H1} and m_{H2}).

The 2D fits provide noticeably improved limits compared to the 1D fits. For instance, combining m_{HH} and X_{Wt} yields the best exclusion, demonstrating the power of exploiting correlated observables. These results motivated further studies exploring the impact of the Higgs self-coupling parameter (κ_λ) on the optimal discriminants and limits.

Across the full range of κ_λ values (-20 to 20), 2D fits provided an average improvement of approximately 9% in the stat-only expected exclusion limits, consistently outperforming

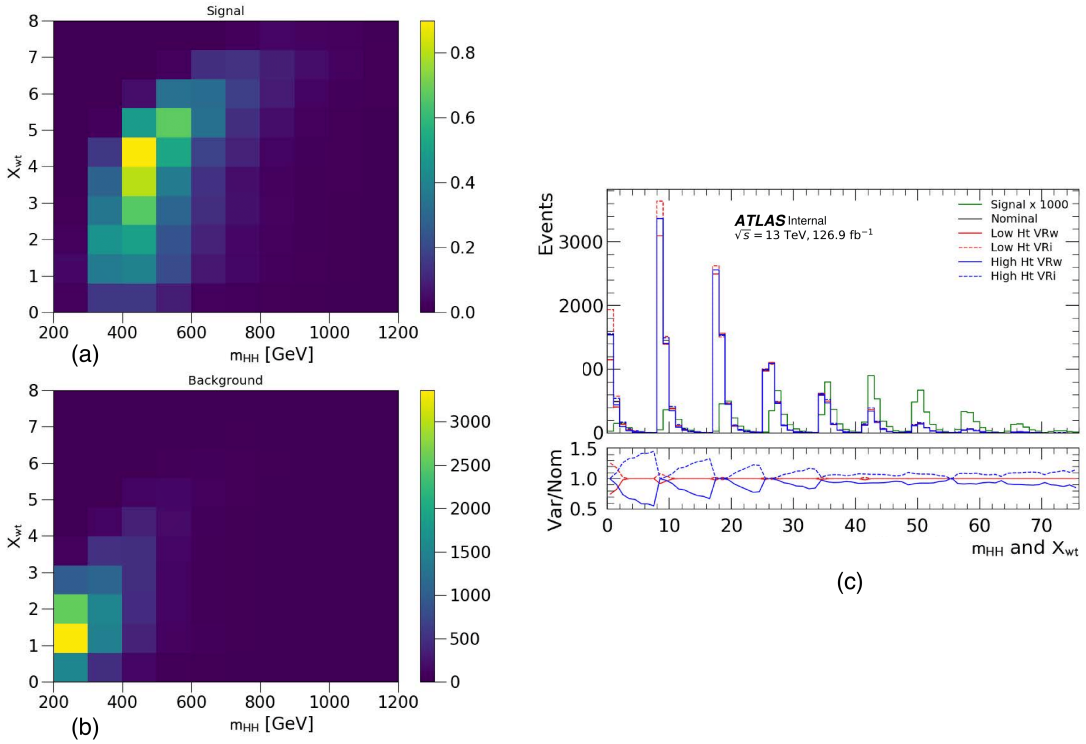


Figure 8.7: The m_{HH} vs X_{Wt} distributions for (a) signal and (b) average estimated background in the SR illustrating the correlations between m_{HH} and X_{Wt}. The transformation into a 1D histogram for fitting purposes (c), shown for these distributions, where the 2D histogram is "rolled out" starting from the bottom-left hand corner to the top-right corner.

the respective 1D fits. This comprehensive study informed the eventual published strategy, which employed categories defined by fine m_{HH} binning, $|\Delta\eta_{HH}|$ (divided into three bins: [0.0, 0.5, 1.0, 1.5]), and X_{HH} (two bins: [0.0, 0.95, 1.6]). This approach maximized the analysis sensitivity while retaining robust modeling and manageable statistical uncertainties, ultimately providing the most reliable constraints on the Higgs self-coupling.

Table 8.3: Statistical-only exclusion limits (95% CL) for all combined years, showing results for single discriminants and their two-dimensional combinations in the SM non-resonant Higgs pair production analysis via gluon-gluon fusion (ggF).

Discriminant(s)	95% CL Exclusion Limit
$\|\Delta\eta_{HH}\|$	11.43
m_{H2}	6.99
m_{H1}	6.91
X_{Wt}	6.90
m_{H1} and m_{H2}	6.81
m_{HH}	6.02
X_{Wt} and $\|\Delta\eta_{HH}\|$	6.45
m_{HH} and $\|\Delta\eta_{HH}\|$	6.36
m_{HH} and m_{H1}	5.81
m_{HH} and X_{Wt}	5.43

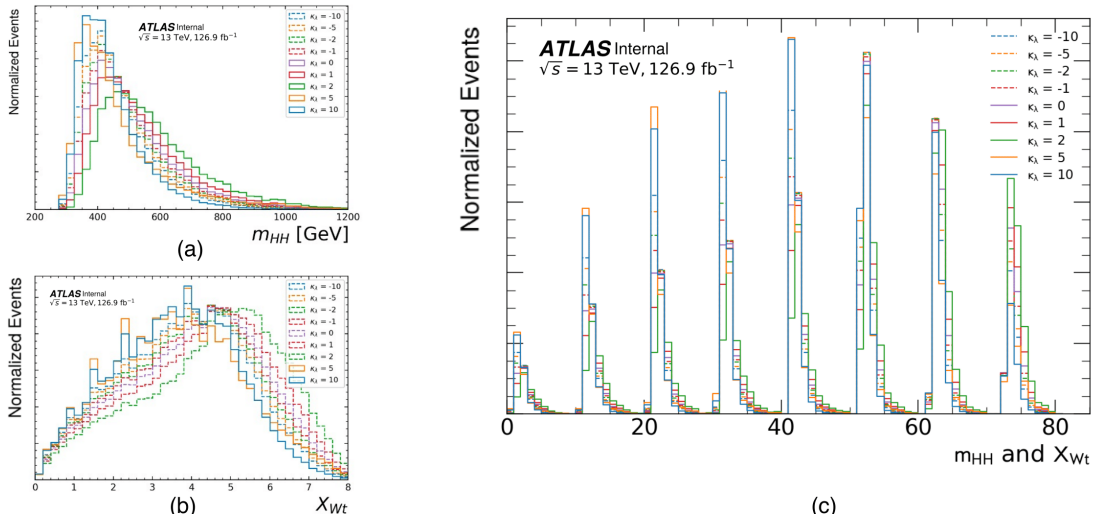


Figure 8.8: Normalized distributions of (a) m_{HH} and (b) X_{Wt} for several representative κ_λ values, and their transformation into a (c) 2D discriminant employed in the statistical fit. The use of a 2D discriminant consistently improves the statistical-only 95% CL limits on κ_λ. The effect of varying κ_λ on the distribution shapes is clearly observed.

8.3 Results

The search for non-resonant $HH \to b\bar{b}b\bar{b}$ production in the resolved channel utilizes a simultaneous maximum likelihood fit to the reconstructed m_{HH} distributions across the six exclusive event categories defined in the ggF signal region, incorporating data from the 2016, 2017, and 2018 LHC running periods. The statistical model comprises Poisson likelihoods for each bin and category, combined with constraint terms representing all relevant systematic uncertainties. Uncertainties from limited data or MC statistics are implemented via Poisson constraints, while experimental and theoretical systematics are incorporated as Gaussian constraints. Key systematics, such as those related to luminosity and signal modeling, are treated as fully correlated across ggF categories and years, while the dominant background modeling uncertainties (e.g., bootstrap and non-closure) are correlated within ggF.

The fit is performed with the signal strength parameter μ_{HH} as the parameter of interest, while all nuisance parameters are profiled. No significant excess is observed. The resulting observed (expected) 95% confidence level (CL) upper limit on the HH production cross section, relative to the Standard Model prediction through ggF and VBF production, is

$$\mu_{HH}^{\text{obs}} = 5.4\ (8.1) \times \text{SM}\,,$$

where the quoted limits include all sources of statistical and systematic uncertainty. These results mark a substantial improvement over the previous iteration of the analysis in [101], benefiting from increased luminosity, advanced event categorization, and improved background modeling techniques.

Additionally, constraints are placed on the Higgs boson trilinear self-coupling modifier, κ_λ. The observed (expected) allowed range at 95% CL is

$$\kappa_\lambda \in [-3.9,\ 11.1] \quad @\ 95\%\ \text{CL} \quad (\text{expected } [-4.6,\ 10.8])\,,$$

with all other Higgs couplings fixed to their Standard Model values. The profile likelihood scan for κ_λ (Figure 8.9) shows that the best-fit value is consistent with the Standard Model, with the observed exclusion limit slightly exceeding the expected due to an upward fluctuation in the low-m_{HH} region.

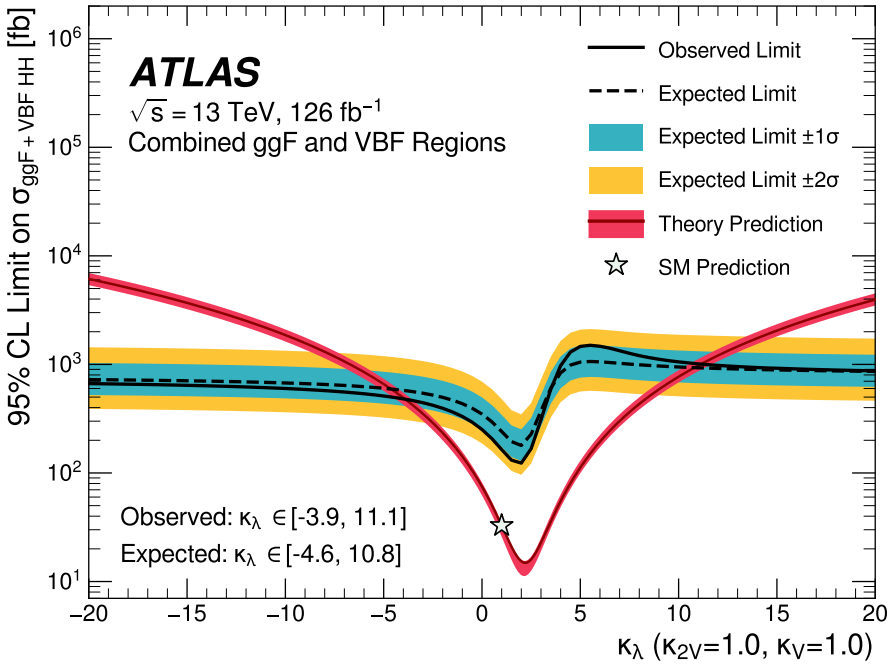


Figure 8.9: Observed and expected 95% confidence level exclusion limits on the non-resonant HH signal strength and constraints on the Higgs boson trilinear self-coupling κ_λ in the resolved $b\bar{b}b\bar{b}$ channel, as published for the Run-2 analysis. The Standard Model value is indicated by the vertical line [102].

All exclusion limits are derived using the CL_s method and account for the full suite of systematic and statistical uncertainties, with the dominant sources being the background modeling uncertainties and the theoretical uncertainties on the signal cross section. The limits presented here are the most stringent to date for this final state with the ATLAS detector, and the observed sensitivity is approaching that predicted by the Standard Model. With the ongoing collection of Run 3 data and the continued refinement of analysis techniques, further improvements in sensitivity are anticipated.

8.4 Areas for Improvement

Despite the significant progress achieved in this analysis, several areas offer clear opportunities for further gains in sensitivity. The dominant contribution to the total uncertainty on the cross section limit arises from the background modeling procedure and theoretical predictions. As a result, experimental efforts are naturally directed toward reducing background-related systematic uncertainties in future analyses.

To address these challenges, new generative background estimation techniques have been developed, as described in Section 6.4 and implemented in the resonant $X \to SH \to b\bar{b}b\bar{b}$ search discussed in Section 7.4. The ensemble-based neural network method previously used relied on two key assumptions: (1) the control region is sufficiently similar in kinematics to the signal region for a reweighting to be valid, and (2) the $2b$-tagged sample provides an adequate basis for training the reweighting. Although these assumptions were validated, the residual background systematics indicate that the method is not fully robust, particularly when control and signal regions differ.

An alternative approach (the GP+Flows method) offers improved flexibility. Unlike the ensemble reweighting, the normalizing flow-based model learns the kinematic correlations directly as a function of input variables, allowing for more accurate extrapolation from the control to the signal region. The flows are conditioned on candidate mass variables predicted by the Gaussian Process, enabling the model to generate background predictions that adapt to changes in the underlying kinematics between regions, thus reducing reliance on the similarity assumption.

Further performance gains have also come from advances in event reconstruction and object identification. The adoption of the GN2 b-tagger (Section 6.1) for small-R jets, replacing the earlier DL1r tagger, provides approximately a factor of three improvement in the rejection of light- and charm-jets, depending on jet p_T and η. Such improvements in b-tagging, together with enhanced jet reconstruction and systematic uncertainty modeling, are expected to yield greater analysis sensitivity.

Another major improvement in Run 3 is the deployment of dedicated asymmetric multi-jet triggers optimized for the $HH \to b\bar{b}b\bar{b}$ final state. Unlike the symmetric jet triggers

used in earlier analyses, these new triggers implement staggered p_T thresholds for the four leading jets, more closely matching the kinematics of non-resonant di-Higgs events where the b-jets are often produced with unequal momenta [156].

The impact is substantial: the new Run 3 trigger strategy increases the inclusive selection efficiency for $HH \to b\bar{b}b\bar{b}$ by approximately 50% compared to the Run 2 triggers, as shown in Figure 8.10, with the largest gains (up to 75%) achieved in the critical low m_{HH} region near the $2m_H$ threshold. The gain in efficiency is driven by improvements in the b-jet identification algorithms that consequently allows loosening jet p_T thresholds. The improvement significantly boosts the analysis sensitivity, ensuring better acceptance of signal events across the full phase space as more data is collected in Run 3.

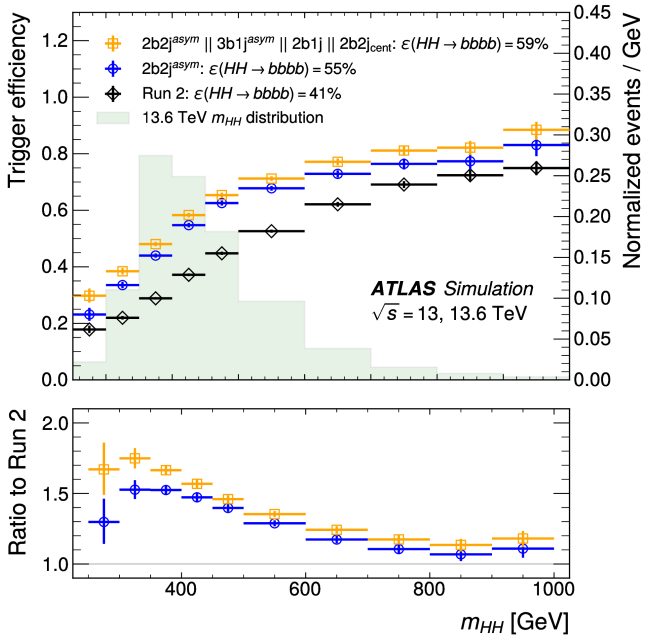


Figure 8.10: Performance of the b-jet trigger HLT selections in simulated $HH \to b\bar{b}b\bar{b}$ events. The trigger efficiency, evaluated relative to all inclusive $HH \to b\bar{b}b\bar{b}$ events, is shown as a function of the true di-Higgs boson invariant mass, m_{HH}. The efficiency is presented for the $2b2j^{asym}$ selection and for the logical OR of the $2b2j^{asym}$, $3b1j^{asym}$, $2b1j$, and $2b2j^{cent}$ triggers and is compared with the trigger strategy employed in Run 2 [156].

Finally, with the start of Run 3, significantly more data are being collected. The partial Run 3 dataset, collected during 2022 and 2023, already includes approximately 70 fb^{-1}, with an anticipated total of 280–300 fb^{-1} expected for all of Run 3. Although the new data alone will not immediately yield a dramatic sensitivity gain compared to the full Run 2 dataset, they represent crucial additional statistics for probing this rare process and are an essential ingredient for future discoveries.

Chapter 9

NON-RESONANT $HH \to b\bar{b}b\bar{b}$ WITH RUN 2 AND PARTIAL RUN 3 DATA

Having presented the search for non-resonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ decay channel with the ATLAS Run 2 dataset in the previous chapter, this chapter extends the analysis to include both Run 2 and partial Run 3 data, with a particular focus on the gluon-gluon fusion production mode in the resolved topology. While the core analysis workflow is largely established, ongoing optimization studies are being conducted at several stages and key decisions regarding the evaluation of background systematics and the choice of final discriminant remain under active investigation. Nevertheless, the current results obtained with the baseline methods are presented in this chapter with an emphasis on my own work.

9.1 Data and Monte Carlo Datasets

The signal Monte Carlo (MC) samples used in this analysis simulate non-resonant Higgs boson pair production (HH) in the $b\bar{b}b\bar{b}$ final state. The ggF signal samples are generated using the Powheg-Box v2 generator interfaced with Pythia 8 for parton showering and hadronization, employing the PDF4LHC21 parton distribution functions. Run 2 samples are produced at a center-of-mass energy of 13 TeV, while Run 3 samples are generated at 13.6 TeV. Detector simulation is performed with the fast simulation framework AF3, with additional jet energy scale (JES) and jet energy resolution (JER) systematic uncertainties included to cover residual differences with the full Geant4-based simulation, as described in Chapter 4.

Samples are produced for several benchmark values of the Higgs boson trilinear self-coupling modifier, κ_λ, which parameterizes deviations from the Standard Model prediction. The following κ_λ values are considered: -1, 0, 1, 2.5, 5, and 10, corresponding to dataset IDs (DSIDs) 603403–603408, respectively.

The data used in this analysis were collected by the ATLAS detector during Run 2 (2016–2018) and Run 3 (2022–2023) of the LHC, in proton-proton collisions at center-of-mass energies of $\sqrt{s} = 13$ TeV and $\sqrt{s} = 13.6$ TeV, respectively. The total integrated luminosity used in the analysis, selected with small-R b-jet triggers, is 177.75 fb^{-1}, representing nearly a 40% increase over Run 2 alone. Data collected in 2015 are excluded due to differences in b-tagging algorithms and negligible impact relative to the required resources. Additionally, a portion of the 2016 dataset is excluded due to b-jet trigger inefficiencies caused by a bug in the online vertex finding algorithm, as described in Section 7.1.

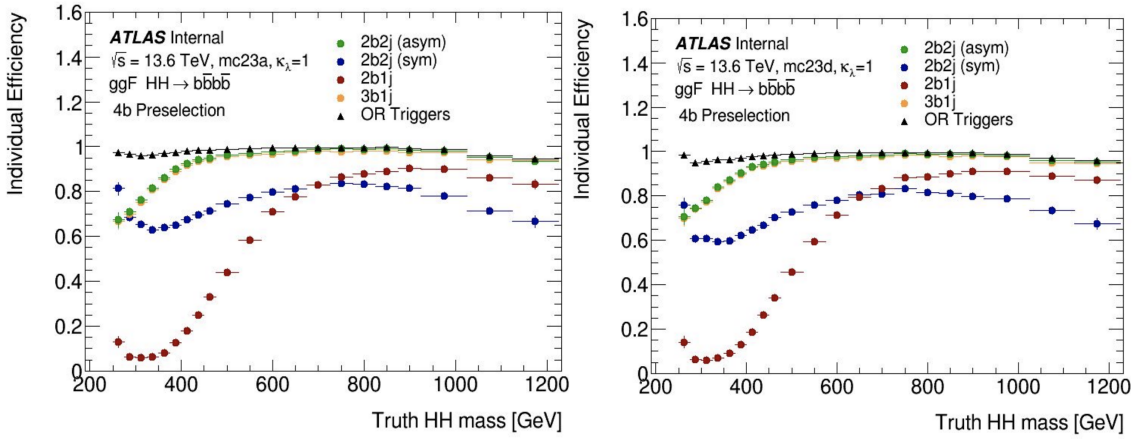
Table 9.1: Integrated luminosities of each data-taking year for the resolved analysis using b-jet triggers.

