

Search for Resonant Production from $HH \rightarrow b\bar{b}b\bar{b}$ Final states in ATLAS Detector using Boosted and Superboosted Analysis

Raif Rafideen Bin Norisam

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
of
University College London.

Department of Physics and Astronomy
University College London

May 30, 2025

Declaration:

I, Raif Rafideen Bin Norisam, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This thesis presents a search for resonant Higgs boson pair production in the $HH \rightarrow b\bar{b}b\bar{b}$ final state via gluon-gluon fusion (ggF), using 139 fb^{-1} of ATLAS Run-2 data. The analysis is performed using two reconstruction strategies: the Boosted and Superboosted approaches. The primary objective is to search for new resonances (X_R) predicted by Beyond the Standard Model (BSM) theories, decaying into pairs of Higgs bosons. Two benchmark BSM theories are considered: excited massive Kaluza–Klein gravitons from the Randall–Sundrum model and heavy neutral scalar bosons from the Two-Higgs-Doublet Model (2HDM), both of which offer potential solutions to the hierarchy problem in the Standard Model (SM).

The resonant HH searches in the $HH \rightarrow b\bar{b}b\bar{b}$ channel are organized into three reconstruction strategies: Resolved, Boosted, and Superboosted. The Resolved strategy targets Higgs boson pairs with mild to intermediate Lorentz boosts, covering resonance masses in the range 251–1600 GeV. The Boosted strategy focuses on intermediate to high-boost Higgs boson pairs, probing masses between 900–5000 GeV. The Superboosted strategy, newly developed for this work, is designed for extremely high-boost Higgs boson pairs, with sensitivity to resonance masses in the 1600–5000 GeV range.

The sensitivity to the process $\sigma(pp \rightarrow X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b})$ is evaluated using the Boosted and Superboosted strategies with 139 fb^{-1} of data. Results are expressed in terms of 95% confidence level (CL) upper limits on the production cross-section as a function of resonance mass. No significant excess is observed, and 95% CL upper limits are established. The combination of the Resolved and Boosted strategies reveals no significant deviation from SM predictions.

Impact Statement

Particle physics is a branch of physics that explores the fundamental nature of the universe and its role in cosmic evolution. The ATLAS experiment at CERN is a large-scale international collaboration involving scientists and engineers from around the world, dedicated to investigating the properties of fundamental particles. This includes the search for physics beyond the Standard Model, such as the existence of extra spatial dimensions and Supersymmetry, which may offer a path toward unifying gravity with the Standard Model. The results presented in this thesis contribute to the development of search strategies for identifying new particles decaying within the ATLAS detector. These contributions are expected to support future ATLAS analyses and enhance the broader understanding of particle physics for the benefit of humanity.

Acknowledgements

I would like to express my deepest gratitude and appreciation to my supervisor, Professor Nikos Konstantinidis, for his unwavering support and guidance throughout my four-year PhD journey, especially during the challenging times of the COVID-19 pandemic.

I would also like to extend my sincere appreciation to my sponsor, His Majesty the Yang di-Pertuan Agong, Sultan Muhammad V the Supreme Head of the Federation and 15th King of Malaysia, as well as the Malaysian Public Service Department, for their generous funding that enabled me to pursue my PhD at UCL and CERN.

My heartfelt thanks go to my colleagues Liao Shan, Lucas Borgna, Alex Emerman, Jana, Santiago Rafael, Jason Veatch, Max Swiatlowski, Rafael, and all other members of the $HH \rightarrow b\bar{b}b\bar{b}$ group for their support and guidance during my research. Special thanks to Chris Malena Delitzch and Steven Schramm for their invaluable mentorship during my qualification task.

To my family—Norisam Selamat, Nasiruddin Ismail, Dayana Isyafiq, Saif Isyaiqal, Nadia Elaiza, Ayman Naqib, Fadhil Nazlan, Athirah Ahmad, and Kamarul Ariffin—I honour you as my closest supporters, who have stood by me through the ups and downs of life and my PhD journey. To my beloved mother, Yusda Yusof, I am especially grateful for her sacrifice, constant support, tireless efforts, and unceasing prayers throughout the course of my thesis writing and doctoral studies.

I also wish to thank my UCL colleagues—Sam Grant, Cathal, Aiden Kelly, Jamie, and Adam for being wonderful friends and companions on this PhD jour-

ney. Last but not least, I am immensely grateful to my friends Royzilee Mustaffa, Norizamel Mustaffa, Nukman Sodiq, Nanang Tukiman, Etty Hastuti, Faizul Nizam, Shazwan Shukor, Siti Munirah Jalil, Dr Hazim Hanif, Dr Hazim Arief, Nadjmi Azamin, Azim Rahman, Hilman Fikri Azman, Balqis Azhar, Syafiq Zolkeply, Badri Kudos, Luqman Basir, Afiq Yunus, Aidil Idham, Firman Mohammed, Adam Junaidi, Saiful Azhar, Iqbal Zhorif, Dr Ali Hamid, Vafa Abadi, Dr Adrees Arbab, Izni Rahim, Hakim Zouargui, Dr. Hairi, Amber Sabbath, Hamish and Simon Gordon, who have continuously supported my mental well-being during my PhD studies in London.

Preface

This thesis presents a search for resonant Higgs boson pair production in the $HH \rightarrow b\bar{b}b\bar{b}$ final state via gluon-gluon fusion (ggF) using the ATLAS detector and 139 fb^{-1} of Run-2 data, using Boosted and Superboosted analyses. The primary goal of this research is to search for new physics beyond the Standard Model (BSM) that could address the Hierarchy Problem. Specifically, we focus on two BSM scenarios: excited Kaluza-Klein Graviton particles from the Randall-Sundrum model (G_{kk}^*) and Heavy Neutral Scalar bosons (X) from the Two-Higgs-Doublet Model (2HDM). These particles are predicted to decay into a pair of Higgs bosons, which subsequently decay into two pairs of $b\bar{b}$ quarks.

The search for resonant HH production in the $b\bar{b}b\bar{b}$ final state is divided into three analysis methods: Resolved, Boosted, and Superboosted. The Resolved analysis targets mild and intermediate Lorentz boost Higgs boson pairs, focusing on resonant masses between 251 GeV and 1600 GeV. The Boosted analysis targets intermediate to high Lorentz boost Higgs boson pairs, with resonant masses between 900 GeV and 5000 GeV. The Superboosted analysis, a newly developed method, targets extremely high Lorentz boost Higgs boson pairs, with resonant masses between 1600 GeV and 5000 GeV.

My contribution to the ATLAS collaboration for this research primarily focused on developing the Superboosted analysis, building upon the framework of the Boosted analysis. The Boosted analysis was developed by Santiago Rafael, William Balunas, and Alex Emerman, and all public results for this analysis were derived from their work. I am also contributed to the Boosted analysis by generating Monte

Carlo Simulation-driven QCD multijet events for background estimation comparison with data driven QCD multijet. In addition, I am also contribute to the Boosted analysis by generating Monte Carlo signal samples for very heavy Kaluza-Klein Graviton (G_{kk}^*) decays into Higgs boson pairs, which were also used in the Superboosted analysis.

For the cross-section limits for resonant Higgs pair production decaying into pairs of $b\bar{b}$ quarks, only the combination of Resolved and Boosted analyses has been performed and published. Due to time constraints, a combined limit for all three analyses (Resolved, Boosted, and Superboosted) was not performed in this thesis. However, combining all three analyses is recommended for future work.

Contents

1	Introduction	25
2	The Standard Model of Particle Physics	27
2.1	Quantum Electrodynamics (QED)	30
2.1.1	Gauge Symmetry in QED	30
2.2	Quantum Chromodynamics (QCD)	32
2.2.1	Gauge Symmetry in QCD	32
2.2.2	Colour Confinement and Asymptotic Freedom	33
2.3	Weak Interactions, Electroweak Theory and the Higgs Mechanism .	35
2.3.1	Parity Violation in Weak Interactions	36
2.3.2	Electroweak Theory	38
2.3.3	Spontaneous Symmetry Breaking and Higgs Mechanism . .	41
2.3.4	Higgs Mechanism for Fermions	43
2.3.5	Higgs Boson	45
2.3.6	Higgs Boson Production, Discovery, and Decay Modes at the LHC	48
2.4	Hierarchy Problem and Physics Beyond the Standard Model (BSM)	53
2.4.1	Hierarchy Problem and Higgs Boson Mass Correction . . .	53
2.4.2	Two Higgs Doublet Model (2HDM)	56
2.4.3	Randall-Sundrum Model	64
2.5	Higgs Boson Pair Production	71
2.5.1	Higgs Boson Pair Production in the ATLAS Experiment . .	71
2.5.2	Resonant Higgs Boson Pair Production Decay Channels . .	72

3	The Large Hadron Collider and ATLAS Detector	75
3.1	The ATLAS Detector	77
3.1.1	ATLAS Coordinate System	77
3.1.2	Inner Detector	78
3.1.3	ATLAS Calorimeters	80
3.1.4	Muon Spectrometer	81
3.1.5	ATLAS Trigger System	82
3.1.6	Luminosity	84
4	Analysis Tools	86
4.1	Monte Carlo Simulation	86
4.1.1	Monte Carlo Simulation Process	86
4.1.2	Monte Carlo Event Generator Examples	87
4.2	Jets	89
4.2.1	Jet Algorithms	89
4.2.2	Jet Mass Reconstruction Tools	91
4.2.3	Jet Calibration	92
4.3	Track jets	95
4.4	b-tagging	97
4.4.1	Low-Level b-tagger	97
4.4.2	High-Level b -tagger (DL1r)	102
4.4.3	b -tagging Efficiency	103
4.4.4	b -tagging Efficiency Calibration	103
5	Search for Resonant production from $pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ in ATLAS De-	
	tector using Boosted and Superboosted Analysis	105
5.1	Analysis methods in the search for resonant production $HH \rightarrow b\bar{b}b\bar{b}$	107
5.1.1	Resolved Analysis	107
5.1.2	Boosted Analysis	108
5.1.3	Superboosted Analysis	109
5.2	Data and Monte Carlo Samples	112

5.2.1	Data Samples	112
5.2.2	Monte Carlo Samples	112
5.3	Object Reconstruction	117
5.3.1	Calorimeter jets	117
5.3.2	Track jets	117
5.3.3	Muons	118
5.4	Event Selection	120
5.4.1	Trigger Selection	120
5.4.2	Kinematic Selection on Calorimeter Jets	121
5.4.3	Collinear (Overlapping) Track Jet Veto	128
5.4.4	Resolved Analysis Veto	129
5.4.5	b -tagged Track Jet Categorisation	130
5.4.6	Signal Selection for $HH \rightarrow b\bar{b}b\bar{b}$	132
5.4.7	Selection for Control and Validation Regions	134
5.5	Background Estimation	137
5.5.1	Kinematic Reweighting	138
5.5.2	Background Modelling	140
5.5.3	$m(HH)$ Smoothing Fit	144
5.6	Systematic Uncertainties	147
5.6.1	Background Normalisation Uncertainty	147
5.6.2	Background Shape Uncertainties	158
5.6.3	Top-quark Pair Production Modelling Uncertainty	166
5.6.4	Luminosity	175
5.6.5	Theoretical Uncertainty for Signal	176
5.6.6	b -tagging Track Jet Scale Factor	177
5.6.7	Jet Systematics	178
5.7	Statistical Analysis	185
5.7.1	Signal Interpolation on X MC Samples	186
5.7.2	Limit Setting Procedure	189
5.7.3	Pull and Impact	195

5.7.4	Signal Region Event Yields	206
5.7.5	Observed Limits and Resolved+Boosted Combination . . .	210
6	Conclusions	215
Appendices		216
A	Ghost Association vs ΔR Track Jet Matching Comparison for Boosted Jets	216
B	Jet Mass Scale (JMS) Uncertainties for Small-R Jets	221
C	Signal Efficiency and Higgs Boson Candidate Mass Resolution Comparison Between Boosted and Superboosted Analyses	223
D	Boosted and Superboosted Background kinematics plots	226

List of Figures

2.1	Feynman diagrams for Higgs boson production modes at the LHC: a) Gluon-gluon fusion (ggF), b) Vector boson fusion (VBF), c) Vector boson associated production (VH), d) Top quark associated production ($t\bar{t}H$) [1]	50
2.2	The Feynman diagram for one-loop Higgs boson mass correction from the top quark, $W^\pm$ boson, Z boson, and the Higgs boson. . . .	54
2.3	Illustration of the Randall–Sundrum model setup.	65
2.4	Examples of leading order Feynman diagrams for non resonant Higgs pair production via gluon-gluon fusion (ggF) and vector boson fusion (VBF) in ATLAS.	72
2.5	Branching ratios for Higgs boson pair production decay channels [2].	73
2.6	Feynman diagrams for two Beyond Standard Model resonant Higgs pair production via gluon-gluon fusion (ggF) in ATLAS.	74
3.1	CERN accelerator complex [3].	76
3.2	The ATLAS Detector [4]	78
3.3	ATLAS Inner Detector system. [5]	80
3.4	The ATLAS Calorimeter System [6]	81
3.5	The ATLAS Muon Subsystem [7].	82
3.6	ATLAS trigger system during Run 2. [8]	84
4.1	Double subjet B-labeling efficiency as a function of the truth Higgs jet p_T for different values of ρ , derived from Monte Carlo simulation [9].	96

4.2	Diagram illustrating the impact parameter of tracks in the d_0 and z_0 directions.	98
4.3	The input variables for the JetFitter algorithm [10].	101
4.4	b -tagging efficiency scale factors as a function of jet p_T at the 77% working point [11].	104
5.1	The topologies of the three analyses used in the search for resonant production of $HH \rightarrow b\bar{b}b\bar{b}$	111
5.2	Trigger efficiency as a function of offline leading jet p_T (top) and mass (bottom) for the Boosted analysis across four data-taking years.	123
5.3	Trigger efficiency as a function of offline leading jet p_T (top) and mass (bottom) for the Superboosted analysis across three data-taking years.	124
5.4	Normalised $ \Delta\eta_{J_1J_2}^b $ distribution for QCD multijet and three different G_{kk}^* Monte Carlo samples before the $ \Delta\eta_{J_1J_2}^b $ cut is applied to calorimeter jets in the Boosted analysis.	125
5.5	Two-dimensional histograms of $\Delta R_{2p}/2$ versus $p_T^{J_1^s}$ in Data and Signal $m_{G_{kk}^*}$ MC Samples.	127
5.6	Configuration of overlapping and non-overlapping track jets inside Boosted (or Superboosted) jets. Overlapping track jets are defined when the collimated (hard) track jets intercept the axes of softer track jets.	128
5.7	Topology of the high-tag and low-tag categories for b -tagged track jet categorisation in the Boosted and Superboosted analyses.	132
5.8	Contours in the $m(H_1^i)-m(H_2^i)$ plane representing the Control, Validation, and Signal regions for the Boosted and Superboosted analyses [12].	135
5.9	Ratio of the relevant kinematic distributions between b -tagged and untagged Higgs boson candidates in data, together with the corresponding splines after 10 iterations of kinematic reweighting in the Boosted analysis.	141

- 5.10 Ratio of the relevant kinematic distributions between b -tagged and untagged Higgs boson candidates in data, together with the corresponding splines after 10 iterations of kinematic reweighting in the Superboosted analysis. 142
- 5.11 Comparison of the multijet $m(HH)_s$ distributions in the $2b$ split region using the Boosted and Superboosted MJ smoothing functions. . 148
- 5.12 The $m(HH)_s$ distributions after applying the smoothing fit to the QCD multijet and $t\bar{t}$ backgrounds in the Signal Region of the Superboosted analysis. The eigenvariation represents the uncertainty of the smoothing, obtained from variations of the p_0 , p_1 , and p_2 parameters. 148
- 5.13 The $m(HH)_b$ distributions after applying the smoothing fit to the QCD multijet and $t\bar{t}$ backgrounds in the Signal Region of the Boosted analysis. The eigenvariation represents the uncertainty of the smoothing, derived from variations of the p_0 , p_1 , and p_2 parameters. 149
- 5.14 The Boosted background kinematic distributions in the Control Region for the $2b$, $3b$, and $4b$ categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic distributions are presented in Appendix D. 150
- 5.15 The Boosted background kinematic distributions in the Validation Region for the $2b$, $3b$, and $4b$ categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic distributions are presented in Appendix D. 151

5.16	Boosted background kinematic distributions in the Signal Region for the $2b$, $3b$, and $4b$ categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic variable distributions are provided in Appendix D.	152
5.17	Superboosted background kinematic distributions in the Control Region for the $2b$ and $3b+$ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.	153
5.18	Superboosted background kinematic distributions in the Validation Region for the $2b$ and $3b+$ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.	154
5.19	Superboosted background kinematic distributions in the Signal Region for the $2b$ and $3b+$ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.	155
5.20	Smoothed QCD multijet $m(HH)_b$ distributions in the Boosted analysis, showing variations due to fitting range and functional form choices.	161
5.21	Smoothed $t\bar{t}$ $m(HH)_b$ distributions in the Boosted analysis, showing variations due to fitting range and functional form choices.	162
5.22	Smoothed QCD multijet $m(HH)_s$ distributions in the Superboosted analysis, showing variations due to fitting range and functional form choices.	163
5.23	Smoothed $t\bar{t}$ $m(HH)_s$ distributions in the Superboosted analysis, showing variations due to fitting range and functional form choices.	164

5.24	Comparison of $m(HH)_b$ distributions between MC-driven and data-driven QCD multijet models in the combined Signal and Validation regions for the Boosted analysis.	167
5.25	Comparison of $m(HH)_s$ distributions between MC-driven and data-driven QCD multijet models in the combined Signal and Validation regions for the Superboosted analysis.	168
5.26	Comparison of $2b$ category $m(HH)_i$ distributions for data- $t\bar{t}$ and data-driven QCD multijet across Control, Validation, and Signal regions in the Boosted and Superboosted analyses.	168
5.27	(a) Non-closure $m(HH)_b$ distribution comparison between data-driven QCD multijet derived from $1b-1f$ in the Signal region and $2b$ data- $t\bar{t}$ in the Validation region. (b) Non-closure $m(HH)_b$ comparison between QCD multijet from the Validation region and $2b$ data- $t\bar{t}$ in the Control region.	169
5.28	(a) Non-closure $m(HH)_s$ comparison between transformed Validation-to-Signal region $2b$ data- $t\bar{t}$ and data-driven QCD multijet derived from $1b-1f$ Signal region. (b) Non-closure $m(HH)_i$ comparison between transformed Control-to-Validation region $2b$ data- $t\bar{t}$ and data-driven QCD multijet from the Validation region. . .	169
5.29	Comparison of the $t\bar{t}$ $m(HH)_b$ distributions between two different matrix element generators (Powheg and aMC@NLO) for the $2b$, $3b$, and $4b$ categories in the Boosted analysis.	170
5.30	Comparison of the $t\bar{t}$ $m(HH)_s$ distributions between two different matrix element generators (Powheg and aMC@NLO) for the $2b$ and $3b+$ categories in the Superboosted analysis.	171
5.31	Comparison of the $t\bar{t}$ $m(HH)_b$ distributions between two different parton shower generators in the Boosted analysis.	171
5.32	Comparison of the $t\bar{t}$ $m(HH)_s$ distributions between two different parton shower generators in the Superboosted analysis.	172

5.33	Comparison of the $t\bar{t} m(HH)_b$ distributions between different renormalisation scale (μ_R) variations in the Boosted analysis.	172
5.34	Comparison of the $t\bar{t} m(HH)_s$ distributions between different renormalisation scale (μ_R) variations in the Superboosted analysis.	173
5.35	Comparison of the $t\bar{t} m(HH)_b$ distributions between different factorisation scale (μ_F) variations in the Boosted analysis.	173
5.36	Comparison of the $t\bar{t} m(HH)_s$ distributions between different factorisation scale (μ_F) variations in the Superboosted analysis.	174
5.37	Comparison of the $t\bar{t} m(HH)_b$ distributions for two different h_{damp} parameters in the Boosted analysis.	174
5.38	Comparison of the $t\bar{t} m(HH)_s$ distributions for two different h_{damp} parameters in the Superboosted analysis.	175
5.39	Comparison of the $t\bar{t} m(HH)_b$ distributions for final-state radiation (FSR) variations in the Boosted analysis.	175
5.40	Comparison of the $t\bar{t} m(HH)_s$ distributions for final-state radiation (FSR) variations in the Superboosted analysis.	176
5.41	Leading Higgs candidate mass $m(H_1^i)$ distributions before and after applying JMR uncertainties at different G_{kk}^* masses. "JMR FF 1up" and "JMR FF 1down" refer to the σ_{FF+} and σ_{FF-} smearing variations from the forward-folding JMR in Boosted jets.	183
5.42	Leading Higgs candidate mass $m(H_1^i)$ distributions in the $2b$ category before and after applying JMR uncertainties at different G_{kk}^* masses. "Boosted Jet JMR FF Mean" refers to the average values of the histogram bins between the σ_{FF+} and σ_{FF-} variations for Boosted jets.	184
5.43	Signal interpolation $m(HH)_b$ distributions for the Boosted analysis for a Heavy Neutral Scalar Boson with $m_X = 3.5$ TeV, together with a cross-check at $m_X = 2.5$ TeV.	187

5.44	Signal interpolation $m(HH)_s$ distributions for the Superboosted analysis for a Heavy Neutral Scalar Boson with $m_X = 3.5$ TeV, together with a cross-check at $m_X = 2.5$ TeV.	188
5.45	Pulls (circles and black lines with bottom scale) and impacts (blue and red bars with top scale) for various G_{kk}^* masses in the Boosted analysis.	198
5.46	Pulls (circles and black lines with bottom scale) and impacts (blue and red bars with top scale) for various X masses in the Boosted analysis.	199
5.47	Pulls (circles and black lines with bottom scale) and impacts for different G_{kk}^* masses in the Superboosted analysis.	200
5.48	Pulls (circles and black lines with bottom scale) and impacts for different X masses in the Superboosted analysis.	201
5.49	Comparison of signal $m(HH)_s$ distributions under $\pm 1\sigma$ variations of a Jet Mass Scale (JMS) uncertainty component for the 5 TeV G_{kk}^* and X MC samples. The X exhibits stronger asymmetric distortion due to its narrower natural width, whereas the broader G_{kk}^* resonance displays a more balanced shift.	202
5.50	Observed limits for the G_{kk}^* and X models in the Boosted analysis with systematic uncertainties.	212
5.51	Observed limits for the G_{kk}^* and X models in the Superboosted analysis with systematic uncertainties.	212
5.52	Observed local p -values for the Excited Kaluza-Klein Graviton (G_{kk}^*) and Heavy Neutral Scalar Boson (X) models in the Boosted and Superboosted analyses.	213
5.53	Expected and observed 95% CL upper limits on the production cross section for resonant HH decays into two $b\bar{b}$ pairs, from the combined Boosted and Resolved analyses [13].	214
5.54	Local p -values as a function of the resonance mass for the G_{kk}^* and X models in the Resolved analysis.	214

A.1	Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ track jet matching in data without a Higgs mass cut.	217
A.2	Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 2$ TeV graviton samples without a Higgs mass cut.	218
A.3	Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 3$ TeV graviton samples without a Higgs mass cut.	219
A.4	Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 4$ TeV samples without a Higgs mass cut.	220
B.1	JMS uncertainties for small- R jets across different jet m/p_T intervals, derived using the R_{trk} method.	222
B.2	JMS uncertainties for small- R jets due to differences between Topocluster and Particle Flow algorithms, shown for various m/p_T ranges.	222
C.1	Signal efficiency comparison between Boosted and Superboosted analyses in the G_{kk}^* and X MC samples. The signal is defined as the number of events passing all high tag categories in the Signal Region.	224
C.2	Comparison of leading Higgs boson candidate mass $m(H_1^i)$ and transverse momentum $p_T(H_1^i)$ between the Boosted and Superboosted analyses for three different m_X values.	225
D.1	The Background kinematics in $2b$ Control Region Boosted Analysis	226
D.2	The Background kinematics in $3b$ Control Region in Boosted Analysis	227
D.3	The Background kinematics in $4b$ Control Region Boosted Analysis	227
D.4	The Background kinematics in $2b$ Validation Region Boosted Analysis	228
D.5	The Background kinematics in $3b$ Validation Region in Boosted Analysis	228

D.6	The Background kinematics in $4b$ Validation Region Boosted Analysis	229
D.7	The Background kinematics in $2b$ Signal Region Boosted Analysis .	229
D.8	The Background kinematics in $3b$ Signal Region in Boosted Analysis	230
D.9	The Background kinematics in $4b$ Signal Region Boosted Analysis .	230
D.10	The Background kinematics in $2b$ Control Region Superboosted Analysis	231
D.11	The Background kinematics in $3b+$ Control Region in Super- boosted Analysis	231
D.12	The Background kinematics in $2b$ Validation Region Superboosted Analysis	232
D.13	The Background kinematics in $3b+$ Validation Region in Super- boosted Analysis	232
D.14	The Background kinematics in $2b$ Signal Region Superboosted Analysis	233
D.15	The Background kinematics in $3b+$ Signal Region in Superboosted Analysis	233

List of Tables

2.1	Three generations of Standard Model quarks [14].	29
2.2	Three generations of Standard Model leptons [14].	29
2.3	Standard Model gauge bosons [14].	29
2.4	Electric charge Q , chirality, weak isospin T_3 , and weak hypercharge Y of particles in the Electroweak Theory.	39
2.5	Feynman rules for Higgs boson self-interactions and interactions with weak bosons. Factorials represent symmetry factors, account- ing for permutations of identical particles in Feynman diagrams. The Minkowski metric is denoted by $\eta_{\mu\nu}$	46
2.6	Feynman rules for Higgs boson couplings to fermions.	47
2.7	Branching ratios and relative uncertainties for several Higgs boson decay channels, listed in descending order [15].	51
2.8	Current status of Higgs boson searches across different decay chan- nels.	52
2.9	The four types of 2HDM models that suppress tree-level FCNCs. In all models, right-handed up-type quarks couple exclusively to the second doublet, while other right-handed fermions may couple to either doublet depending on the model.	63
2.10	Main decay channels of Higgs boson pair production targeted by ATLAS.	73
5.1	Details of the $t\bar{t}$ MC samples used to assess systematic uncertainties in the modelling of top quark pair production.	115

5.2	Summary of the offline triggers used in the Boosted and Superboosted analyses during the 2015–2018 ATLAS Run-2 data-taking period.	122
5.3	Definitions of the Signal, Validation, and Control regions based on the reconstructed Higgs boson candidate masses.	136
5.4	Normalisation scale factors, their corresponding uncertainties, and correlations for each high-tag category in the Boosted and Superboosted analyses.	144
5.5	Smoothing fit ranges for the QCD multijet and $t\bar{t} m(HH)_i$ distributions in each high-tag category for the Boosted and Superboosted analyses. The smoothing procedure is applied only in the Validation and Signal regions.	146
5.6	Control Region variation uncertainties for each high-tag category in the Boosted and Superboosted analyses.	157
5.7	Extrapolation uncertainties for each high-tag category in the Boosted and Superboosted analyses.	158
5.8	Functional forms used for assessing background smoothing uncertainties.	160
5.9	Target m_X values for signal interpolation and their corresponding MC simulated resonant mass bounds.	186
5.10	Summary of systematic uncertainties, uncertainty types, and corresponding constraint terms in the profile likelihood.	190
5.11	Tag categories for each resonant mass range used in the combined limit calculation for the Boosted and Superboosted analyses.	194
5.12	Summary of abbreviations for nuisance parameter names used in the pull and impact plots for the Boosted and Superboosted analyses.	203
5.13	Relative impact on expected limits for each systematic uncertainty in the Heavy Neutral Scalar boson (X) model in the Boosted analysis.	204

5.14	Relative impact on expected limits for each systematic uncertainty in the Excited Kaluza-Klein Graviton (G_{kk}^*) model in the Boosted analysis.	205
5.15	Relative impact on expected limits for each systematic uncertainty in the Heavy Neutral Scalar boson (X) model in the Superboosted analysis.	205
5.16	Relative impact on expected limits for each systematic uncertainty in the Excited Kaluza-Klein Graviton (G_{kk}^*) model in the Superboosted analysis.	206
5.17	Event yields for the selection of individual masses of the Heavy Neutral Scalar Boson X in the Boosted analysis.	207
5.18	Event yields for the selection of individual masses of the Excited Kaluza-Klein Graviton G_{kk}^* in the Boosted analysis.	208
5.19	Event yields for the selection of individual masses of the Heavy Neutral Scalar Boson X in the Superboosted analysis.	208
5.20	Event yields for the selection of individual masses of the Excited Kaluza-Klein Graviton G_{kk}^* in the Superboosted analysis.	209

Chapter 1

Introduction

Since the discovery of the Higgs boson in 2012 [16, 17], numerous experiments have been conducted in the search for physics Beyond the Standard Model (BSM), notably by the ATLAS, CMS, and LHCb collaborations at CERN. Within the ATLAS experiment, Higgs boson pair production has become a major focus due to its unique physical processes, which offer a window into both Standard Model (SM) and BSM phenomena.

A key aspect of Higgs boson pair production is the non-resonant mode, which involves measuring the Higgs self-coupling. This search is of particular importance as it seeks to confirm the Higgs boson's self-interaction, a fundamental prediction of the Higgs mechanism within the Standard Model. In addition to non-resonant production, Higgs boson pairs can also be produced via resonant processes predicted by various BSM theories, such as those involving Kaluza–Klein gravitons or heavy neutral scalar bosons in the Two-Higgs-Doublet Model (2HDM). These models are of special interest due to their potential to address the hierarchy problem in the Standard Model, such as the corrections to the Higgs boson mass.

This thesis is dedicated to the search for resonant Higgs boson pair production in the $HH \rightarrow b\bar{b}b\bar{b}$ decay channel within the ATLAS experiment, focusing specifically on resonance masses in the multi-TeV range. The choice of the $b\bar{b}b\bar{b}$ final state is motivated by its large branching ratio, which constitutes approximately 33% of all Higgs boson pair decays. Additionally, targeting multi-TeV resonance

masses exploits the characteristic decay of high-energy Higgs bosons into highly Lorentz-boosted $b\bar{b}$ pairs. This feature enhances the search sensitivity due to reduced contamination from multijet background events in this energy regime. Accordingly, this thesis employs two complementary analysis strategies: the *Boosted* and *Superboosted* analyses.

The Boosted analysis focuses on Higgs boson pairs with high Lorentz boosts decaying into the $b\bar{b}b\bar{b}$ final state. In contrast, the Superboosted analysis extends this approach to probe even more extreme Lorentz boosts of the Higgs boson decay products, requiring specialized techniques for reconstructing highly collimated jets.

Both the Boosted and Superboosted analyses are designed to target the $HH \rightarrow b\bar{b}b\bar{b}$ process in the TeV-scale resonance mass range. These methods are expected to improve sensitivity in this region, where BSM signals may be more easily distinguished from Standard Model backgrounds.

The remainder of this thesis is organized as follows: Chapter 2 provides an overview of the Standard Model of particle physics, including the Higgs mechanism, the Higgs boson, the hierarchy problem, BSM theories, and Higgs boson pair production. Chapter 3 introduces the Large Hadron Collider and describes the components of the ATLAS detector. Chapter 4 discusses the analysis tools used in Chapter 5, which presents the main search for resonant $HH \rightarrow b\bar{b}b\bar{b}$ production using the Boosted and Superboosted analyses.

Chapter 2

The Standard Model of Particle Physics

The Standard Model (SM), as described in [18] and [19], serves as the theoretical framework that governs the fundamental interactions among elementary particles in the universe, excluding gravity. This framework is constructed from four foundational theories: Quantum Field Theory (QFT), Electroweak Theory [20], Quantum Chromodynamics (QCD), and the Higgs mechanism [21, 22, 23, 24].

The Electroweak Theory unifies the weak and electromagnetic interactions, while QCD governs the strong interaction. The Higgs mechanism, embedded within the Electroweak Theory, provides an explanation for the origin of mass via spontaneous symmetry breaking.

Within the Standard Model, elementary particles are categorised into two main groups: bosons and fermions. Bosons possess integer spin and obey Bose-Einstein statistics. Fermions, on the other hand, have half-integer spin and follow Fermi-Dirac statistics, which include the Pauli exclusion principle.

The current formulation of the SM includes 12 known fermions and 5 known bosons. Fermions are further divided into leptons and quarks, each classified into three generations. The lepton family [18] consists of the electron (e), muon (μ), and tau (τ), each accompanied by a corresponding neutrino. These charged leptons share the same electric charge but differ in mass across generations. Neutrinos

possess extremely small masses and interact only via the weak force, making detection particularly difficult. Each lepton is accompanied by an anti-lepton with identical mass and opposite electric charge and lepton number (with the exception of potential neutrino mass-related considerations).

Quarks are distinguished from leptons by their participation in strong interactions. Each generation includes an “up-type” quark with a charge of $+\frac{2}{3}e$ and a “down-type” quark with a charge of $-\frac{1}{3}e$. All quarks carry colour charge and interact via QCD, while also participating in weak and electromagnetic interactions. Each quark is paired with a corresponding anti-quark possessing opposite electric charge and quantum numbers.

Quarks form composite particles called hadrons, which are subdivided into mesons and baryons. Mesons consist of a quark–antiquark pair, such as the neutral pion formed from either an up quark and anti-up quark or a down quark and anti-down quark. Baryons consist of three quarks, with the proton comprising two up quarks and one down quark. Antiparticles such as the antiproton are composed of the corresponding anti-quarks.

The SM bosons include the photon (γ), gluon (g), $W^\pm$ bosons, Z^0 boson, and the Higgs boson (H). The photon mediates electromagnetic interactions between charged particles, while gluons mediate the strong force among quarks. The $W^\pm$ and Z^0 bosons are responsible for weak interactions [19]. The Higgs boson, discovered in 2012, provides mass to fundamental particles through the Higgs mechanism.

Tables 2.1, 2.2, and 2.3 list all known SM particles along with their key properties.

Name	Symbol	Spin $[\frac{h}{2\pi}]$	Charge [e]	Mass [GeV]
up	u	1/2	+2/3	0.002
down	d	1/2	-1/3	0.005
charm	c	1/2	+2/3	1.28
strange	s	1/2	-1/3	0.095
top	t	1/2	+2/3	173.1
bottom	b	1/2	-1/3	4.18

Table 2.1: Three generations of Standard Model quarks [14].

Name	Symbol	Spin $[\frac{h}{2\pi}]$	Charge [e]	Mass [GeV]
electron	e	1/2	-1	5×10^{-4}
electron neutrino	ν_e	1/2	0	$< 0.8 \times 10^{-10}$ [25]
muon	μ	1/2	-1	0.106
muon neutrino	ν_μ	1/2	0	$< 2 \times 10^{-4}$
tau	τ	1/2	-1	1.8
tau neutrino	ν_τ	1/2	0	$< 2 \times 10^{-2}$

Table 2.2: Three generations of Standard Model leptons [14].

Name	Symbol	Spin $[\frac{h}{2\pi}]$	Charge [e]	Mass [GeV]
photon	γ	1	0	0
gluon	g	1	0	0
$W^\pm$	$W^\pm$	1	± 1	80.4
Z	Z^0	1	0	91.2
Higgs boson	H	0	0	125.1

Table 2.3: Standard Model gauge bosons [14].

2.1 Quantum Electrodynamics (QED)

In classical electrodynamics, electromagnetic interactions between charged particles and currents arise from their respective electric and magnetic fields. These interactions are described by Coulomb's law, Ampère's law, and Maxwell's equations. According to classical theory, the strength of the electromagnetic interaction depends on the distance between two charged particles or currents: the closer the separation, the stronger the interaction.

At the quantum scale, however, classical electrodynamics becomes inadequate due to the inherently probabilistic nature of particle behavior. To address this limitation, physicists developed Quantum Electrodynamics (QED), a branch of Quantum Field Theory that accurately describes electromagnetic interactions between charged particles, particularly at high energies.

In QED, electromagnetic interactions are mediated by the exchange of virtual photons between charged particles. The strength of this interaction is determined by the fine-structure constant α_{em} , which governs the coupling between virtual photons and charged particles. Unlike the classical inverse-square law, QED predicts more complex behavior. In particular, the interaction strength increases as the separation between charged particles becomes extremely small or as their energies increase substantially. This effect originates from vacuum polarization, where virtual electron-positron pairs in the quantum vacuum screen the bare electric charge. Consequently, charged particles experience an *effective charge* that differs from the bare charge. These effective charges, which modify the interaction strength, can be estimated through charge renormalisation techniques.

2.1.1 Gauge Symmetry in QED

QED is formulated to respect gauge invariance under local phase transformations of the photon and fermion fields in the Lagrangian density. Gauge symmetry is essential for ensuring the conservation of physical quantities such as electric charge. The theory exhibits a $U(1)$ gauge symmetry, which guarantees charge conservation in electromagnetic interactions. The Lagrangian density $\mathcal{L}$ for QED is given by:

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\psi}(i\gamma^\mu D_\mu - m)\psi \quad (2.1)$$

Here, $F_{\mu\nu}$ is the electromagnetic field tensor, ψ and $\bar{\psi}$ denote the fermion field and its Dirac adjoint, γ^μ are the gamma matrices, m is the fermion mass, and D_μ is the covariant derivative defined as:

$$D_\mu = \partial_\mu + ieA_\mu. \quad (2.2)$$

In this expression, A_μ represents the photon field, and e is the electric charge.

Under the $U(1)$ gauge symmetry, the fermion and photon fields undergo local phase transformations that leave the Lagrangian invariant:

$$\psi \rightarrow e^{i\theta(x)}\psi \quad (2.3)$$

$$A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\theta(x) \quad (2.4)$$

In equations 2.3 and 2.4, $\theta(x)$ is the local phase function, and $e^{i\theta(x)}$ denotes the corresponding complex phase transformation. These transformations collectively define the $U(1)$ gauge symmetry of the QED Lagrangian.

2.2 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD), as discussed in references [26] and [27], is a branch of Quantum Field Theory that describes strong interactions among quarks and gluons. The QCD framework introduces a property known as colour charge, which is represented by Red (R), Green (G), and Blue (B) charges. Each colour has a corresponding anti-colour: anti-Red ($\bar{R}$), anti-Green ($\bar{G}$), and anti-Blue ($\bar{B}$). Colour-neutral states can be formed either by combining three distinct colour charges or by pairing a colour with its corresponding anti-colour, yielding configurations such as RGB or $G\bar{G}$.

2.2.1 Gauge Symmetry in QCD

Strong interactions among quarks and gluons are governed by the QCD Lagrangian density:

$$\mathcal{L} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a + \bar{\psi}(i\gamma^\mu D_\mu^g - m)\psi \quad (2.5)$$

Here, ψ and $\bar{\psi}$ denote the quark field and its adjoint, γ^μ are the gamma matrices, m represents the quark mass, and $F_{\mu\nu}^a$ is the gluon field strength tensor, given by

$$F_{\mu\nu}^a = \partial_\mu \Lambda_\nu^a - \partial_\nu \Lambda_\mu^a + g_s f^{abc} \Lambda_\mu^b \Lambda_\nu^c \quad (2.6)$$

In this expression, $\Lambda_\mu^a(x)$ denotes the gluon field, which comprises eight components corresponding to colour combinations, and g_s is the strong coupling constant. The structure constants f^{abc} are associated with the $SU(3)$ gauge group. The covariant derivative is defined as:

$$D_\mu^g = \partial_\mu - ig_s T^a \Lambda_\mu^a(x) \quad (2.7)$$

The $SU(3)$ generators T^a are expressed in terms of the eight Gell-Mann matrices λ^a . Their inclusion ensures gauge invariance under the following local transformations:

$$\psi \rightarrow U(x)\psi \quad (2.8)$$

$$\Lambda_\mu^a(x) \rightarrow U(x)\Lambda_\mu^a(x)U^\dagger(x) + \frac{i}{g_s} \left(U(x)\partial_\mu U^\dagger(x) \right) \quad (2.9)$$

Here, $U(x)$ is an infinitesimal $SU(3)$ gauge transformation matrix, constructed from eight phase functions $\theta^a(x)$ and the Gell-Mann matrices λ^a :

$$U(x) = \exp \left(-i \frac{\theta^a(x)\lambda^a}{2} \right) \approx 1 - \frac{i}{2} \theta^a(x)\lambda^a \quad (2.10)$$

Unlike the $U(1)$ symmetry in QED, the $SU(3)$ symmetry in QCD does not correspond to a conserved physical quantity. However, its non-Abelian nature—characterised by the non-commutativity of the T^a generators—leads to features such as gluon self-interactions. This property underpins fundamental phenomena, including colour confinement and asymptotic freedom, discussed in Section 2.2.2.

2.2.2 Colour Confinement and Asymptotic Freedom

Colour confinement refers to the inability to observe individual quarks in isolation, as they are permanently bound within colour-neutral hadrons.

When external energy is applied to separate a quark–antiquark pair, the resulting gluon field stretches and accumulates energy. Once a critical threshold is surpassed, new quark–antiquark pairs are spontaneously created, maintaining colour neutrality. These newly formed pairs combine with the original quarks, forming hadrons. This mechanism continues until all coloured objects are confined within hadronic states.

Experimental evidence of colour confinement is observed in high-energy hadron collider experiments, where quark and gluon production leads to fragmentation and hadronisation. The resulting parton showers manifest as hadronic jets, offering insight into the dynamics of QCD.

Asymptotic freedom, described in Refs. [28, 29], is another defining feature of QCD. At low energies, strong coupling confines quarks within hadrons. However, as energy and momentum transfer increase, the strength of the strong interaction decreases, allowing quarks and gluons to behave nearly as free particles. This regime

enables the use of perturbative methods in QCD for theoretical predictions under extreme conditions.

At energy scales near ~ 0.5 GeV, the strong coupling constant (α_s)¹ becomes large, rendering perturbative techniques ineffective. Although α_s can be modelled as a function of energy, its precise behaviour across all regimes remains non-trivial to determine.

To address this challenge, Parton Distribution Functions (PDFs) are employed to describe the probability of finding a parton (quark or gluon) carrying a specific momentum fraction x within a hadron. PDFs are essential for characterising hadronic structure and are widely utilised in high-energy processes such as deep inelastic scattering. In such processes, the energy scale Q defines the hardness of the interaction. By tuning PDFs to experimental data, reliable predictions of hadronic processes can be achieved in both perturbative and non-perturbative domains.

¹Here, α_s denotes the strong coupling constant, defined as $g_s^2/4\pi$. The evolution of α_s with the energy scale μ is governed by the renormalisation group equation and depends on the ratio μ/Λ_{QCD} . The QCD scale Λ_{QCD} delineates the boundary between the perturbative and non-perturbative regimes. For $\mu \lesssim \Lambda_{\text{QCD}} \sim 0.5$ GeV, non-perturbative effects such as confinement become dominant.

2.3 Weak Interactions, Electroweak Theory and the Higgs Mechanism

The weak interaction, one of the four fundamental forces described by the Standard Model, governs processes involving flavour-changing transitions in elementary particles and is mediated by the weak bosons. A well-known manifestation of this interaction is β^- -decay in atomic nuclei. In this process, a neutron decays into a proton, accompanied by the emission of an electron and an antineutrino. At the quark level, a "down" quark undergoes a flavour transition into an "up" quark, mediated by the emission of a W^- boson. This boson subsequently decays into an electron and an antineutrino. Conversely, β^+ -decay involves the transformation of an "up" quark into a "down" quark, with the emission of a W^+ boson, which decays into a positron and an electron neutrino.

In addition to the charged weak bosons, $W^\pm$, the neutral weak boson, Z^0 , also mediates weak interactions. The Z^0 boson can be produced and studied at high-energy particle colliders, such as the Large Hadron Collider (LHC) and the Large Electron–Positron (LEP) collider. The weak interaction is characterised by its extremely short range and brief lifetime, attributed to the large masses of the weak bosons.

At low energies (i.e. $E \ll 100$ GeV), the effective strength of the weak interaction is approximately 10^{-4} times that of the electromagnetic interaction, primarily due to suppression from the large masses of the $W^\pm$ and Z^0 bosons. As a result, the weak interaction is significantly weaker than both the electromagnetic and strong interactions, though it remains much stronger than gravity.

At higher energies (i.e. $E \gtrsim 100$ GeV), the weak and electromagnetic interactions unify into a single electroweak interaction. In this regime, the mass suppression associated with the $W^\pm$ and Z^0 bosons becomes negligible, and the electroweak interaction exhibits a strength comparable to that of the electromagnetic interaction. A more detailed discussion of the electroweak interaction is provided in Section 2.3.2.

2.3.1 Parity Violation in Weak Interactions

A distinctive feature of the weak interaction is its violation of parity symmetry. Parity refers to a spatial inversion transformation, where the coordinates of a particle's wavefunction are reflected. Under a parity transformation, the wavefunction $\psi(x, y, z)$ transforms as:

$$P : \psi(x, y, z) \rightarrow \psi(-x, -y, -z) \quad (2.11)$$

Here, P denotes the parity operator. Parity violation in weak interactions was first demonstrated in the experiment conducted by Chien-Shiung Wu and collaborators [30]. In this experiment, the directional emission of electrons in the β -decay of cobalt-60 nuclei was measured. When the nuclei were polarised, the majority of emitted electrons exhibited spins anti-aligned with the nuclear spin direction, thereby providing direct evidence for parity violation in weak processes.

This discovery introduced the fundamental concepts of helicity and chirality. Helicity is defined as the projection of a particle's spin onto its direction of motion. A particle with spin aligned (anti-aligned) with its momentum is said to have right-handed (left-handed) helicity. However, helicity is not Lorentz-invariant for massive particles, as it can be reversed under a suitable Lorentz boost.

In contrast, chirality is a Lorentz-invariant quantity defined via chiral projection operators:

$$P_{L/R} = \frac{1}{2}(1 \mp \gamma^5) \quad (2.12)$$

where $P_{L/R}$ projects out the left- or right-chiral component of a fermion field, and γ^5 is the fifth Dirac matrix. In the massless limit, helicity and chirality coincide; however, for massive fermions, they are distinct.

Only left-chiral fermions and right-chiral antifermions participate in weak interactions. Right-chiral fermions and left-chiral antifermions do not couple to the

weak interaction. The chiral decomposition of a Dirac spinor ψ is given by:

$$\psi_L = \frac{1}{2}(1 - \gamma^5)\psi, \quad \psi_R = \frac{1}{2}(1 + \gamma^5)\psi \quad (2.13)$$

where ψ_L and ψ_R denote the left- and right-handed components, respectively. These components satisfy the identity:

$$\bar{\psi}\psi = \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R \quad (2.14)$$

where the overbar denotes the Dirac adjoint. The discovery of parity violation and the chiral nature of weak interactions laid the groundwork for the unification of the electromagnetic and weak forces, which culminated in the electroweak theory.

2.3.2 Electroweak Theory

In the early 1970s, Abdus Salam, Sheldon Glashow, and Steven Weinberg [23] proposed the Electroweak Theory to unify the electromagnetic and weak interactions. They formulated the electroweak Lagrangian, $\mathcal{L}_{ew}$, in a massless form as follows:

$$\mathcal{L}_{ew} = \mathcal{L}_{symm} \quad (2.15)$$

Here, $\mathcal{L}_{symm}$ describes the interaction between fermions and electroweak gauge bosons and is written as:

$$\mathcal{L}_{symm} = -\frac{1}{4} \sum_k F_k^{\mu\nu} F_{\mu\nu}^k - \frac{1}{4} B^{\mu\nu} B_{\mu\nu} + \bar{\Psi}_L (i\gamma^\mu D_\mu) \Psi_L + \bar{\Psi}_R (i\gamma^\mu D_\mu) \Psi_R \quad (2.16)$$

In this expression, $F_k^{\mu\nu}$ ($k = 1, 2, 3$) and $B^{\mu\nu}$ are the field strength tensors corresponding to the weak isospin and weak hypercharge gauge fields, respectively. The left-handed fermion doublets Ψ_L and right-handed singlets Ψ_R are defined as:

$$\Psi_L = \begin{bmatrix} \nu_k \\ l_k \end{bmatrix}_L \quad (\text{Leptons}) \quad (2.17)$$

$$\Psi_L = \begin{bmatrix} u_k \\ d_k \end{bmatrix}_L \quad (\text{Quarks}) \quad (2.18)$$

$$\Psi_R = l_{R_k}, u_{R_k}, d_{R_k} \quad (2.19)$$

Here, l_k , u_k , and d_k represent the three generations of charged leptons, up-type quarks, and down-type quarks, respectively. The covariant derivative D_μ governing electroweak interactions is defined as:

$$\begin{aligned} D_\mu &= \partial_\mu + ig' \frac{Y}{2} B_\mu + ig T_k W_\mu^k \quad (\text{Left-Handed Fermions}) \\ D_\mu &= \partial_\mu + ig' \frac{Y}{2} B_\mu \quad (\text{Right-Handed Fermions}) \end{aligned} \quad (2.20)$$

Here, W_μ^k and B_μ denote the gauge fields associated with the weak isospin and weak hypercharge interactions, respectively. The weak isospin generators T_k are related to the Pauli matrices σ_k as:

$$T_k = \frac{1}{2} \sigma_k \quad (2.21)$$

The weak hypercharge Y is connected to the electric charge Q and the third component of weak isospin T_3 via the Gell-Mann–Nishijima relation:

$$Q = T_3 + \frac{Y}{2} \quad (2.22)$$

The values of T_3 and Y vary depending on a particle's electric charge and chirality, as shown in Table 2.4.

Particle Type	Chirality	Q	T_3	Y
Up-type quark, u_k	Left	$+\frac{2}{3}$	$+\frac{1}{2}$	$+\frac{1}{3}$
Down-type quark, d_k	Left	$-\frac{1}{3}$	$-\frac{1}{2}$	$+\frac{1}{3}$
Lepton neutrino, ν_k	Left	0	$+\frac{1}{2}$	-1
Lepton, l_k	Left	-1	$-\frac{1}{2}$	-1
Up-type quark, u_k	Right	$+\frac{2}{3}$	0	$+\frac{4}{3}$
Down-type quark, d_k	Right	$-\frac{1}{3}$	0	$-\frac{2}{3}$
Lepton, l_k	Right	-1	0	-2
$W^\pm$ Boson	-	± 1	± 1	0
Z Boson	-	0	0	0
Photon, γ	-	0	0	0
Gluon, g	-	0	0	0

Table 2.4: Electric charge Q , chirality, weak isospin T_3 , and weak hypercharge Y of particles in the Electroweak Theory.

The constants g and g' denote the coupling strengths for the weak isospin and weak hypercharge interactions, respectively. These constants are related through the weak mixing angle, or Weinberg angle θ_W , by:

$$\cos \theta_W = \frac{g'}{\sqrt{g'^2 + g^2}}, \quad \sin \theta_W = \frac{g}{\sqrt{g'^2 + g^2}} \quad (2.23)$$

This angle determines how the W_μ^3 and B_μ fields mix to form the physical photon (γ) and Z boson fields. The mixing can be expressed in matrix form as:

$$\begin{bmatrix} \gamma \\ Z \end{bmatrix} = \begin{bmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{bmatrix} \begin{bmatrix} B \\ W^3 \end{bmatrix} \quad (2.24)$$

The Electroweak Lagrangian possesses an $SU(2)_L \otimes U(1)_Y$ gauge symmetry. However, it presents two key limitations:

- It does not incorporate fermion masses (with the exception of neutrinos).
- It fails to account for the short range of the weak interaction, implying that the weak bosons must be massive.

To resolve these issues, the gauge symmetry of $\mathcal{L}_{\text{symm}}$ must be spontaneously broken. This is achieved via the Higgs mechanism [21, 22].

2.3.3 Spontaneous Symmetry Breaking and Higgs Mechanism

The Higgs mechanism constitutes a fundamental component of the Electroweak Theory, providing a consistent framework for the generation of mass for weak bosons and fermions. This mechanism introduces a scalar field doublet described by the Lagrangian $\mathcal{L}_{\text{Higgs}}$, which is incorporated into the electroweak Lagrangian $\mathcal{L}_{\text{ew}}$ as follows:

$$\mathcal{L}_{\text{ew}} = \mathcal{L}_{\text{symm}} + \mathcal{L}_{\text{Higgs}}, \quad (2.25)$$

with

$$\mathcal{L}_{\text{Higgs}} = (D_\mu \Phi)^\dagger (D^\mu \Phi) - V(\Phi^\dagger \Phi) \quad (2.26)$$

and

$$V(\Phi^\dagger \Phi) = -\mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2. \quad (2.27)$$

Here, Φ denotes a complex scalar doublet expressed as:

$$\Phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}, \quad (2.28)$$

where ϕ^0 and ϕ^+ correspond to the neutral and charged components of the scalar field, respectively. The scalar potential $V(\Phi^\dagger \Phi)$ depends on two real constants, μ and λ . Spontaneous symmetry breaking (SSB) occurs when the scalar field Φ acquires a non-zero vacuum expectation value (VEV), with only the neutral component contributing:

$$\Phi_{\text{vev}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix}, \quad v = \sqrt{\frac{\mu^2}{\lambda}}, \quad Y_\Phi = 1, \quad T_\Phi^3 = -\frac{1}{2}. \quad (2.29)$$

Here, v represents the non-zero VEV, valid for $\mu^2 > 0$. Its numerical value, $v = 246 \text{ GeV}$, is determined from the weak interaction coupling strength and the Fermi constant G_F :

$$v = \sqrt{\frac{1}{\sqrt{2}G_F}} \simeq 246 \text{ GeV}, \quad (2.30)$$

where $G_F \approx 1.166 \times 10^{-5} \text{ GeV}^{-2}$.