Year	Luminosity (fb^{-1})
2016	24.56
2017	43.65
2018	57.70
2022	26.07
2023	25.77
Total	**177.75**

9.2 Triggers

For Run 2, the same trigger bucketing strategy as in the previous non-resonant $HH \to b\bar{b}b\bar{b}$ and the resonant $X \to SH \to b\bar{b}b\bar{b}$ analyses is employed, as illustrated in Figure 7.5. Events are divided into two buckets based on offline jet kinematics and trigger requirements. The first bucket uses the **2b+1j** trigger and requires the leading jet to have $p_{\mathrm{T}} > 170$ GeV and the third jet $p_{\mathrm{T}} > 70$ GeV, implicitly setting a threshold on the second jet. Events not passing this selection but firing the **2b+2j** trigger are assigned to the second bucket. This approach ensures efficient triggering across the full phase space by combining the strengths of both triggers.

In Run 3, the trigger strategy for the resolved $HH \to b\bar{b}b\bar{b}$ analysis is streamlined to use a single asymmetric multi-jet trigger per data-taking year. For 2022 and 2023, unique **3b+1j** asymmetric triggers are used. The 2022 3b+1j asymmetric trigger requires four jets with transverse momenta of at least 80, 55, 28, and 20 GeV, with three of these jets passing the b-tagging requirement with the DL1d tagging algorithm at the 70% efficiency working point [156, 157]. The 2023 trigger is similar but sets jet p_T thresholds at 75, 50, 25, and 20 GeV, again requiring at least three b-tagged jets with the GN1 tagging algorithm also at the 70% efficiency working point. As demonstrated in Figure 9.1, the **2b+2j** asymmetric trigger actually has the highest individual trigger efficiency for the signal across the kinematic phase space, the **3b+1j** has nearly the same efficiency while offering a reduced background making it the ideal candidate as the single choice of trigger.



(a) (b)

Figure 9.1: Individual trigger efficiencies for non-resonant $HH \to b\bar{b}b\bar{b}$ production in the (a) mc23a and (b) mc23d simulated samples, shown after applying a $4b$ preselection. The preselection requires four jets with $p_T > 25$ GeV, $|\eta_{\text{jet}}| < 2.5$, and GN2 b-tagging at the 77% working point [158].

9.3 Event Selection and Reconstruction

After applying the trigger bucketing scheme for Run 2 or the single 3b+1j trigger for Run 3, the resolved analysis event selection requires at least four jets that satisfy the following criteria:

- $p_T > 40$ GeV for Run 2, $p_T > 20$ GeV for Run 3
- $|\eta| < 2.5$
- b-tagged with the GN2v01 tagger at the 77% working point
- Passes the neural network jet vertex tagger (NN JVT)

The lower p_T cut in Run 3 is there to take advantage of the lower p_T threshold in the trigger while all other aspects are the same between Run 2 and Run 3.

9.3.1 Min ΔR Pairing

In the non-resonant $HH \to b\bar{b}b\bar{b}$ analysis, reconstructing the two Higgs candidates from the four leading jets uses the minimum ΔR pairing algorithm. For each event, all three possible ways to pair the four jets into two Higgs candidates are considered. For each pairing configuration, the algorithm calculates the angular separation ΔR_{jj} between the two jets in each Higgs candidate. The pairing chosen is the one where the leading Higgs candidate (the candidate with the higher transverse momentum) has the smallest ΔR_{jj} between its two jets. In other words, the algorithm selects the jet pairing that minimizes ΔR_{jj}^{HC1}, where $HC1$ denotes the leading Higgs candidate.

This strategy is motivated by the expected kinematics of Higgs boson decays. When the Higgs bosons are highly boosted, their b-jet decay products tend to be more collimated. As a result, focusing on the smallest ΔR in the leading candidate suppresses background events with low invariant mass, which would otherwise accumulate in the signal region if alternative pairing schemes were used. Although this method may result in reduced signal efficiency for low m_{HH}, particularly away from the Standard Model value of κ_λ, it

significantly improves the smoothness of the background mass distribution. Furthermore, it minimizes kinematic biases between the control and signal regions, supporting a more reliable background estimation. The choice of min ΔR pairing is therefore well-suited to maximize sensitivity for signals with higher masses, while reducing the risk of background sculpting.

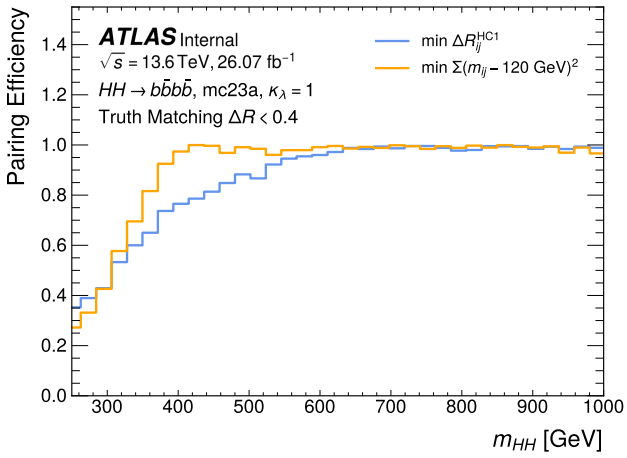
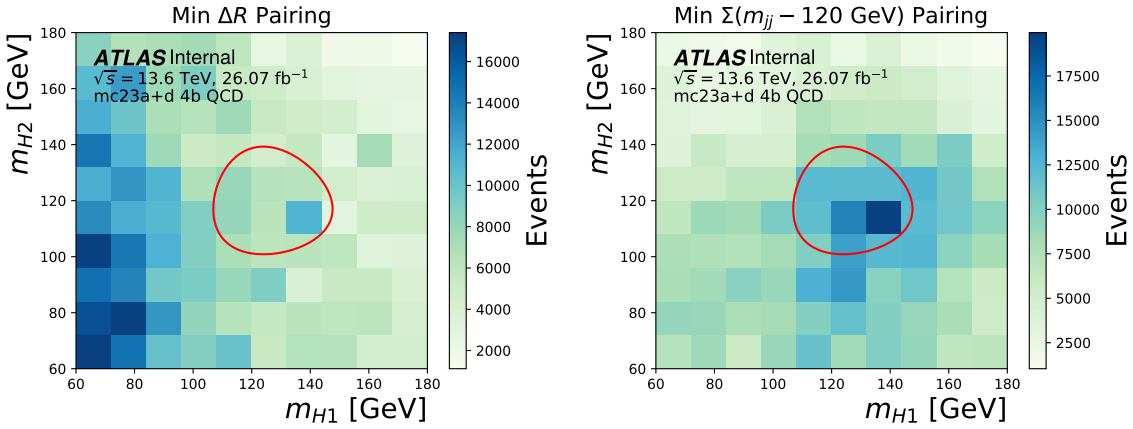


Figure 9.2: Pairing efficiency as a function of the reconstructed m_{HH} for the min ΔR and min $\chi^2 = \sum(m_{jj} - m_X)^2$ pairing algorithms. The min ΔR pairing method, which is the nominal pairing method, is poor at low masses, but very accurate for $m_{HH} > 500$ GeV.

Alternative pairing methods, such as using machine learning techniques or selecting the combination that minimizes the invariant mass error ($\chi^2 = \sum(m_{jj} - m_X)^2$ with $m_X = 120$ GeV), have also been studied. These methods can enhance the signal efficiency, especially at low m_{HH}, but this gain comes at the cost of moving a larger fraction of the multi-jet QCD background into the signal region. This effect is illustrated in Figure 9.3, where simulated $4b$ multi-jet QCD events using the mass-based pairing method become kinematically similar to signal events, leading to a less robust background estimate. Since the analysis sensitivity is driven primarily by the high-mass region, where the min ΔR algorithm achieves a high pairing accuracy and background contamination is low, this approach remains the preferred choice.

With the new triggers introduced in Run 3, which increase sensitivity to low-mass di-Higgs production, the possibility of leveraging alternative pairing strategies may warrant further investigation in future analyses. Ongoing studies on optimizing the low-mass selection are discussed in Appendix A.



(a) (b)

Figure 9.3: Distributions of simulated $4b$ QCD events for mc23a+d in the $m_{H1} - m_{H2}$ mass plane for the (a) min ΔR pairing and the (b) mass-based pairing. The min ΔR approach leads to much less background in the signal region (red contour).

9.3.2 Kinematic Regions

The definition of the signal region (SR) for the Run 2 analysis was re-evaluated for the current study, and it was confirmed to remain optimal for the non-resonant $HH \to b\bar{b}b\bar{b}$ search. The SR is defined by $X_{HH} < 1.6$ and centered at $(m_{H1}, m_{H2}) = (124\,\text{GeV},\ 117\,\text{GeV})$, where

$$X_{HH} = \sqrt{\left(\frac{m_{H1} - 124\,\text{GeV}}{0.1\,m_{H1}}\right)^2 + \left(\frac{m_{H2} - 117\,\text{GeV}}{0.1\,m_{H2}}\right)^2}, \tag{9.1}$$

which reflects the expected resolutions of the reconstructed Higgs candidate masses.

With the introduction of more flexible background modeling—based on normalizing flows and Gaussian processes, as discussed in the following section—it became necessary to

update the control region (CR) definition. Unlike the elliptical SR, the CR is now defined as a rectangular box in the (m_{H1}, m_{H2}) plane, with boundaries calculated as follows:

$$r_{\mathrm{CR},x} = \begin{cases} \alpha_{\mathrm{CR}}\, x \left(\frac{1}{0.84} - \frac{1}{1.16}\right) & \text{if } |x - 125| < 25 \\ \alpha_{\mathrm{CR}}\, 124 \left(\frac{1}{0.84} - \frac{1}{1.16}\right) & \text{otherwise} \end{cases} \tag{9.2}$$

$$x_{\min} = x(1+\varepsilon) - r_{\mathrm{CR},x} \tag{9.3}$$

$$x_{\max} = x(1+\varepsilon) + r_{\mathrm{CR},x} \tag{9.4}$$

$$r_{\mathrm{CR},y} = \begin{cases} \alpha_{\mathrm{CR}}\, y \left(\frac{1}{0.84} - \frac{1}{1.16}\right) & \text{if } |y - 125| < 25 \\ \alpha_{\mathrm{CR}}\, 117 \left(\frac{1}{0.84} - \frac{1}{1.16}\right) & \text{otherwise} \end{cases} \tag{9.5}$$

$$y_{\min} = y(1+\varepsilon) - r_{\mathrm{CR},y} \tag{9.6}$$

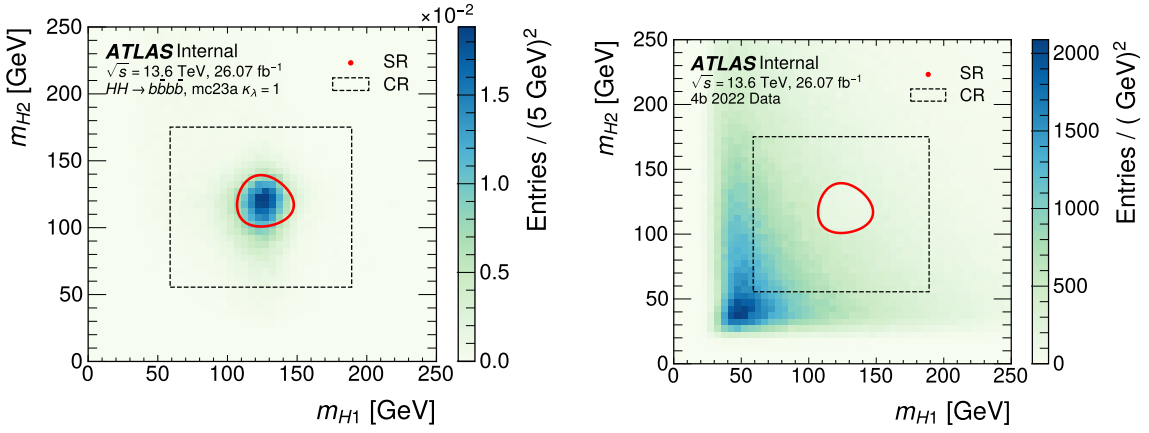
$$y_{\max} = y(1+\varepsilon) + r_{\mathrm{CR},y} \tag{9.7}$$

Here, α_{CR} sets the size of the CR window, and ε allows for an optional shift of the CR center relative to the SR. For this analysis, the values $\alpha_{\mathrm{CR}} = 1.6$ and $\varepsilon = 0.0$ are used to maximize statistical significance while preserving robust background modeling. This results in the CR spanning from $(x_{\min}, y_{\min}) = (58.85\ \mathrm{GeV},\ 55.48\ \mathrm{GeV})$ to $(x_{\max}, y_{\max}) = (189.16\ \mathrm{GeV},\ 178.52\ \mathrm{GeV})$.

The kinematic boundaries of both the signal and control regions are illustrated in Figure 9.4, which shows the (m_{H1}, m_{H2}) mass plane for both simulated SM signal (left) and blinded 2022 4b data (right), with the SR indicated by the red ellipse and the CR by the dashed black rectangle. The Higgs candidate masses are reconstructed using the min ΔR pairing method as discussed previously.

9.3.3 Background Reduction

Following the discussion of backgrounds in the four-b final state in Section 6.3, the kinematic selections established for the Run 2 analysis are retained for Run 3, as they continue to provide strong discrimination between signal and background. As illustrated in Figure 9.5, the requirement $|\Delta\eta_{HH}| < 1.5$ efficiently selects signal events, maintaining a high signal



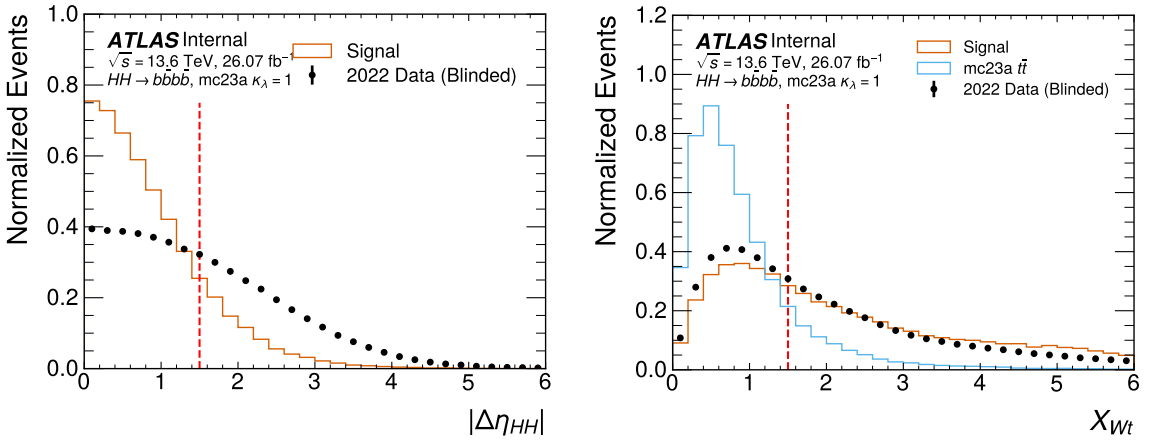
(a) (b)

Figure 9.4: The (m_{H1}, m_{H2}) mass planes for (a) simulated SM signal (mc23a) and (b) blinded $4b$ data from 2022, showing the signal region (red ellipse) and control region (dashed black box) defined for the analysis. Higgs candidate masses are reconstructed using the min ΔR pairing.

efficiency while substantially reducing the multijet background in data. Additionally, the top veto cut, $X_{Wt} < 1.5$, remains highly effective, removing approximately 85% of the $t\bar{t}$ background, which is the sub-dominant source of contamination after basic selection. These selections collectively optimize the sensitivity of the analysis by preserving signal yield and suppressing key backgrounds.

9.4 Background Estimation

The large data-driven background systematic uncertainties observed in the Run 2 analysis (see Section 6.4) motivated a transition from the NN-based reweighting approach to a generative modeling strategy based on normalizing flows and Gaussian processes. This new framework was designed specifically to bypass the extrapolation uncertainty associated with the previous control region (CR) to signal region (SR) transfer, while improved multivariate modeling offers the potential to reduce statistical uncertainties and even eliminate non-



(a) (b)

Figure 9.5: Distributions of the simulated SM mc23a signal and blinded four-b data from 2022, where events in the signal region (SR) are masked, for (a) $|\Delta\eta_{HH}|$ and (b) X_{Wt}. The red vertical lines indicate the kinematic selection cuts: $|\Delta\eta_{HH}| < 1.5$ and $X_{Wt} < 1.5$. The X_{Wt} distribution also shows the simulated $t\bar{t}$ background, demonstrating the effectiveness of the veto in suppressing this background while preserving signal events.

closure uncertainties.

The implementation closely follows the strategy used in the $X \to SH \to b\bar{b}b\bar{b}$ analysis, with the central objective being to predict the joint multivariate distribution of background events in the blinded signal region (SR) using only observed data from the rectangular control region (CR). This generative modeling proceeds in two integrated stages. First, the two-dimensional distribution of the Higgs candidate masses, (m_{H1}, m_{H2}), in the SR is inferred by interpolating from the CR using Gaussian process (GP) regression. The GP is trained on binned event yields in the (m_{H1}, m_{H2}) plane, with careful masking of bins that overlap with the SR to avoid statistical bias.

Once the GP provides an estimate for the SR mass plane, a set of conditional normalizing flows is trained to model the full joint probability density of the key kinematic variables, $\vec{x}$, conditioned on the Higgs candidate masses and data-taking year, $\vec{y} = (m_{H1}, m_{H2}, \text{year})$.

These flows are designed to capture all correlations among the variables of interest, and their conditioning on both mass values and year makes this modeling task more complex than in previous analyses. Hyperparameters and training options for the flows, such as the number of layers, hidden dimensions, and set of input variables, can be tuned, allowing flexible adaptation of the model architecture and training strategy.

The variables used in flow training, conditioning, and downstream reconstruction are summarized in Table 9.2. Specifically, the flows are trained to model the conditional density

$$p(\vec{x} \mid \vec{y}),$$

where $\vec{x}$ includes $p_{T,H1}$, $p_{T,H2}$, η_{H1}, η_{H2}, $\Delta\phi_{HH}$, and X_{Wt}, while $\vec{y}$ comprises m_{H1}, m_{H2}, and the data-taking year. Event-level observables such as m_{HH}, $p_{T,HH}$, and $\Delta\eta_{HH}$ are then reconstructed from the sampled outputs using four-vector algebra.

Figures 9.6 and 9.7 show example distributions for various kinematic variables in the CR, comparing the 4b data and the normalizing flows prediction. The method provides excellent modeling of all input and reconstructed variables across all data-taking years.

The two-dimensional Gaussian process interpolation for the SR, as performed for the 2022 data, is illustrated in Figure 9.8. To ensure robustness of the background prediction, an ensemble approach is employed: 25 independent trainings are performed for both the normalizing flows and the Gaussian process, and the spread among these trainings is used as a systematic uncertainty. The mean of the ensemble is adopted as the nominal prediction, while the standard deviation is visualized as an uncertainty band in the relevant one-dimensional histograms.