The non-zero VEV enables interaction between the scalar field and the weak isospin gauge fields W_μ^i , along with the hypercharge gauge field B_μ , through the kinetic term involving the covariant derivative D_μ . This interaction results in mass generation for the weak bosons, while the photon remains massless, as demonstrated in equations 2.31–2.33:

$$(D_\mu \Phi)^\dagger (D^\mu \Phi) = \left| \left(\partial_\mu + i \frac{g'}{2} B_\mu + i \frac{g}{2} \sigma_k W_\mu^k \right) \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v \end{bmatrix} \right|^2 \quad (2.31)$$

$$= \frac{1}{2} \left| \left(\partial_\mu + i \frac{g'}{2} B_\mu + i \frac{g}{2} \sigma_k W_\mu^k \right) \begin{bmatrix} 0 \\ v \end{bmatrix} \right|^2 \quad (2.32)$$

$$= \frac{v^2}{8} \left[g^2 \left((W_\mu^1)^2 + (W_\mu^2)^2 \right) + \left(g W_\mu^3 - g' B_\mu \right)^2 \right] \quad (2.33)$$

$$= \frac{1}{2} m_W^2 W_\mu^\dagger W^\mu + \frac{1}{2} m_Z^2 Z_\mu Z^\mu, \quad (2.34)$$

where $W_\mu^\pm$ and Z_μ denote the massive weak bosons, while A_μ represents the massless photon. Their definitions and corresponding masses are given by:

$$\begin{aligned} W_\mu^\pm &\equiv \frac{1}{\sqrt{2}} \left(W_\mu^1 \mp i W_\mu^2 \right), \quad m_W = \frac{gv}{2}, \\ Z_\mu &\equiv \frac{1}{\sqrt{g^2 + g'^2}} \left(g W_\mu^3 - g' B_\mu \right), \quad m_Z = \frac{v}{2} \sqrt{g^2 + g'^2}, \\ A_\mu &\equiv \frac{1}{\sqrt{g^2 + g'^2}} \left(g W_\mu^3 + g' B_\mu \right), \quad m_A = 0. \end{aligned} \quad (2.35)$$

Equation 2.35 shows that the masses of the weak bosons are directly proportional to the corresponding coupling constants and the VEV. Consequently, the Higgs mechanism spontaneously breaks the $SU(2)_L \otimes U(1)_Y$ gauge symmetry of the Electroweak Theory to $U(1)_{\text{em}}$, preserving only the electromagnetic gauge sym-

metry.

2.3.4 Higgs Mechanism for Fermions

In addition to the weak bosons, fermions also acquire mass via the Higgs mechanism. This occurs through interactions between the scalar field and fermion fields, described by Yukawa interactions within the $\mathcal{L}_{ew}$ term during spontaneous symmetry breaking (SSB). The full Lagrangian takes the form:

$$\mathcal{L}_{ew} = \mathcal{L}_{symm} + \mathcal{L}_{yuk} + \mathcal{L}_{Higgs} \quad (2.36)$$

The Yukawa Lagrangian, $\mathcal{L}_{yuk}$, encodes the interactions between the fermion fields and the scalar field doublet, and is given by:

$$\mathcal{L}_{yuk} = -\sum_k \left(y_{e_k} \bar{\mathbf{L}}_k \Phi e_{R_k} + y_{d_k} \bar{\mathbf{Q}}_k \Phi d_{R_k} + y_{u_k} \bar{\mathbf{Q}}_k \tilde{\Phi} u_{R_k} \right) + h.c. \quad , \quad k = \{1, 2, 3\} \quad (2.37)$$

Here, $\tilde{\Phi}$ denotes the conjugate scalar field, expressed as:

$$\tilde{\Phi} = \begin{bmatrix} \phi^{0*} \\ \phi^{+*} \end{bmatrix} \quad (2.38)$$

The quantity $\bar{\mathbf{L}}_k$ denotes the adjoint of the left-handed lepton field doublet:

$$\bar{\mathbf{L}}_k = [\bar{\nu}_{L_k} \quad \bar{e}_{L_k}] \quad (2.39)$$

In this expression, e_{L_k} and ν_{L_k} denote the left-handed charged lepton and neutrino fields, respectively, while e_{R_k} denotes the corresponding right-handed lepton singlet. Similarly, $\bar{\mathbf{Q}}_k$ represents the adjoint of the left-handed quark field doublet:

$$\bar{\mathbf{Q}}_k = [\bar{u}_{L_k} \quad \bar{d}_{L_k}] \quad (2.40)$$

Here, u_{L_k} and d_{L_k} denote the left-handed up-type and down-type quark fields,

and u_{R_k} and d_{R_k} represent the right-handed singlets. The index $k = \{1, 2, 3\}$ refers to the three fermion generations. The parameters y_{e_k} , y_{u_k} , and y_{d_k} are the corresponding Yukawa coupling constants.

Upon spontaneous symmetry breaking, the vacuum expectation value (VEV) of the scalar field, Φ_{vev} , induces fermion masses. Equation 2.37 then becomes:

$$\begin{aligned} \mathcal{L}_{yuk} = - & \left(\frac{y_{e_k}}{\sqrt{2}} \begin{bmatrix} \bar{\nu}_{L_k} & e_{L_k} \end{bmatrix} \begin{bmatrix} 0 \\ v \end{bmatrix} e_{R_k} + \frac{y_{d_k}}{\sqrt{2}} \begin{bmatrix} \bar{u}_{L_k} & \bar{d}_{L_k} \end{bmatrix} \begin{bmatrix} 0 \\ v \end{bmatrix} d_{R_k} \right. \\ & \left. + \frac{y_{u_k}}{\sqrt{2}} \begin{bmatrix} \bar{u}_{L_k} & \bar{d}_{L_k} \end{bmatrix} \begin{bmatrix} v \\ 0 \end{bmatrix} u_{R_k} \right) + h.c. \end{aligned} \quad (2.41)$$

This expression simplifies to:

$$\begin{aligned} \mathcal{L}_{yuk} = - & \left(\frac{y_{e_k} v}{\sqrt{2}} e_{L_k} e_{R_k} + \frac{y_{d_k} v}{\sqrt{2}} \bar{d}_{L_k} d_{R_k} + \frac{y_{u_k} v}{\sqrt{2}} \bar{u}_{L_k} u_{R_k} \right. \\ & \left. + \frac{y_{e_k} v}{\sqrt{2}} e_{R_k} e_{L_k} + \frac{y_{d_k} v}{\sqrt{2}} \bar{d}_{R_k} d_{L_k} + \frac{y_{u_k} v}{\sqrt{2}} \bar{u}_{R_k} u_{L_k} \right) \end{aligned} \quad (2.42)$$

This can be re-expressed in terms of the fermion masses:

$$\begin{aligned} \mathcal{L}_{yuk} = - & (m_{e_k} (e_{L_k} e_{R_k} + e_{R_k} e_{L_k}) + m_{d_k} (\bar{d}_{L_k} d_{R_k} + \bar{d}_{R_k} d_{L_k}) \\ & + m_{u_k} (\bar{u}_{L_k} u_{R_k} + \bar{u}_{R_k} u_{L_k})) \end{aligned} \quad (2.43)$$

From equations 2.42 and 2.43, the masses of the fermions are seen to be proportional to their respective Yukawa coupling constants and the scalar field VEV, v , as follows:

$$m_{e_k} = \frac{y_{e_k} v}{\sqrt{2}} \quad , \quad m_{u_k} = \frac{y_{u_k} v}{\sqrt{2}} \quad , \quad m_{d_k} = \frac{y_{d_k} v}{\sqrt{2}} \quad (2.44)$$

Thus, the Higgs mechanism provides a coherent and consistent framework for explaining the origin of mass for both weak bosons and fermions within electroweak

theory. Additionally, this mechanism predicts the existence of a massive scalar particle—the Higgs boson—whose properties will be addressed in Section 2.3.5.

2.3.5 Higgs Boson

The Higgs boson emerges as a bosonic excitation of the scalar field around the non-zero vacuum expectation value (VEV) induced by spontaneous symmetry breaking (SSB). This configuration is expressed by including the Higgs field H in Equation 2.29, resulting in:

$$\Phi_{\text{vev},H} = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + H \end{bmatrix} \quad (2.45)$$

Substituting this expression into Equations 2.26 and 2.27, and retaining only the terms involving H , the Lagrangian describing the Higgs boson becomes:

$$\mathcal{L}_H = \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{m_H^2}{2} H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 + \mathcal{L}_{H,int} \quad (2.46)$$

Here, $\mathcal{L}_H$ denotes the Lagrangian for the Higgs boson, where $m_H = \sqrt{2\lambda}v$ is the Higgs boson mass. The term $\mathcal{L}_{H,int}$ encapsulates the interaction terms between the Higgs boson and the weak gauge bosons, explicitly written as:

$$\mathcal{L}_{H,int} = \frac{2m_W^2}{v} H W_\mu^\dagger W^\mu + \frac{m_Z^2}{v} H Z_\mu Z^\mu + \frac{m_W^2}{v^2} H H W_\mu^\dagger W^\mu + \frac{m_Z^2}{2v^2} H H Z_\mu Z^\mu \quad (2.47)$$

The complete Lagrangian describing the Higgs boson, including self-interactions and couplings with the weak bosons, is:

$$\begin{aligned} \mathcal{L}_H = & \frac{1}{2} \partial_\mu H \partial^\mu H - \frac{m_H^2}{2} H^2 - \lambda v H^3 - \frac{\lambda}{4} H^4 \\ & + \frac{2m_W^2}{v} H W_\mu^\dagger W^\mu + \frac{m_Z^2}{v} H Z_\mu Z^\mu + \frac{m_W^2}{v^2} H H W_\mu^\dagger W^\mu + \frac{m_Z^2}{2v^2} H H Z_\mu Z^\mu \end{aligned} \quad (2.48)$$

In this expression, the first term describes the kinetic energy of the Higgs boson, the second term represents the mass term, the third and fourth terms account for self-interactions, and the final four terms describe couplings to the $W^\pm$ and Z

bosons.

The Higgs mechanism consequently predicts a massive scalar boson exhibiting both self-interactions and interactions with the weak bosons. The associated Feynman rules are summarised in Table 2.5.

Interaction Vertex	Coupling Variable	Coupling Term
HHH	c_{HHH}	$-i\lambda v \cdot 3! = -3i\frac{m_H^2}{v}$
$HHHH$	c_{HHHH}	$-i\frac{\lambda}{4} \cdot 4! = -3i\frac{m_H^2}{v^2}$
$HW_\mu^+ W_\nu^-$	c_{HWW}	$2i\frac{m_W^2}{v} \eta_{\mu\nu}$
$HHW_\mu^+ W_\nu^-$	c_{HHWW}	$2! \cdot i\frac{m_W^2}{v^2} \eta_{\mu\nu}$
$HZ_\mu Z_\nu$	c_{HZZ}	$2! \cdot i\frac{m_Z^2}{v} \eta_{\mu\nu}$
$HHZ_\mu Z_\nu$	c_{HHZZ}	$2! \cdot 2! \cdot i\frac{m_Z^2}{2v^2} \eta_{\mu\nu}$

Table 2.5: Feynman rules for Higgs boson self-interactions and interactions with weak bosons. Factorials represent symmetry factors, accounting for permutations of identical particles in Feynman diagrams. The Minkowski metric is denoted by $\eta_{\mu\nu}$.

In addition to self-interactions and couplings with weak bosons, the Higgs boson also interacts with fermions through Yukawa interactions. The Higgs-dependent terms are extracted from the Yukawa Lagrangians in Equations 2.37 and 2.41:

$$\begin{aligned} \mathcal{L}_{\text{yuk},H} = - \left(\frac{y_{e_k} H}{\sqrt{2}} \bar{e}_{L_k} e_{R_k} + \frac{y_{d_k} H}{\sqrt{2}} \bar{d}_{L_k} d_{R_k} + \frac{y_{u_k} H}{\sqrt{2}} \bar{u}_{L_k} u_{R_k} \right. \\ \left. + \frac{y_{e_k} H}{\sqrt{2}} \bar{e}_{R_k} e_{L_k} + \frac{y_{d_k} H}{\sqrt{2}} \bar{d}_{R_k} d_{L_k} + \frac{y_{u_k} H}{\sqrt{2}} \bar{u}_{R_k} u_{L_k} \right) \end{aligned} \quad (2.49)$$

$$\begin{aligned} \mathcal{L}_{\text{yuk},H} = - \left(\frac{m_{e_i}}{v} H (\bar{e}_{L_k} e_{R_k} + \bar{e}_{R_k} e_{L_k}) + \frac{m_{d_i}}{v} H (\bar{d}_{L_k} d_{R_k} + \bar{d}_{R_k} d_{L_k}) \right. \\ \left. + \frac{m_{u_i}}{v} H (\bar{u}_{L_k} u_{R_k} + \bar{u}_{R_k} u_{L_k}) \right) \end{aligned} \quad (2.50)$$

The coupling strength between the Higgs boson and each fermion species scales proportionally with the fermion mass, as summarised in Table 2.6.

Despite strong couplings to massive particles, the Higgs boson does not couple

Interaction Vertex	Coupling Term
Higgs to lepton: $He_i e_i$	$g_{Hee_i} = -i \frac{m_{e_i}}{v}$
Higgs to down-type quark: $Hd_i d_i$	$g_{Hdd_i} = -i \frac{m_{d_i}}{v}$
Higgs to up-type quark: $Hu_i u_i$	$g_{Huu_i} = -i \frac{m_{u_i}}{v}$

Table 2.6: Feynman rules for Higgs boson couplings to fermions.

to gluons, because gluons interact exclusively with particles carrying color charge, such as quarks and other gluons. No direct interaction exists between the photon and the Higgs boson either, since the Electroweak Lagrangian lacks such a term after SSB, as evident in Equation 2.47. Nevertheless, loop-induced processes allow indirect interactions of the Higgs boson with gluons and photons, mediated by virtual loops of massive particles such as quarks, leptons, and $W^\pm$ bosons. No analogous Higgs-photon interaction arises via Z boson loops due to the absence of a direct Z -photon coupling.

Although indirect interactions extend the Higgs boson's influence throughout the Standard Model, the discovery of the physical Higgs boson in 2012 marks a significant milestone, as it validates the Higgs mechanism. Further details on the discovery are presented in Section 2.3.6.

2.3.6 Higgs Boson Production, Discovery, and Decay Modes at the LHC

The Higgs mechanism provides an elegant framework for spontaneous symmetry breaking (SSB) within the electroweak theory. Prior to the operation of the LHC, the Higgs boson had not been observed in nature. The first confirmed observation occurred at the LHC, currently the only collider capable of producing Higgs bosons at high rates, although previous colliders have also contributed to the theoretical and experimental groundwork in Higgs boson studies.

2.3.6.1 Higgs Boson Production at the LHC

At the LHC, Higgs bosons are produced through four primary mechanisms [31]. The relative contributions of these production channels depend on various factors, including interaction strengths and the parton distribution functions (PDFs) of the proton constituents. Figure 2.1 and the descriptions below outline the Feynman diagrams and physical processes associated with the dominant Higgs boson production modes.

- Gluon-gluon fusion (ggF)

Gluon-gluon fusion represents the dominant Higgs production mechanism at the LHC, contributing approximately 87% of the total production rate. In this process, gluons from the colliding protons interact via a virtual top-quark loop to yield a Higgs boson. Despite proceeding at the loop level, the process is significantly enhanced by the strong Yukawa coupling between the Higgs boson and the top quark. Additionally, gluons constitute the most abundant partons in the proton, and the high centre-of-mass energy of the LHC ensures frequent gluon interactions, further amplifying the ggF cross-section.

- Vector boson fusion (VBF)

Vector boson fusion is the second most probable production mode, responsible for roughly 7% of Higgs boson events at the LHC. This mechanism involves the emission of virtual weak bosons (W or Z) from incoming quarks,

which subsequently fuse to produce a Higgs boson. Although weak bosons exhibit sizeable couplings to the Higgs due to their mass, their emission occurs less frequently than that of gluons, primarily because of the smaller electroweak coupling constants. As a result, the VBF contribution remains subdominant relative to ggF .

- Vector boson associated production (VH), or “Higgsstrahlung”

In vector boson associated production, which contributes approximately 4% of the total, a quark–antiquark pair annihilates to produce a W or Z boson that subsequently emits a Higgs boson. This process is suppressed by limited phase space, owing to the large invariant mass of the final-state system. Consequently, the VH production rate remains lower than those of ggF and VBF .

- Top quark associated production ($t\bar{t}H$)

Top quark associated production constitutes the rarest among the leading production channels, with a contribution of about 1%. In this process, two gluons interact to produce a top–antitop quark pair alongside a Higgs boson. The substantial mass of the final-state particles constrains the available phase space, resulting in a comparatively low production probability. Despite the small cross-section, this channel offers a direct probe of the top–Higgs Yukawa coupling and is therefore of considerable interest in precision studies of the Higgs sector.

Although gluon-gluon fusion dominates Higgs boson production, this mode presents significant challenges for experimental analyses due to large backgrounds arising from QCD jets. To enhance signal sensitivity, both the ATLAS and CMS collaborations have developed strategies that integrate multiple production and decay channels. This inclusive approach was instrumental in the discovery of the Higgs boson in 2012. A detailed discussion of the discovery and the associated decay channels is provided in Section 2.3.6.2.

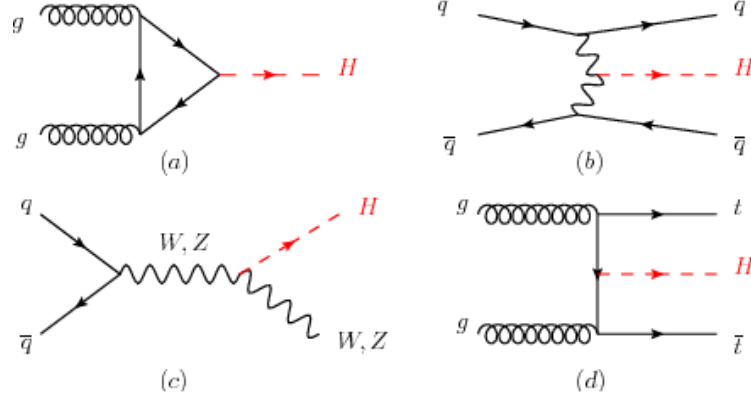


Figure 2.1: Feynman diagrams for Higgs boson production modes at the LHC: a) Gluon-gluon fusion (ggF), b) Vector boson fusion (VBF), c) Vector boson associated production (VH), d) Top quark associated production ($t\bar{t}H$) [1]

2.3.6.2 Higgs Boson Discovery and Decay Channels at the LHC

In July 2012, the ATLAS and CMS experiments at the LHC announced the discovery of the Higgs boson, with a measured mass of approximately 125.09 ± 0.21 (stat) ± 0.11 (syst) GeV, as reported in references [16, 17]. These results align with Standard Model predictions. The discovery combined results from all four Higgs boson production mechanisms and multiple decay channels, including $H \rightarrow \bar{b}b$, $H \rightarrow \tau^+\tau^-$, $H \rightarrow WW^* \rightarrow l^+\nu l^-\nu$, $H \rightarrow ZZ^* \rightarrow 4l$, and $H \rightarrow \gamma\gamma$. Here, l denotes electrons or muons, while ν represents neutrinos.

For the $H \rightarrow WW^* \rightarrow l^+\nu l^-\nu$ ² and $H \rightarrow ZZ^*$ channels, one or more of the W or Z bosons may be produced off-shell, possessing a virtual mass different from the physical mass in order to conserve momentum and energy during the decay process.

Although the combination of all channels enabled the discovery, the most significant contributions originated from the $H \rightarrow ZZ^* \rightarrow 4l$ and $H \rightarrow \gamma\gamma$ channels. These channels exhibited excesses of events with statistical significances of 3.6σ and 4.6σ in ATLAS, and 3.2σ and 4.1σ in CMS, respectively. As a result, the ATLAS and CMS collaborations have pursued more refined searches in individ-

²Neutrinos (ν) traverse the detector material without direct observation due to their extremely weak interaction with matter. Therefore, the ATLAS and CMS experiments employ the missing transverse energy (MET) variable to account for undetected energy resulting from neutrinos in $H \rightarrow WW^*$ decays. MET provides an estimate of the likelihood of neutrino presence originating from WW^* decays.

ual decay channels, each characterised by distinct branching ratios, as presented in Table 2.7.

Higgs Boson Decay Channel	Branching Ratio	Relative Uncertainty
$H \rightarrow \bar{b}b$	5.82×10^{-1}	+1.2% / -1.3%
$H \rightarrow WW^*$	2.14×10^{-1}	$\pm 1.5\%$
$H \rightarrow \tau^+\tau^-$	6.27×10^{-2}	$\pm 1.6\%$
$H \rightarrow \bar{c}c$	2.89×10^{-2}	+5.5% / -2.0%
$H \rightarrow ZZ^*$	2.62×10^{-2}	$\pm 1.5\%$
$H \rightarrow \gamma\gamma$	2.27×10^{-3}	$\pm 2.1\%$
$H \rightarrow Z\gamma$	1.53×10^{-3}	$\pm 5.8\%$
$H \rightarrow \mu^+\mu^-$	2.18×10^{-4}	$\pm 1.7\%$

Table 2.7: Branching ratios and relative uncertainties for several Higgs boson decay channels, listed in descending order [15].

During Run 1 and Run 2 of the LHC, several Higgs boson decay channels were conclusively observed by the CMS and ATLAS experiments. The status of each decay channel is summarised in Table 2.8.

Higgs Boson Decay Channel	Search Status
$H \rightarrow \bar{b}b$	Discovery [32]
$H \rightarrow WW^*$	Discovery [33]
$H \rightarrow \tau^+\tau^-$	Discovery [34]
$H \rightarrow \bar{c}c$	No Evidence Yet [35]
$H \rightarrow ZZ^*$	Discovery [36]
$H \rightarrow \gamma\gamma$	Discovery [36]
$H \rightarrow Z\gamma$	Evidence [37]
$H \rightarrow \mu^+\mu^-$	Evidence [38]

Table 2.8: Current status of Higgs boson searches across different decay channels.

Although several decay channels remain unobserved, the discoveries to date have opened new directions for investigating Higgs boson processes. One such avenue involves Higgs boson pair production, particularly in the $HH \rightarrow \bar{b}b\bar{b}b$ final state, as discussed in Section 2.5. Overall, the discovery of the Higgs boson provides compelling validation of the Higgs mechanism as the explanation for electroweak symmetry breaking within the Standard Model framework.

2.4 Hierarchy Problem and Physics Beyond the Standard Model (BSM)

Sections 2.1 to 2.3 have described how fundamental particles and their interactions are explained by the Standard Model, covering Quantum Electrodynamics (QED), Quantum Chromodynamics (QCD), Electroweak Theory, and the Higgs mechanism, which accounts for spontaneous symmetry breaking (SSB) in the electroweak theory and mass generation for all massive fundamental particles. However, the Standard Model has limitations. The framework does not incorporate gravity, one of the four fundamental forces governing the universe, making the unification of all four forces particularly challenging. One of the major obstacles to achieving unification is known as the hierarchy problem.

2.4.1 Hierarchy Problem and Higgs Boson Mass Correction

The hierarchy problem in particle physics arises from the mismatch between the energy scale of quantum gravity and the physics described by the Standard Model [39]. Quantum gravity aims to describe gravity within the framework of quantum mechanics. While the Standard Model successfully describes particle physics up to the TeV scale (corresponding to the energy scale of electroweak interactions), the framework begins to break down near the Planck scale, where quantum gravitational effects become significant (i.e., $M_p \sim 10^{19}$ GeV).

A prominent manifestation of the hierarchy problem is found in the corrections to the Higgs boson mass. The bare mass of the Higgs boson represents the mass in the absence of interactions with other particles. In physical reality, the Higgs boson interacts with various particles, and these interactions contribute to the physical mass. Such contributions are evaluated through radiative corrections, employing loop integrals involving virtual massive particles interacting with the Higgs boson (as illustrated by the Feynman diagram in Figure 2.2), which can be mathematically expressed as follows [39]:

$$m_{H,corr}^2 = m_H^2 + \Delta m_H^2 \quad (2.51)$$

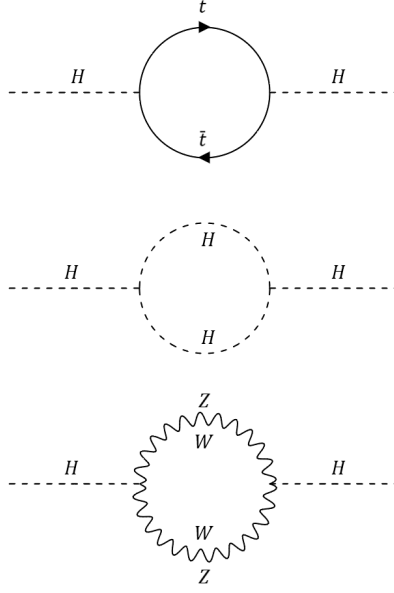


Figure 2.2: The Feynman diagram for one-loop Higgs boson mass correction from the top quark, $W^\pm$ boson, Z boson, and the Higgs boson.

In this expression, $m_{H,corr}^2$ denotes the corrected Higgs boson mass, corresponding to the experimentally observed value. The term m_H^2 represents the bare Higgs boson mass, while Δm_H^2 accounts for the radiative corrections arising from virtual loops of massive particles, such as the Higgs boson itself, weak bosons, and top quarks. This correction term can be expressed as:

$$\Delta m_H^2 = f(m_H, m_Z, m_W, m_t, \Lambda) \quad (2.52)$$

Here, f denotes a function of the parameters resulting from the loop integrals. The parameter Λ refers to the cutoff scale, which defines the energy scale at which the loop integrals are regularised to prevent divergence. For the case of a one-loop integral, Equation 2.52 becomes:

$$\Delta m_H^2 = \frac{3\Lambda^2}{8\pi^2 v^2} (m_H^2 + 2m_W^2 + m_Z^2 - 4m_t^2) + \dots \quad (2.53)$$

The first four terms represent the leading-order corrections, while the ellipsis indicates the presence of logarithmic contributions from virtual massive particles in the loop, resulting from the one-loop integral. However, the logarithmic terms grow

much more slowly and become less significant compared to the leading-order terms at very large values of Λ . For simplicity, only the leading-order contribution to Δm_H^2 is considered here. Equation 2.53 highlights the critical role played by the cutoff scale in the Higgs boson mass correction, demonstrating strong sensitivity to the energy scale involved. The hierarchy problem becomes apparent when the energy scale of quantum gravity is introduced, i.e., $\Lambda = M_p$. Under such conditions, the corrected Higgs boson mass exceeds the bare mass by approximately 17 orders of magnitude. This discrepancy necessitates an "unnatural" (or extreme) degree of fine-tuning for m_H^2 in order to reproduce the observed Higgs boson mass. Consequently, numerous theoretical models have been proposed to address this issue, including Supersymmetry and models involving extra spatial dimensions. These solutions are collectively categorised as theories Beyond the Standard Model (BSM). This thesis focuses on two such BSM models: the Two-Higgs-Doublet Model (2HDM) and the Randall-Sundrum Model.

2.4.2 Two Higgs Doublet Model (2HDM)

One of the common Beyond the Standard Model (BSM) frameworks is the two Higgs doublet model (2HDM) [40]. Recognised as one of the simplest extensions of the Standard Model (SM) Higgs mechanism, the model introduces two Higgs doublets instead of the single doublet found in the SM. The 2HDM plays a pivotal role in the Minimal Supersymmetric Standard Model (MSSM), a specific theoretical framework that realises Supersymmetry (SUSY) [41]. This necessity stems from the requirement for the MSSM Lagrangian to remain invariant under SUSY transformations, while simultaneously enabling mass generation for both up-type and down-type quarks [41]³.

2.4.2.1 Supersymmetry

Supersymmetry represents a BSM theoretical framework that proposes fermionic or bosonic superpartners for the particles in the SM. These superpartners arise from the action of the supersymmetry operator Q applied to the fermionic and bosonic fields, namely

$$Q|\text{Boson}\rangle = |\text{Fermion}\rangle, \quad Q|\text{Fermion}\rangle = |\text{Boson}\rangle \quad (2.54)$$

where $|\text{Fermion}\rangle$ and $|\text{Boson}\rangle$ denote the fermionic and bosonic fields, respectively. The predicted superpartners are expected to interact with the Higgs field, effectively cancelling the contributions from each massive SM particle to the Higgs boson mass correction, as described in equation 2.53. Such cancellations offer a resolution to the "unnatural" fine-tuning problem associated with the Higgs mass across energy scales, thereby addressing the hierarchy problem.

Detecting these superpartners remains a significant challenge, as their expected

³In the MSSM, the supersymmetric nature of the Lagrangian imposes the holomorphy condition on the superpotential, thereby excluding complex conjugate or adjoint terms. Yukawa interactions are embedded within the superpotential, and to preserve supersymmetry while facilitating mass generation for up-type quarks, the usual Higgs conjugate terms employed in the Standard Model cannot be utilised. Consequently, a second Higgs doublet is introduced. This leads to the presence of two Higgs doublets in the MSSM: H_u , which couples to up-type quarks, and H_d , which couples to down-type quarks and charged leptons. The inclusion of both Higgs doublets ensures that all Yukawa interactions remain holomorphic, preserving supersymmetry within the theory.

masses could lie beyond the reach of current collider energies, such as those at the Large Hadron Collider (LHC). Despite these obstacles, the ATLAS and CMS experiments have conducted extensive searches for SUSY particles at the LHC, yet no evidence has been observed to date [42]. Although supersymmetry itself does not mandate the presence of two Higgs doublets, the MSSM's incorporation of such a structure underscores the importance of exploring the phenomenology of 2HDM in testing MSSM predictions. Consequently, the ATLAS and CMS experiments have prioritised 2HDM as a central focus within their BSM search programmes at the LHC.

2.4.2.2 The Physics of 2HDM

The physics of the Two Higgs Doublet Model (2HDM) [40] involves the introduction of two scalar doublets, Φ_1 and Φ_2 , into the Higgs Lagrangian, denoted as $\mathcal{L}_{2HDM}$, such that

$$\Phi \rightarrow \begin{bmatrix} \Phi_1 \\ \Phi_2 \end{bmatrix} \quad (2.55)$$

where $\mathcal{L}_{2HDM}$ can be expressed as follows:

$$\mathcal{L}_{2HDM} = (D_\mu \Phi_1)^\dagger (D^\mu \Phi_1) + (D_\mu \Phi_2)^\dagger (D^\mu \Phi_2) - V(\Phi_1, \Phi_2) \quad (2.56)$$

Here, the general form of the 2HDM potential is given by

$$\begin{aligned} V(\Phi_1, \Phi_2) = & m_1^2 \Phi_1^\dagger \Phi_1 + m_2^2 \Phi_2^\dagger \Phi_2 + m_{12}^2 \Phi_1^\dagger \Phi_2 + m_{12}^{2*} \Phi_2^\dagger \Phi_1 \\ & + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1) (\Phi_2^\dagger \Phi_2) + \lambda_4 (\Phi_2^\dagger \Phi_2) (\Phi_1^\dagger \Phi_1) \\ & + \frac{\lambda_5}{2} (\Phi_1^\dagger \Phi_2)^2 + \frac{\lambda_5^*}{2} (\Phi_2^\dagger \Phi_1)^2 + \left\{ \left[\lambda_6 (\Phi_1^\dagger \Phi_1) + \lambda_7 (\Phi_2^\dagger \Phi_2) \right] (\Phi_1^\dagger \Phi_2) + \text{h.c.} \right\} \end{aligned} \quad (2.57)$$

Here, m_1 and m_2 are the mass parameters associated with the first and second doublets, respectively. The terms m_{12} and m_{12}^* represent the complex mass parameter and its conjugate, corresponding to the mixing between the two doublets. The parameters λ_{1-4} are real-valued, while λ_{5-7} are generally complex. Each doublet

contains charged and neutral components, as follows:

$$\Phi_1 = \begin{bmatrix} \phi_1^+ \\ \phi_1^0 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} \phi_2^+ \\ \phi_2^0 \end{bmatrix} \quad (2.58)$$

Here, ϕ_1^+ and ϕ_2^+ are the charged components of the doublets, while ϕ_1^0 and ϕ_2^0 are the neutral components.

For most phenomenological studies of the 2HDM, the parameters m_{12} and λ_{5-7} are typically assumed to be real. This assumption ensures that the 2HDM Lagrangian preserves CP symmetry, meaning that interactions involving the two doublets do not violate charge or parity transformations. In addition to CP conservation, it is also common to assume the absence of flavour-changing neutral currents (FCNCs) at tree level. This can be achieved by imposing a softly broken Z_2 symmetry on the Lagrangian, under which

$$\Phi_1 \rightarrow \Phi_1, \quad \Phi_2 \rightarrow -\Phi_2 \quad (2.59)$$

Flavour-changing neutral currents (FCNCs) are highly suppressed processes in the Standard Model (SM), primarily due to the GIM mechanism [43][44]. However, in the 2HDM, FCNC interactions can occur at tree level⁴.

With these assumptions in place, the potential $V(\Phi_1, \Phi_2)$ simplifies to

$$\begin{aligned} V(\Phi_1, \Phi_2) = & m_1^2 \Phi_1^\dagger \Phi_1 + m_2^2 \Phi_2^\dagger \Phi_2 - m_{12}^2 \left[\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1 \right] \\ & + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 + \lambda_3 (\Phi_1^\dagger \Phi_1) (\Phi_2^\dagger \Phi_2) \\ & + \lambda_4 (\Phi_2^\dagger \Phi_2) (\Phi_1^\dagger \Phi_1) + \frac{\lambda_5}{2} \left[\left(\Phi_1^\dagger \Phi_2 \right)^2 + \left(\Phi_2^\dagger \Phi_1 \right)^2 \right] \end{aligned} \quad (2.60)$$

⁴FCNC interactions involve quarks or leptons changing flavour without changing their electric charge. In the SM, such processes are forbidden at tree level and are extremely suppressed at higher orders. For instance, in the quark sector, this suppression arises from the structure of the electroweak Lagrangian and the Glashow-Iliopoulos-Maiani (GIM) mechanism. The electroweak Lagrangian ensures flavour conservation in neutral current interactions at tree level, while the GIM mechanism further suppresses FCNC interactions at loop level. This suppression results from cancellations among contributions from different quark flavours, aided by the small mass differences of lighter quarks and the specific structure of the CKM matrix, leading to extremely small transition amplitudes. However, in the 2HDM, the structure of the fermion mass matrix permits tree-level FCNC interactions.

where m_{12} is the soft term that breaks the Z_2 symmetry⁵. The minima of this potential can be determined from the vacuum expectation values (VEVs) of each doublet, given by

$$\Phi_{1,\text{vev}} = \begin{bmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{bmatrix}, \quad \Phi_{2,\text{vev}} = \begin{bmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{bmatrix} \quad (2.61)$$

where v_1 and v_2 represent the VEVs of each doublet. Expanding around these minima, each doublet can be expressed in terms of four scalar fields as

$$\Phi_a = \begin{bmatrix} \phi_a^+ \\ \frac{v_a + \rho_a + i\eta_a}{\sqrt{2}} \end{bmatrix}, \quad a = 1, 2 \quad (2.62)$$

where ϕ_a^+ corresponds to the charged component of the doublet fields. The neutral component of each doublet consists of three terms: the vacuum expectation value v_a , and two scalar fields, ρ_a and η_a .

By substituting Eq. 2.62 into Eq. 2.56, the 2HDM Lagrangian contains five distinct scalar particles: two charged Higgs bosons $H^\pm$, a pseudoscalar boson A , a heavy neutral scalar X , and the SM-like Higgs boson H . The mass terms and their corresponding eigenstates for each of these particles in the Lagrangian can be expressed as follows:

$$\begin{aligned} \mathcal{L}_{H^\pm, \text{mass}} &= m_{H^\pm}^2 H^+ H^-, \quad \mathcal{L}_{A, \text{mass}} = m_A^2 A^2 \\ \mathcal{L}_{X, \text{mass}} &= m_X^2 X^2, \quad \mathcal{L}_{H, \text{mass}} = m_H^2 H^2 \end{aligned} \quad (2.63)$$

The corresponding eigenstates are related by the following transformations:

$$\begin{bmatrix} G^\pm \\ H^\pm \end{bmatrix} = \begin{bmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{bmatrix} \begin{bmatrix} \phi_1^\pm \\ \phi_2^\pm \end{bmatrix} \quad (2.64)$$

⁵The Z_2 symmetry refers to the Lagrangian that remains invariant under the sign flip of the scalar fields. In the case of 2HDM, only the m_{12} term explicitly breaks the Z_2 symmetry.

$$\begin{bmatrix} G^0 \\ A \end{bmatrix} = \begin{bmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad (2.65)$$

$$\begin{bmatrix} X \\ H \end{bmatrix} = \begin{bmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \quad (2.66)$$

The masses of the five Higgs bosons are obtained by diagonalising the mass matrices M_{ij}^k , where $k = \phi^\pm, \eta, \rho$, corresponding to the scalar fields in the doublets. These can be expressed as:

$$\phi_i^- M_{ij}^{\pm} \phi_j^+ = [m_{12}^2 - (\lambda_4 + \lambda_5)v_1 v_2] \begin{bmatrix} \phi_1^- \\ \phi_2^- \end{bmatrix} \begin{bmatrix} \tan \beta & -1 \\ -1 & \frac{1}{\tan \beta} \end{bmatrix} \begin{bmatrix} \phi_1^+ \\ \phi_2^+ \end{bmatrix} \quad (2.67)$$

$$\eta_i M_{ij}^{\eta} \eta_j = [m_{12}^2 - 2\lambda_5 v_1 v_2] \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \begin{bmatrix} \tan \beta & -1 \\ -1 & \frac{1}{\tan \beta} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad (2.68)$$

$$\rho_i M_{ij}^{\rho} \rho_j = \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \begin{bmatrix} m_{12}^2 \tan \beta + v_1 \lambda_1 & -m_{12}^2 + \lambda_{345} v_1 v_2 \\ -m_{12}^2 + \lambda_{345} v_1 v_2 & m_{12}^2 + \frac{1}{\tan \beta} v_2 \lambda_2 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} \quad (2.69)$$

$$\lambda_{345} = \lambda_3 + \lambda_4 + \lambda_5 \quad (2.70)$$

The mass eigenvalues are then given by:

$$\begin{aligned}
m_{H^\pm}^2 &= \left(\frac{m_{12}^2}{v_1 v_2} - \lambda_4 - \lambda_5 \right) v_{2\text{HDM}}^2 \\
m_A^2 &= \left(\frac{m_{12}^2}{v_1 v_2} - 2\lambda_5 \right) v_{2\text{HDM}}^2 \\
m_{H,h}^2 &= \frac{1}{2} \left[M_{11}^\rho + M_{22}^\rho \pm \sqrt{(M_{11}^\rho - M_{22}^\rho)^2 + 4M_{12}^\rho} \right]
\end{aligned} \tag{2.71}$$

Here, α is the mixing angle between the two scalar fields ρ_i . The term $v_{2\text{HDM}}^2$ represents the sum of the squares of the two VEVs, while $\tan\beta$ is defined as the ratio between these two VEVs:

$$\begin{aligned}
v_{2\text{HDM}}^2 &= v_1^2 + v_2^2 \\
\tan\beta &= \frac{v_2}{v_1}, \quad \sin\beta = \frac{v_2}{v_{2\text{HDM}}}, \quad \cos\beta = \frac{v_1}{v_{2\text{HDM}}}
\end{aligned} \tag{2.72}$$

From the equations 2.64-2.70 above, the charged and neutral Goldstone boson terms (i.e., G^0 , $G^\pm$) are absorbed by the weak boson interaction terms, resulting in no individual Goldstone boson mass terms present in the 2HDM Lagrangian. In order to validate the 2HDM model, six essential free parameters must be measured: the four masses of the different Higgs bosons, along with α and $\tan\beta$.

2.4.2.3 Mass Generation to Weak Bosons and Fermions

While the 2HDM features a rich spectrum of eigenstates leading to five distinct Higgs bosons, it must also provide masses to the weak bosons and fermions to enable the spontaneous symmetry breaking (SSB) of $SU(2)_L \otimes U(1)_Y$ into $U(1)_{em}$. The masses of the weak bosons are generated via the kinetic term of the 2HDM in Eq. 2.56, where the mass terms of the weak bosons are given by

$$\mathcal{L}_{2\text{HDM}, m_{W^\pm, Z}} = \frac{1}{2} g^2 (v_1^2 + v_2^2) W_\mu^\dagger W^\mu + \frac{1}{2} (g'^2 + g^2) (v_1^2 + v_2^2) Z_\mu Z^\mu. \tag{2.73}$$

Here, the squared masses of the weak bosons are proportional to the sum of the squares of the vacuum expectation values (VEVs) of the two Higgs doublets, with

$v_{2HDM} = \sqrt{v_1^2 + v_2^2}$ set to 246 GeV. Consequently, in the 2HDM, the weak boson mass terms incorporate contributions from both doublet VEVs while retaining the same coupling constants. The ratio of the VEVs, $\tan \beta = v_2/v_1$, can be determined experimentally.

In the case of fermions, within the general CP-conserving 2HDM, fermion masses arise from the Yukawa interactions of the two doublets, $\mathcal{L}_{2HDM,yuk}$, expressed as

$$\mathcal{L}_{2HDM,yuk} = - \sum_{n=1}^2 \left(y_{ij}^{n,e} \bar{\mathbf{L}}_i \Phi_n e_{Rj} + y_{ij}^{n,d} \bar{\mathbf{Q}}_i \Phi_n d_{Rj} + y_{ij}^{n,u} \bar{\mathbf{Q}}_i \tilde{\Phi}_n u_{Rj} \right) + \text{h.c.} \quad (2.74)$$

Here, Φ_n and $\tilde{\Phi}_n$ represent the two scalar doublets and their conjugates, respectively. The Yukawa couplings, y_{ij}^1 and y_{ij}^2 , can be expressed in terms of the fermion mass matrix M_{ij}^k as

$$M_{ij}^k = y_{ij}^{1,k} \frac{v_1}{\sqrt{2}} + y_{ij}^{2,k} \frac{v_2}{\sqrt{2}}, \quad k = u, d, e, \quad (2.75)$$

where i and j denote the generation indices for fermions. In the SM, the fermion mass matrices can be diagonalized to align with the fermion flavor basis. However, in the general 2HDM, the two Yukawa coupling matrices cannot be diagonalized simultaneously, leading to mixing between fermion generations and resulting in tree-level flavor-changing neutral currents (FCNCs). To suppress such interactions, the 2HDM introduces four specific model types: Type-I, Type-II, Lepton-Specific, and Flipped, which differ in the coupling assignments of the fermions and Higgs doublets, as summarized in Table 2.9.

These four configurations are derived by incorporating the softly broken Z_2 symmetry, given in Eq. 2.59, into Eq. 2.74. Although all models suppress tree-level FCNCs, Type-I and Type-II are the most extensively studied in the context of LHC phenomenology.

2.4.2.4 Heavy Neutral Scalar and SM Higgs Boson interaction

Within the 2HDM, the five physical Higgs bosons can interact with each other, offering insights into their self-couplings. Following the discovery of the SM Higgs

Model Type	u_R^k coupling	d_R^k coupling	e_R^k coupling
Type-I	Φ_2	Φ_2	Φ_2
Type-II	Φ_2	Φ_1	Φ_1
Lepton-Specific	Φ_2	Φ_2	Φ_1
Flipped	Φ_2	Φ_1	Φ_2

Table 2.9: The four types of 2HDM models that suppress tree-level FCNCs. In all models, right-handed up-type quarks couple exclusively to the second doublet, while other right-handed fermions may couple to either doublet depending on the model.

boson, probing the interactions between the heavy neutral scalar X and the SM Higgs boson H becomes particularly compelling. This provides an opportunity to investigate the existence of the heavy neutral scalar boson while simultaneously measuring its coupling strength to the SM Higgs boson. The corresponding Lagrangian $\mathcal{L}_{XHH,\text{int}}$ governing this interaction is given by [45]:

$$\begin{aligned}
\mathcal{L}_{XHH,\text{int}} &= c_{XHH} XHH \\
c_{XHH} &= \frac{\cos(\alpha - \beta)}{\sin(2\beta)} \left[\sin(2\alpha)(2m_H^2 + m_X^2) - \frac{2m_{12}^2}{\sin(2\beta)} (3\sin(2\alpha) - \sin(2\beta)) \right]
\end{aligned}
\tag{2.76}$$

Here, c_{XHH} denotes the trilinear coupling between the heavy neutral scalar boson and two SM Higgs bosons. As indicated by equation 2.76, the heavy neutral scalar boson can decay into a pair of SM Higgs bosons, a process that can be explored at high-energy colliders such as the LHC via resonant production through gluon-gluon fusion. However, gluons interact with X only indirectly through virtual heavy quark loops, as the Lagrangian contains no direct coupling terms between X and gluons, given that X carries no color charge.

2.4.3 Randall-Sundrum Model

The Randall-Sundrum (RS) model is another BSM theory extensively studied by both theorists and experimentalists. Inspired by string theory, this model offers a potential solution to the hierarchy problem [46].

2.4.3.1 String Theory and Brane World

String theory, a Beyond the Standard Model (BSM) framework, proposes that the SM particles correspond to vibrating strings existing at the Planck scale, i.e., at the size of the Planck length. Distinct vibrational modes of these strings give rise to the variety of particles observed in the Standard Model. However, string theory necessitates ten dimensions of spacetime⁶. To reconcile this with the observable three spatial dimensions, the theory suggests that the six additional spatial dimensions are compactified to extremely small scales, on the order of the Planck length. Unfortunately, directly testing string theory remains a challenge, as it predicts phenomena at energy scales vastly exceeding those accessible by current particle colliders such as the LHC—by approximately 16 orders of magnitude.

Nevertheless, certain BSM models inspired by string theory offer indirect avenues for experimental investigation at accessible energy scales. One such example is the brane world model [47], which proposes that the universe resides on a three-dimensional brane embedded within a higher-dimensional bulk space. In this framework, gravity can propagate into the higher-dimensional bulk, leading to the observed weakness of gravitational interactions in our universe. The brane world concept encompasses several sub-models, including the Large Extra Dimensions scenario and the Randall-Sundrum model. This thesis focuses on the Randall-Sundrum model in detail.

⁶String theory encompasses various formulations, some requiring different numbers of spacetime dimensions. For instance, bosonic string theory requires 26 dimensions, whereas the more common superstring theory involves 10 spacetime dimensions (including one time dimension and nine spatial dimensions). The string theory referenced in this thesis pertains to superstring theory.

2.4.3.2 The Setup of the RS Model

The setup of the RS model involves two branes: the first contains the universe where the SM particles reside (commonly referred to as the TeV brane), while the second brane (known as the Planck brane), where quantum gravity effects dominate, is located at the boundary of the extra dimension, as illustrated in Figure 2.3. These branes are separated by a warped extra dimension, meaning that the distance between them depends on their position within the extra dimension. The metric, ds^2 , describing the geometry of this warped space is given by

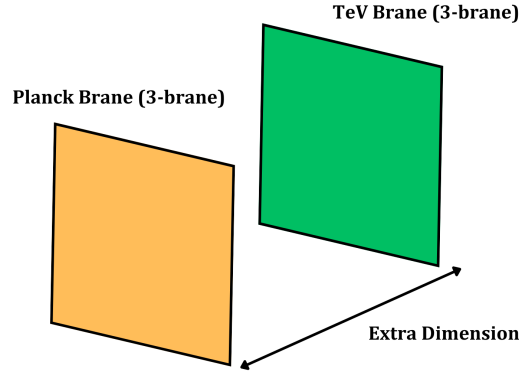


Figure 2.3: Illustration of the Randall–Sundrum model setup.

$$ds^2 = e^{-2kr_c|\phi|} \eta_{\mu\nu} dx^\mu dx^\nu + r_c^2 d\phi^2 \quad \mu, \nu = 0, 1, 2, 3 \quad (2.77)$$

Here, $\eta_{\mu\nu}$ denotes the Minkowski metric, which is warped by the exponential factor $e^{-2kr_c|\phi|}$. The parameter k represents the curvature scale, r_c is the compactification radius, and ϕ is the coordinate of the extra dimension, spanning the range $0 \leq |\phi| \leq \pi$. This metric is constructed to preserve Poincaré invariance⁷ on the 3-branes, where each brane possesses the following induced metric:

$$g_{\mu\nu}^{\text{hid}}(x^\mu) \equiv g_{MN}(x^\mu, \phi = 0), \quad g_{\mu\nu}^{\text{vis}}(x^\mu) \equiv g_{MN}(x^\mu, \phi = \pi) \quad (2.78)$$

⁷Poincaré invariance refers to the property that the laws of physics remain invariant under certain transformations of spacetime, such as translations, rotations, and Lorentz boosts.

In this expression, g_{MN} with indices $M, N = \{\mu, \phi\}$ refers to the five-dimensional metric tensor. Specifically, $g_{\mu\nu}^{\text{vis}}$ denotes the metric tensor on the TeV brane, while $g_{\mu\nu}^{\text{hid}}$ corresponds to the metric on the Planck brane. The curvature scale k is determined by solving the five-dimensional Einstein field equations using the metric in Eq. (2.77) and the configuration of the two 3-branes. The solution requires a negative cosmological constant to yield a real value of k and to balance the brane tensions⁸ between the branes. As a result, this setup corresponds to an Anti-de Sitter (AdS) space between the branes, characterised by its negative curvature.

2.4.3.3 Solution to the Hierarchy Problem

Within the RS model, the hierarchy problem is addressed not through the cancellation of loop-induced mass corrections, but rather by embedding the four-dimensional metric on the TeV brane, Eq. (2.78), into the action of the SM Higgs field, S_{Higgs} , as follows:

$$\begin{aligned} S_{\text{Higgs}} &= \int d^4x \sqrt{-g^{\text{vis}}} \left[g_{\mu\nu}^{\text{vis}} D^\mu H^\dagger D^\nu H - \lambda \left(H^\dagger H - v^2 \right)^2 \right] \\ &= \int d^4x e^{-4kr_c\pi} \left[e^{2kr_c\pi} \eta_{\mu\nu} D^\mu H^\dagger D^\nu H - \lambda \left(H^\dagger H - v^2 \right)^2 \right] \end{aligned} \quad (2.79)$$

Here, g^{vis} represents the determinant of the TeV brane metric. By normalising the Higgs field $\tilde{H}$ with respect to the TeV brane, namely

$$\tilde{H} = e^{-kr_c\pi} H, \quad (2.80)$$

the action in Eq. (2.79) can be rewritten as

$$S_{\text{Higgs}} = \int d^4x \left[\eta_{\mu\nu} D^\mu \tilde{H}^\dagger D^\nu \tilde{H} - \lambda \left(\tilde{H}^\dagger \tilde{H} - e^{-2kr_c\pi} v^2 \right)^2 \right] \quad (2.81)$$

From the vacuum expectation value (VEV) term in Eq. (2.81), it is evident that

⁸In the braneworld scenario, brane tensions are associated with the energy densities of the branes, which influence their expansion or contraction. In the RS model, the TeV brane tension is set to be positive, while the Planck brane tension is negative, preventing the branes from collapsing into one another.

the VEV of the Higgs field undergoes an exponential suppression along the distance to the TeV brane in the extra dimension. The effective VEV, v_{eff} , for the Higgs field on the TeV brane becomes

$$v_{\text{eff}} = e^{-kr_c\pi} v \quad (2.82)$$

Assuming the fundamental VEV scale v is of the order of M_{P} , the observed VEV within the SM is significantly suppressed relative to M_{P} due to the negative exponential factor. In order to reproduce the electroweak scale, the value of $kr_c\pi$ must satisfy

$$kr_c\pi \sim \ln 10^{15} \sim 35 \quad (2.83)$$

Thus, this mechanism naturally resolves the fine-tuning problem associated with the Higgs boson mass. While the Higgs mass remains large on the Planck brane, it is exponentially reduced to the observed value on the TeV brane owing to the warped geometry of the extra dimension.