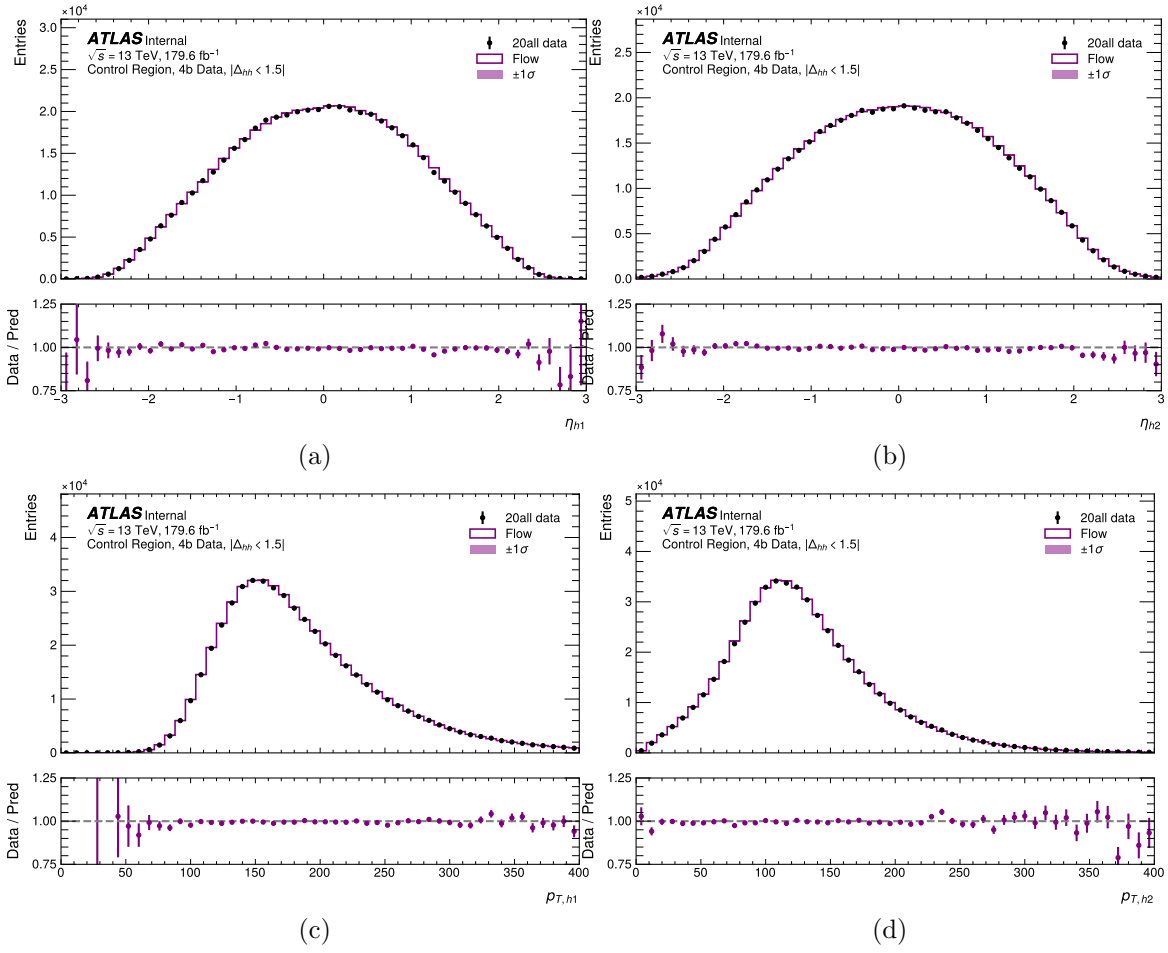


Figure 9.6: Distributions of 4b data and normalizing flow predictions for all data-taking years (2016–2018, 2022–2023) in the control region: (a) η_{H1}, (b) η_{H2}, (c) $p_{T,H1}$, and (d) $p_{T,H2}$. There is overall great closure between data and prediction.

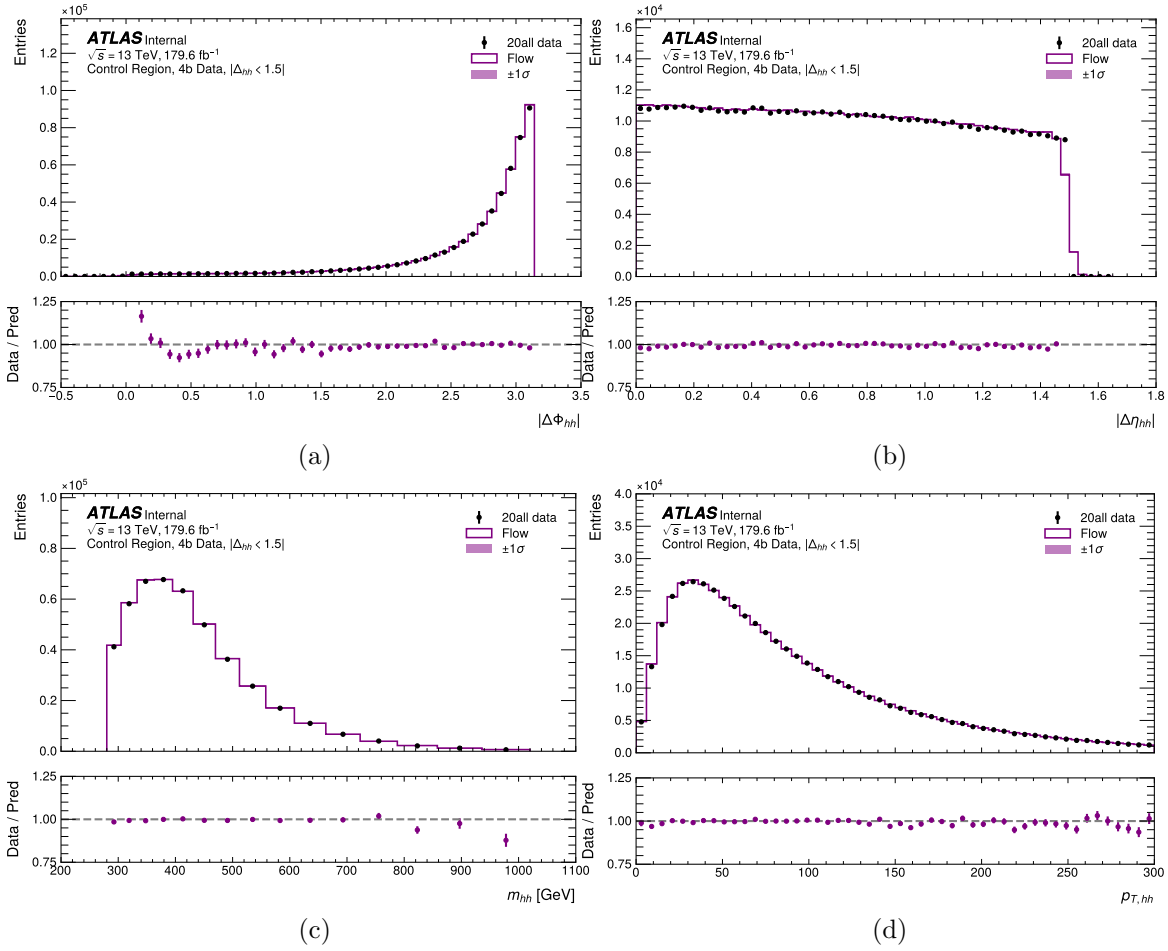
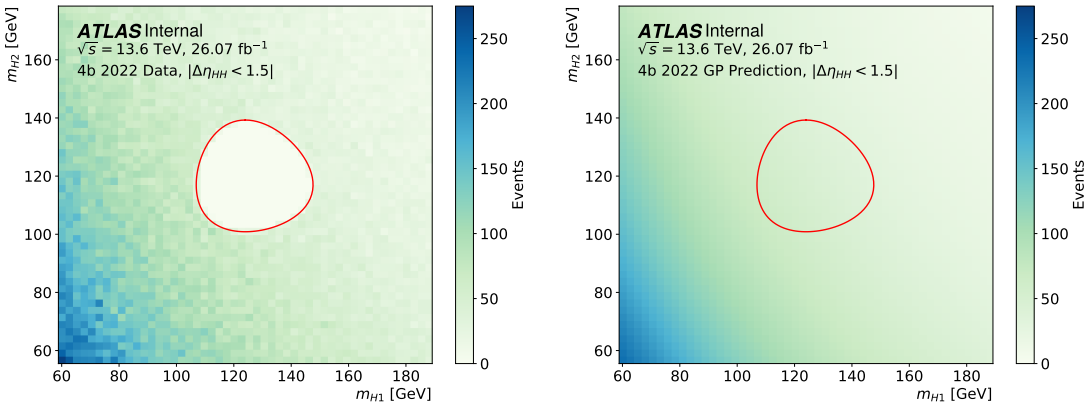


Figure 9.7: Distributions of 4b data and normalizing flow predictions for all data-taking years (2016–2018, 2022–2023) in the control region: (a) $|\Delta\phi_{HH}|$, (b) $|\Delta\eta_{HH}|$, (c) m_{HH}, and (d) $p_{T,HH}$. There is overall great closure between data and prediction outside of the boundary of the $|\Delta\eta_{HH}|$ cut.

Table 9.2: Kinematic variables used in normalizing flow training, conditioning, and downstream reconstruction for the background estimate.

Flow Training Variables	**Description**
$p_{T,H1}$	Transverse momentum of leading Higgs candidate
$p_{T,H2}$	Transverse momentum of subleading Higgs candidate
η_{H1}	Pseudorapidity of leading Higgs candidate
η_{H2}	Pseudorapidity of subleading Higgs candidate
$\Delta\phi_{HH}$	Azimuthal angle difference between Higgs candidates
X_{Wt}	Top veto discriminant
Flow Conditioning Variables	**Description**
m_{H1}	Leading Higgs candidate mass
m_{H2}	Subleading Higgs candidate mass
Year	Data collection year
Reconstructed Variables	**Description**
m_{HH}	HH system mass
$p_{T,HH}$	HH system transverse momentum
$\Delta\eta_{HH}$	Pseudorapidity difference between Higgs candidates



(a) (b)

Figure 9.8: The distribution of the (a) data and the (b) average GP prediction over the (m_{H1}, m_{H2}) mass plane for the 4b data in 2022 for events where $|\Delta\eta_{HH}| < 1.5$. The 4b data is blinded in the SR (red) of (a).

9.4.1 Background Validation

Because the analysis remains blinded in the SR to prevent potential bias, the robustness and accuracy of the background estimation are established through rigorous testing in well-defined validation regions constructed by systematic modifications to the nominal event selection.

Establishing these validation regions is essential for assessing the quality of the background model. Five principal configurations are considered, as illustrated in Figure 9.9. These regions are obtained by inverting a key kinematic selection (requiring $|\Delta\eta_{HH}| > 1.5$ instead of < 1.5), altering the b-tag multiplicity (e.g., selecting exactly two or three b-tagged jets), or combining both approaches. These regions serve as critical cross-checks of the method in data samples with kinematic and flavor compositions distinct from the SR.

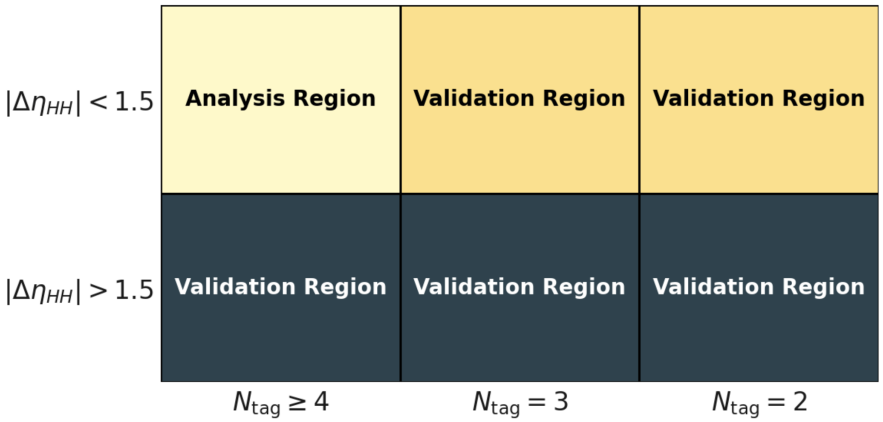


Figure 9.9: The additional validation regions used to test the data-driven background estimate. These regions are obtained by inverting the $|\Delta\eta_{HH}| < 1.5$ kinematic cut, altering the b-tag multiplicity (N_{tag}), or both.

While extensive studies have been performed for all these validation regions, we focus here on the $3b$ region with $|\Delta\eta_{HH}| < 1.5$ since it is kinematically most similar to the SR. The performance of the Gaussian process in this region is exemplary: it achieves very good closure in the total yield in the SR and accurately models the (m_{H1}, m_{H2}) mass plane, as

evidenced by the ratio of data to prediction and the difference between data and prediction normalized to the data's statistical uncertainty, as shown in Figure 9.10.

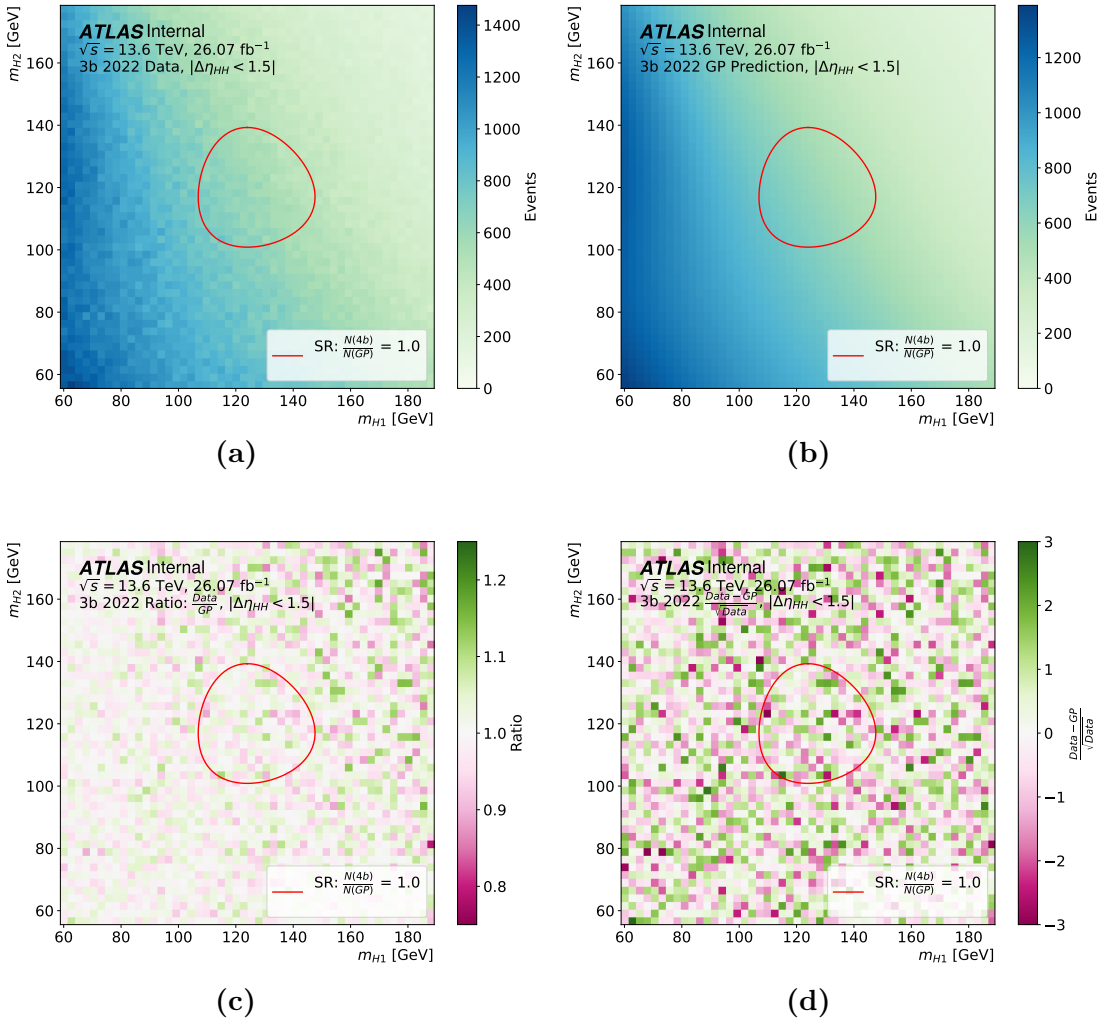
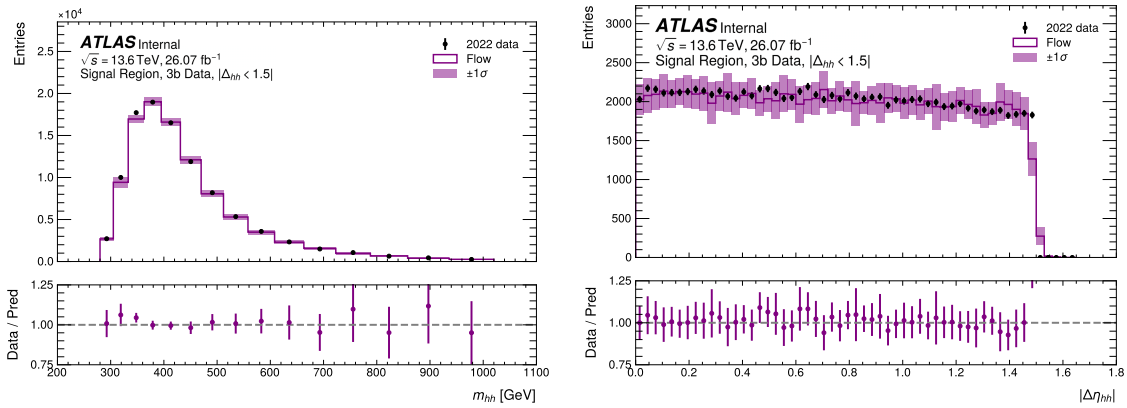


Figure 9.10: Binned distributions in the (m_{H1}, m_{H2}) mass plane for the $3b$ validation region with $|\Delta\eta_{HH}| < 1.5$ in 2022: (a) observed data, (b) GP-based background prediction, (c) ratio of data to prediction, and (d) the difference between data and prediction normalized to the statistical uncertainty in the data. The agreement between data and prediction is excellent across the phase space. The signal region ($X_{HH} < 1.6$) is indicated by the red contour.

To further validate the background modeling, we examine the performance of the combined normalizing flows and Gaussian processes for additional key kinematic variables in the SR, particularly m_{HH} and $|\Delta\eta_{HH}|$, shown in Figure 9.11. The agreement is very good, except near the boundary imposed by the $|\Delta\eta_{HH}| < 1.5$ selection.



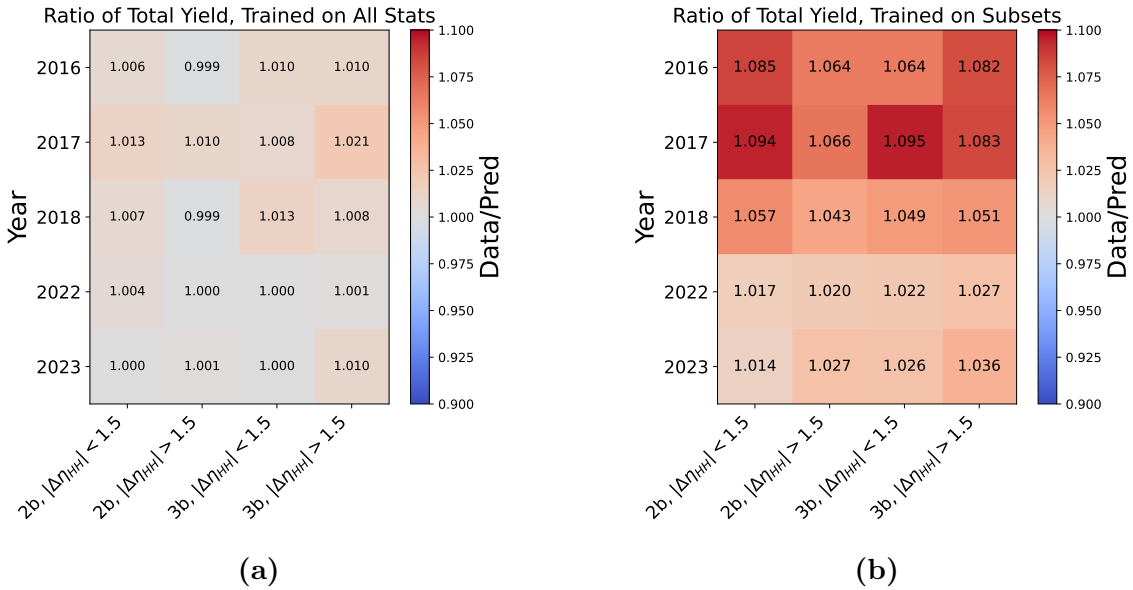
(a) (b)

Figure 9.11: The distribution of the data and the background prediction for (a) m_{HH} and (b) $|\Delta\eta_{HH}|$ for the $3b$ validation region with $|\Delta\eta_{HH}| < 1.5$ in 2022 in the SR. The modeling is generally well within the uncertainty of the standard deviation of the 25 predictions outside of the kinematic boundary.