2.4.3.4 Kaluza-Klein Gravitons in RS Model

The RS model offers a simple solution to the hierarchy problem. However, the phenomenological aspects of the model require thorough study and experimental verification. A prominent phenomenological feature of this model involves the search for massive Kaluza-Klein gravitons (G_{kk}^* particles).

In the RS model, gravity in the universe is mediated by the Kaluza-Klein graviton [46]. The Kaluza-Klein graviton⁹ is a spin-2 boson that arises from quantum fluctuations of the metric tensor h_{MN} in the RS metric, expressed as follows:

$$ds^2 = e^{-2kr_c|\phi|} (\eta_{\mu\nu} + h_{\mu\nu}) dx^\mu dx^\nu + r_c^2 d\phi^2 \quad (2.84)$$

⁹The term Kaluza-Klein refers to the Kaluza-Klein theory [48], which seeks to unify general relativity and electromagnetism. This theory proposes the existence of an additional spatial dimension, compactified to a small circle on the order of the Planck length. Within this framework, the momentum of a particle in the fifth dimension manifests as an effective mass in the four-dimensional spacetime. This procedure is known as Kaluza-Klein decomposition.

Here, the tensor fields h_{MN} are assumed to have influence only on the 3-branes, leading to $h_{M5} = h_{5M} = 0$. Consequently, h_{MN} reduces to $h_{\mu\nu}$, which corresponds to the four-dimensional G_{kk}^* tensor fields. The Kaluza-Klein decomposition can thus be written as:

$$h_{\mu\nu}(x, z) = \sum_n^{\infty} h_{\mu\nu}^n(x) \psi_n(z) \quad (2.85)$$

In this expression, $h_{\mu\nu}^n(x)$ represents the massive G_{kk}^* tensor field modes, with their masses given by:

$$m_{G_{kk}^*}^n = k e^{-kr_c \pi} j_n \quad (2.86)$$

Here, j_n denotes the n th zero of the first-order Bessel function of the first kind J_1 , satisfying $J_1(j_n) = 0$. The function $\psi_n(z)$ describes the wavefunction modes of G_{kk}^* in the extra dimension, expressed in terms of the conformally flat coordinate z defined by:

$$|z| = \frac{1}{k} (e^{kr_c |\phi|} - 1) \quad (2.87)$$

The wavefunction $\psi_n(z)$ is composed of the second-order Bessel functions of the first kind J_2 and the second kind Y_2 . It can be obtained by solving the Schrödinger-like equation derived from the five-dimensional Einstein field equations in the conformally flat RS metric 2.84:

$$-\frac{d^2 \psi_n}{dz^2} + \left(\frac{9}{4} \left[\frac{dA(z)}{dz} \right]^2 - \frac{3}{2} \frac{d^2 A(z)}{dz^2} \right) \psi_n(z) = (m_{G_{kk}^*}^n)^2 \psi_n(z) \quad (2.88)$$

$$A(z) = \ln(k|z| + 1)$$

The conformal RS metric is written as:

$$ds^2 = e^{-2A(z)} ([\eta_{\mu\nu} + h_{\mu\nu}] dx^\mu dx^\nu + dz^2) \quad (2.89)$$

From this mathematical formulation, the G_{kk}^* particle in the RS model exhibits a rich spectrum of mass modes, while simultaneously mediating gravity at the TeV brane.

In addition to its mass modes, the nature of the G_{kk}^* particle in the RS model provides an explanation for the observed weakness of gravity relative to the other fundamental forces, as elaborated in Section 2.4.3.5.

2.4.3.5 Weakness of Gravity in RS Model

Although the solution for the G_{kk}^* tensor field in the RS model includes massive modes, the long-range nature of gravity necessitates that the mediator particle be massless. To accommodate this requirement, the RS model sets the zeroth mode (ground state) of $h_{\mu\nu}^n$ to be massless:

$$m_{G_{kk}^*}^0 = 0 \quad (2.90)$$

Accordingly, the zeroth mode wavefunction solution $\psi_0(z)$ from equation 2.88 behaves as a rational function:

$$\psi_0(z) = \frac{1}{(k|z| + 1)^{3/2}} \quad (2.91)$$

Equation 2.91 indicates that the G_{kk}^* particles are localised near the Planck brane (at $z = 0$) and diminish in density towards the bulk (at $|z| > 0$). Consequently, the TeV brane contains a significantly lower concentration of G_{kk}^* particles compared to the Planck brane. This localisation provides an explanation for the apparent weakness of gravity in the observable universe.

2.4.3.6 Kaluza-Klein Graviton, Higgs Boson, and Gluon Interaction in the RS Model

In the RS model, the G_{kk}^* tensor field influences the geometry of space-time, as shown in equation 2.84. This property enables the G_{kk}^* particles to interact with SM particles on the TeV brane [49]. One example involves the interaction between the G_{kk}^* and the SM Higgs boson, which can be obtained from the G_{kk}^* -Higgs interaction term in the Lagrangian, $\mathcal{L}_{H,grav,int}$:

$$\mathcal{L}_{H,grav,int} = -c_{GHH} \partial_\mu H \partial_\nu H h^{\mu\nu} \quad (2.92)$$

This interaction involves a single G_{kk}^* and a pair of SM Higgs bosons, with the trilinear coupling strength given by

$$c_{GHH} = \frac{2}{M_p \pi r_c \mu_{\text{TeV}}} \quad (2.93)$$

Here, μ_{TeV} denotes the cutoff energy scale of the model. This energy scale is proportional to the first mass mode of the G_{kk}^* , whose predicted mass lies approximately within the GeV to TeV range. In addition to the Higgs boson interaction, the G_{kk}^* couples directly to gluons, with the corresponding Lagrangian $\mathcal{L}_{g,grav,int}$ expressed as

$$\mathcal{L}_{g,grav,int} = -c_{ggG} h_{\mu\nu} F_a^{\mu\rho} F_\rho^{\nu a} \quad (2.94)$$

Here, $F_a^{\mu\rho}$ denotes the gluon field tensor, and c_{ggG} represents the coupling strength between the G_{kk}^* and two gluons:

$$c_{ggG} \approx \frac{0.47}{(\pi^2 k r_c^2) M_p \mu_{\text{TeV}}} \quad (2.95)$$

This indicates that two gluons can fuse directly to produce a massive G_{kk}^* , which subsequently decays into a pair of Higgs bosons. Such processes can be investigated at high-energy colliders, such as the LHC, through resonant production, for example via gluon-gluon fusion.

2.5 Higgs Boson Pair Production

Section 2.4 has examined the nature of the Higgs boson and the implications of its mass for the hierarchy problem, motivating solutions proposed by theories Beyond the Standard Model (BSM). This section focuses on the phenomenology of Higgs boson pair production at the ATLAS experiment at the LHC, along with the connection to probing BSM models and Higgs boson self-interactions.

2.5.1 Higgs Boson Pair Production in the ATLAS Experiment

Higgs boson pair production involves the creation of two Higgs bosons during proton-proton collisions at the LHC. At ATLAS, this process is predominantly initiated through the gluon-fusion mechanism (ggF) and the vector boson fusion (VBF) mode [50, 51]. A primary motivation for this search lies in probing Higgs boson self-interactions, which enables measurement of the coupling strength c_{HHH} between Higgs bosons [50]. In addition to the Higgs boson self-coupling, the investigation includes the coupling between two Higgs bosons and two weak bosons, denoted as c_{2V} , as illustrated in Figure 2.4. This coupling is accessible via the VBF mode, where V refers to the W or Z bosons.

Beyond the study of c_{HHH} [52] and c_{2V} , Higgs boson pair production provides an opportunity to search for particles predicted by BSM theories that decay into Higgs boson pairs. ATLAS categorises the searches for Higgs boson pair production into two main analyses: non-resonant and resonant [53].

Non-resonant analysis focuses on exploring Higgs boson pair production across a range of modified values for c_{HHH} and c_{2V} . These values are expressed in terms of coupling modifiers, κ_λ and κ_{2V} , defined as

$$\kappa_\lambda = \frac{c_{HHH}}{c_{HHH}^{SM}}, \quad \kappa_{2V} = \frac{c_{2V}}{c_{2V}^{SM}} \quad (2.96)$$

where c_{HHH}^{SM} and c_{2V}^{SM} represent the Standard Model predictions for these couplings. By definition, $\kappa_{\lambda/2V} = 1$ indicates alignment with the Standard Model ex-

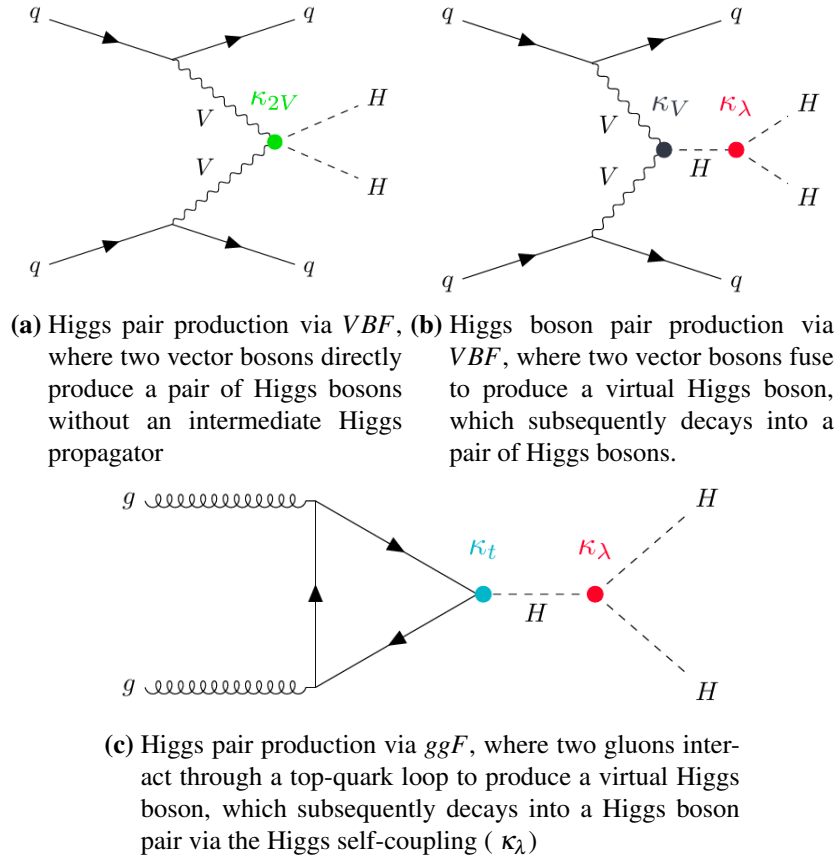


Figure 2.4: Examples of leading order Feynman diagrams for non resonant Higgs pair production via gluon-gluon fusion (ggF) and vector boson fusion (VBF) in ATLAS.

pectation.

In contrast, the resonant analysis targets the detection of a massive resonance predicted by BSM theories, decaying into Higgs boson pairs. At ATLAS, two primary BSM candidates are considered in this analysis: the excited state of a massive Kaluza-Klein graviton, G_{kk}^* , predicted by the Randall–Sundrum model, and a heavy neutral scalar boson, X , arising from the Two-Higgs-Doublet Model (2HDM).

2.5.2 Resonant Higgs Boson Pair Production Decay Channels

In the resonant analysis, Higgs boson pairs may decay into various combinations of Higgs boson decay products. Approximately fifteen decay channels arise from these combinations, as illustrated in Figure 2.5. The ATLAS experiment concentrates on five principal decay channels, listed in Table 2.10, prioritising those with

higher branching ratios and better prospects for background suppression. This thesis focuses on resonant Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state via the gluon-fusion mode, depicted in Figure 2.6. Chapter 5 will provide a detailed discussion of this analysis.

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	34%				
WW	25%	4.6%			
$\tau\tau$	7.3%	2.7%	0.39%		
ZZ	3.1%	1.1%	0.33%	0.069%	
$\gamma\gamma$	0.26%	0.10%	0.028%	0.012%	0.0005%

Figure 2.5: Branching ratios for Higgs boson pair production decay channels [2].

Higgs boson pair decay channels	Branching ratio
$HH \rightarrow b\bar{b}b\bar{b}$ [13]	0.34
$HH \rightarrow b\bar{b}WW^*$ [54]	0.25
$HH \rightarrow b\bar{b}\tau\tau$ [55]	0.073
$HH \rightarrow b\bar{b}\gamma\gamma$ [56]	0.0026
$HH \rightarrow \gamma\gamma WW^*$ [57]	0.0010

Table 2.10: Main decay channels of Higgs boson pair production targeted by ATLAS.

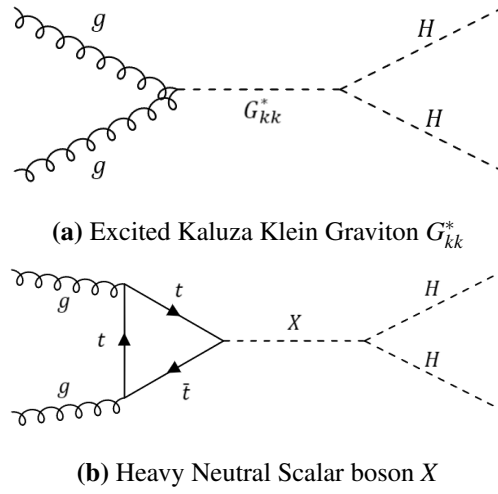


Figure 2.6: Feynman diagrams for two Beyond Standard Model resonant Higgs pair production via gluon-gluon fusion (ggF) in ATLAS.

Chapter 3

The Large Hadron Collider and ATLAS Detector

The Large Hadron Collider (LHC) is a 27-kilometer circular hadron collider located near Geneva, Switzerland. Designed as a synchrotron, the LHC collides protons (and lead ions) at a rate of approximately $O(10^9)$ particles per second. The collider reaches a peak center-of-mass energy of $\sqrt{s} = 14$ TeV and delivers an instantaneous luminosity exceeding $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$.

Construction of the LHC included four main detectors dedicated to distinct experimental goals: ATLAS (A Toroidal LHC Apparatus) [58], CMS (Compact Muon Solenoid) [59], LHCb, and ALICE (A Large Ion Collider Experiment) [60]. The LHC ring and associated detectors are situated underground, at depths ranging from 50 to 175 meters.

Since its commissioning, the LHC has completed three operational phases: Run 1, Run 2, and Run 3. Run 1, spanning from 2009 to 2013, delivered proton-proton collisions at center-of-mass energies of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV, yielding a total integrated luminosity of $L = 20.3 \text{ fb}^{-1}$, as recorded by the ATLAS detector. Run 2, conducted between 2015 and 2018, increased the center-of-mass energy to $\sqrt{s} = 13$ TeV, achieving an integrated luminosity of $L = 139 \text{ fb}^{-1}$ [61]. The ongoing Run 3 is expected to deliver a total integrated luminosity approximately three times greater than that of Run 2, with a center-of-mass energy of

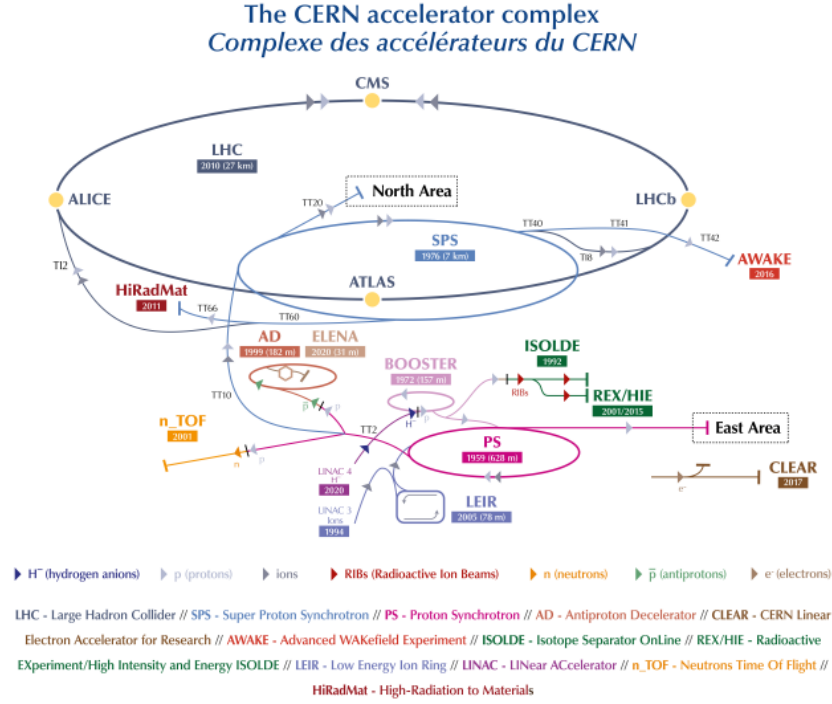


Figure 3.1: CERN accelerator complex [3].

$\sqrt{s} = 13.6$ TeV [62]. The dataset analysed in this thesis corresponds to Run 2.

Figure 3.1 presents the layout of the LHC accelerator complex. Proton production begins with the ionisation of hydrogen gas. The resulting protons are injected into the linear accelerator (Linac-2), where they are accelerated to 50 MeV. The 50 MeV proton beams from Linac-2 are subsequently transferred to the Proton Synchrotron Booster (PSB) for further acceleration to 1.4 GeV. Following this, the protons are injected into the Proton Synchrotron, reaching an energy of 25 GeV, before being transferred to the Super Proton Synchrotron, where they are accelerated to 450 GeV. The final stage of acceleration takes place in the LHC ring, increasing the proton beam energy to 13 TeV prior to collision within the detectors.

3.1 The ATLAS Detector

The ATLAS detector ranks among the three primary experiments (alongside CMS and LHCb) that utilise proton beam collisions for physics research. This massive apparatus weighs 7000 tonnes and spans 25 meters in diameter and 44 meters in length. The detector comprises three main sub-detectors: the Inner Detector (ID), the Calorimeter (Cal), and the Muon Spectrometer (MS). Each sub-detector plays a crucial role in measuring the kinematics of particles. This section provides a detailed description of each sub-detector.

3.1.1 ATLAS Coordinate System

The ATLAS detector employs a coordinate system specifically designed for physics analyses involving relativistic particle collisions. The origin of this coordinate system is positioned at the geometric centre of the detector. The x -axis points towards the centre of the LHC ring, the y -axis extends vertically upwards, and the z -axis runs along the beam pipe in a counterclockwise direction.

To measure the four-momentum vectors of particle trajectories, the x - y plane is defined as the "transverse" plane, while the z -axis along the beam pipe is designated as the "longitudinal" direction. The ATLAS detector utilises the spherical coordinate system to describe particle directions from the origin. The azimuthal angle, ϕ , measures the rotation from the x -axis within the transverse plane, whereas the polar angle, θ , quantifies the angle from the transverse plane towards the longitudinal axis.

The detector also employs a variable known as pseudorapidity, η , which depends on the particle's momentum and is defined as:

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_L}{|\mathbf{p}| - p_L} \right) = \operatorname{arctanh} \left(\frac{p_L}{|\mathbf{p}|} \right), \quad (3.1)$$

where $|\mathbf{p}|$ represents the magnitude of the particle's three-momentum, and p_L denotes the component of the momentum along the longitudinal direction. The

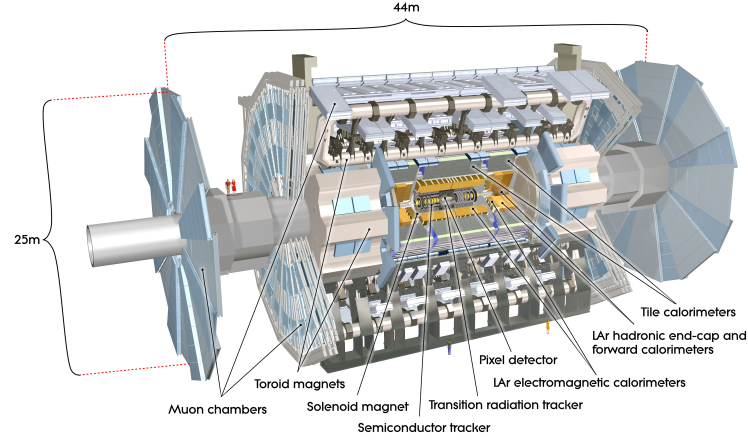


Figure 3.2: The ATLAS Detector [4]

pseudorapidity can also be expressed in terms of the polar angle, as commonly used in ATLAS, as shown in Eq. 3.2.

$$\eta = -\frac{1}{2} \ln \left(\tan \frac{\theta}{2} \right) \quad (3.2)$$

Particles in the region around low $|\eta|$ values are referred to as the "central" region, while those at high $|\eta|$ values correspond to the "forward" region.

3.1.2 Inner Detector

The Inner Detector (ID) constitutes the innermost component of the ATLAS Detector, dedicated to measuring the trajectories of charged particles (tracks) emerging from the interaction point. The ID consists of three sub-components: the silicon pixel detector, the semiconductor tracker (SCT), and the transition radiation tracker (TRT). A magnetic field of 2 T, produced by solenoid coils surrounding the ID, assists in determining the trajectories of charged particles. The ID provides coverage over the pseudorapidity range of $|\eta| < 2.5$.

3.1.2.1 Silicon Pixel Detector

The silicon pixel detector is positioned closest to the interaction point, encompassing the entire outer surface of the beam pipe along the longitudinal span of the ATLAS detector. This detector is optimised for measuring particles that decay in close

proximity to the interaction point. The pixel detector comprises approximately 92 million pixels, arranged in four barrel layers surrounding the beam pipe and three layers on each end cap. Each pixel measures $50 \times 400 \mu\text{m}^2$, providing a resolution of $10 \mu\text{m}$ in the transverse plane and $115 \mu\text{m}$ in the longitudinal plane.

3.1.2.2 Semiconductor Tracker

The subsequent sub-detector is the Semiconductor Tracker (SCT), covering the outer layer of the pixel detector. The SCT incorporates approximately 6 million silicon strip sensors, distributed across four barrel layers encircling the pixel detector and three layers on each end cap. Each barrel layer employs double-sided sensors, enabling readout on both the front and back sides of the strips, while the end cap layers utilise single-sided sensors operating on the front side. The SCT design facilitates track measurements in both the longitudinal (η) and transverse (ϕ) directions. Each strip has an average length of $80 \mu\text{m}$ and achieves a resolution of up to $25 \mu\text{m}$ in both directions. The pseudorapidity coverage of the SCT extends to $|\eta| < 2.5$.

3.1.2.3 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) forms the outermost layer of the ID. The TRT consists of approximately 300,000 drift tubes filled with xenon gas, yielding a total of 350,000 readout channels and delivering a detector resolution of $130 \mu\text{m}$ in the transverse direction.

The ID additionally provides impact parameter (IP) resolution, defined as the closest distance between the reconstructed tracks and the primary vertex of the decay particles at the interaction point. This characteristic is essential for identifying tracks originating from b-hadrons. In Run 2 data, the ID achieved an IP resolution of approximately $10 \mu\text{m}$ in the transverse direction (d_0) and $80 \mu\text{m}$ in the longitudinal direction (z_0).

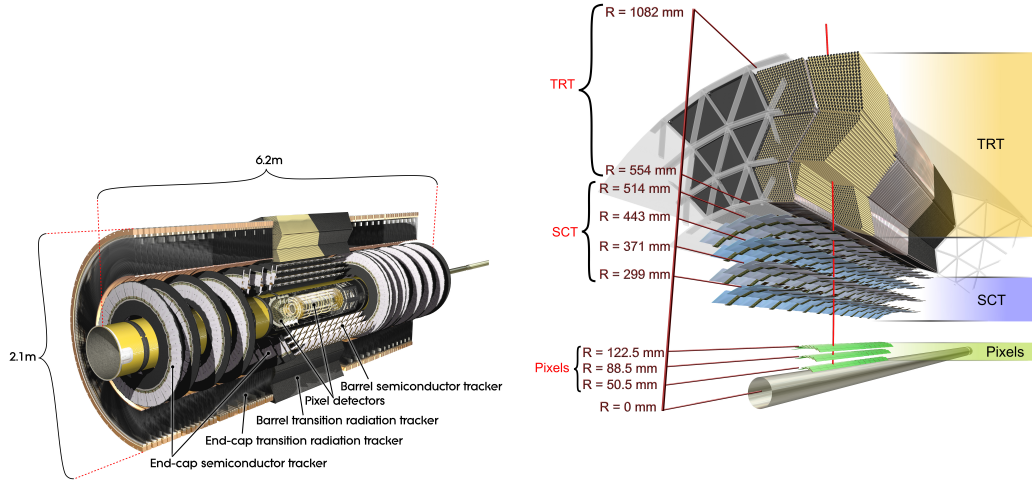


Figure 3.3: ATLAS Inner Detector system. [5]

3.1.3 ATLAS Calorimeters

The ATLAS calorimeters are designed to measure the energies of particles interacting via electromagnetic and strong forces. Two distinct types constitute the system: the Electromagnetic Calorimeter (ECAL) and the Hadronic Calorimeter (HCAL).

3.1.3.1 Electromagnetic Calorimeter (ECAL)

Positioned inside the ATLAS detector, surrounding the solenoids that encompass the Inner Detector, the ECAL comprises alternating layers of stainless steel and lead, filled with liquid argon. The passive materials (stainless steel and lead) induce electromagnetic showers from incoming high-energy particles such as electrons, charged hadrons, and photons. Meanwhile, the active material (liquid argon) measures the energy deposited by these showers. The ECAL covers a barrel region ($|\eta| < 1.475$) and an endcap region ($1.375 < |\eta| < 3.2$). The approximate energy resolution of the ECAL is given by $\sigma(E)/E \sim 10\%/\sqrt{E} \oplus 0.2\%$.

3.1.3.2 Hadronic Calorimeter (HCAL)

Situated as the outer part of the ATLAS calorimeter system, the HCAL is dedicated to measuring the energies of hadrons. The HCAL comprises three distinct regions: the central barrel, the endcap, and the forward region, each constructed from dif-

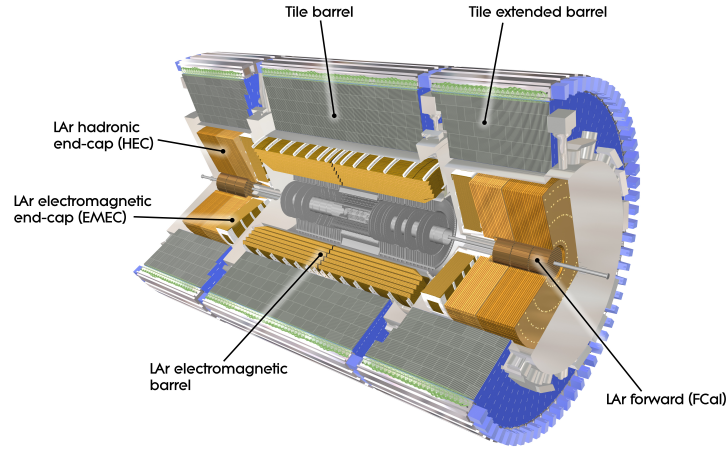


Figure 3.4: The ATLAS Calorimeter System [6]

ferent materials. The central barrel region ($|\eta| < 1.0$) employs stainless steel as the passive material to initiate hadronic showers. These showers, formed by interactions of incoming hadrons (such as pions, kaons, neutrons, and protons) with nuclei via strong interactions, deposit energy in plastic scintillators serving as the active material.

The extended part of the central barrel ($0.8 < |\eta| < 1.7$) utilises iron plates as the passive material to induce hadronic showers, with liquid argon employed as the active medium for energy measurement. In the endcap region ($1.5 < |\eta| < 3.2$), copper plates serve as the passive material for triggering hadronic showers, while liquid argon functions as the active medium. For the forward region calorimeter ($3.1 < |\eta| < 4.9$), copper is used in the electromagnetic section as the passive material, and tungsten is employed in the HCAL section. The energy resolution of the HCAL is approximately $\sigma(E)/E \sim 50\%/\sqrt{E} \oplus 3\%$, deteriorating to $\sigma(E)/E \sim 100\%/\sqrt{E} \oplus 10\%$ in the forward regions. Figure 3.4 illustrates the components of the ATLAS calorimeter system.

3.1.4 Muon Spectrometer

The Muon Spectrometer (MS) comprises two primary detector technologies: the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). The MDT system covers the pseudorapidity range of $|\eta| < 2.0$, providing an average spa-

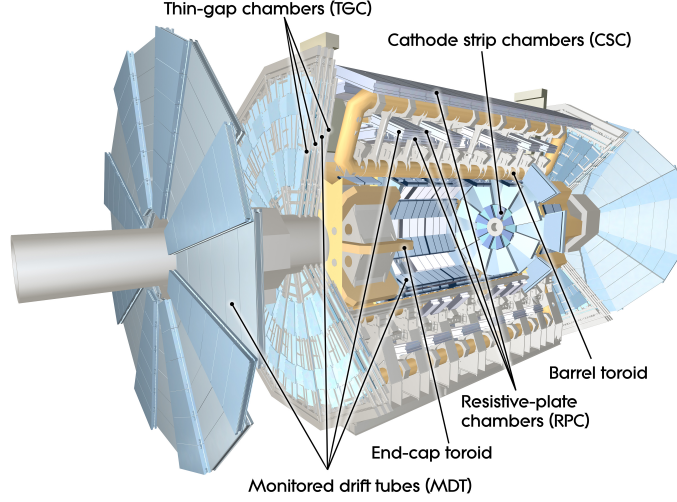


Figure 3.5: The ATLAS Muon Subsystem [7].

tial resolution of $80 \mu\text{m}$ in the transverse direction and $35 \mu\text{m}$ in the longitudinal direction per chamber. The CSC system, operating in the forward region of $2.0 < |\eta| < 2.7$, achieves an average spatial resolution of 5 mm in the transverse direction and $40 \mu\text{m}$ in the longitudinal direction.

The MS includes dedicated trigger detectors to select events containing muons within $|\eta| < 2.4$. Resistive Plate Chambers (RPC) are employed in the barrel region ($|\eta| < 1.05$), while Thin Gap Chambers (TGC) are used in the endcap region ($1.05 < |\eta| < 2.4$). To facilitate the measurement of muon momentum, the MS is equipped with three large air-core toroidal magnets: one in the barrel and one in each endcap. The momentum resolution of the MS reaches approximately 3–10% for muons with transverse momenta between 100 GeV and 1 TeV.

3.1.5 ATLAS Trigger System

During LHC runs, proton beams travel in bunches separated by 25 nanoseconds. This results in a bunch crossing rate of 40 megahertz, leading to an enormous number of collisions within a short period. To manage this, the ATLAS experiment employs a trigger system that selects only events of interest for data acquisition and physics analysis. The ATLAS trigger system consists of two levels: the Level-1 Trigger (L1) and the High Level Trigger (HLT) [8].

3.1.5.1 Level-1 Trigger

The Level-1 Trigger (L1) [8] is a hardware-based system that processes reduced-granularity information from the calorimeter and muon detectors. The L1 Calorimeter Trigger (L1Calo) extracts signals from the calorimeter detectors, calibrates them via the preprocessor, and forwards the data to the Cluster Processor (CP) and Jet/Energy-sum Processor (JEP). The CP identifies electron, photon, and τ -lepton candidates, while the JEP selects jet candidates and computes the total transverse energy and missing transverse energy (E_T^{miss}) recorded by the calorimeter.

The L1 Muon Trigger (L1Muon) processes signals from hits in the RPCs and TGCs, measuring the deviation between the observed muon hit pattern and the expected pattern for muons with infinite momentum. The output from L1Muon, transmitted via the Muon Central Trigger Processor Interface (MUCTPI), together with data from L1Calo and the L1 Topological Processor (L1Topo), serves as input for the Central Trigger Processor (CTP).

The L1Topo processes signals from L1Muon and L1Calo to perform topological selections critical for physics analyses. These selections include invariant mass (m), missing transverse energy (E_T^{miss}), total transverse momentum sum (H_T), and object separations in pseudorapidity ($\Delta\eta$) and azimuthal angle ($\Delta\phi$).

The Central Trigger Processor (CTP) imposes a deadtime limit to constrain the number of L1-accepted events within the detector readout latency window. For each L1-accepted event, the Front-End (FE) detector electronics read out data from all detectors and transmit them to the ReadOut Drivers (RODs) for initial formatting and processing. The RODs forward this data to the ReadOut System (ROS) for buffering. Data from the ROS is then transferred to the data collection network before being processed by the High Level Trigger (HLT).

Additionally, the L1 Trigger identifies Regions of Interest (RoI) in the η and ϕ coordinates of the detector, guiding the subsequent event selection performed by the HLT. The L1 Trigger operates with a maximum readout rate of 100 kHz and a latency limit of 2 μ s.

3.1.5.2 High Level Trigger

The High Level Trigger (HLT) [8] is a software-based system that processes events accepted by the L1 Trigger with full detector granularity for comprehensive event reconstruction and selection. Dedicated CPU farms execute fast decision algorithms to determine which events are retained. Events accepted by the HLT are subsequently stored in CERN's Tier-0 data repository. The maximum physics output rate of the HLT is 1.2 kHz.

Two categories of triggers operate at both L1 and HLT stages: prescaled and unprescaled triggers. Prescaled triggers process only a fraction of the input events, as defined by a preassigned scaling factor, whereas unprescaled triggers process the full set of input events. Figure 3.6 illustrates the ATLAS trigger system during Run 2.

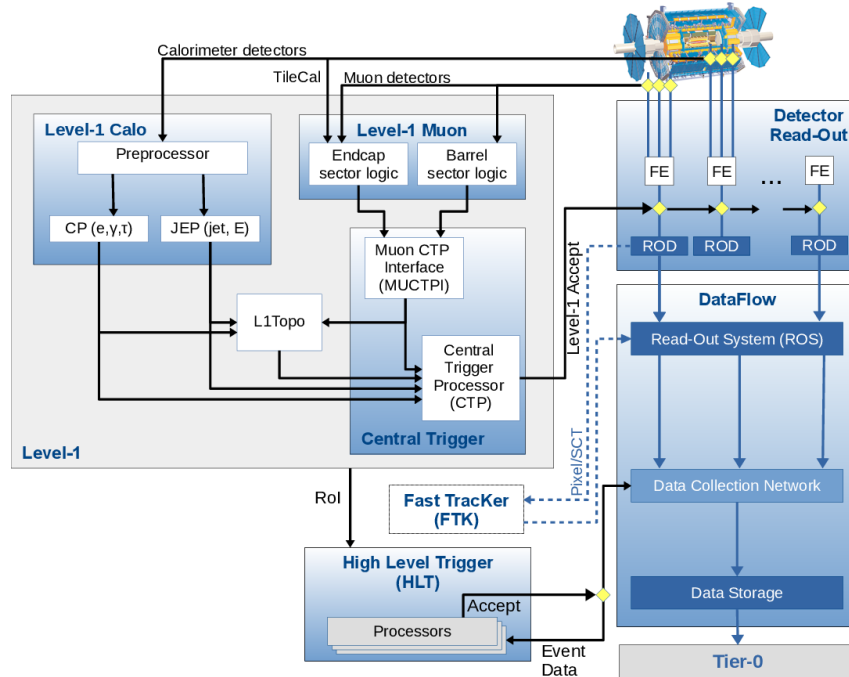


Figure 3.6: ATLAS trigger system during Run 2. [8]

3.1.6 Luminosity

In ATLAS, the rate of particle collisions at the interaction point is determined using the concept of luminosity. The introduction of luminosity enables the estimation of

the number of events for a given physics process during the pp collisions at the interaction point. Two definitions are employed to describe luminosity: instantaneous luminosity and integrated luminosity.

Instantaneous luminosity, denoted as $\mathcal{L}$, represents the rate of inelastic proton collisions per unit area at the interaction point during a given runtime t , and is expressed as follows:

$$\mathcal{L} = \frac{n_b f_r n_p^2}{2\pi \sigma_x \sigma_y} \quad (3.3)$$

Here, f_r refers to the rotational frequency of the beam, n_b corresponds to the number of bunches per beam, and n_p indicates the number of protons per bunch. The terms σ_x and σ_y represent the beam widths in the horizontal and vertical directions, respectively [63].

Integrated luminosity, $\mathcal{L}_{\text{int}}$, describes the total number of inelastic proton collisions per unit area at the interaction point over a given runtime t , and is given by:

$$\mathcal{L}_{\text{int}} = \int \mathcal{L} dt \quad (3.4)$$

Luminosity measurements are performed using the Luminosity Cherenkov Integrating Detector (LUCID) [64], positioned on both sides of the forward region of the ATLAS detector. The calibration of these measurements is carried out through the Van der Meer scan technique.

The Van der Meer scan serves as a method to precisely determine the luminosity of colliding beams in particle accelerators. During this procedure, the two beams are separated in both the horizontal and vertical directions while recording the collision rate. The collision rate reaches its maximum when the beams are perfectly overlapped. This maximum rate is directly linked to the luminosity, quantifying the number of particle collisions per unit area and per unit time.

Chapter 4

Analysis Tools

This chapter presents the physics analysis tools developed by ATLAS for the search for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks. Four main analysis tools are utilised in this study, namely Monte Carlo simulation tools, jets, track jets, and b -tagging techniques.

4.1 Monte Carlo Simulation

Monte Carlo simulations [26] serve as essential tools for simulating events observed in the detector, providing a basis for comparison with experimental data. ATLAS employs the Monte Carlo (MC) method to generate events based on random numerical values sampled from probability distributions.

4.1.1 Monte Carlo Simulation Process

The MC simulation process in ATLAS consists of several key stages:

- **Hard Process Generation:**

The initial stage involves generating the hard scattering process of particle interactions using Leading Order (LO) or Next-to-Leading Order (NLO) matrix elements, combined with Parton Distribution Functions (PDFs). The cross-section is computed through numerical integration over the available phase space.

- **Parton Shower (PS):**

Following the hard process, the Parton Shower algorithm simulates gluon

radiation from the colour charge of interacting quarks, which subsequently leads to the production of quark–antiquark pairs. These parton showers evolve into colour-neutral hadrons. The algorithm successively adds partons to the event, decreasing their energy to approximately 1 GeV.

- **Underlying Event (UE):**

The underlying event accounts for additional soft interactions between remnants of the incoming protons, enriching the parton showers with further particle activity.

- **Detector Simulation:**

The final stage models the response of the ATLAS detector using the GEANT4 toolkit [65, 66]. The complete output of the MC simulation is subsequently compared with experimental measurements from ATLAS to facilitate physics analyses.

4.1.2 Monte Carlo Event Generator Examples

Several Monte Carlo event generators are employed in the analysis described in Chapter 5, including:

- **Pythia 8**

PYTHIA 8 [67] operates as a multi-leg Leading Order (LO) matrix element generator. Parton showers are simulated using transverse momentum (p_T) ordering, while hadronisation is modelled with the Lund string framework [68].

- **Herwig 7**

HERWIG 7 [69] provides multi-leg event generation capabilities at both Leading Order (LO) and Next-to-Leading Order (NLO). Parton showers are simulated using the dipole formalism, treating each branching as a parton pair (dipole) rather than a single emitter. For hadronisation, the cluster model [70] is employed.

- **Powheg**

POWHEG [71] generates events at NLO accuracy, incorporating additional

parton interactions at the hard-scattering stage. In ATLAS, either PYTHIA 8 or HERWIG 7 is utilised for the parton shower and hadronisation steps.

- **MadGraph5_aMC@NLO**

MADGRAPH5_AMC@NLO [72] provides event generation at both LO and NLO precision, relying on PYTHIA 8 for parton showering and hadronisation. The generator distinguishes between matrix element orders by nomenclature: “MadGraph” denotes Leading Order (LO) configurations, whereas “aMC@NLO” refers to Next-to-Leading Order (NLO) configurations.

4.2 Jets

Jets are collections of hadrons originating from the fragmentation of quarks and gluons. In ATLAS, jets are typically associated with the hadronisation of quarks and gluons produced in partonic interactions between protons (or heavy ions) at the LHC. Jets are measured using the ATLAS calorimeter, with tracking information from the Inner Detector (ID) and Muon Spectrometer (MS) incorporated for jet calibration and flavour tagging, which identifies jets containing b -hadrons, c -hadrons, and light-flavour jets. This section provides further details on the jet reconstruction techniques employed in ATLAS.

4.2.1 Jet Algorithms

In ATLAS, jets are reconstructed by clustering energy deposits in calorimeter cells into a single energy measurement using dedicated jet reconstruction algorithms. The anti- k_t algorithm is one of the most commonly applied algorithms in ATLAS. This algorithm clusters the nearest jet constituents with the highest transverse momenta (p_T) within a predefined radius [73]. The anti- k_t algorithm proceeds as follows:

1. **Nearest Neighbour:** Identify the two nearest jet constituents with the highest p_T . Compute the distance between them, d_{ij} , as given by Equation 4.1:

$$d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}, \quad (4.1)$$

where:

- p_{Ti} and p_{Tj} are the transverse momenta of the jet constituents i and j .
- ΔR_{ij} denotes the distance between constituents i and j in the η - ϕ plane, defined by Equation 4.2.
- R is the distance parameter that determines the jet size.

$$\Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \quad (4.2)$$

2. **Clustering:** Merge the pair of jet constituents with the smallest d_{ij} value into a single jet.
3. **Iteration:** Repeat steps 1 and 2 until all jet constituents have been clustered into jets.
4. **Final Jet Four-Momentum:** The final four-momentum of the jet is obtained by summing the four-momenta of all constituent particles.

The analysis in Chapter 5 employs two principal jet algorithms: the anti- k_t algorithm with $R = 0.4$, referred to as small-radius or small- R jets, and the anti- k_t algorithm with $R = 1.0$, referred to as large-radius or large- R jets. In addition to these primary reconstruction algorithms, several topological clustering techniques are utilised to enhance jet kinematic measurements from the ATLAS calorimeter:

- **Topocluster Jets (Topo):**

The Topocluster algorithm [74] forms topologically connected clusters in calorimeter cells exhibiting high signal-to-noise ratios. While predominantly used in earlier ATLAS analyses, the Topocluster algorithm contributes in this study to the derivation of jet mass scale (JMS) uncertainties for small- R jets.

- **Particle Flow (PFlow):**

The Particle Flow algorithm [75] integrates the Topocluster approach with calorimeter information and tracking data from the Inner Detector. By removing calorimeter energy deposits that correspond to reconstructed tracks, double-counting of jet energies is mitigated, and jet energy resolution at low p_T is improved. PFlow serves as the primary topological cluster algorithm for small- R jets in ATLAS Run 2.

- **Local Hadronic Cell Weighting (LCW):**

The LCW algorithm [74] applies weights to calorimeter cell energy measurements to account for the differing responses to hadronic and electromagnetic showers. These weights, determined from simulations, balance the measurement of these signals. Furthermore, LCW compensates for signal losses in-

duced by the Topocluster algorithm and inactive calorimeter material. LCW is the primary topological cluster algorithm adopted in ATLAS Run 2 for large- R jets.

4.2.2 Jet Mass Reconstruction Tools

When highly Lorentz-boosted massive parent particles, such as W and Z bosons or top quarks, undergo decay, their products tend to travel in a collimated configuration. To search for such particles, ATLAS employs jet mass reconstruction tools to measure the jet mass, aiding the discrimination between continuum multijet backgrounds and the decays of massive particles. The method of jet mass reconstruction depends on the jet size.

4.2.2.1 Small- R jets

For small- R jets, the jet mass is calculated directly from the calorimeter information, as shown in Equation 4.3:

$$m_J^{calo} = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \mathbf{p}_i\right)^2} \quad (4.3)$$

where E_i and $\mathbf{p}_i$ represent the energies and three-momenta of the jet constituents within the clustered jet J [76].

4.2.2.2 Large- R jets

For large- R jets, jet mass reconstruction incorporates additional information from tracks matched to trimmed calorimeter jets, a technique described further in Section 4.2.3.2, resulting in improved mass resolution [76]. The combined jet mass, m_J^{comb} , is computed using Equation 4.4:

$$m_J^{comb} = w_{calo} \times m_J^{calo} + w_{TA} \times m^{TA} \quad (4.4)$$

Here, m_J^{comb} denotes the combined jet mass for large- R jets, while m_J^{calo} corresponds to the calorimeter jet mass as defined in Equation 4.3. The term m^{TA} refers to the track-assisted mass, calculated according to Equation 4.5.

$$m^{TA} = \frac{p_T^{calo}}{p_T^{track}} m^{track} \quad (4.5)$$

The definitions of the terms are as follows:

- p_T^{calo} and p_T^{track} represent the transverse momenta of the calorimeter jet and the matched tracks, respectively.
- m^{track} corresponds to the mass of the matched tracks, where the individual track mass is assigned according to the pion mass hypothesis.
- w_{calo} and w_{TA} are weights derived from the inverse quadrature of the mass resolutions, as given in Equation 4.6.

$$w_{TA} = \frac{\sigma_{TA}^{-2}}{\sigma_{TA}^{-2} + \sigma_{calo}^{-2}} ; w_{calo} = \frac{\sigma_{calo}^{-2}}{\sigma_{TA}^{-2} + \sigma_{calo}^{-2}} \quad (4.6)$$

Here, σ_{calo} and σ_{TA} denote the resolutions for the calorimeter jet mass and the track-assisted mass, respectively.

4.2.3 Jet Calibration

Jet calibration in ATLAS involves correcting the measured energy, momentum, and mass of jets to enhance their accuracy. The calibration procedure varies according to the jet size.

4.2.3.1 Origin Correction (For small-R jets only)

For small- R jets, the jet origin is corrected to align with the interaction point. The primary vertex is reconstructed based on the transverse momenta of tracks associated with the interaction point, as shown in Equation 4.7:

$$p_T^{vertex} = \sum_{i < n_{track}} p_T^{track,i} \quad (4.7)$$

Here, $p_T^{track,i}$ denotes the transverse momentum of the i -th track, p_T^{vertex} represents the transverse momentum of the vertex, and n_{track} corresponds to the number of tracks associated with the vertex [77].

4.2.3.2 Pileup Mitigation

Pileup, referring to additional energy deposits from other proton-proton collisions, affects jet measurements. Two types of pileup are considered: in-time and out-of-time. In-time pileup originates from simultaneous collisions, whereas out-of-time pileup results from collisions occurring immediately before or after the collision of interest. Pileup mitigation strategies differ between small- R and large- R jets.

- **Small- R Jets:**

The area-based and residual-based pileup subtraction method is applied, accounting for pileup density (ρ), the number of primary vertices (N_{PV}), and the average number of interactions per bunch crossing (μ), as expressed in Equation 4.8:

$$p_T^{corr} = p_T^{reco} - \rho \times A - \alpha(N_{PV} - 1) - \beta\langle\mu\rangle \quad (4.8)$$

In this expression, p_T^{reco} is the reconstructed jet transverse momentum prior to correction, p_T^{corr} is the corrected transverse momentum, A is the jet area, and α and β are empirical constants [77].

- **Large- R Jets:**

Jet trimming is employed to mitigate pileup in large- R jets [78]. The jet is reclustered into smaller subjets using the anti- k_t algorithm with a radius parameter of $R = 0.2$, and soft constituents with transverse momenta below 5% of the parent jet are removed. The surviving subjets are subsequently reclustered to form the final jet.

4.2.3.3 Absolute Monte Carlo Simulation Based Calibration

Following pileup mitigation, the subsequent step in jet calibration involves adjusting the detector response using Monte Carlo simulation [77][79]. This procedure

focuses on three primary variables: jet energy (Jet Energy Scale or JES), jet mass (Jet Mass Scale or JMS), and jet pseudorapidity (η).

To perform this calibration, the ratio of reconstructed jet mass (or energy) to the true jet mass (or energy) is evaluated as a function of the true jet transverse momentum, p_T^{truth} (or detector jet energy, E_{reco} , for energy calibration). For jet pseudorapidity calibration, the difference between the reconstructed jet pseudorapidity (η_{reco}) and the true jet pseudorapidity (η_{truth}) is determined as a function of η_{det} , which denotes the jet η relative to the detector's geometric center. The calibrated jet responses are subsequently applied to reconstructed jets to obtain their updated four-momenta. Further refinement of this Monte Carlo-based calibration is achieved by comparing the simulation to real data, as described in Section 4.2.3.5.

4.2.3.4 Global Sequential Calibration

The Global Sequential Calibration (GSC) [77] adjusts the detector response according to different particle types, including hadrons, leptons, and photons. This calibration comprises three categories:

- **Non-compensation:**

Corrects the detector response for hadrons, leptons, and photons.

- **Flavour dependence:**

Adjusts the detector response for quarks and gluons.

- **Punch through:**

Calibrates the response for jets that extend beyond the ATLAS calorimeter.

The GSC procedure utilises information from tracks in the Inner Detector (ID), muon tracks in the Muon Spectrometer (MS), and energy deposits in the calorimeter to derive a smooth fit for the jet response as a function of p_T and η (or E and η for punch through calibration). This smoothing process mitigates fluctuations in the jet response arising from different particle types while preserving the average jet energy.

4.2.3.5 *In-situ* Jet Calibration

The final step in jet calibration addresses discrepancies between the jet response in data and in Monte Carlo simulation. The *in-situ* calibration encompasses both jet energy scale (JES) [77] and jet mass scale (JMS) [79, 77]. JES calibration comprises three categories:

- η -intercalibration:
Balances the momentum of forward jets recoiling against central jets.
- Z/γ +jet balance:
Balances the momentum of jets recoiling against well-calibrated Z bosons or photons.
- Multijet balance:
Balances the momentum of a single high- p_T jet recoiling against a system of well-calibrated low- p_T jets.

For the *in-situ* JMS calibration, the method differs from JES calibration by comparing the calorimeter-to-track jet mass response, denoted as R_{trk} . These R_{trk} values are obtained as functions of the transverse momenta and mass of calorimeter jets, based on measurements from both real data and Monte Carlo simulation [79].

4.3 Track jets

In ATLAS, track jets are collections of tracks from the Inner Detector (ID) clustered into jets. Track jets are reconstructed using the anti- k_t algorithm with a variable radius, where the radius parameter R depends on the track jet transverse momentum p_T^j , as shown in Equation 4.9:

$$R \propto \frac{\rho_{const}}{p_T^j} ; R \in [0.02, 0.4] \quad (4.9)$$

Here, ρ_{const} denotes a constant term, whose value is optimised to resolve two b -tagged track jets within a single jet at high transverse momentum p_T (as illus-

trated in Figure 4.1)¹. For physics analyses, track jets in ATLAS are typically matched with calorimeter jets using either the Ghost Association method or the $\Delta R(i, j)$ matching technique.

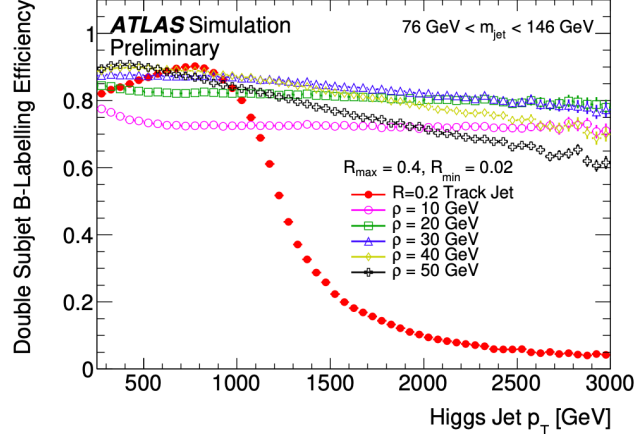


Figure 4.1: Double subject B-labeling efficiency as a function of the truth Higgs jet p_T for different values of ρ , derived from Monte Carlo simulation [9].

The Ghost Association method assigns infinitesimally small momentum components to track jets and compares their direction with the untrimmed calorimeter jet constituents. Upon finding a match between the direction of the track jet and the calorimeter jet constituent, the four-momentum of the untrimmed calorimeter jet constituent is adopted as the track jet’s four-momentum [80].

In the case of $\Delta R(i, j)$ matching, the angular separation between the track jet and the calorimeter jet in the η – ϕ plane is used, with the $\Delta R(i, j)$ defined as follows:

$$\Delta R(i, j) = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2} \quad (4.10)$$

Here, $\Delta R(i, j)$ represents the angular distance between two physics objects (i, j) as used in ATLAS analyses. The variables $\eta_{i/j}$ and $\phi_{i/j}$ denote the pseudorapidity

¹The Higgs jet in Figure 4.1 refers to the truth jet originating from the Higgs boson decay into $b\bar{b}$. The double subject B-labeling efficiency corresponds to the probability of two subjets within a single calorimeter jet being identified as b -jets, which match the truth b -jets from Higgs boson decay. In this context, B-labeling on subjets typically involves b -tagged track jets due to their kinematic association with these subjets via ghost association. A higher double subject B-labeling efficiency indicates effective resolution of two separate b -tagged track jets within a single calorimeter jet. As shown in Figure 4.1, variable radius track jets achieve high double subject B-labeling efficiency across a wide range of high p_T Higgs jets.

and azimuthal coordinates of the respective physics objects. In this case, i and j correspond to the calorimeter jet and the track jet, respectively. The matching criterion for $\Delta R(i, j)$ is typically required to fall within the size of the calorimeter jet.