While these results are extremely promising, it must be emphasized that they are obtained with the full statistics available in the $3b$ validation region with $|\Delta\eta_{HH}| < 1.5$. Because machine learning algorithms typically perform better with greater available statistics, it is not guaranteed that the modeling in the nominal $4b$ region, where the available statistics are about one-tenth as large, will be as strong. To address this, the normalizing flows and Gaussian processes were also trained using downsampled datasets, where the $3b$ ($2b$) regions are trained on only 10% (1%) of their available statistics, with the resulting predictions normalized back to the full dataset yields. This procedure was repeated for 10 unique subsets of each validation region, resulting in a total of 250 separate background

predictions per region (10 subsets, each with 25 independent trainings).

Rather than plot all 250 predictions, the evaluation of closure is performed using the average of the 25 predictions within each subset, and then averaging these 10 subset results. Figure 9.12 shows the overall ratio of data to prediction for the SR in each 2b and 3b validation region, comparing the models trained with full statistics (panel a) to those trained on the downsampled subsets (panel b). The Run 2 years show a notable under-prediction in the total yield by up to 10% in the most extreme cases, while the modeling for Run 3 is improved though still not perfect.



(a) **(b)**

Figure 9.12: Ratio of the total data yield to the predicted background yield in the signal region for each $2b$ and $3b$ validation region, shown separately for all data-taking years. Panel (a) shows the result when the normalizing flows and Gaussian processes are trained on the full statistics, while panel (b) presents the average ratio from models trained on 10 downsampled subsets (each averaged over 25 seeds). The comparison illustrates the degree of non-closure in each region, with a pronounced under-prediction of the total yield for Run 2 years, reaching up to 10% in the most extreme case, while the agreement is notably improved for the Run 3 years though still not perfect.

The observed non-closure in the total yield between the data and background prediction

is visible across the entire phase space for the kinematic variables in the validation regions. This outcome suggests that further optimization of the machine learning hyperparameters is required, particularly for bins with low statistics, as is the case in the nominal $4b$ region. Several improvements are currently underway, including widening the binning of the mass plane to increase statistical power in each bin and the application of an Anscombe transformation as a variance-stabilizing transformation for Poisson-distributed data. The Anscombe transformation is designed to render the variance approximately constant across different count values, thereby improving the reliability of the GP interpolation and normalizing flow training in low-statistics regimes [159]. Such modifications are expected to improve both the closure and the uncertainty estimates in the most challenging regions.

9.5 Preliminary Results

It is important to emphasize that the analysis is still in progress and the evaluation of systematic uncertainties is ongoing. While theoretical uncertainties are already available, the implementation of experimental systematics for both background and signal is still under development. Background systematics are of particular concern, as the previous Run 2 analysis showed that they were the dominant source of experimental uncertainty. Consequently, significant effort is being dedicated to establishing a systematic uncertainty treatment that can both accurately quantify and, where possible, reduce the total background uncertainty in the statistical fit. This priority has directly motivated the background modeling improvements discussed in the preceding sections.

9.5.1 Final Discriminant

The current baseline for the final discriminant follows the Run 2 analysis by categorizing events in m_{HH} according to their X_{HH} and $|\Delta\eta_{HH}|$ values. This strategy enhances sensitivity by partitioning the analysis phase space such that regions with different signal-to-background ratios contribute most effectively to the overall search.

Six analysis categories are defined for the m_{HH} distribution:

- Category 0: $0.0 < |\Delta\eta_{HH}| < 0.5$, $0.0 < X_{HH} < 0.95$
- Category 1: $0.5 < |\Delta\eta_{HH}| < 1.0$, $0.0 < X_{HH} < 0.95$
- Category 2: $1.0 < |\Delta\eta_{HH}| < 1.5$, $0.0 < X_{HH} < 0.95$
- Category 3: $0.0 < |\Delta\eta_{HH}| < 0.5$, $0.95 < X_{HH} < 1.6$
- Category 4: $0.5 < |\Delta\eta_{HH}| < 1.0$, $0.95 < X_{HH} < 1.6$
- Category 5: $1.0 < |\Delta\eta_{HH}| < 1.5$, $0.95 < X_{HH} < 1.6$

In parallel, a transformer-based multivariate discriminant has been developed to distinguish signal events from the dominant backgrounds, namely multi-jet QCD and $t\bar{t}$ production. The model was trained using the same input variables listed in Table 9.2 that are used for the data-driven background estimation. This choice is essential, as the generative background modeling with normalizing flows and Gaussian processes only allows reliable estimation of variables included in the training and prediction stages within the signal region. While the transformer shows promising potential, its implementation and optimization are not the focus of this thesis and are subject to further development by other members of the analysis team; as such, the exact details are not presented here. Current studies are focused on comparing the performance of the transformer discriminant with the categorized m_{HH} approach, assessing both statistical sensitivity and robustness to systematic uncertainties. The transformer, in particular, requires additional validation to ensure compatibility with the background estimation framework, while the bin boundaries for the categorized strategy are being revisited and optimized for the current dataset rather than relying solely on Run 2 definitions.

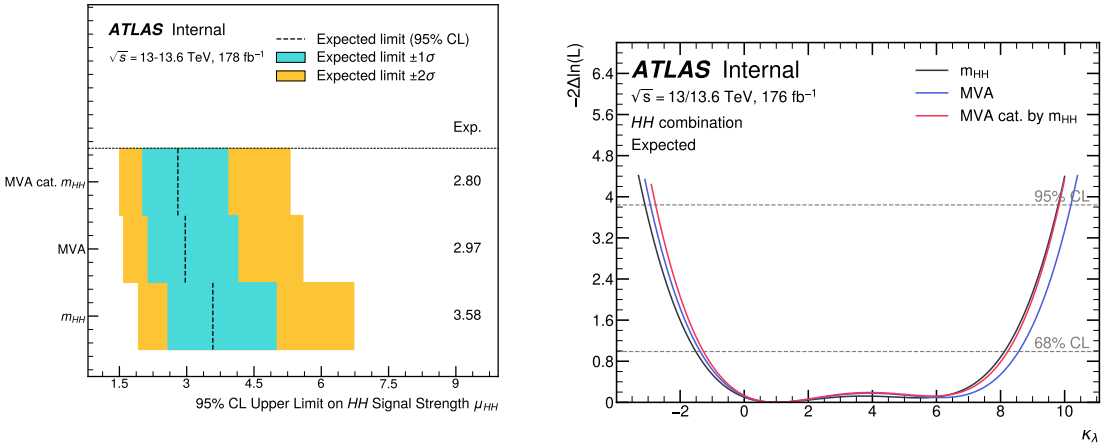
9.5.2 Statistical Limits

The improvement in the expected 95% confidence level upper limits on the signal strength of non-resonant $HH \to 4b$ production, from 6.02 with m_{HH} as the discriminant in the Run 2 analysis, see Table 8.3, to 3.58 with categorized m_{HH} as the discriminant discriminant in the combined Run 2 and partial Run 3 analysis, demonstrates the substantial gains achieved through methodological advancements. With only a 40% increase in integrated luminosity (which would naively reduce the limit to about 4.8), the observed improvement is much greater. This enhanced sensitivity is the result of multiple factors, including improved jet energy and mass resolution, the deployment of the state-of-the-art GN2 b-tagging algorithm, and a more accurate background estimation framework using normalizing flows and Gaussian processes. Collectively, these developments have more than doubled the analysis sensitivity compared to what would be expected from luminosity scaling alone, highlighting the critical impact of continued innovation in experimental techniques.

It is important to note, however, that the Run 2 limit referenced above was used m_{HH} without categorization as the sole discriminant. Using the categorized m_{HH} discriminant would have yielded a stronger limit in the study presented in Section 8.2 as well, so direct comparison between the two results should be interpreted with care.

Preliminary studies with the transformer-based MVA discriminant suggest an additional 17% improvement in the statistics-only limit, reaching a 95% CL upper limit of 2.97. Figure 9.13 compares the expected limits and κ_λ constraints from the categorized m_{HH} and the transformer-based MVA. While the details of the MVA approach are still under development and are beyond the scope of this thesis, the results are encouraging. If systematic uncertainties can be controlled with this discriminant, further improvements over the Run 2-only results appear attainable. There is even additional 5.7% sensitivity gain if the MVA discriminant is categorized into a low-mass and high-mass region of m_{HH} with the boundary at 450 GeV.

As these results are preliminary, with systematics still being incorporated and some workflow steps undergoing further optimization, additional changes are expected before the final analysis. Nonetheless, the outlook is highly promising. Further studies targeting sensitivity improvements in a full Run 3 analysis, particularly in the low-mass kinematic regime, are presented in Appendix A.



(a) (b)

Figure 9.13: (a) Expected 95% CL upper limits on the signal strength for non-resonant $HH \to b\bar{b}b\bar{b}$ production via ggF in the resolved channel, shown for three discriminants: categorized m_{HH}, MVA, and MVA categorized in m_{HH}. The median expected limit is presented along with the $\pm 1\sigma$ and $\pm 2\sigma$ uncertainty bands, assuming no signal contribution. (b) Expected 68% and 95% confidence level upper limits on non-resonant $HH \to b\bar{b}b\bar{b}$ production as a function of κ_λ, for each discriminant strategy with all other couplings fixed to their Standard Model values.

Chapter 10

CONCLUSION

10.1 Summary

This thesis presents a comprehensive body of work spanning jet calibration and the development and execution of Higgs boson pair production searches within the ATLAS Experiment. In the jet definitions and calibrations subgroup, I developed dedicated calibrations for small-R UFO jets in AF3 samples, as well as systematic uncertainty prescriptions for both small-R and large-R PFlow jets in AF3 samples for Run 2 and Run 3. These efforts have been essential for ensuring the reliability and precision of reconstructed jet observables in the fast simulation for current and future analyses.

My contributions to the Run 2 non-resonant $HH \to b\bar{b}b\bar{b}$ analysis were instrumental in defining the background systematic uncertainties and informing the selection of the final discriminant for the resolved ggF channel. This analysis ultimately delivered the third most sensitive decay channel in the ATLAS combination, advancing the long-term goal of probing and measuring the Higgs boson self-coupling.

As a primary analyst for the Run 2 resonant $X \to SH \to b\bar{b}b\bar{b}$ search, I was responsible for a broad range of critical tasks, including Monte Carlo validation, jet pairing, optimization of kinematic selections, and the estimation and validation of backgrounds. In addition, I served as an editor for the internal analysis note, organizing and documenting the workflow throughout the process. This coordinated effort culminated in the first ATLAS results for this final state. Although no significant excess was observed within the phase space examined in the resolved topology, the analysis provides a solid foundation for future searches in this channel.

Finally, I contributed substantially to the background validation and systematic uncertainty assessments of the non-resonant $HH \to b\bar{b}b\bar{b}$ search using the Run 2 and early Run 3 data. The results obtained from the statistical-only limits are particularly promising,

indicating a substantial improvement over Run 2 and suggesting even greater sensitivity upon completion. My studies targeting enhanced sensitivity to low-mass events have laid the groundwork for future iterations of the analysis with the complete Run 3 dataset, as discussed in Appendix A.

10.2 Future Outlook

Though no significant global excesses have been observed in Run 2 analyses, the sensitivity of both non-resonant and resonant searches for Higgs pair production is expected to increase substantially with the accumulation of more data along with ongoing improvements in reconstruction, machine learning techniques, and analysis strategies. The broader significance of Higgs physics is illustrated in Figure 10.1, which highlights the central role the Higgs boson plays in addressing fundamental open questions in particle physics. These include the origin of mass, the structure of the scalar sector, the stability of the vacuum, connections to dark matter, and the potential for discovering supersymmetric or composite states.

In preparation for the HL-LHC era, substantial efforts have been made by the high energy physics community to project future sensitivities as input to the 2026 European Strategy Update for Particle Physics. Of particular relevance to this thesis are projections for non-resonant Higgs boson pair production which is expected to yield approximately 10^5 di-Higgs events. Extrapolating from Run 2 results, and assuming a dataset of 3000 fb^{-1} at a center-of-mass energy of $\sqrt{s} = 14$ TeV (representing an increase of 18% and 19% in the gluon fusion and vector boson fusion production cross sections, respectively), the ATLAS combination analysis anticipates a significance of 4.26σ for probing the Higgs self-coupling under a baseline systematic scenario with moderate improvements over Run 2 uncertainties. Specifically, the $b\bar{b}b\bar{b}$ final state is projected to yield a 1.0σ sensitivity to the trilinear coupling modifier κ_λ with expected exclusion limits in the range $[1.6, 7.5]$ at 95% confidence level [160].

An observation of non-resonant Higgs boson pair production as expected in the SM is anticipated with the combination of channels across ATLAS and CMS once the integrated luminosity surpasses 2000 fb^{-1}. With the full HL-LHC dataset of 3000 fb^{-1}, the combined ATLAS and CMS analyses are expected to achieve a 7.2σ significance and constrain κ_λ with

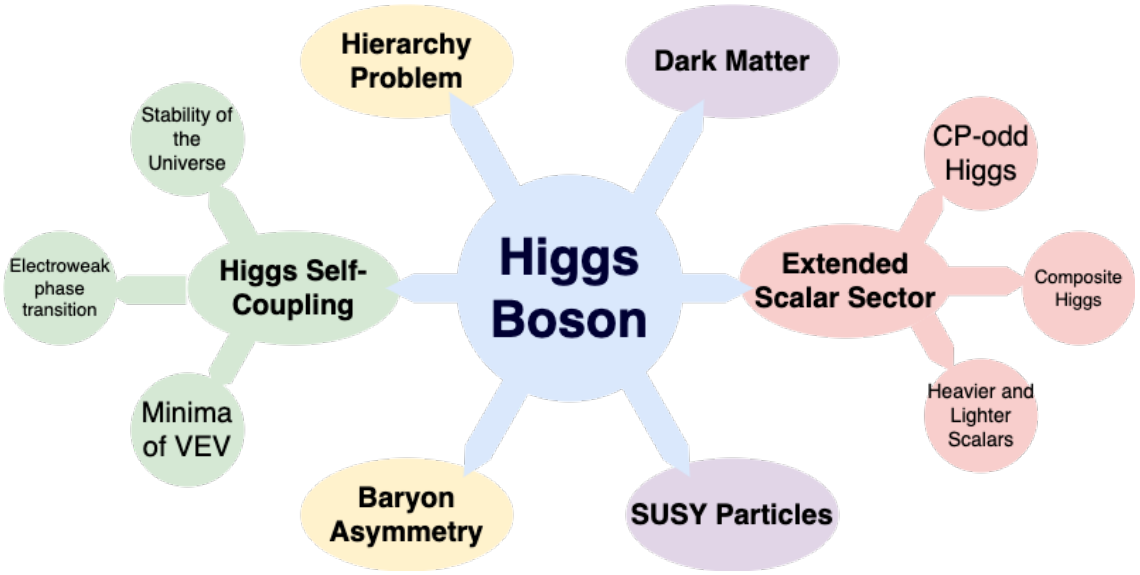


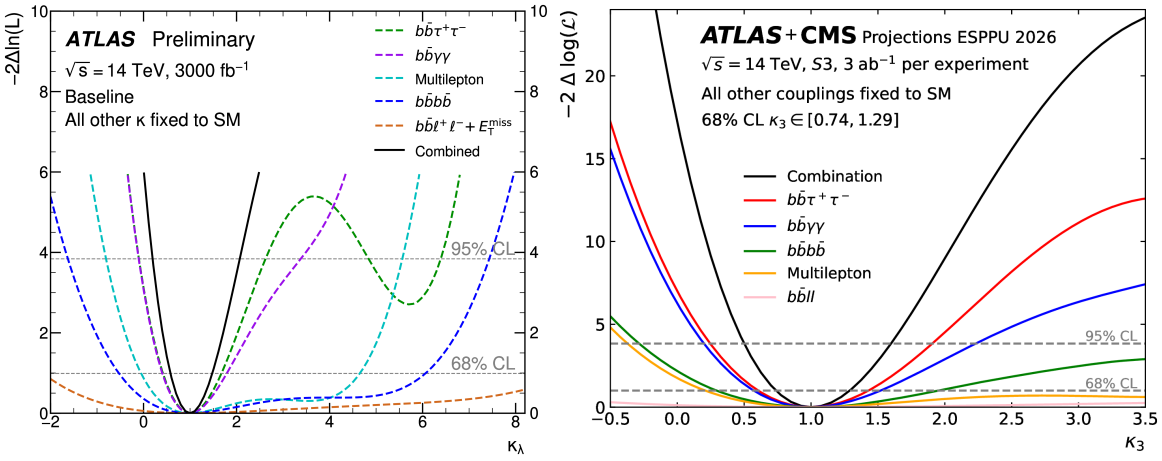
Figure 10.1: The central role of the Higgs boson in modern particle physics. The Higgs field not only gives mass to elementary particles via electroweak symmetry breaking but also connects to a broad array of theoretical and experimental frontiers, including Higgs self-coupling, dark matter, baryon asymmetry, supersymmetric particles, the stability of the vacuum, composite Higgs models, and extensions to the scalar sector. Each branch highlights a key area where the Higgs boson could provide insight into physics beyond the Standard Model.

a precision better than 30%, as illustrated in Figure 10.2 [161].

10.2.1 Possible Improvements for 4b Analyses

Future advances in the sensitivity of $4b$ analyses will depend on a combination of increased data, technical upgrades, and innovative analysis techniques. Several areas stand out as promising directions for improvement, many of which are active fields of research within ATLAS and the wider LHC community:

- **Advanced Machine Learning for Background Estimation:** The challenge of accurately modeling the overwhelming multi-jet QCD background in the $4b$ final state



(a) (b)

Figure 10.2: Expected values of the negative log-likelihood ratio as a function of κ_λ for an integrated luminosity of 3000 fb-1 at $\sqrt{s} = 14$ TeV for the ATLAS (ATLAS+CMS) combination of different channels and their statistical combination in the baseline scenario in Figure 10.2a (10.2b) [160, 161].

remains central. Building on recent progress, even more sophisticated ML models, such as graph generative models and foundational models, can be leveraged for data-driven background estimation. These approaches have the potential to learn complex, high-dimensional correlations directly from data, reducing modeling uncertainties and improving background rejection.

- **Statistical Uncertainty Reduction:** Statistical uncertainties in both background and signal modeling can be mitigated by utilizing larger training samples from fast simulation frameworks (like AF3), which enable the production of high-statistics MC samples with reduced computational cost. This is particularly relevant for training ML-based background models or calibrating rare signal processes.

- **More Precise Simulation and Data-Driven Methods:** Improvements in theoretical calculations, parton shower modeling, and detector simulation fidelity will fur-

ther reduce systematic uncertainties. Enhanced MC event generators and increased use of data-driven techniques (e.g., control regions, sidebands, and hybrid MC/data approaches) can yield more accurate background estimates and validation of the modeling.

- **Improved b-Tagging and Jet Calibration:** Upgrades to b-tagging algorithms include exploiting deep learning, high-granularity tracking, and new detector layers, which will boost signal efficiency and purity. Parallel improvements in jet energy scale and resolution calibrations, especially with ML-driven or in-situ methods, are expected to directly reduce uncertainties and increase search sensitivity [162, 163].

- **Trigger Innovations:** Lowering jet p_T thresholds in high-rate environments is a major challenge. Recent studies demonstrate that ML-based triggers, including online deep learning pre-selection, can maintain high signal efficiency while controlling rates, thereby enabling the retention of more signal-like events in future runs.

- **Systematic Uncertainty Reduction:** Targeted studies and calibration campaigns can further reduce the dominant systematic uncertainties, particularly those related to jet energy scale, b-tagging, and background modeling. As uncertainties shrink, analyses will become increasingly limited by statistical rather than systematic effects.

- **Broader and More Precise BSM Searches:** For BSM searches and interpretations, several directions are actively being explored:

  - Expanding resonant scalar searches to cover wider resonance mass and width hypotheses, including $X \to SS$ and non-minimal scalar scenarios.
  - Incorporating model-dependent width effects, such as large or variable resonance widths, to improve sensitivity to non-narrow signals.
  - Leveraging full Run 3 and HL-LHC datasets to probe more subtle BSM effects and rare signatures.

Together, these improvements, supported by ongoing hardware upgrades, theoretical developments, and methodological innovation, will continue to drive the $4b$ analyses toward higher precision and greater discovery potential in both Standard Model and BSM searches for Higgs boson pair production.

10.3 Contributions

The following subsections present the publications that my work has contributed towards, along with the talks, posters, conferences, and workshops in which I have been involved.

10.3.1 Publications

- ***Expected***: *ATLAS* Collaboration. Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using the Run 2 and partial Run 3 pp collisions by the ATLAS detector. Late 2025.

- ***Expected***: *ATLAS* Collaboration. Search for new resonances decaying to a scalar and a Standard Model-like Higgs boson in the $b\bar{b}b\bar{b}$ final state using the Run 2 pp collisions by the ATLAS detector. Fall 2025.

- ***Expected***: *ATLAS* Collaboration. Data and Monte Carlo studies of final states containing jets using pp collisions at $\sqrt{s} = 13.6$ TeV with the ATLAS detector. Fall 2025.

- *ATLAS* Collaboration. Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Physical Review D*, 2023.

10.3.2 Talks, Posters, Conferences, Workshops

- *Jet and missing transverse momentum performance at ATLAS and CMS*, Higgs Pairs 2025, Isola d'Elba, Italy, 2025.05.

- *Jet Definitions and Calibrations (Current Calibrations)*, Session Co-Chair, Hadronic Calibration Workshop, Ottawa, Canada, 2024.08.

- *Search for a massive scalar resonance decaying to a lighter scalar and a Higgs boson in the 4b final state with the resolved topology*, HDBS Workshop, Nikhef, Amsterdam, Netherlands, 2024.07.

- *Ongoing Run 2 Efforts in the X to SH to bbbb Analysis*, ATLAS Di-Higgs Workshop, CERN, Meyrin, Switzerland, 2024.01.

- *Calibration of Small-R Jets in R22 for AtlFast3*, Hadronic Calibration Workshop, Valencia, Spain, 2023.09.

BIBLIOGRAPHY

[1] W. N. Cottingham and D. A. Greenwood. *An Introduction to the Standard Model of Particle Physics*. 2nd. Cambridge University Press, 2007. ISBN: 9780521852494.

[2] Mary K. Gaillard et al. "The Standard Model of Particle Physics". In: *Reviews of Modern Physics* 71.2 (Mar. 1999), S96–S111. ISSN: 1539-0756. DOI: 10.1103/revmodphys.71.s96. URL: http://dx.doi.org/10.1103/RevModPhys.71.S96.

[3] John R. Ellis. *Limits of the Standard Model*. Tech. rep. Geneva: CERN, 2002. URL: https://cds.cern.ch/record/591472.

[4] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. URL: https://arxiv.org/abs/1207.7214.

[5] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. URL: https://arxiv.org/abs/1207.7235.

[6] I. Gil-Botella. *The Deep Underground Neutrino Experiment (DUNE) program*. 2024. arXiv: 2412.14941 [hep-ex]. URL: https://arxiv.org/abs/2412.14941.

[7] Christopher W. Walter. "The Super-Kamiokande Experiment". In: *Neutrino Oscillations*. WORLD SCIENTIFIC, Mar. 2008, pp. 19–43. ISBN: 9789812771971. DOI: 10.1142/9789812771971_0002. URL: http://dx.doi.org/10.1142/9789812771971_0002.

[8] C. La Licata. "Belle II Prospects for CP-violation Measurements". In: *Acta Physica Polonica B* 49.6 (2018), p. 1229. ISSN: 1509-5770. DOI: 10.5506/aphyspolb.49.1229. URL: http://dx.doi.org/10.5506/APhysPolB.49.1229.

[9] LHCb Collaboration. *Observation of charge-parity symmetry breaking in baryon decays*. Tech. rep. Submitted to Nature. Geneva: CERN, 2025. arXiv: 2503.16954. URL: https://cds.cern.ch/record/2927827.

[10] Muon g-2 Collaboration. "Measurement of the Positive Muon Anomalous Magnetic Moment to 0.20 ppm". In: *Physical Review Letters* 131.16 (Oct. 2023). ISSN: 1079-7114. DOI: 10.1103/physrevlett.131.161802. URL: http://dx.doi.org/10.1103/PhysRevLett.131.161802.

[11] Douglas Clowe et al. "A Direct Empirical Proof of the Existence of Dark Matter". In: *The Astrophysical Journal* 648.2 (Aug. 2006), pp. L109–L113. ISSN: 1538-4357. DOI: 10.1086/508162. URL: http://dx.doi.org/10.1086/508162.

[12] Gianfranco Bertone and Dan Hooper. "History of dark matter". In: *Reviews of Modern Physics* 90.4 (Oct. 2018). ISSN: 1539-0756. DOI: 10.1103/revmodphys.90.045002. URL: http://dx.doi.org/10.1103/RevModPhys.90.045002.

[13] S. F. Novaes. *Standard Model: An Introduction*. 2000. arXiv: hep-ph/0001283 [hep-ph]. URL: https://arxiv.org/abs/hep-ph/0001283.

[14] Fernando Quevedo and Andreas Schachner. *Cambridge Lectures on The Standard Model*. 2024. arXiv: 2409.09211 [hep-th]. URL: https://arxiv.org/abs/2409.09211.

[15] A.B. Arbuzov. "Quantum Field Theory and the Electroweak Standard Model". In: (2018). 35 pages, pp. 1–34. DOI: 10.23730/CYRSP-2017-004.1. arXiv: 1801.05670. URL: https://cds.cern.ch/record/2315477.

[16] Zeinab Dehghan and Manfried Faber. "What do we know about the confinement mechanism?" In: *arXiv preprint arXiv:2412.10767* (2024). arXiv: 2412.10767 [hep-lat].

[17] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[18] Steven Weinberg. "A Model of Leptons". In: *Physical Review Letters* 19.21 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[19] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[20] Claudio Giganti et al. "Neutrino oscillations: the rise of the PMNS paradigm". In: *Progress in Particle and Nuclear Physics* 98 (2018), pp. 1–54. DOI: 10.1016/j.ppnp.2017.10.001. arXiv: 1710.00715 [hep-ex].

[21] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Physical Review Letters* 13.16 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://doi.org/10.1103/PhysRevLett.13.508.

[22] François Englert and Robert Brout. "Broken Symmetry and the Mass of Gauge Bosons". In: *Physical Review Letters* 13.9 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://doi.org/10.1103/PhysRevLett.13.321.

[23] ATLAS Collaboration. "Evidence for light-by-light scattering in heavy-ion collisions with the ATLAS detector at the LHC". In: *Nature Phys.* 13 (2017), p. 852. DOI: 10.1038/nphys4208. arXiv: 1702.01625 [hep-ex].

[24] V H Satheesh Kumar and P K Suresh. *Gravitons in Kaluza-Klein Theory*. 2006. arXiv: gr-qc/0605016 [gr-qc]. URL: https://arxiv.org/abs/gr-qc/0605016.

[25] A. Zee. *Group Theory in a Nutshell for Physicists*. Princeton, NJ: Princeton University Press, 2016. ISBN: 9780691162690.

[26] Howard Georgi. *Lie Algebras in Particle Physics: From Isospin to Unified Theories*. Boca Raton, FL: CRC Press, 2018. ISBN: 9780429499210.

[27] Mark Srednicki. *Quantum Field Theory*. An introduction to quantum field theory. Cambridge, UK: Cambridge University Press, 2007. ISBN: 9780521864497.

[28] Mark Thomson. *Modern Particle Physics*. Cambridge, UK: Cambridge University Press, 2013. ISBN: 9781107034266.

[29] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge, UK: Cambridge University Press, 2014. ISBN: 9781107034730.

[30] J. J. Aubert et al. "Experimental Observation of a Heavy Particle J". In: *Phys. Rev. Lett.* 33 (1974), pp. 1404–1406. DOI: 10.1103/PhysRevLett.33.1404.

[31] J. E. Augustin et al. "Discovery of a Narrow Resonance in e^+e^- Annihilation". In: *Phys. Rev. Lett.* 33 (1974), pp. 1406–1408. DOI: 10.1103/PhysRevLett.33.1406.

[32] S. W. Herb et al. "Observation of a Dimuon Resonance at 9.5 GeV in 400 GeV Proton-Nucleus Collisions". In: *Phys. Rev. Lett.* 39 (1977), pp. 252–255. DOI: 10.1103/PhysRevLett.39.252.

[33] PLUTO Collaboration. "Observation of Narrow Resonances at 9.46 and 10.02 GeV in e^+e^- Annihilation at DORIS". In: *Phys. Lett. B* 78 (1978). PLUTO and DASP2 at DORIS, pp. 361–364.

[34] G. Arnison et al. "Experimental observation of isolated large transverse energy electrons with associated missing energy at $\sqrt{s} = 540$ GeV". In: *Phys. Lett. B* 122 (1983), pp. 103–116. DOI: 10.1016/0370-2693(83)91177-2.

[35] M. Banner et al. "Observation of Single Isolated Electrons of High Transverse Momentum in Events with Missing Transverse Energy at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 122 (1983), pp. 476–485. DOI: 10.1016/0370-2693(83)91605-2.

[36] G. Arnison et al. "Experimental observation of lepton pairs of invariant mass around 95 GeV/c^2 at the CERN SPS collider". In: *Phys. Lett. B* 126 (1983), pp. 398–410. DOI: 10.1016/0370-2693(83)90188-0.

[37] P. Bagnaia et al. "Evidence for $Z^0 \to e^+e^-$ at the CERN $\bar{p}p$ Collider". In: *Phys. Lett. B* 129 (1983), pp. 130–140. DOI: 10.1016/0370-2693(83)90744-X.

[38] R. Brandelik et al. "Evidence for Planar Events in e^+e^- Annihilation at High Energies". In: *Phys. Lett. B* 86 (1979), pp. 243–249. DOI: 10.1016/0370-2693(79)90534-9.

[39] D0 Collaboration. "Observation of the Top Quark". In: *Physical Review Letters* 74.14 (Apr. 1995), pp. 2632–2637. ISSN: 1079-7114. DOI: 10.1103/physrevlett.74.2632. URL: http://dx.doi.org/10.1103/PhysRevLett.74.2632.

[40] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett. B* 716 (2012), p. 1. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[41] CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett. B* 716 (2012), p. 30. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[42] CERN. *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*. en. 2017. DOI: 10.23731/CYRM-2017-002. URL: https://e-publishing.cern.ch/index.php/CYRM/issue/view/32.

[43] ATLAS Collaboration. "A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 52–59. DOI: 10.1038/s41586-022-04893-w. arXiv: 2207.00092 [hep-ex].

[44] CMS Collaboration. "A portrait of the Higgs boson by the CMS experiment ten years after the discovery". In: *Nature* 607 (2022), pp. 60–68. DOI: 10.1038/s41586-022-04892-x. arXiv: 2207.00043 [hep-ex].

[45] ATLAS Collaboration. "Differential cross-section measurements of Higgs boson production in the $H \to \tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: (2024). arXiv: 2407.16320 [hep-ex].

[46] ATLAS Collaboration. "Interpretations of the ATLAS measurements of Higgs boson production and decay rates and differential cross-sections in pp collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 11 (2024), p. 097. DOI: 10.1007/JHEP11(2024)097. arXiv: 2402.05742 [hep-ex].

[47] CMS Collaboration. "Measurement of the $t\bar{t}H$ and tH production rates in the $H \to b\bar{b}$ decay channel using proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$". In: (2024). arXiv: 2407.10896 [hep-ex].

[48] Csaba Balazs et al. "A Primer on Dark Matter". In: (Nov. 2024). arXiv: 2411.05062 [astro-ph.CO].

[49] Julius Wess and Bruno Zumino. "Supergauge Transformations in Four Dimensions". In: *Nuclear Physics B* 70 (1974), pp. 39–50. DOI: 10.1016/0550-3213(74)90355-1.

[50] ATLAS Collaboration. *ATLAS Luminosity Public Results for Run 2*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. Accessed: 2024-06-07. 2024.

[51] ATLAS Collaboration. *ATLAS Luminosity Public Results for Run 3*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3. Accessed: 2024-06-07. 2024.

[52] B. Abbott et al. "Production and integration of the ATLAS Insertable B-Layer". In: *JINST* 13 (2018), T05008. DOI: 10.1088/1748-0221/13/05/T05008. arXiv: 1803.00844 [physics.ins-det].

[53] ATLAS Collaboration. "Track reconstruction performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV". In: *ATL-PHYS-PUB-2015-018* (2015). https://cds.cern.ch/record/2037683.

[54] ATLAS Collaboration. "A study of the material in the ATLAS inner detector using secondary hadronic interactions". In: *JINST* 7 (2012), P01013. DOI: 10.1088/1748-0221/7/01/P01013. arXiv: 1110.6191 [hep-ex].

[55] ATLAS Collaboration. "Study of the material of the ATLAS inner detector for Run 2 of the LHC". In: *JINST* 12 (2017), P12009. DOI: 10.1088/1748-0221/12/12/P12009. arXiv: 1707.02826 [hep-ex].

[56] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[57] ATLAS LAr Collaboration and Margherita Spalla. *ATLAS LAr Calorimeter Performance in LHC Run 2*. Tech. rep. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2707741.

[58] ATLAS Collaboration. "Operation and performance of the ATLAS Tile Calorimeter in Run 2". In: *JINST* 19.03 (2024), P03011. DOI: 10.1088/1748-0221/19/03/P03011.

[59] ATLAS Collaboration. "Resolution of the ATLAS muon spectrometer monitored drift tubes in LHC Run 2". In: *JINST* 14 (2019), P09011. DOI: 10.1088/1748-0221/14/09/P09011. arXiv: 1906.12226 [hep-ex].

[60] ATLAS Collaboration. "Performance of the ATLAS RPC detector and Level-1 muon barrel trigger at $\sqrt{s} = 13\,\text{TeV}$". In: *JINST* 16 (2021), P07029. DOI: 10.1088/1748-0221/16/07/P07029. arXiv: 2103.01029 [physics.ins-det].

[61] ATLAS Collaboration. *ATLAS Liquid Argon Calorimeter Phase-I Upgrade: Technical Design Report*. ATLAS-TDR-022; CERN-LHCC-2013-017. 2013. URL: https://cds.cern.ch/record/1602230.

[62] ATLAS Collaboration. *ATLAS Forward Proton Phase-I Upgrade: Technical Design Report*. ATLAS-TDR-024; CERN-LHCC-2015-009. 2015. URL: https://cds.cern.ch/record/2017378.

[63] ATLAS Collaboration. *ATLAS Phase-II Upgrade Scoping Document*. CERN-LHCC-2015-020, LHCC-G-166. 2015. URL: https://cds.cern.ch/record/2055248.

[64] ATLAS Collaboration. *ATLAS Muon Spectrometer Phase-II Upgrade: Technical Design Report*. ATLAS-TDR-026; CERN-LHCC-2017-017. 2017. URL: https://cds.cern.ch/record/2285580.

[65] ATLAS Collaboration. *ATLAS LAr Calorimeter Phase-II Upgrade: Technical Design Report*. ATLAS-TDR-027; CERN-LHCC-2017-018. 2017. URL: https://cds.cern.ch/record/2285582.

[66] ATLAS Collaboration. *A High-Granularity Timing Detector for the ATLAS Phase-II Upgrade: Technical Design Report*. ATLAS-TDR-031; CERN-LHCC-2020-007. 2020. URL: https://cds.cern.ch/record/2719855.

[67] ATLAS Collaboration. "Operation of the ATLAS trigger system in Run 2". In: *JINST* 15 (2020), P10004. DOI: 10.1088/1748-0221/15/10/P10004. arXiv: 2007.12539 [physics.ins-det].

[68] ATLAS Collaboration. "The ATLAS trigger system for LHC Run 3 and trigger performance in 2022". In: *JINST* 19 (2024), P06029. DOI: 10.1088/1748-0221/19/06/P06029. arXiv: 2401.06630 [hep-ex].

[69] A. Rimoldi et al. "The simulation for the ATLAS experiment: Present status and outlook". In: *14th International Conference on Computing in High-Energy and Nuclear Physics*. 2005, pp. 227–230.

[70] Andy Buckley et al. *Monte Carlo event generators for high energy particle physics event simulation*. Tech. rep. Monte Carlo Community input to European Strategy Update. 2019. arXiv: 1902.01674. URL: https://cds.cern.ch/record/2657127.

[71] A. Rimoldi and A. Dell'Acqua. "The Full detector simulation for the ATLAS experiment: Status and outlook". In: *eConf* C0303241 (2003), TUMT001. arXiv: physics/0306086.

[72] ATLAS Collaboration. "Software and computing for Run 3 of the ATLAS experiment at the LHC". In: *Eur. Phys. J. C* 85 (2025), p. 234. DOI: 10.1140/epjc/s10052-024-13701-w. arXiv: 2404.06335 [hep-ex].

[73] ATLAS Collaboration. *Jet Reconstruction*. https://opendata.atlas.cern/docs/documentation/physic_objects/jets_reco. Accessed: 2025-05-06.

[74] ATLAS Collaboration. "Performance of pile-up mitigation techniques for jets in pp collisions at $\sqrt{s} = 8\,\text{TeV}$ using the ATLAS detector". In: *Eur. Phys. J. C* 76 (2016), p. 581. DOI: 10.1140/epjc/s10052-016-4395-z. arXiv: 1510.03823 [hep-ex].

[75] ATLAS Collaboration. *Constituent-level pile-up mitigation techniques in ATLAS*. ATLAS-CONF-2017-065. 2017. URL: https://cds.cern.ch/record/2281055.

[76] Burkhard Schmidt. "The High-Luminosity upgrade of the LHC: Physics and Technology Challenges for the Accelerator and the Experiments". In: *J. Phys.: Conf. Ser.* 706.2 (2016), p. 022002. DOI: 10.1088/1742-6596/706/2/022002. URL: https://cds.cern.ch/record/2263093.

[77] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[78] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *Eur. Phys. J. C* 77 (2017), p. 490. DOI: 10.1140/epjc/s10052-017-5004-5. arXiv: 1603.02934 [hep-ex].

[79] ATLAS Collaboration. "Jet reconstruction and performance using particle flow with the ATLAS Detector". In: *Eur. Phys. J. C* 77 (2017), p. 466. DOI: 10.1140/epjc/s10052-017-5031-2. arXiv: 1703.10485 [hep-ex].

[80] ATLAS Collaboration. "Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions". In: *Eur. Phys. J. C* 81 (2021), p. 334. DOI: 10.1140/epjc/s10052-021-09054-3. arXiv: 2009.04986 [hep-ex].

[81] ATLAS Collaboration. *Improving jet substructure performance in ATLAS using Track-CaloClusters*. ATL-PHYS-PUB-2017-015. 2017. URL: https://cds.cern.ch/record/2275636.