4.4 **b**-tagging

While $HH \rightarrow b\bar{b}b\bar{b}$ features the highest branching ratio among all decay channels, the b and $\bar{b}$ quarks from the Higgs boson decays hadronise to form composite particles known as b -hadrons, which typically consist of b quarks bound with lighter quarks. These b -hadrons usually exhibit lifetimes on the order of 10^{-12} seconds. During this short lifetime, high-energy b -hadrons are expected to travel several millimetres before decaying into lighter quarks, producing jets or leptons. Consequently, certain tracks, potentially originating from b -hadrons, are observed to be displaced from the primary vertex of the jets.

To identify jets containing b -hadrons, referred to as b -jets, the ATLAS experiment employs the b -tagging algorithm, which integrates information from the tracks in the Inner Detector (ID), the calorimeter, and the muon chambers to construct a final discriminant separating b -jets, c -jets, and light-jets.

The b -tagging procedure utilises top quark pair production events ($t\bar{t} \rightarrow W^+bW^-\bar{b}$) with dileptonic decays, obtained from both data and Monte Carlo (MC) simulation samples. These samples are selected due to the high purity of the b -jet content. Subsequently, the data and MC samples undergo kinematic selection cuts on several physics objects, including jets, tracks, and muons, which are then provided as inputs to the b -taggers. The b -tagging framework consists of two levels: low-level and high-level b -taggers.

4.4.1 Low-Level **b**-tagger

Low-level b -taggers rely on direct detector information, such as the kinematic properties of tracks, jets, and muons, to distinguish between b -jets and jets of other flavours, including those containing charm hadrons (c -jets), light-jets, and jets from

τ decays. These taggers operate independently and encompass a variety of algorithms, including the Impact Parameter-based algorithm, the Recurrent Neural Network Impact Parameter (RNNIP), the Secondary Vertex algorithm, JetFitter, and the Soft Muon Tagger.

4.4.1.1 Impact Parameter-Based Algorithm

The Impact Parameter-based algorithm measures the displacement of tracks from the primary vertex of the jet in both the longitudinal (z_0) and transverse (d_0) directions, as illustrated in Figure 4.2. This algorithm is categorised into IP2D, which employs only d_0 , and IP3D, which utilises both d_0 and z_0 [81].

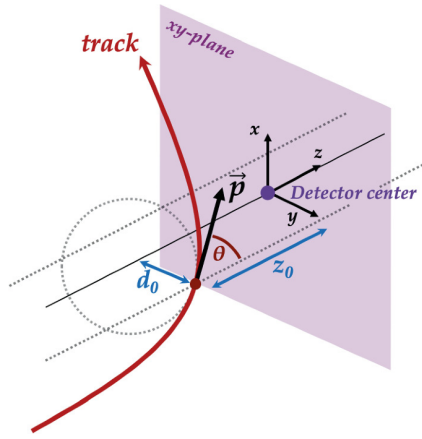


Figure 4.2: Diagram illustrating the impact parameter of tracks in the d_0 and z_0 directions.

The algorithm employs probability density functions (PDFs) for tracks associated with *b*-jets, *c*-jets, and light-jets. These PDFs, derived from Monte Carlo simulations, are assumed to be uncorrelated across jet flavours. Furthermore, the PDFs depend on the impact parameter significance, defined as the ratio of the impact parameter measurement to its uncertainty. Tracks from *b*-jets generally exhibit higher impact parameter significance relative to tracks from lighter-flavour jets.

The algorithm calculates the Log-Likelihood Ratio (LLR) as the discriminant for distinguishing tracks originating from *b*-jets, *c*-jets, and light-jets. The LLR is defined as the sum of the per-track log-probability ratios between the jet flavour hypotheses within the jet [81].

4.4.1.2 Recurrent Neural Network Impact Parameter (RRNIP)

To improve the discrimination between tracks originating from *b*-jets and those from light jets, the Recurrent Neural Network Impact Parameter (RRNIP) algorithm has been developed. This RNN incorporates three input features into the feedforward neural network:

- impact parameter significance S ,
- fractional transverse momentum of the tracks relative to the jet, p_T^{frac} ,
- angular separation between the tracks and the jet, $\Delta R(\text{track}, \text{jet})$.

The output node of the neural network yields the probabilities for the tracks to be associated with *b*-jets, *c*-jets, light jets, or τ -jets, respectively [81]. The *b*-jet discriminant, $D_{\text{RNN}}(b)$, is defined according to Equation 4.11:

$$D_{\text{RNN}}(b) = \ln \frac{p_b}{f_c p_c + f_\tau p_\tau + (1 - f_c - f_\tau) p_{\text{light}}} \quad (4.11)$$

Here,

- p_b , p_c , p_τ , and p_{light} represent the probabilities of the tracks being associated with *b*-jets, *c*-jets, τ -jets, and light jets, respectively.
- f_c and f_τ denote the fractions of *c*-jets and τ -jets, respectively. These fractions are determined from the composition of events containing such jets in $t\bar{t} \rightarrow W^- b W^+ \bar{b}$ decays, as extracted from both Monte Carlo simulation and experimental data.

4.4.1.3 Secondary Vertex Finder (SV1) Algorithm

B-hadrons and their decay products (*b*-jets) often overlap with other hadronic particles within the jet axis, complicating their identification. To improve this, the Secondary Vertex Finder (SV1) algorithm is employed [82].

The SV1 algorithm reconstructs two-track pairs within the jet axis as secondary vertices. It removes two-track pairs that are likely to originate from the decay of

long-lived particles (such as K and Λ baryons), photon conversions, or hadronic interactions with the detector material. This procedure iteratively eliminates incompatible two-track pairs until the χ^2 track-to-vertex fit satisfies the acceptance criteria and the vertex mass remains below 6 GeV, or no further tracks are available. The associated tracks are subsequently compared to models representing b -jets, c -jets, and light jets using the LLR discriminant.

The inputs to the SV1 algorithm are summarised as follows:

- mass of the secondary vertex, m_{SV} ,
- fraction of the energy from the secondary vertex relative to the primary vertex, f_{SVE} ,
- number of tracks associated with the secondary vertex, $N_{trksAtVtx}^{SV}$,
- distance and transverse distance from the primary to the secondary vertex, L_{xyz} and L_{xy} .

4.4.1.4 JetFitter

The JetFitter algorithm [83] reconstructs the decay chain of b - and c -hadrons within a single cascade of b -hadron decays in the jet. It exploits tracks to build multiple vertices (secondary and tertiary) aligned along the jet axis. The performance of the JetFitter algorithm is assessed using the LLR discriminant, which is derived from probability density functions (PDFs) of kinematic variables, including the decay length significance, vertex mass, and energy fraction.

Figure 4.3 presents the set of input variables for the JetFitter algorithm.

4.4.1.5 Soft Muon Tagger

The Soft Muon Tagger (SMT) [84] serves as another low-level b -tagger that exploits muon information to identify semi-leptonic decays of b - and c -hadrons. The branching ratios for b - and c -quark decays into muons are approximately 11% and 10%, respectively. In cases where b -jets or c -jets decay into muons, these muons typically exhibit high transverse momentum relative to the jet axis (p_T^{rel}).

JETFITTER	$m(\text{JF})$	Invariant mass of tracks from displaced vertices
	$f_E(\text{JF})$	Energy fraction of the tracks associated with the displaced vertices
	$\Delta R(\vec{p}_{\text{jet}}, \vec{p}_{\text{vtx}})(\text{JF})$	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
	$S_{xyz}(\text{JF})$	Significance of the average distance between PV and displaced vertices
	$N_{\text{TkAlVtx}}(\text{JF})$	Number of tracks from multi-prong displaced vertices
	$N_{2\text{TkVtx}}(\text{JF})$	Number of two-track vertex candidates (prior to decay chain fit)
	$N_{1\text{-trk vertices}}(\text{JF})$	Number of single-prong displaced vertices
	$N_{\geq 2\text{-trk vertices}}(\text{JF})$	Number of multi-prong displaced vertices
	$L_{xyz}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Distance of 2nd or 3rd vertex from PV
	$L_{xy}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Transverse displacement of the 2nd or 3rd vertex
JETFITTER <i>c</i> -tagging	$m_{\text{Tk}}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Invariant mass of tracks associated with 2nd or 3rd vertex
	$E_{\text{Tk}}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Energy fraction of the tracks associated with 2nd or 3rd vertex
	$f_E(2\text{nd}/3\text{rdvtx})(\text{JF})$	Fraction of charged jet energy in 2nd or 3rd vertex
	$N_{\text{TkAlVtx}}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Number of tracks associated with 2nd or 3rd vertex
	$Y_{\text{trk}}^{\min}, Y_{\text{trk}}^{\max}, Y_{\text{trk}}^{\text{avg}}(2\text{nd}/3\text{rdvtx})(\text{JF})$	Min., max. and avg. track rapidity of tracks at 2nd or 3rd vertex

Figure 4.3: The input variables for the JetFitter algorithm [10].

The SMT employs an angular distance parameter of $\Delta R < 0.4$ to associate muons with jets. Information from the matched muons, including p_T^{rel} , d_0 , and ΔR , is extracted alongside three muon quality variables:

- **Scattering Neighbour Significance S :**

This variable quantifies the significance of kinks along the track, calculated as:

$$S = q \times \sum_i \frac{\Delta\phi_{\text{scat}}^i}{\sigma_{\delta\phi_{\text{scat}}^i}} \quad (4.12)$$

where q denotes the particle charge, $\Delta\phi_{\text{scat}}^i$ represents the angular difference between two half-tracks ending or starting at adjacent hits along the muon trajectory, and $\sigma_{\delta\phi_{\text{scat}}^i}$ corresponds to the associated uncertainty of $\Delta\phi_{\text{scat}}^i$.

- **Momentum Balance Significance M :**

This variable measures the difference between muon momentum measurements in the Inner Detector (ID) and the Muon Spectrometer (MS), expressed as:

$$M = \frac{p^{ID} - p_{\text{ext}}^{MS}}{\sigma_{E_{\text{loss}}}} \quad (4.13)$$

where p^{ID} refers to the muon momentum measurement in the ID, p_{ext}^{MS} indicates the momentum measurement in the MS extrapolated to the primary vertex, and $\sigma_{E_{\text{loss}}}$ denotes the uncertainty associated with the energy loss measured by the calorimeter.

- **Track Curvature Ratio R :**

This variable expresses the ratio of track curvature measurements between the ID and MS, defined as:

$$R = \frac{(q/p)_{ID}}{(q/p)_{MS}} \quad (4.14)$$

where $(q/p)_{ID}$ and $(q/p)_{MS}$ represent the track curvature measurements in the ID and MS, respectively.

These muon quality variables provide an effective means to distinguish muons originating from *b*- or *c*-hadron decays from those produced by π or *K* decays.

4.4.2 High-Level *b*-tagger (DL1r)

The DL1r algorithm [10] serves as a high-level *b*-tagger, employing a Deep Learning Neural Network. DL1r utilises input variables from low-level *b*-tagging algorithms, which are processed through a deep feedforward neural network. The output of DL1r provides the probabilities of tracks originating from *b*-jets, *c*-jets, and light-flavour jets. The *b*-jet discriminant $D(b)_{DL1r}$ in DL1r is defined by Equation 4.15:

$$D(b)_{DL1r} = \ln \frac{p_b}{f_c p_c + (1 - f_c) p_{light}} \quad (4.15)$$

Here,

- p_b , p_c , and p_{light} represent the probabilities of tracks originating from *b*-jets, *c*-jets, and light-flavour jets, respectively.
- f_c denotes the fraction of *c*-flavour jets, derived from the composition of events produced in $t\bar{t} \rightarrow W^- b W^+ \bar{b}$ processes, based on Monte Carlo simulation and data measurements.

The DL1r algorithm constitutes the principal *b*-tagging method employed in the analysis dedicated to the search for resonant production of $HH \rightarrow b\bar{b}b\bar{b}$, as described in Chapter 5.

4.4.3 *b*-tagging Efficiency

In addition to the *b*-tagger, ATLAS defines *b*-tagging efficiency working points for use in physics analyses involving *b*-hadrons. A *b*-tagging efficiency working point corresponds to the threshold applied to the output of the *b*-tagger (in this case, the DL1r algorithm) to classify jets as *b*-jets.

A higher *b*-tagging efficiency working point increases the likelihood of tagging all jets as *b*-jets, including those originating from light-flavor quarks. Conversely, a lower working point reduces the probability of misidentifying light-flavor jets but increases the risk of failing to identify genuine *b*-jets. Most physics analyses at ATLAS adopt a working point of approximately 77% to balance efficiency and purity in *b*-tagging. For the $HH \rightarrow b\bar{b}b\bar{b}$ analysis presented in Chapter 5, the working point is set to 77%.

4.4.4 *b*-tagging Efficiency Calibration

The subsequent step in the *b*-tagging procedure involves calibrating the *b*-tagging efficiency between Monte Carlo simulation and data [11]. This calibration introduces a scale factor, SF_b , defined as the ratio of the efficiency measured in data (ϵ_{data}^b) to that in Monte Carlo simulation (ϵ_{MC}^b), as shown in Equation 4.16:

$$SF_b = \frac{\epsilon_{data}^b}{\epsilon_{MC}^b} \quad (4.16)$$

The scale factors are applied as event weights in both data and simulation, based on relevant kinematic variables associated with *b*-tagging, such as the transverse momentum of the jets (p_T), as illustrated in Figure 4.4. The uncertainty in the *b*-tagging scale factors comprises four components: MC generator modeling, normalization, experimental effects, and extrapolation uncertainties. These components contribute to the systematic uncertainties considered in the analysis discussed in Chapter 5.

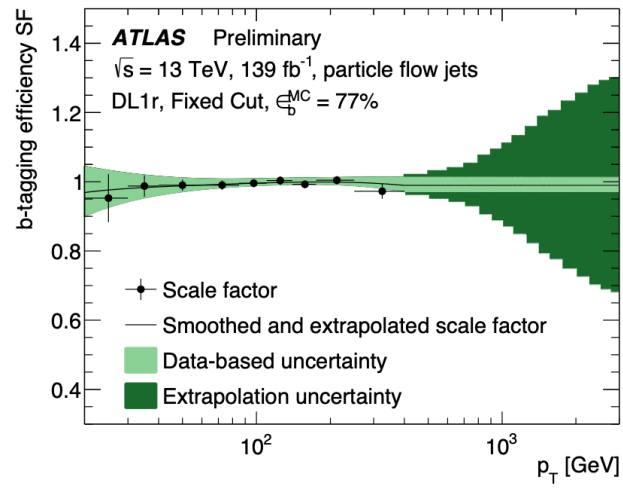


Figure 4.4: *b*-tagging efficiency scale factors as a function of jet p_T at the 77% working point [11].

Chapter 5

Search for Resonant production from $pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ in ATLAS Detector using Boosted and Superboosted Analysis

The previous three chapters have established a foundation relevant to the main topic of this thesis. Chapter two provided an overview of the Standard Model (SM) of particle physics, highlighting its connection to the hierarchy problem. This unresolved issue within the SM motivates the search for physics beyond the Standard Model (BSM), including the investigation of resonant Higgs boson pair production. Chapter three presented the technical details of the ATLAS detector, a sophisticated instrument designed to study particle interactions arising from high-energy proton collisions at the LHC. Chapter four examined the analysis tools employed in the ATLAS experiment, along with their key components and functionalities pertinent to the search for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks.

This final chapter presents an in-depth discussion of the search for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks within the ATLAS detector. A general overview of the different analysis methods applied in searches targeting this decay channel will be provided. Subsequently, the focus

will turn to the specific analysis strategies implemented in this research, namely the Boosted and Superboosted analyses. These sections will cover the data and Monte Carlo simulation samples, object reconstruction techniques, event selection criteria, background estimation strategies, a comprehensive examination of systematic uncertainties associated with each analysis, and finally, the statistical procedures used to set limits on the production cross-section of resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks.

5.1 Analysis methods in the search for resonant production $HH \rightarrow b\bar{b}b\bar{b}$

The current ATLAS searches for resonant $HH \rightarrow b\bar{b}b\bar{b}$ production focus on identifying mass resonances of BSM particles that decay into Higgs boson pairs. The BSM particles considered in this search include Massive Excited Kaluza-Klein Gravitons, G_{kk}^* , predicted by the Randall-Sundrum model, and a heavy neutral scalar boson, X , predicted by the Two-Higgs-Doublet Model (2HDM). These BSM resonances are primarily expected to be produced via gluon-gluon fusion¹. Upon production, these BSM particles decay into Higgs boson pairs, with each Higgs boson subsequently decaying into a pair of $b\bar{b}$ quarks, which then hadronise into jets containing b -hadrons within the ATLAS detector.

To address this decay channel, ATLAS employs three distinct analysis methods: the Resolved analysis, the Boosted analysis, and the Superboosted analysis². These three analysis strategies correspond to different event topologies, each optimised to probe a specific mass range of the BSM resonances. Beginning with Section 5.2, this chapter will focus exclusively on the Boosted and Superboosted analyses.

5.1.1 Resolved Analysis

The resonance of a BSM particle with a lower mass results in the decay into Higgs boson pairs with moderate Lorentz boosts. These moderately boosted Higgs bosons subsequently decay into two pairs of $b\bar{b}$ quarks, which hadronise into four separate jets containing b -hadrons. To identify such events, the Resolved analysis selects events featuring four small-radius jets originating from the interaction point, with at least three of these jets required to be b -tagged, indicating the presence of b -hadrons. ATLAS also incorporates selected charged particle tracks and muon tracks to account for energy loss due to semi-leptonic decays of b -hadrons. The b -tagged small-radius jets are subsequently grouped into two pairs, each representing the

¹These resonances can also be produced via vector boson fusion; however, this thesis considers only gluon-gluon fusion.

²For published results, ATLAS only includes the Resolved and Boosted analyses. The Superboosted analysis remains internal and unpublished.

decay products of a Higgs boson candidate.

The invariant mass of each Higgs boson candidate is reconstructed from the four-momenta of the corresponding pair of b -tagged jets. In contrast to searches targeting a fixed-mass, high-momentum Higgs boson, the Resolved analysis considers all possible combinations of the two reconstructed Higgs boson masses within a defined signal region. This region, optimised for maximum sensitivity, is typically represented as an ellipse in a two-dimensional plane of the two Higgs boson candidate masses.

The four-momenta of the two Higgs boson candidate pairs are then combined to form the final discriminating variable, which is the invariant mass of the system corresponding to the resonant particle decaying into Higgs boson pairs, denoted as $m(HH)$. This variable is utilised to set upper limits on the cross-section of the resonant particle decaying into Higgs boson pairs, which subsequently decay into two pairs of $b\bar{b}$ quarks. The analysis explores various hypotheses for the mass of the resonant particle within the range $251 \text{ GeV} < m_R < 1600 \text{ GeV}$, where $R = \{G_{kk}^*, X\}$. This mass range corresponds to the region where the Resolved analysis achieves the highest sensitivity. Furthermore, the $m(HH)$ distribution is employed to estimate background contributions under different event selection criteria.

5.1.2 Boosted Analysis

In contrast to the Resolved analysis, which is optimised for lower-mass resonances producing moderately boosted Higgs bosons, the Boosted analysis targets higher-mass resonances decaying into highly boosted Higgs bosons. These highly boosted Higgs bosons decay into two pairs of b -quarks, which hadronise into four collimated, high-momentum jets containing b -hadrons. Due to the close proximity of these jets, the Resolved analysis technique becomes inefficient for event identification.

To overcome this limitation, the Boosted analysis employs two large-radius jets originating from the interaction point. Each large-radius jet is expected to contain

the collimated decay products (two b -hadrons) of a highly boosted Higgs boson. The identification strategy involves applying kinematic selections to each large-radius jet, alongside the use of one or two track jets and muon tracks contained within them. These track jets are required to be b -tagged to confirm the presence of b -hadrons.

The invariant mass of each Higgs boson candidate is reconstructed by combining the four-momenta of the corresponding b -tagged track jets, muon tracks, and the associated large-radius jet. Similar to the Resolved analysis, the Boosted analysis defines a signal region represented by an elliptical selection in the two-dimensional plane of the reconstructed Higgs boson masses. The specific shape of this ellipse is optimised according to the kinematic features of highly boosted Higgs bosons, ensuring maximal signal acceptance predicted by Monte Carlo simulations while effectively suppressing background contributions. The four-momenta of the Higgs boson candidate pairs are then combined to form the $m(HH)$ variable, which is used to characterise both signal and background distributions and to derive upper limits on the cross-section for the production of a resonant particle decaying into Higgs boson pairs, which subsequently decay into two pairs of $b\bar{b}$ quarks. The Boosted analysis focuses on resonances within the mass range $900 \text{ GeV} < m_R < 5 \text{ TeV}$, where the sensitivity to the targeted signal topology is maximised.

5.1.3 Superboosted Analysis

The Superboosted analysis represents the third approach employed in searches for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks, $HH \rightarrow b\bar{b}b\bar{b}$. This method targets even heavier resonances compared to the Boosted analysis. Such very heavy resonances decay into Higgs bosons with extremely high Lorentz boosts. While travelling at these high boosts, the Higgs bosons decay into two pairs of $b\bar{b}$ quarks, which subsequently hadronise into jets containing b -hadrons. Due to the extreme boost of the parent Higgs bosons, the resulting jets become even more collimated than those encountered in the Boosted analysis.

The primary challenge for the Boosted analysis in this regime lies in jet mass

resolution. Although both the Boosted and Superboosted analyses utilise similar b -tagging techniques, the large-radius jets adopted in the Boosted analysis tend to suffer a loss in resolution when capturing the decay products of highly boosted Higgs bosons. These decay products become extremely collimated, travelling within a very narrow cone. The limited resolution of large-radius jets in such scenarios hinders the accurate separation of individual particles within the jet and complicates the reconstruction of the Higgs boson mass from their combined momenta.

The Superboosted analysis addresses this limitation by employing two small-radius jets originating from the interaction point. Each of these small-radius jets undergoes kinematic selection and incorporates the selected b -tagged track-jets and muon tracks located near the jet axis. While capturing a smaller volume compared to large-radius jets, these small-radius jets offer potential advantages in jet mass resolution, particularly for jets with very high transverse momentum, p_T . Studies based on Monte Carlo simulations [85] have demonstrated that small-radius jets can provide modest improvements in nominal jet mass resolution relative to large-radius jets with comparable mass and p_T , especially for jets with p_T exceeding 650 GeV. This improvement in jet mass resolution proves crucial for the Superboosted analysis, enabling a more precise reconstruction of the Higgs boson mass, even when the decay products are extremely collimated due to the high Lorentz boost. Moreover, an independent study presented in Appendix C compares the signal acceptance between the Boosted and Superboosted analyses. This study indicates that the Superboosted approach may offer additional signal acceptance for very heavy resonance decays. Together, these findings strongly motivate the use of the Superboosted analysis within this specific search regime.

The invariant mass of each Higgs boson candidate in the Superboosted analysis is reconstructed by combining the four-momenta of the corresponding b -tagged track jets, muon tracks, and the associated small-radius jet. As in the Boosted analysis, the Superboosted approach currently employs the same signal region to capture candidate signal events³. As in the other analyses, the four-momenta of the Higgs

³It is noted that this region has not yet been optimised for the Superboosted analysis due to time

boson pair are combined to form $m(HH)$, which serves as the final discriminating variable. This variable is used to derive both the background and signal distributions, which are subsequently utilised to set upper limits on the cross-section of the hypothesised resonant particle decaying to Higgs boson pairs, with each Higgs boson decaying into two $b\bar{b}$ quark pairs. The Superboosted analysis targets the resonance mass range of $1600 \text{ GeV} < m_R < 5 \text{ TeV}$, where it offers the highest sensitivity for the corresponding analysis topology. Figure 5.1 illustrates the topologies of the three different analyses.

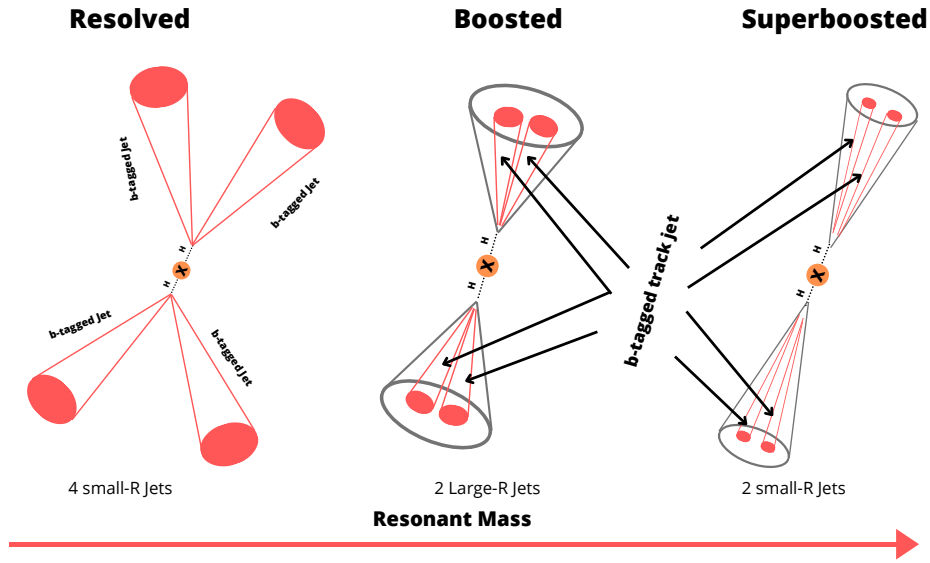


Figure 5.1: The topologies of the three analyses used in the search for resonant production of $HH \rightarrow b\bar{b}b\bar{b}$.

constraints. Further optimisation is recommended to potentially enhance signal acceptance in future work.

5.2 Data and Monte Carlo Samples

Section 5.1 has provided an overview of the analysis methods employed in the search for resonant $HH \rightarrow b\bar{b}b\bar{b}$ production. This section presents a detailed description of the data and Monte Carlo samples utilised in the Boosted and Superboosted analyses.