[82] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "SoftKiller, a particle-level pileup removal method". In: *Eur. Phys. J. C* 75.2 (2015), p. 59. DOI: 10.1140/epjc/s10052-015-3267-2. arXiv: 1407.0408 [hep-ph].

[83] ATLAS Collaboration. *The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim*. ATL-PHYS-PUB-2010-013. 2010. URL: https://cds.cern.ch/record/1300517.

[84] ATLAS Collaboration. "AtlFast3: The Next Generation of Fast Simulation in ATLAS". In: *Comput. Softw. Big Sci.* 6 (2022), p. 7. DOI: 10.1007/s41781-021-00079-7. arXiv: 2109.02551 [hep-ex].

[85] S. Agostinelli et al. "GEANT4 – a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250. DOI: 10.1016/S0168-9002(03)01368-8.

[86] A. Ribon et al. “Status of Geant4 hadronic physics for the simulation of LHC experiments at the start of the LHC physics program”. In: (2010). URL: https://lcgapp.cern.ch/project/docs/noteStatusHadronic2010.pdf.

[87] ATLAS Collaboration. *Towards tuning electromagnetic shower properties to data with AtlFast3*. ATL-SOFT-PUB-2021-004. 2021. URL: https://cds.cern.ch/record/2790872.

[88] ATLAS Collaboration. “Software and computing for Run 3 of the ATLAS experiment at the LHC”. In: (2024). arXiv: 2404.06335 [hep-ex].

[89] Evangelos Kourlitis. “HL-LHC and Beyond Computing Challenges”. In: *PoS* LHCP2023 (2024), p. 112. DOI: 10.22323/1.450.0112.

[90] *ATLAS Software and Computing HL-LHC Roadmap*. Tech. rep. Geneva: CERN, 2022. URL: https://cds.cern.ch/record/2802918.

[91] Martina Javurkova and the ATLAS Collaboration. “The Fast Simulation Program of ATLAS at the LHC”. In: (2024). URL: https://cds.cern.ch/record/2905117.

[92] ATLAS Collaboration. “Software and Computing for Run 3 of the ATLAS Experiment”. In: *arXiv preprint arXiv:2404.06335* (2024). URL: https://arxiv.org/abs/2404.06335.

[93] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector”. In: *Eur. Phys. J. C* 81 (2021), p. 689. DOI: 10.1140/epjc/s10052-021-09402-3. arXiv: 2007.02645 [hep-ex].

[94] ATLAS Collaboration. “New techniques for jet calibration with the ATLAS detector”. In: *Eur. Phys. J. C* 83 (2023), p. 761. DOI: 10.1140/epjc/s10052-023-11837-9. arXiv: 2303.17312 [hep-ex].

[95] ATLAS Collaboration. *In-situ measurements of the ATLAS large-radius jet response in 13 TeV pp collisions*. ATLAS-CONF-2017-063. 2017. URL: https://cds.cern.ch/record/2275655.

[96] ATLAS Collaboration. *Monte Carlo Calibration and Combination of In-situ Measurements of Jet Energy Scale, Jet Energy Resolution and Jet Mass in ATLAS*. ATLAS-CONF-2015-037. 2015. URL: https://cds.cern.ch/record/2044941.

[97] ATLAS Collaboration. *Jet Mass Resolutions in ATLAS using Run 2 Monte CarloSimulation*. Tech. rep. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2631339.

[98] Steven D. Bass et al. "The Higgs boson implications and prospects for future discoveries. The Higgs boson – its implications and prospects for future discoveries". In: *Nature Rev. Phys.* 3.9 (2021). Preprint of article submitted to Nature Reviews Physics, 25 pages, 8 figures, pp. 608–624. DOI: 10.1038/s42254-021-00341-2. arXiv: 2104.06821. URL: https://cds.cern.ch/record/2764481.

[99] ATLAS Collaboration. *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel using 13 TeV pp collision data from the ATLAS detector*. ATLAS-CONF-2021-030. 2021. URL: https://cds.cern.ch/record/2777236.

[100] ATLAS Collaboration. "Measurements of gluon-gluon fusion and vector-boson fusion Higgs boson production cross-sections in the $H \to WW^* \to e\nu\mu\nu$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *Phys. Lett. B* 789 (2019), p. 508. DOI: 10.1016/j.physletb.2018.11.064. arXiv: 1808.09054 [hep-ex].

[101] ATLAS Collaboration. *Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment*. ATLAS-CONF-2019-049. 2019. URL: https://cds.cern.ch/record/2693958.

[102] ATLAS Collaboration. *Search for non-resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. ATLAS-CONF-2022-035. 2022. URL: https://cds.cern.ch/record/2811390.

[103] Elisabeth Petit. "Higgs self-couplings: Experimental vision". In: (2019). URL: https://cds.cern.ch/record/2704639.

[104] ATLAS Collaboration. "Summary of non-resonant and resonant Higgs boson pair searches from the ATLAS experiment". In: (2021).

[105] ATLAS Collaboration. "Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 105 (2022), p. 092002. DOI: 10.1103/PhysRevD.105.092002. arXiv: 2202.07288 [hep-ex].

[106] T. D. Lee. "A Theory of Spontaneous T Violation". In: *Phys. Rev. D* 8 (4 Aug. 1973), pp. 1226–1239. DOI: 10.1103/PhysRevD.8.1226. URL: https://link.aps.org/doi/10.1103/PhysRevD.8.1226.

[107] G. C. Branco et al. "Theory and phenomenology of two-Higgs-doublet models". In: *Physics Reports* 516.1-2 (July 2012), pp. 1–102. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2012.02.002. arXiv: 1106.0034.

[108] A. Djouadi et al. *The Minimal supersymmetric standard model: Group summary report*. 1998. arXiv: hep-ph/9901246.

[109] ATLAS Collaboration. "Combination of Searches for Resonant Higgs Boson Pair Production Using pp Collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS Detector". In: *Phys. Rev. Lett.* 132 (2024), p. 231801. DOI: 10.1103/PhysRevLett.132.231801. arXiv: 2311.15956 [hep-ex].

[110] Thomas Biekötter et al. "Possible indications for new Higgs bosons in the reach of the LHC: N2HDM and NMSSM interpretations". In: *European Physical Journal C* 82 (2 Sept. 2021). DOI: 10.1140/epjc/s10052-022-10099-1. URL: http://arxiv.org/abs/2109.01128.

[111] Ulrich Ellwanger, Cyril Hugonie, and Ana M. Teixeira. "The Next-to-Minimal Supersymmetric Standard Model". In: *Physics Reports* 496.1-2 (Oct. 2009), pp. 1–77. DOI: 10.1016/j.physrep.2010.07.001. arXiv: 0910.1785v5. URL: http://arxiv.org/abs/0910.1785.

[112] M. Gomez-Bock et al. "Concepts of Electroweak Symmetry Breaking and Higgs Physics". In: *2007 CERN-CLAF School of High-Energy Physics, CLASHEP 2007 - Proceedings* (Dec. 2007), pp. 177–238. arXiv: 0712.2419. URL: https://arxiv.org/abs/0712.2419v1.

[113] Sebastian Baum and Nausheen R. Shah. "Two Higgs Doublets and a Complex Singlet: Disentangling the Decay Topologies and Associated Phenomenology". In: *Journal of High Energy Physics* 2018 (12 Aug. 2018). DOI: 10.1007/JHEP12(2018)044. URL: http://arxiv.org/abs/1808.02667.

[114] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. "Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios". In: *European Physical Journal C* 80 (2 Aug. 2019). DOI: 10.1140/epjc/s10052-020-7655-x. URL: http://arxiv.org/abs/1908.08554.

[115] ATLAS Collaboration. "Search for nonresonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 108 (2023), p. 052003. DOI: 10.1103/PhysRevD.108.052003. arXiv: 2301.03212 [hep-ex].

[116] ATLAS Collaboration. "Search for the non-resonant production of Higgs boson pairs via gluon fusion and vector-boson fusion in the $b\bar{b}\tau^+\tau^-$ final state in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *Phys. Rev. D* 110 (2024), p. 032012. DOI: 10.1103/PhysRevD.110.032012. arXiv: 2404.12660 [hep-ex].

[117] ATLAS Collaboration. "Studies of new Higgs boson interactions through nonresonant HH production in the $b\bar{b}\gamma\gamma$ final state in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 01 (2024), p. 066. DOI: 10.1007/JHEP01(2024)066. arXiv: 2310.12301 [hep-ex].

[118] ATLAS Collaboration. "Search for non-resonant Higgs boson pair production in the $2b + 2\ell + E_{\text{T}}^{\text{miss}}$ final state in pp collisions at $\sqrt{s} = 13\text{TeV}$ with the ATLAS detector". In: *JHEP* 02 (2024), p. 037. DOI: 10.1007/JHEP02(2024)037. arXiv: 2310.11286 [hep-ex].

[119] ATLAS Collaboration. "Search for non-resonant Higgs boson pair production in final states with leptons, taus, and photons in pp collisions at $\sqrt{s} = 13\,\text{TeV}$ with the ATLAS detector". In: *JHEP* 08 (2024), p. 164. DOI: 10.1007/JHEP08(2024)164. arXiv: 2405.20040 [hep-ex].

[120] ATLAS Collaboration. *Combination of searches for Higgs boson pair production in pp collisions at $\sqrt{s} = 13\,TeV$ with the ATLAS detector*. ATLAS-CONF-2024-006. 2024. URL: https://cds.cern.ch/record/2898670.

[121] CMS Collaboration. "Search for nonresonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 04 (2019), p. 112. DOI: 10.1007/JHEP04(2019)112. arXiv: 1810.11854 [hep-ex].

[122] CMS Collaboration. "Search for nonresonant Higgs boson pair production in final state with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *Phys. Lett. B* 842 (2023), p. 137531. DOI: 10.1016/j.physletb.2022.137531. arXiv: 2206.09401 [hep-ex].

[123] CMS Collaboration. "Search for nonresonant Higgs boson pair production in final states with two bottom quarks and two photons in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 03 (2021), p. 257. DOI: 10.1007/JHEP03(2021)257. arXiv: 2011.12373 [hep-ex].

[124] CMS Collaboration. "Search for Higgs boson pair production in the $b\bar{b}W^+W^-$ decay mode in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 07 (2024), p. 293. DOI: 10.1007/JHEP07(2024)293. arXiv: 2403.09430 [hep-ex].

[125] CMS Collaboration. "Search for Higgs boson pairs decaying to WW^*WW^*, $WW^*\tau\tau$, and $\tau\tau\tau\tau$ in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *JHEP* 07 (2023), p. 095. DOI: 10.1007/JHEP07(2023)095. arXiv: 2206.10268.

[126] CMS Collaboration. *Search for the nonresonant and resonant production of a Higgs boson in association with an additional scalar boson in the $\gamma\gamma\tau\tau$ final state*. Tech. rep. CMS-PAS-HIG-22-012. https://cds.cern.ch/record/2848400. CERN, 2022.

[127] CMS Collaboration. "Search for nonresonant Higgs boson pair production in the four leptons plus two b jets final state in proton–proton collisions at $\sqrt{s} = 13\,\text{TeV}$". In: *JHEP* 06 (2023), p. 130. DOI: 10.1007/JHEP06(2023)130. arXiv: 2206.10657 [hep-ex].

[128] CMS Collaboration. *Search for nonresonant Higgs boson pair production in the $WW\gamma\gamma$ channel in pp collisions at $\sqrt{s} = 13$ TeV*. Tech. rep. CMS-PAS-HIG-21-014. https://cds.cern.ch/record/2823866. CERN, 2022.

[129] CMS Collaboration. *Combination of searches for nonresonant Higgs boson pair production in proton–proton collisions at $\sqrt{s} = 13$ TeV*. CMS Physics Analysis Summary CMS-PAS-HIG-20-011. Revised version dated 16 November 2024. CERN, Nov. 2024.

[130] ATLAS Collaboration. "Search for new resonances in events with one or more jets and a pair of tau leptons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 2023.10 (2023), p. 009. DOI: 10.1007/JHEP10(2023)009. URL: https://link.springer.com/article/10.1007/JHEP10(2023)009.

[131] ATLAS Collaboration. "Search for heavy resonances decaying into a Higgs boson and a new scalar particle in the $VV\gamma\gamma$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 2024.11 (2024), p. 047. DOI: 10.1007/JHEP11(2024)047. URL: https://link.springer.com/article/10.1007/JHEP11(2024)047.

[132] ATLAS Collaboration. "Search for heavy resonances decaying into a Higgs boson and a new scalar particle in the $b\bar{b}\gamma\gamma$ final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 2024.10 (2024), p. 104. DOI: 10.1007/JHEP10(2024)104. URL: https://link.springer.com/article/10.1007/JHEP10(2024)104.

[133] CMS Collaboration. "Search for a massive scalar resonance decaying to a light scalar and a Higgs boson in the four b quarks final state with boosted topology". In: *Phys. Lett. B* 842 (2023), p. 137392. DOI: 10.1016/j.physletb.2022.137392. arXiv: 2204.12413 [hep-ex].

[134] *Search for a new scalar resonance decaying to a Higgs boson and a new scalar with two bottom quarks and two photons in the final state in proton-proton collisions at*

$\sqrt{s} = 13\,\text{TeV}$. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2929445.

[135] CMS Collaboration. "Searches for Higgs boson production through decays of heavy resonances". In: (2024). arXiv: 2403.16926 [hep-ex].

[136] CMS Collaboration. *Search for a new heavy scalar boson decaying into a Higgs boson and a new scalar particle in the four b-quarks final state using proton-proton collisions at sqrts = 13 TeV*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2914871.

[137] ATLAS Collaboration. "ATLAS flavour-tagging algorithms for the LHC Run 2 *pp* collision dataset". In: *Eur. Phys. J. C* 83 (2023), p. 681. DOI: 10.1140/epjc/s10052-023-11699-1. arXiv: 2211.16345 [physics.data-an].

[138] ATLAS Collaboration. "Calibration of the light-flavour jet mistagging efficiency of the b-tagging algorithms with Z+jets events using $139\,\text{fb}^{-1}$ of ATLAS proton–proton collision data at $\sqrt{s} = 13\,\text{TeV}$". In: *Eur. Phys. J. C* 83 (2023), p. 728. DOI: 10.1140/epjc/s10052-023-11736-z. arXiv: 2301.06319 [hep-ex].

[139] ATLAS Collaboration. *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*. ATL-PHYS-PUB-2022-027. 2022. URL: https://cds.cern.ch/record/2811135.

[140] ATLAS Collaboration. "Configuration and performance of the ATLAS b-jet triggers in Run 2". In: *Eur. Phys. J. C* 81 (2021), p. 1087. DOI: 10.1140/epjc/s10052-021-09775-5. arXiv: 2106.03584 [hep-ex].

[141] Maxence Draguet and ATLAS Collaboration. *Flavour Tagging with Graph Neural Network at ATLAS*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2912358.

[142] Lukas Heinrich, Matthew Feickert, and Giordon Stark. *pyhf: v0.7.6*. Version 0.7.6. https://github.com/scikit-hep/pyhf/releases/tag/v0.7.6. DOI: 10.5281/zenodo.1169739. URL: https://doi.org/10.5281/zenodo.1169739.

[143] A L Read. "Presentation of search results: the CLs technique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), p. 2693. DOI: 10.1088/0954-3899/28/10/313. URL: https://dx.doi.org/10.1088/0954-3899/28/10/313.

[144] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[145] ATLAS Collaboration. *Search for new resonances decaying to a scalar and a Standard Model-like Higgs boson in the 4b final state using the Run 2 pp collision data collected by the ATLAS detector*. Tech. rep. Geneva: CERN, 2024. URL: https://cds.cern.ch/record/2898531.

[146] Avrim Blum, Adam Kalai, and John Langford. "Beating the hold-out: bounds for K-fold and progressive cross-validation". In: *Proceedings of the Annual ACM Conference on Computational Learning Theory* (1999), pp. 203–208.

[147] ATLAS Collaboration. *Measurement of t-Channel Single-Top-Quark Production in pp Collisions at* $\sqrt{s} = 5.02$ *TeV with the ATLAS Detector*. ATLAS-CONF-2023-033. 2023. URL: https://cds.cern.ch/record/2862022.

[148] Conor Durkan et al. "Neural Spline Flows". In: *arXiv e-prints*, arXiv:1906.04032 (June 2019), arXiv:1906.04032. DOI: 10.48550/arXiv.1906.04032. arXiv: 1906.04032 [stat.ML].

[149] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. *Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles*. 2017. arXiv: 1612.01474 [stat.ML].

[150] Stanislav Fort, Huiyi Hu, and Balaji Lakshminarayanan. *Deep Ensembles: A Loss Landscape Perspective*. 2020. arXiv: 1912.02757 [stat.ML].

[151] Jonathon Shlens. "A Tutorial on Principal Component Analysis". In: (2014). arXiv: 1404.1100 [cs.LG].

[152] Jonas Wittbrodt. *ScannerS*. https://gitlab.com/jonaswittbrodt/ScannerS. 2025.

[153] Takafumi Kanamori, Shohei Hido, and Masashi Sugiyama. "A Least-squares Approach to Direct Importance Estimation". In: *Journal of Machine Learning Research* 10 (July 2009), pp. 1391–1445. DOI: 10.1145/1577069.1755831.

[154] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[155] Nicole Michelle Hartman. *A Search for Non-Resonant HH→4b at s = 13 TeV with the ATLAS detector – or – 2b, and then another 2b... now that's the thesis question.* Oct. 2023.

[156] ATLAS Collaboration. "Configuration, Performance, and Commissioning of the ATLAS *b*-jet Triggers for the 2022 and 2023 LHC data-taking periods". In: *JINST* 20.3 (2025). 47 pages in total, author list starting page 30, 8 figures, 4 tables, published in JINST. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUP 2022-02/, P03002. DOI: 10.1088/1748-0221/20/03/P03002. arXiv: 2501.11420. URL: https://cds.cern.ch/record/2921892.

[157] ATLAS Collaboration. "Fast b-tagging at the high-level trigger of the ATLAS experiment in LHC Run 3". In: *JINST* 18.11 (2023). 37 pages in total, author list starting page 20, 5 figures, 2 tables. All figures including auxiliary figures are available at https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/TRIG-2022-03, P11006. DOI: 10.1088/1748-0221/18/11/P11006. arXiv: 2306.09738. URL: https://cds.cern.ch/record/2862499.

[158] Morvan Vincent. *Run 3 Trigger Efficiency Studies and Bucket Strategy*. Presentation at HH4b Analysis Meeting, 30 January. Available at: https://indico.cern.ch/event/1508003/#80-run-3-trigger-efficiency-st. 2025.

[159] Arie Rond, Raja Giryes, and Michael Elad. *Poisson Inverse Problems by the Plug-and-Play scheme*. 2015. arXiv: 1511.02500 [cs.CV]. URL: https://arxiv.org/abs/1511.02500.

[160] ATLAS Collaboration. *Projected sensitivity of measurements of Higgs boson pair production with the ATLAS experiment at the HL-LHC*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2925853.

[161] ATLAS and CMS Collaborations. *Highlights of the HL-LHC physics projections by ATLAS and CMS*. Tech. rep. Geneva: CERN, 2025. URL: https://cds.cern.ch/record/2928907.

[162] ATLAS Collaboration. "Simultaneous energy and mass calibration of large-radius jets with the ATLAS detector using a deep neural network". In: *Mach. Learn. Sci. Tech.* 5 (2024), p. 035051. DOI: 10.1088/2632-2153/ad611e. arXiv: 2311.08885 [hep-ex].

[163] ATLAS Collaboration. *Transformer networks for constituent-based b-jet calibration with the ATLAS detector*. ATL-PHYS-PUB-2024-015. 2024. URL: https://cds.cern.ch/record/https://cds.cern.ch/record/2905688.

Appendix A

LOW-MASS EVENTS IN THE $HH \to b\bar{b}b\bar{b}$ ANALYSIS

A.1 Motivation

The non-resonant $HH \to b\bar{b}b\bar{b}$ analyses from Run 2, as well as preliminary results from Run 3, have established this channel as one of the most powerful for probing the Higgs boson self-coupling. Given the substantial achievements so far, the analysis has been extensively optimized for overall sensitivity—leaving limited room for incremental gains using standard approaches. Nevertheless, the regime of low-mass m_{HH} events remains comparatively under-explored and may present new opportunities for sensitivity gains, provided that novel analytical strategies can be developed.

Several longstanding challenges have hindered sensitivity to low-mass events. These include the trigger turn-on behavior near threshold, difficulties in accurately reconstructing signal events, and the overwhelming multi-jet QCD background at low masses. However, recent advancements—such as asymmetric triggers specifically developed for this analysis in Run 3—have greatly improved the ability to select low-mass events in both signal and background. This motivates a focused effort on developing reconstruction and analysis methods tailored to this previously inaccessible region.

A.2 Pairing

As discussed in Section 9.3.1, the minimum ΔR pairing algorithm is highly effective for medium- and high-mass m_{HH} events, where pairing efficiency becomes nearly perfect above 500 GeV. However, its performance degrades for low-mass events. When m_{HH} is small, the Higgs bosons tend to have lower transverse momentum, resulting in b-jets that are widely separated in angle rather than collimated. Under such circumstances, a pairing algorithm based on minimal angular distance loses discriminating power. This effect is clearly demonstrated in Figure A.1, which shows the angular separation between b-quarks

from leading and subleading Higgs bosons, comparing correct and incorrect pairings for both low-mass ($m_{HH} < 450$ GeV) and high-mass ($m_{HH} > 450$ GeV) events.

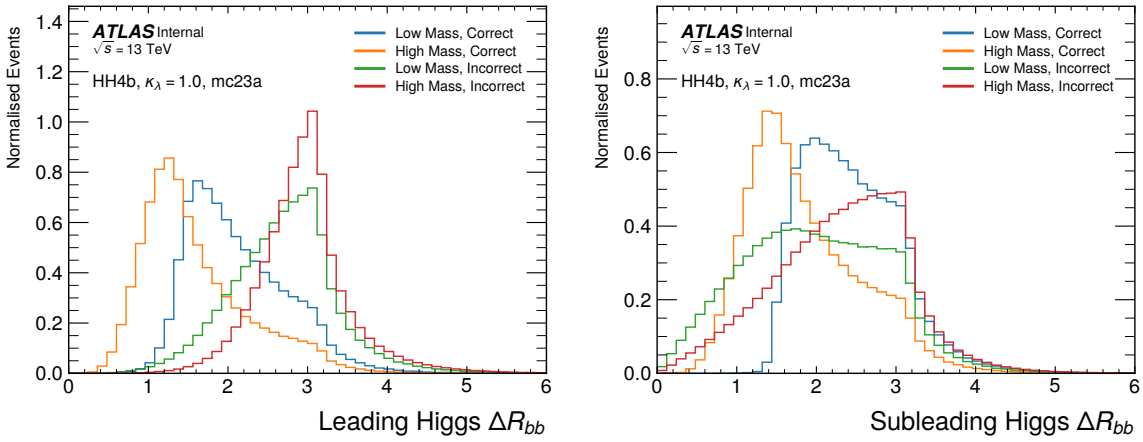


Figure A.1: Truth distribution of the angular distance between the b-quarks for the (a) leading and (b) subleading Higgs bosons in the mc23a signal, showing both correct and incorrect pairings for low-mass ($m_{HH} < 450$ GeV) and high-mass ($m_{HH} > 450$ GeV) events.

The definition of "low-mass" is itself somewhat ambiguous, with the boundary depending on the analysis context. Figure A.2 extends the previous comparison to three ranges of m_{HH}, further illustrating how kinematics evolve with m_{HH}. These results suggest that the low-mass regime should be defined as m_{HH} below approximately 400–450 GeV.

Pairing is particularly problematic for low-mass events because the combinatorics of jet assignments in both signal and QCD background can produce mass configurations that mimic signal-like features. Figure A.3 demonstrates this point for $3b$ data with $m_{HH} <$ 400 GeV, where each possible pairing yields substantial populations in both the signal and control regions.

To overcome these challenges, future analyses seeking to exploit low-mass events might consider strategies that circumvent explicit pairing altogether—such as defining kinematic regions in the space of multivariate discriminants trained to separate signal from background. This approach could avoid the loss in efficiency associated with traditional jet

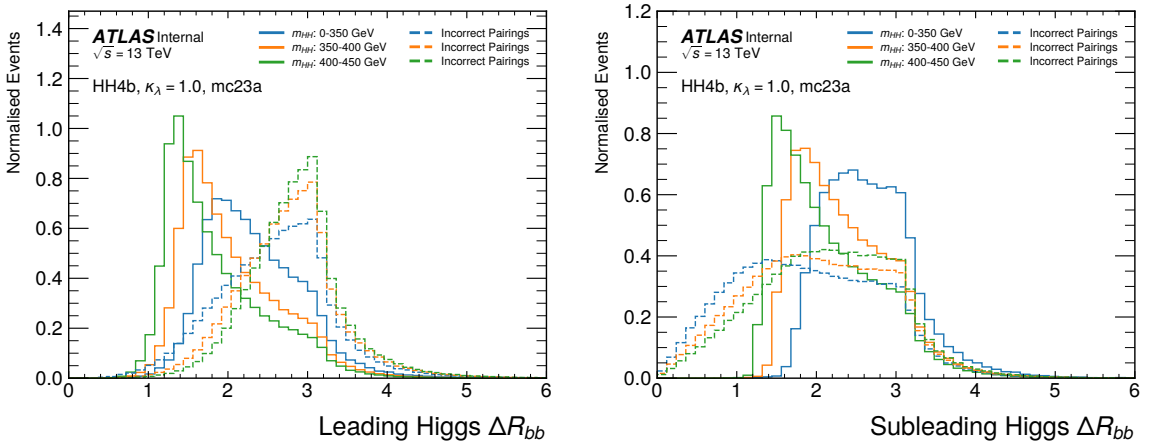


Figure A.2: Truth distribution of the angular distance between the b-quarks for the (a) leading and (b) subleading Higgs bosons in the mc23a signal, shown for three m_{HH} intervals: [0, 250], [350, 400], and [400, 450] GeV.

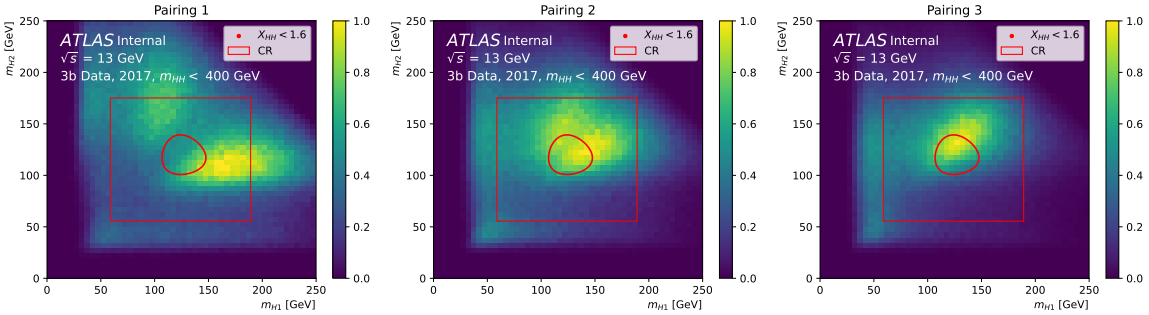


Figure A.3: Distributions of $3b$ events with $m_{HH} < 400$ GeV from 2017 data in the m_{H1}-m_{H2} plane for each of the three possible jet pairings. With jets ordered by decreasing p_T and indices [1,2,3,4]: pairing 1 is [(1,2), (3,4)], pairing 2 is [(1,3), (2,4)], and pairing 3 is [(1,4), (2,3)].

pairing, particularly in the resolved topology.

A.3 Potential Sensitivity Gain

It is instructive to quantify the potential improvements in analysis sensitivity that could result from more effective use of low-mass events. To estimate this, the existing Run 2 workspace can be used to perform "what-if" studies, where the signal yield is artificially increased or the background decreased in the low-mass region. This allows one to directly observe the impact on the expected (statistical only) 95% confidence level limits. Three boundaries are considered, based on the workspace binning: 396, 433, and 470 GeV, with $m_{HH} < 433$ GeV serving as a reasonable working definition.

For example, Figure A.4 shows that reducing the background by a factor of three for events with $m_{HH} < 433$ GeV improves the expected stat-only limit by approximately 7%. Conversely, simultaneously increasing the signal and reducing the background by a factor of three in the same region yields a much larger improvement of about 36%, as seen in Figure A.5. While such gains are optimistic, they illustrate the scale of enhancement possible if new methods can substantially improve signal efficiency and/or background rejection.

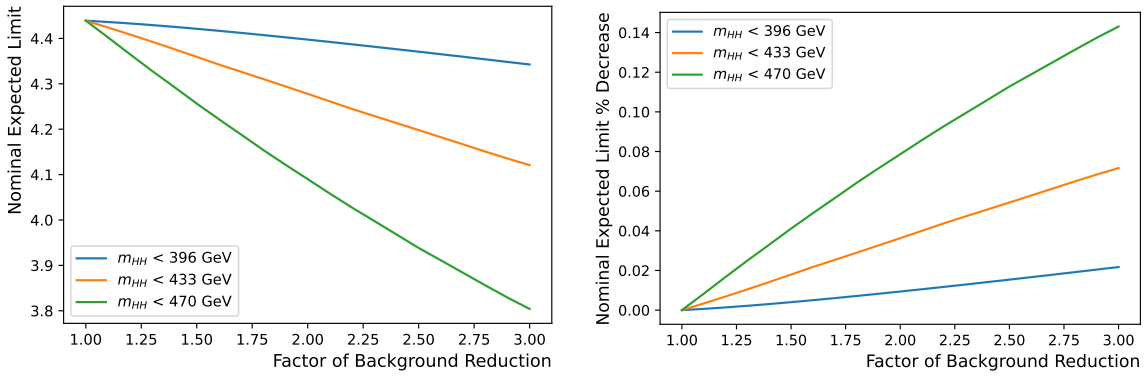


Figure A.4: (a) Stat-only 95% CL limits and (b) their percentage improvement when reducing the low-mass background by a factor of up to three, for several boundaries of m_{HH}.

Current results show that completely removing events with $m_{HH} < 433$ GeV would degrade the stat-only limit by only about 5%, which is a testament to the strength of the existing analysis for high-mass events. To capitalize on the potential of low-mass events

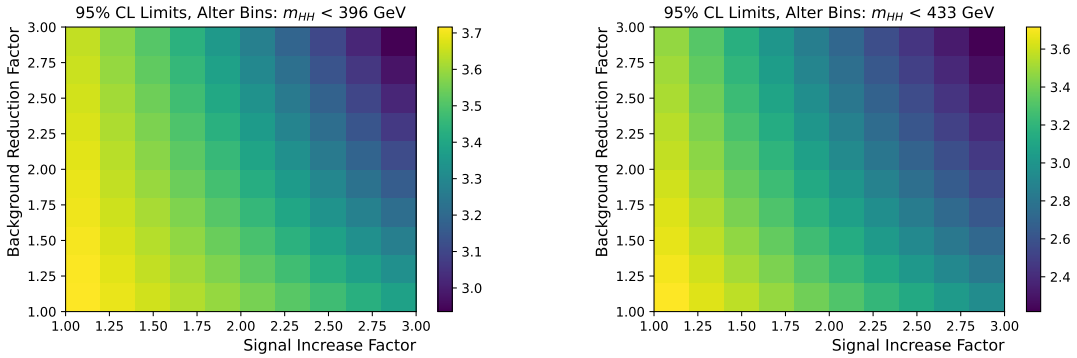


Figure A.5: Stat-only 95% CL limits when both increasing the signal and reducing the background by factors between 1 and 3 for low-mass events, with boundaries at (a) $m_{HH} <$ 396 GeV and (b) $m_{HH} < 433$ GeV.

without sacrificing performance in the high-mass regime, a promising strategy would be to treat these regions independently, possibly as separate sub-analyses.

A.4 NN Reweighting

In an effort to improve background modeling, a neural network-based reweighting procedure—similar to that described in Section 6.4—was applied to the low-mass region. Here, the neural network is trained to reweight simulated QCD events to better match the data distributions in the control region, using the same density ratio estimation method as previously established. Unlike the background estimation, which reweights $2b$ to $4b$ data, this study focuses on reweighting simulated QCD to match data in each n_{tag} category.

This NN-based reweighting shows excellent closure for the $2b$ sample across several kinematic variables in the signal region, as shown in Figure A.6.

However, for the $3b$ and $4b$ samples, the NN reweighting does not yield the same level of performance. Figure A.7 shows that, although some improvement is observed for 3b events, significant non-closure persists in the SR. The problem is even more pronounced for 4b events, as seen in Figure A.8, where a lack of smooth modeling is evident—even in the CR where the network was trained. The limited statistics in the simulated 4b QCD sample

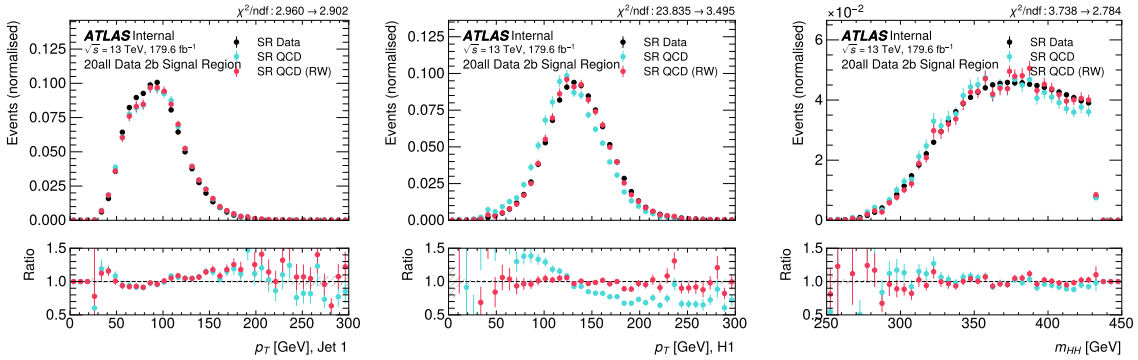


Figure A.6: Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 2b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the SR. Good closure is observed across the kinematic phase space.

exacerbate this challenge, as individual events with large weights can dominate the training process.

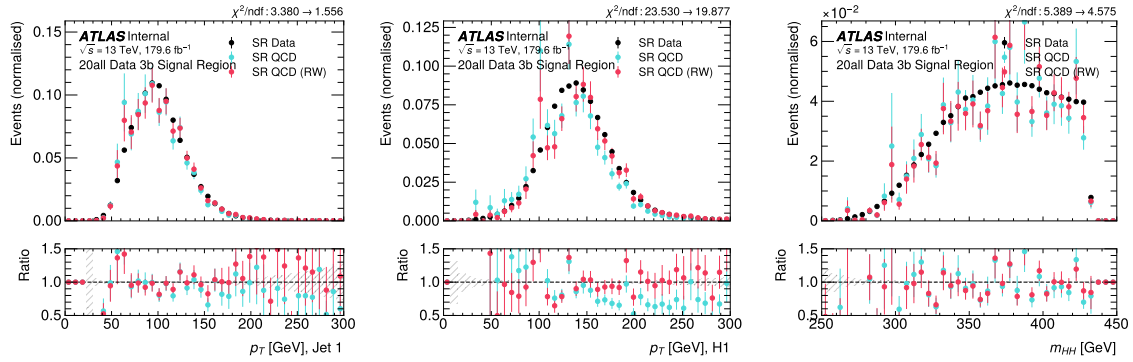


Figure A.7: Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 3b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the SR. While closure improves, significant modeling challenges remain.

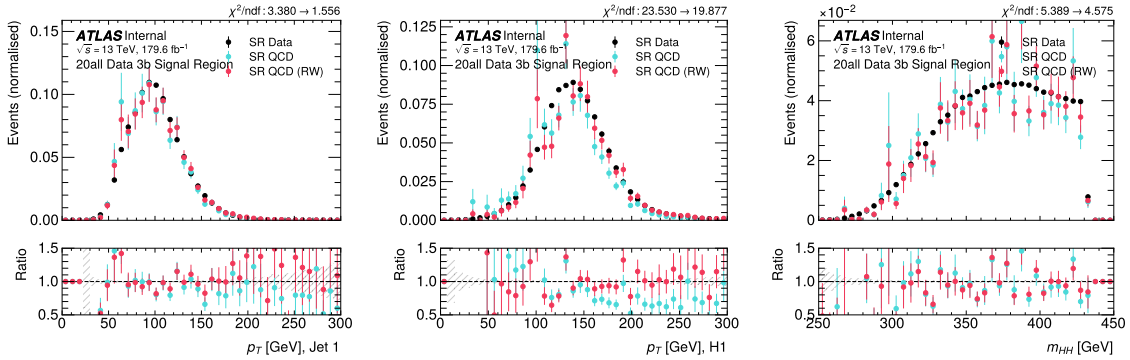


Figure A.8: Distributions of (a) the p_T of the leading Higgs candidate jet 1, (b) the p_T of the leading Higgs candidate, and (c) m_{HH} for low-mass 4b events from simulated multi-jet QCD, after NN reweighting to match Run 2 and partial Run 3 data in the CR. While the modeling shows some improvement, significant discrepancies remain, particularly in the control region used for training.

A.4.1 Shortcomings

One of the primary limitations encountered in these studies is the paucity of simulated QCD statistics, especially for the 4b sample. Modern machine learning methods perform best when trained on large, representative datasets. Possible remedies include switching to faster simulation tools such as AF3 or implementing more efficient event filtering to avoid simulating events that will not be used. Care must be taken, however, to ensure that such optimizations do not distort the relevant kinematic distributions.

Another issue relates to the definition of kinematic regions in these studies. The SR and CR used here were optimized for the full analysis, with an emphasis on the high-mass region. Therefore, a future analysis focused on low-mass events may benefit from redefining the signal and control regions to better suit this phase space.

A.5 Outlook

Recent advances in triggers and analysis methodology have opened the door to improved sensitivity from low-mass $HH \to b\bar{b}b\bar{b}$ events. To fully exploit this opportunity, it may

be necessary to develop dedicated reconstruction, background estimation, and discriminant strategies tailored to the low-mass regime. While improving jet pairing algorithms could yield benefits, more transformative approaches, such as using multivariate discriminants and redefining kinematic regions, may offer even greater gains.

Despite longstanding difficulties in modeling multi-jet QCD background, recent developments in NN-based reweighting are promising, especially as simulation samples improve in both size and fidelity. This, in turn, could enable more robust data-driven background estimation and improved training for multivariate discriminants. Ultimately, fully leveraging low-mass events may require treating them as a distinct analysis component, with specialized tools for maximizing their contribution to overall sensitivity.

Appendix B

BDT FOR PAIRING IN $X \to SH \to b\bar{b}b\bar{b}$

While the primary focus of my work was on the development of a neural network-based pairing method, the BDT pairing algorithm remains the nominal approach used throughout the $X \to SH \to b\bar{b}b\bar{b}$ analysis. This appendix summarizes the BDT strategy, training procedure, and key performance validations, complementing the more detailed discussion of pairing methodology found in the main body of the thesis.

B.1 BDT Pairing and Training

The BDT is trained using all available signal MC samples with m_X between 200 and 3000 GeV. To avoid statistical bias, a two-fold cross-validation scheme is used: events are split into odd and even event numbers, and each subset is used both for training and independent testing. Two separate BDT trainings are performed, one for compressed and one for non-compressed signal topologies, reflecting the distinct kinematic regimes.

The BDT is implemented with TMVA and is trained on the same set of events as used for the NN-based method, with three possible jet pairings considered per event. The input variable set is optimized for kinematic discrimination, as summarized in Table B.1. These include six angular variables for the two jet pairs, two p_T variables, and two mass-related variables, the invariant mass of all four b-jets, the difference between the four-jet mass, and the sum of the pair masses.

For each event, the BDT evaluates all three pairings, and the assignment with the highest BDT score is chosen for signal reconstruction. Figure B.1 shows the BDT output distributions for both training and test samples in the compressed and non-compressed regimes. The excellent agreement confirms the absence of overtraining.

Figure B.2 summarizes the number of events passing selection used in training and the resulting BDT pairing efficiency for different mass points. No significant dependence of

Table B.1: Input variables used to train the BDT for jet pairing in SH4b. For each event, all three pairings of the four selected b-jets (indices i, j, k, l) are evaluated.

Input Variable	**Description**
$\Delta\eta_{ij}$	Pseudorapidity difference between i-th and j-th b-jet
$\Delta\phi_{ij}$	Azimuthal angle difference between i-th and j-th b-jet
ΔR_{ij}	Angular distance between i-th and j-th b-jet
$\Delta\eta_{kl}$	Pseudorapidity difference between k-th and l-th b-jet
$\Delta\phi_{kl}$	Azimuthal angle difference between k-th and l-th b-jet
ΔR_{kl}	Angular distance between k-th and l-th b-jet
$p_T(ij)$	p_T of the ij b-jet pair
$p_T(kl)$	p_T of the kl b-jet pair
m_{ijkl}	Invariant mass of all four b-jets
$m_{ijkl} - (m_{ij} + m_{kl})$	Difference between the four-jet mass and the sum of the two pair masses

efficiency on event count is observed, indicating stable BDT performance across the signal phase space.

B.2 Reconstructed Mass Resonances

The BDT pairing is further validated by comparing reconstructed mass distributions for X, S, and H using the BDT pairing, a cut-based min ΔR pairing, and truth-matched assignments. As shown in Figure B.3, the BDT pairing preserves the resonance structure and does not introduce significant mass bias. The use of a dedicated BDT for the compressed regime provides additional gains for those signal models.

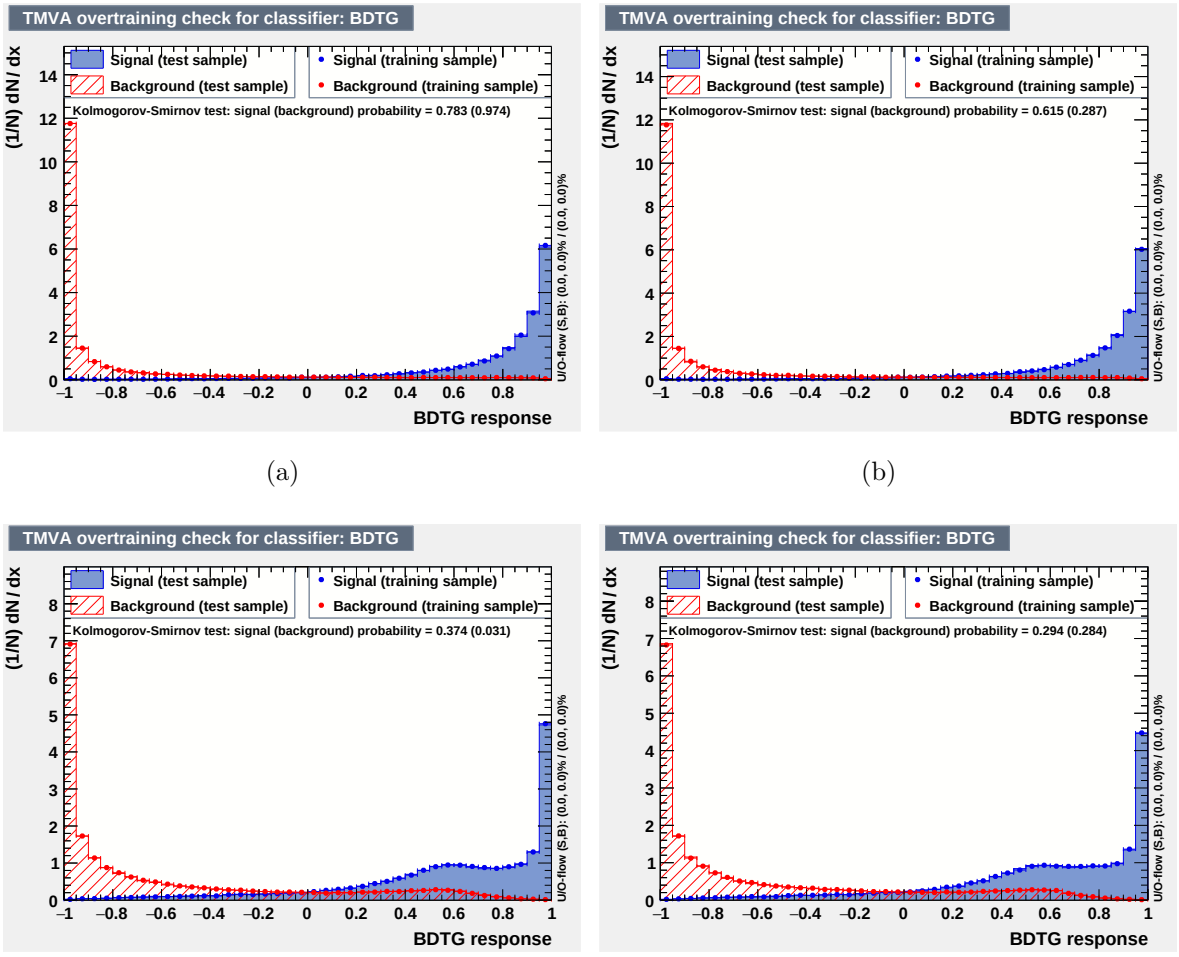
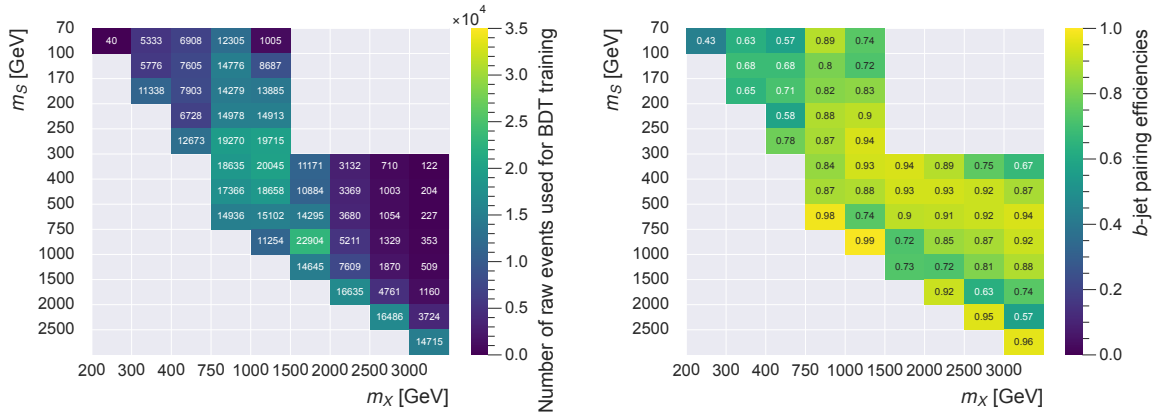
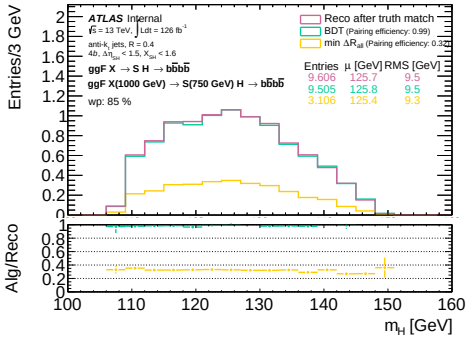


Figure B.1: Distributions of the BDT output for signal (correct pairing, score near 1) and background (incorrect pairing, score near -1) for (a) odd and (b) even events in the compressed region, and (c) odd and (d) even events in the non-compressed region. The agreement between training and test samples indicates no overtraining. See SH4b Internal Note [145].

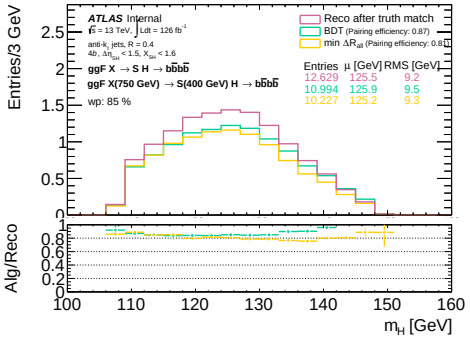


(a) (b)

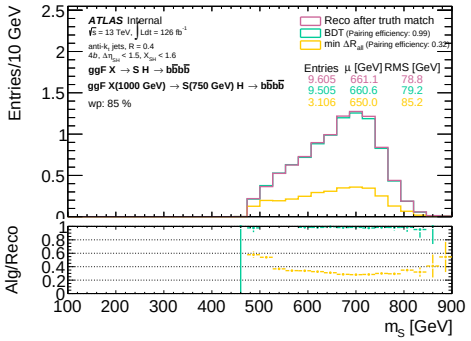
Figure B.2: (a) Number of simulated events passing selection in the training dataset. (b) BDT pairing efficiencies for different signal mass points, defined as the weighted fraction of selected events with the correct pairing identified. See SH4b Internal Note [145].



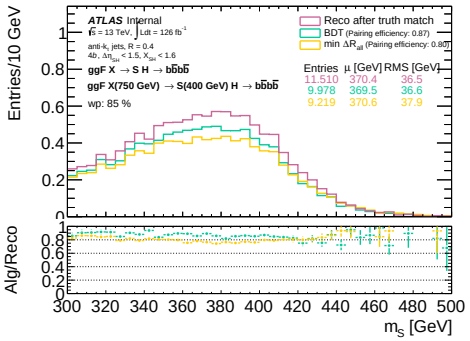
(a)



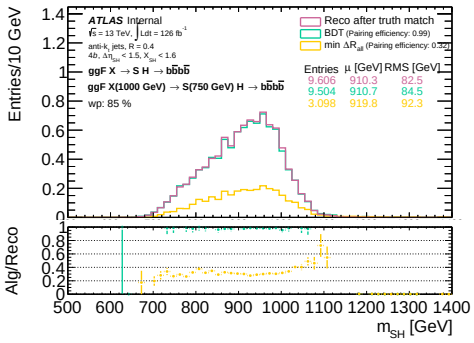
(b)



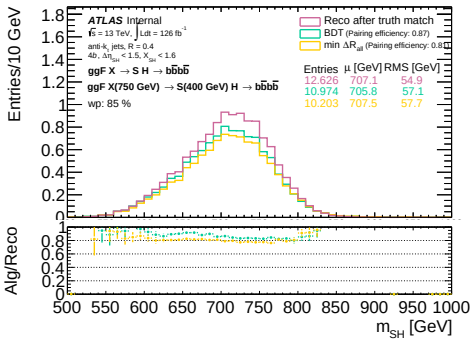
(c)



(d)



(e)



(f)

Figure B.3: Reconstructed mass distributions for H ((a), (b)), S ((c), (d)), and X ((e), (f)) for a compressed signal ($m_X = 1000$ GeV, $m_S = 750$ GeV; left column) and a non-compressed signal ($m_X = 750$ GeV, $m_S = 400$ GeV; right column). Results are shown for BDT pairing (green), min ΔR pairing (yellow), and truth-matching (pink). See SH4b Internal Note [145].

Appendix C

COMPARISON OF ASYMMETRIC AND SYMMETRIC SCALAR RESONANCE SEARCHES

This thesis presents the first ATLAS results from a search for scalar resonances in the asymmetric decay $X \to SH \to b\bar{b}b\bar{b}$, as detailed in Section 7.6.2. While no previous ATLAS results exist for this specific decay topology, a comparison to the search for heavy scalar resonances in the symmetric Higgs boson pair decay, $X \to HH \to b\bar{b}b\bar{b}$, offers valuable context on the achieved sensitivity. The results of the symmetric search are shown in Figure C.1.

A simple linear interpolation of the observed and expected limits from the asymmetric analysis, with m_S values between 100 GeV and 170 GeV, suggests that the results at $m_S \approx$ 125 GeV are comparable to those from the resolved channel of the resonant $HH \to b\bar{b}b\bar{b}$ analysis for $m_X \leq 1000$ GeV. This comparison is visualized in Figure C.2.

Despite the use of improved analysis techniques and modern object reconstruction, the asymmetric search is noticeably less sensitive than the symmetric search for m_X values of 400 GeV, 750 GeV, and 1000 GeV. This difference can be attributed to several inherent complexities of the asymmetric decay topology, which present greater challenges to the analysis.

First, the correct pairing of the four b-jets is significantly more complicated in the asymmetric case. Unlike the symmetric search where both Higgs bosons have a well-defined mass of 125 GeV, the asymmetric search has only one known mass (m_H) and a variable, unknown mass (m_S) that can range from 70 GeV to 2500 GeV. This wide kinematic variation makes it difficult for a single pairing algorithm to perform optimally across all signal samples. Additionally, even if the jets are paired correctly, the correct assignment of each pair to either the H or S boson must still be determined, which can result in a loss of signal efficiency ranging from 25% to 35% when the masses are similar.

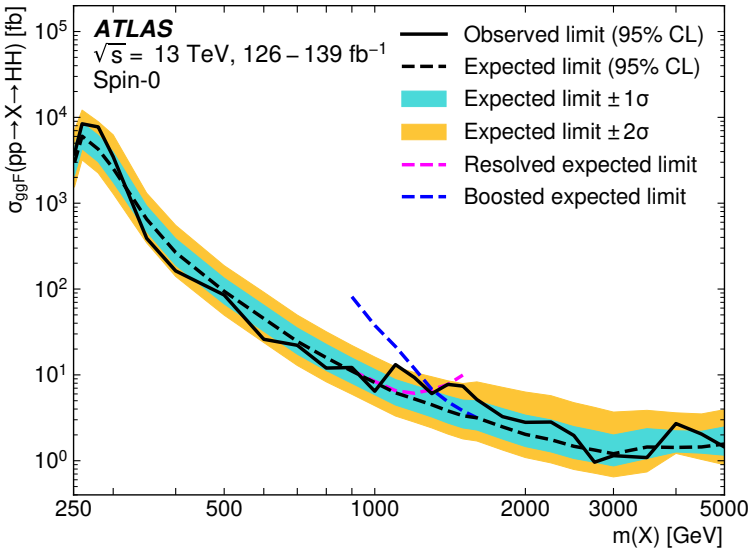


Figure C.1: Expected (dashed black lines) and observed (solid black lines) 95% CL upper limits on the cross section times the branching ratio of resonant HH production in the spin-0 model. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties. Expected limits using each of the resolved and boosted channels individually (dashed colored lines) are shown [105].

Second, the signal region definition and subsequent background estimation are more challenging. The symmetric search benefits from a single, well-defined signal region where both di-jet pairs are required to have an invariant mass close to m_H. This allows for the use of control regions that are immediately adjacent to the signal region, simplifying the data-driven background estimation. In contrast, the asymmetric search requires a different analysis for each specific m_S hypothesis. While the full signal region is blinded, each individual pseudo-signal region for a given (m_X, m_S) mass pair is not directly adjacent to the overall control region. This topological difference in the signal and control regions complicates the background estimation and reduces the statistical power of the analysis.

In summary, while the sensitivity of the asymmetric analysis is lower for most benchmark points compared to the symmetric search, the results are nevertheless promising. This

work represents a crucial first step in exploring a challenging decay channel and lays the foundation for future improvements in this unique search topology.

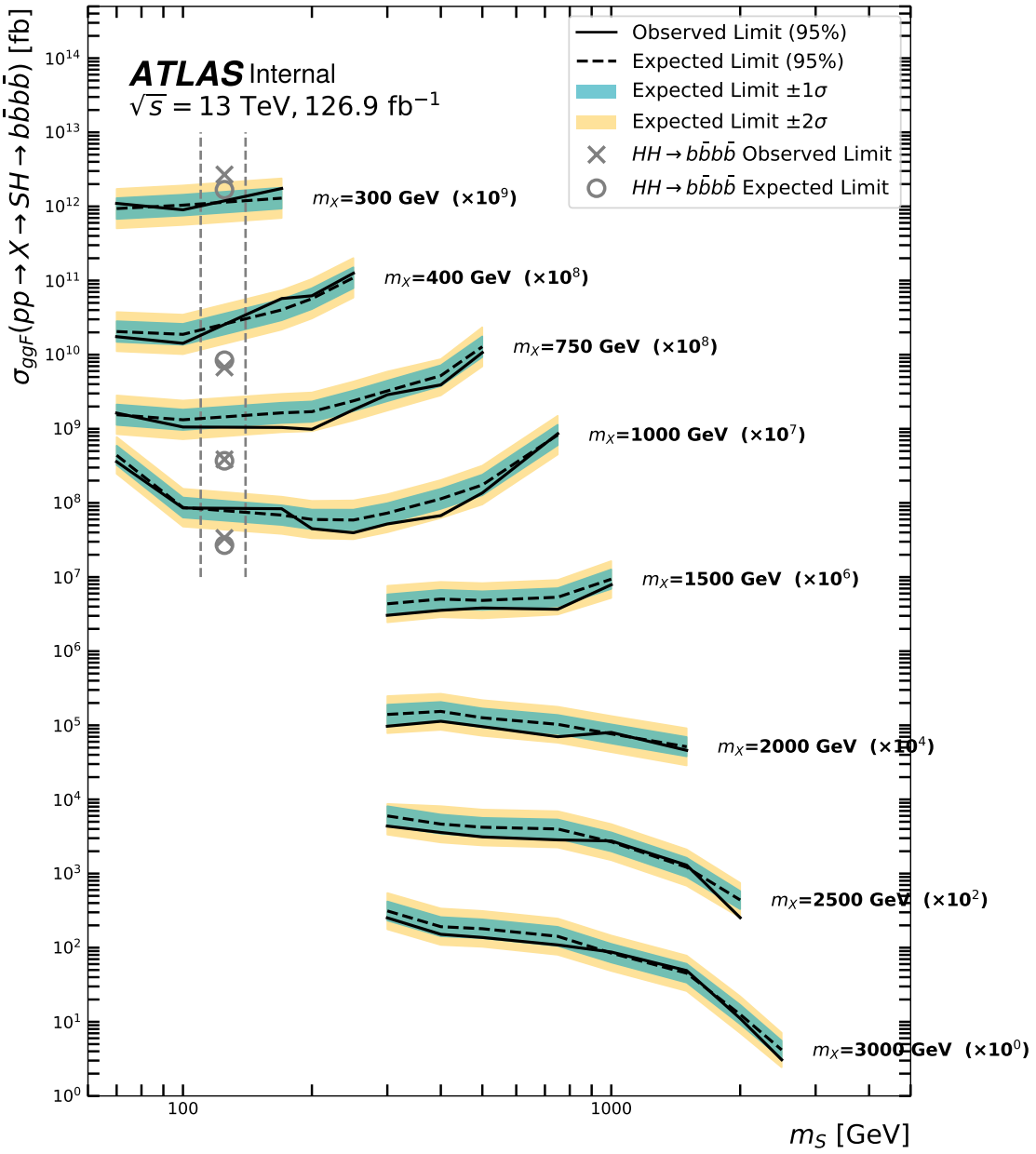


Figure C.2: Observed (solid black line) and expected (dashed black line) 95% CL upper limits on the cross section times the branching ratio. For each unique m_X value, a scaling factor is applied to the limits so that all mass pairs can be visualized together. The inner green (outer yellow) bands represent the 1σ and 2σ uncertainties, respectively. The observed and expected limits on the cross section times the branching ratio for the resonant $HH \to b\bar{b}b\bar{b}$ results are depicted (approximately) in gray.