5.2.1 Data Samples

The Boosted and Superboosted analyses employ 139 fb^{-1} of data collected by the ATLAS detector during the 2015–2018 data-taking period at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$. These data samples serve two primary purposes: the search for resonant $HH \rightarrow b\bar{b}b\bar{b}$ signals and the modelling of background contributions from multi-jet events, which arise from the hadronisation of quark and gluon interactions between the partons in the colliding protons (QCD multi-jet background). The QCD multi-jet events represent the largest production cross-section and are expected to contain jets with misidentified b -tagged track jets, thereby leading to a final state that closely resembles the $HH \rightarrow b\bar{b}b\bar{b}$ signal. The procedure for background derivation will be described in detail in Section 5.5.

5.2.2 Monte Carlo Samples

The Boosted and Superboosted analyses utilise three categories of Monte Carlo (MC) samples: signal samples (G_{kk}^* and X), $t\bar{t}$ samples, and QCD multi-jet samples. These samples serve three main purposes: modelling of the signal, modelling of the top quark pair production background, and cross-checking the data-driven estimation of the QCD multi-jet background.

In addition to the $t\bar{t}$ background modelling, several supplementary $t\bar{t}$ MC samples are incorporated into the Boosted and Superboosted analyses. These additional samples are employed to evaluate systematic uncertainties associated with the modelling of the $t\bar{t}$ background in each analysis. The following subsections provide a detailed description of all relevant Monte Carlo samples used in the Boosted and Superboosted analyses.

5.2.2.1 Signal

The first MC signal sample employed in the Boosted and Superboosted analyses corresponds to the massive excited Kaluza-Klein graviton G_{kk}^* from the Randall–Sundrum model, which decays into pairs of Higgs bosons, subsequently decaying into two $b\bar{b}$ pairs, i.e., $pp \rightarrow G_{kk}^* \rightarrow HH \rightarrow b\bar{b}b\bar{b}$. This process is produced via gluon-gluon fusion (ggF). In modelling this process, ATLAS fixes the value of the warp factor k , setting the ratio between k and the reduced Planck mass $\bar{M}_p$ to $k/\bar{M}_p = 1.0$, where $\bar{M}_p = 2.4 \times 10^{18}$ GeV. In addition to fixing the warp factor, ATLAS generates two sets of G_{kk}^* MC samples. The first set, dedicated to the Boosted analysis, comprises 18 MC samples with different G_{kk}^* masses ranging from $900 \text{ GeV} < m_{G_{kk}^*} < 5 \text{ TeV}$. The second set, prepared for the Superboosted analysis, contains 11 MC samples covering the mass range of $1600 \text{ GeV} < m_{G_{kk}^*} < 5 \text{ TeV}$. Each sample is generated at leading order (LO) in α_s using the MadGraph 5-aMC@NLO (MadGraph) 2.2.2 matrix element generator, interfaced with Pythia v8.186 for the modelling of the underlying event, parton showering, and hadronisation. The NNPDF 2.3 LO PDF set is used for these samples [86][87].

The second MC signal sample utilised in both the Boosted and Superboosted analyses corresponds to the heavy neutral scalar boson X from the 2HDM, decaying into Higgs boson pairs, which subsequently decay into two $b\bar{b}$ pairs, i.e., $pp \rightarrow X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$. This process is also produced via gluon-gluon fusion (ggF). Similar to the treatment of G_{kk}^* , ATLAS prepares two sets of MC samples for the heavy neutral scalar boson X . The first set, designed for the Boosted analysis, contains 14 MC samples with X masses ranging from $900 \text{ GeV} < m_X < 5 \text{ TeV}$. The second set, intended for the Superboosted analysis, comprises 7 MC samples with masses spanning $1600 \text{ GeV} < m_X < 5 \text{ TeV}$. Each of these samples is generated at leading order (LO) in α_s using MadGraph 5-aMC@NLO (MadGraph) 2.6.1, coupled with Herwig 7.1.3 for the modelling of the underlying event, parton showering, and hadronisation. The MMHT2014LO PDF set is used for these samples [88]. Despite the large gaps between resonant masses in the X MC samples for

$m_X > 2$ TeV, both Boosted and Superboosted analyses employ a signal interpolation method to obtain signal distributions (including associated systematic uncertainties) for $m_X = \{2250, 2750, 3500, 4500\}$ GeV. This approach ensures alignment of the resonant mass points for X with those of G_{kk}^* in the cross-section limit setting procedure. Further details of the signal interpolation method are provided in Section 5.7.1.

5.2.2.2 Top quark pair production

The second type of MC sample used in the Boosted and Superboosted analyses corresponds to top quark pair production. This category models the background contribution arising from $t\bar{t}$ production, which proceeds through the decay chain $t\bar{t} \rightarrow bW^- \bar{b}W^+$, i.e., $pp \rightarrow t\bar{t} \rightarrow b\bar{b}W^-W^+$. Top quark pair production represents the second largest contribution to background events due to the sizeable production cross section. The decay products, including hadronic W decays and b -quark decays, yield a combination of genuine and misidentified b -jets, potentially mimicking the final state of $HH \rightarrow b\bar{b}b\bar{b}$.

For the modelling of $t\bar{t}$ production, the nominal MC samples are generated based on top quark pair decays into bW^- and $\bar{b}W^+$, with the process initiated via gluon-gluon fusion (ggF). Two distinct decay modes are considered in this context: all-hadronic and non-all-hadronic. The all-hadronic decay mode involves both $W^\pm$ bosons decaying into two light quarks each, whereas the non-all-hadronic mode refers to configurations in which at least one of the $W^\pm$ bosons decays into leptons.⁴

These two distinct final states are modelled using separate MC samples and subsequently combined to represent the full $t\bar{t}$ production. The procedure for combining these samples is described in Section 5.5. Each MC sample used in the Boosted and Superboosted analyses is generated at NLO in α_s using the POWHEG-BOX v2 matrix element generator. The underlying event simulation, parton showering, and hadronization are performed using PYTHIA v8.230. For the PDF set, the

⁴Non-all-hadronic $t\bar{t}$ decays include multiple combinations of hadronic and leptonic final states. To account for all such configurations, ATLAS simulates events corresponding to each final state, weighted according to their respective branching ratios.

NNPDF 2.3 LO [86, 87] is employed. Furthermore, the MC samples incorporate the damping parameter h_{damp} , which regulates the contribution of hard radiation (particles with large transverse momentum, p_T) in the parton shower [89, 90]. The value of this parameter is set to $h_{\text{damp}} = 1.5 m_{\text{top}}$, where m_{top} denotes the mass of the top quark.

In addition to the nominal $t\bar{t}$ samples, three alternative sets of $t\bar{t}$ MC samples are utilised to evaluate systematic uncertainties associated with the matrix element generator, parton shower modelling, and the choice of h_{damp} parameter. Each alternative set contains two MC samples covering the all-hadronic and non-all-hadronic decay modes, with all samples generated from the $pp \rightarrow t\bar{t} \rightarrow b\bar{b}W^-W^+$ process via ggF . Table 5.1 summarises the details of these $t\bar{t}$ MC samples employed for systematic uncertainty evaluation.

$t\bar{t}$ systematics	Matrix Element	Parton Shower	PDF	h_{damp}
Nominal	Powheg-Box v2	Pythia v8.230	NNPDF 2.3 LO	$1.5 m_{\text{top}}$
h_{damp} parameter	Powheg-Box v2	Pythia v8.230	NNPDF 2.3 LO	$3 m_{\text{top}}$
Parton Shower	Powheg-Box v2	Herwig 7.1.3	NNPDF 2.3 LO	$1.5 m_{\text{top}}$
Matrix Element	MadGraph 5-aMC@NLO	Pythia v8.230	NNPDF 2.3 LO	$1.5 m_{\text{top}}$

Table 5.1: Details of the $t\bar{t}$ MC samples used to assess systematic uncertainties in the modelling of top quark pair production.

5.2.2.3 QCD multijet

The third type of MC sample employed in the Boosted and Superboosted analyses is the QCD multijet sample. Although a data-driven method is primarily used to model the QCD multijet background, this MC sample serves to perform closure checks against the QCD multijet background derived from data. The discrepancies between the data-driven and MC-based backgrounds are incorporated as uncer-

tainties associated with potential mismodelling between the two approaches. The procedure for deriving this uncertainty will be further detailed in Section 5.6.

The QCD multijet Monte Carlo (MC) sample is based on $2 \rightarrow 2$ scattering processes, wherein two partons interact to produce two outgoing partons. These high-energy partons subsequently hadronise into jets. In the context of searches for Higgs boson pair production decaying into two pairs of $b\bar{b}$, some of these jets may be misidentified as b -jets, resulting in a final state that resembles the signal of interest. This QCD multijet MC sample is generated using PYTHIA v8.235, which provides the pure QCD $2 \rightarrow 2$ Leading Order (LO) α_s matrix element calculation, underlying event modelling, parton shower, and hadronisation. The parton distribution function (PDF) set used for this sample is NNPDF 2.3 LO [86, 87].

5.3 Object Reconstruction

Building upon the data and Monte Carlo samples described in Section 5.2, this section discusses the reconstruction of physics objects used in the Boosted and Superboosted analyses.

These analyses rely on three main objects: calorimeter jets, track jets, and muon tracks. These objects are matched to each other using dedicated selection criteria and matching methods. Once matched, their corresponding four-momenta are combined and used for event selection to identify candidate Higgs boson pairs. The reconstruction details for each of these objects are described in this section.

5.3.1 Calorimeter jets

To reconstruct calorimeter jets for the Boosted analysis, two calibrated large- R jets originating from the interaction point are required. These jets are referred to as Boosted jets and denoted as J_b . Following the convention of the Boosted analysis, the Boosted jet with the higher transverse energy in each event is labeled the Leading Boosted jet (J_b^1), and the other is labeled the Subleading Boosted jet (J_b^2).

Similarly, for the Superboosted analysis, two calibrated small- R jets originating from the interaction point are utilised. These jets are referred to as Superboosted jets and denoted as J_s . Following the same convention as in the Boosted analysis, the Superboosted jet with the higher transverse energy in each event is labeled the Leading Superboosted jet (J_s^1), and the other is labeled the Subleading Superboosted jet (J_s^2).

5.3.2 Track jets

The track jets used in the Boosted and Superboosted analyses are reconstructed from tracks in the Inner Detector (ID) using dedicated jet clustering algorithms.

For the Boosted analysis, the track jets associated with the Boosted jets are denoted as j_b . These j_b are reconstructed using the anti- k_t algorithm with a variable radius parameter R_b , where the size of R_b is inversely proportional to the track jet transverse momentum $p_T^{j_b}$, as described by Equation 4.9. For the Boosted analysis,

ρ_{const} is set to 30 GeV, and the size of these track jets varies within the range $0.02 \leq R_b \leq 0.4$.

To associate the track jets with the Boosted jets, each track jet is required to have $p_T^{j_b} > 5$ GeV and $|\eta^{j_b}| < 2.5$, and must be matched to a Boosted jet using the Ghost Association method [80]. Furthermore, any matched track jet considered as b -tagged must satisfy the 77% b -tagging efficiency working point using the DL1r algorithm. This specific working point is selected to optimise the balance between background rejection (mistagging of light- and charm-jets) and b -tagging efficiency.

For the Superboosted analysis, the track jets associated with the Superboosted jets are denoted as j_s . Similar to the Boosted analysis, the j_s are reconstructed using the anti- k_t algorithm with a variable radius parameter R_s , with ρ_{const} set to 30 GeV. The size of these track jets varies within the range $0.02 \leq R_s \leq 0.4$.

To associate the j_s with the Superboosted jets, each j_s must have $p_T^{j_s} > 5$ GeV and $|\eta^{j_s}| < 2.5$, and must be located within $\Delta R(J_s, j_s) < 0.4$ ⁵. For b -tagging, any matched track jet is required to pass the 77% b -tagging efficiency working point using the DL1r algorithm to be considered as b -tagged. The working point choice in the Superboosted analysis also aims to optimise the balance between background rejection and b -tagging efficiency.

5.3.3 Muons

In both the Boosted and Superboosted analyses, muon reconstruction is derived from the matching of muon tracks from the Muon Spectrometer (MS) to track jets. The purpose of matching muon tracks from the MS is to account for missing energy from semi-leptonic b -decays and to apply kinematic corrections to the Boosted and Superboosted jets.

For the Boosted analysis, muon tracks associated with the Boosted jets are denoted as mu_b . To be included in the Boosted jets, these muons must first satisfy a

⁵The Ghost Association method is commonly used for matching track jets to large- R jets. However, within ATLAS, this method is not currently developed for matching track jets to small- R jets. Therefore, the Superboosted analysis utilises the ΔR method for matching track jets to Superboosted jets.

”medium” quality working point. Each mu_b is required to have a transverse momentum of at least $p_T^{mu_b} > 4$ GeV and a pseudorapidity $|\eta^{mu_b}| < 2.5$. Furthermore, each mu_b must be located within the proximity of a b -tagged track jet, with the distance between them satisfying $\Delta R(mu_b, j_b) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{mu_b})$. The $\Delta R(mu_b, j_b)$ is calculated using Equation 4.2. Once these criteria are fulfilled, the four-momentum of the muon is combined with the corresponding Boosted jet. The combined mass of the Boosted jet, $m_{j_b}^{\text{comb}}$, is then recalculated using Equation 4.4.

For the Superboosted analysis, muon tracks associated with the Superboosted jets are denoted as mu_s . To be included in the Superboosted jets, these muons must also satisfy the ”medium” quality working point. Each mu_s must have a transverse momentum of at least $p_T^{mu_s} > 4$ GeV and a pseudorapidity $|\eta^{mu_s}| < 2.5$. Additionally, each mu_s must be located within the proximity of a b -tagged track jet, requiring $\Delta R(mu_s, j_s) < \min(0.4, 0.04 + 10 \text{ GeV}/p_T^{mu_s})$. Similar to the Boosted analysis, $\Delta R(mu_s, j_s)$ is calculated using Equation 4.2. Once all criteria are met, the four-momentum of the muon is combined with the corresponding Superboosted jet.

5.4 Event Selection

Section 5.3 discussed the reconstruction of physics objects for the Boosted and Superboosted analyses. This section describes the application of event selection criteria on these physics objects to search for Higgs boson pairs decaying into two pairs of $b\bar{b}$ quarks.

The event selection criteria for the Boosted and Superboosted analyses consist of eight steps: trigger selection, kinematic selection of calorimeter jets, collinearity of track jets, Resolved analysis veto, tag categories, signal selection for $HH \rightarrow b\bar{b}b\bar{b}$, and the selection of control and validation regions for background modelling. Details for each of these selection criteria are discussed in Sections 5.4.1-5.4.7.

5.4.1 Trigger Selection

The first selection criterion for identifying the signal is the trigger selection. All reconstructed physics objects for both the Boosted and Superboosted analyses are required to pass several unprescaled High Level Triggers (HLT). Each of these triggers imposes specific requirements on parameters such as the transverse energy of the calorimeter jets (E_T^J), the mass of the calorimeter jets (m^J), the size and type of the calorimeter jets, the topological jet clustering algorithms, and the data-taking period.

For the Boosted analysis, all Leading Boosted jets are required to pass one of the following offline unprescaled High Level Triggers:

- **HLT_j360_a10_lcw_sub_L1J100**

This trigger requires untrimmed large-radius calorimeter jets (**a10**) to have $E_T^J > 360$ GeV at the HLT and $E_T^J > 100$ GeV at the L1Topo trigger. The untrimmed large- R calorimeter jets must be reconstructed using the LCW algorithm.

- **HLT_j420_a10_lcw_L1J100**

Similar to the first trigger, this trigger requires the transverse energy of untrimmed large-radius calorimeter jets to satisfy $E_T^J > 420$ GeV at the HLT and $E_T^J > 100$ GeV at the L1Topo trigger. The jets are reconstructed using

the LCW algorithm.

- **HLT_j420_a10t_lcw_jes_40smcINF_L1J100**

This trigger requires trimmed large-radius calorimeter jets (**a10t**) to satisfy $E_T^J > 420$ GeV and $m^J > 40$ GeV at the HLT, and $E_T^J > 100$ GeV at the L1Topo trigger. In addition to the kinematic requirements, the trimmed large-radius calorimeter jets must be reconstructed using the LCW algorithm with subsequent jet energy scale (JES) calibration.

- **HLT_j420_a10t_lcw_jes_35smcINF_L1J100**

Similar to the third trigger, this trigger requires trimmed large-radius calorimeter jets to satisfy $E_T^J > 420$ GeV and $m^J > 35$ GeV at the HLT, and $E_T^J > 100$ GeV at the L1Topo trigger. The jets must also be reconstructed using the LCW algorithm with JES calibration.

For the Superboosted analysis, all Leading Superboosted jets are required to pass one of the following offline unprescaled High Level Triggers:

- **HLT_j380**

This trigger requires the transverse energy of small-radius calorimeter jets to satisfy $E_T^J > 380$ GeV at the HLT.

- **HLT_j420**

This trigger requires the transverse energy of small-radius calorimeter jets to satisfy $E_T^J > 420$ GeV at the HLT.

Table 5.2 and Figures 5.2 and 5.3 summarize the triggers used for specific data-taking years and their corresponding trigger efficiencies.

5.4.2 Kinematic Selection on Calorimeter Jets

The second selection criterion for identifying Higgs boson pair candidates involves applying kinematic requirements to all calorimeter jets in the Boosted and Super-

Year	Boosted Jet Trigger	Superboosted Jet Trigger
2015	HLT_j360_a10_lcw_sub_L1J100	HLT_j380
2016	HLT_j420_a10_lcw_L1J100	HLT_j380
2017	HLT_j420_a10t_lcw_jes_40smcINF_L1J100	HLT_j420
2018	HLT_j420_a10t_lcw_jes_35smcINF_L1J100	HLT_j420

Table 5.2: Summary of the offline triggers used in the Boosted and Superboosted analyses during the 2015–2018 ATLAS Run-2 data-taking period.

boosted analyses. Five different kinematic selections are applied to these calorimeter jets: mass m^J , transverse momentum p_T^J , pseudorapidity η^J , separation between two jets $|\Delta\eta_{J_1J_2}|$, and the mass-to-transverse momentum ratio m^J/p_T^J .

5.4.2.1 Selection on Mass, Transverse Momentum, and Pseudorapidity

In the Boosted analysis, the two Boosted jets in each event are required to satisfy $p_T^{J^{1,2}} > 250$ GeV, $m^{J^{1,2}} > 50$ GeV, and $|\eta^{J^{1,2}}| < 2.0$. Additionally, the leading Boosted jet must fulfill $p_T^{J^1} > 450$ GeV to ensure operation at very high trigger efficiency (i.e., at least 99%).

In the Superboosted analysis, the two Superboosted jets in each event must satisfy $p_T^{J_s^{1,2}} > 250$ GeV, with their corresponding masses and pseudorapidities required to fulfill $m^{J_s^{1,2}} > 50$ GeV and $|\eta^{J_s^{1,2}}| < 2.0$, respectively. Furthermore, the leading Superboosted jet must have $p_T^{J_s^1} > 450$ GeV to maintain an operational trigger efficiency of at least 99%.

5.4.2.2 Selection on $|\Delta\eta_{J_1J_2}|$

The next selection applied to the calorimeter jets concerns $|\Delta\eta_{J_1J_2}|$. This variable is defined as the absolute difference between the pseudorapidities of the leading and subleading jets:

$$|\Delta\eta_{J_1J_2}^i| = |\eta^{J_i^1} - \eta^{J_i^2}| \quad (5.1)$$

where i refers to $i = \{s, b\}$. The application of this cut aims to reduce the QCD multijet background while retaining high signal acceptance in both analyses.

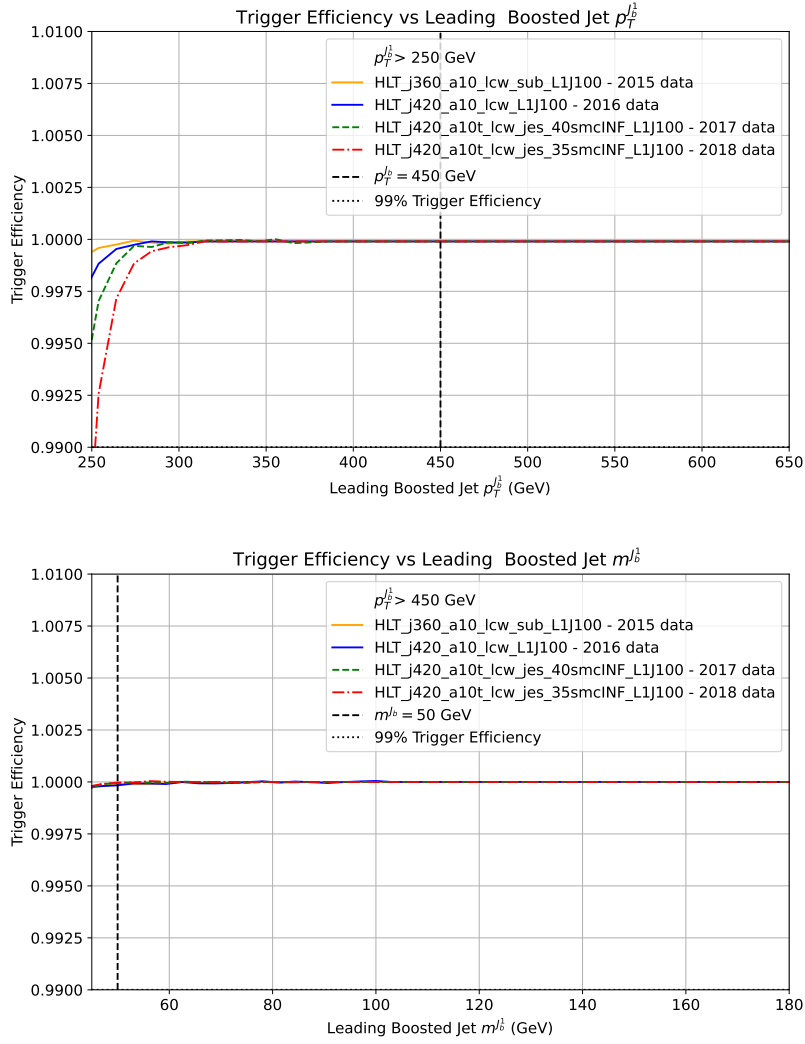


Figure 5.2: Trigger efficiency as a function of offline leading jet p_T (top) and mass (bottom) for the Boosted analysis across four data-taking years.

In the Boosted and Superboosted analyses, the G_{kk}^* decay process is modeled through the s -channel. In this channel, two gluons fuse to form a G_{kk}^* resonance, which subsequently decays into a pair of Higgs bosons traveling back-to-back in the transverse plane. For a sufficiently heavy G_{kk}^* , these Higgs bosons are highly boosted and decay into collimated $b\bar{b}$ pairs. Owing to momentum conservation in the two-body decay of G_{kk}^* , constrained by $m_{G_{kk}^*}$, the resulting jets tend to have a symmetric distribution in pseudorapidity and are concentrated around the central region of the ATLAS detector.

In contrast, QCD multijet events are predominantly produced via t -channel and

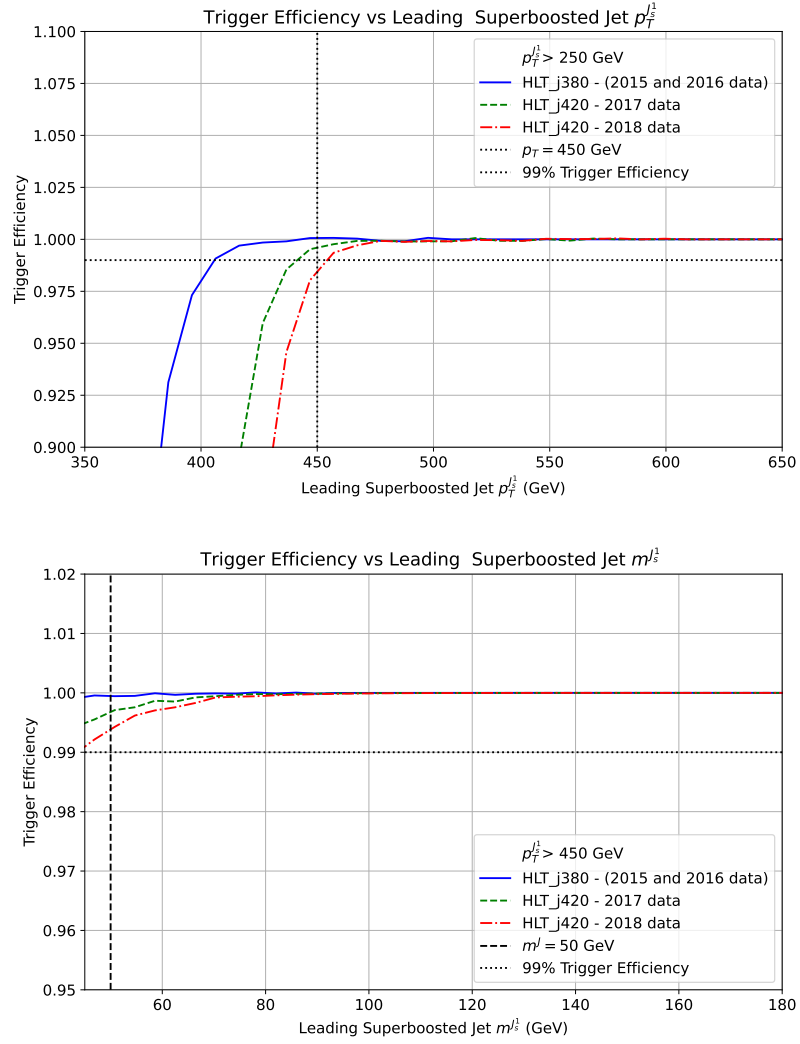


Figure 5.3: Trigger efficiency as a function of offline leading jet p_T (top) and mass (bottom) for the Superboosted analysis across three data-taking years.

u -channel processes, which involve scattering. These channels yield a broader range of momentum transfers among the incoming particles, resulting in a wider distribution of scattering angles. Consequently, QCD multijet events tend to produce jets that are more forward in the ATLAS detector compared to those originating from heavy G_{kk}^* decays. This difference leads to a broader $|\Delta\eta_{J_1J_2}^i|$ distribution for QCD multijet events, as illustrated in Figure 5.4.

To suppress the QCD multijet background, the Boosted and Superboosted analyses optimise the selection cut on $|\Delta\eta_{J_1J_2}^i|$ based on the regions of the normalised $|\Delta\eta_{J_1J_2}^i|$ distributions where a higher signal-to-background ratio is observed, as

shown in Figure 5.4. For both analyses, the selection requirement is set to $|\Delta\eta_{J_1 J_2}^i| < 1.3$.

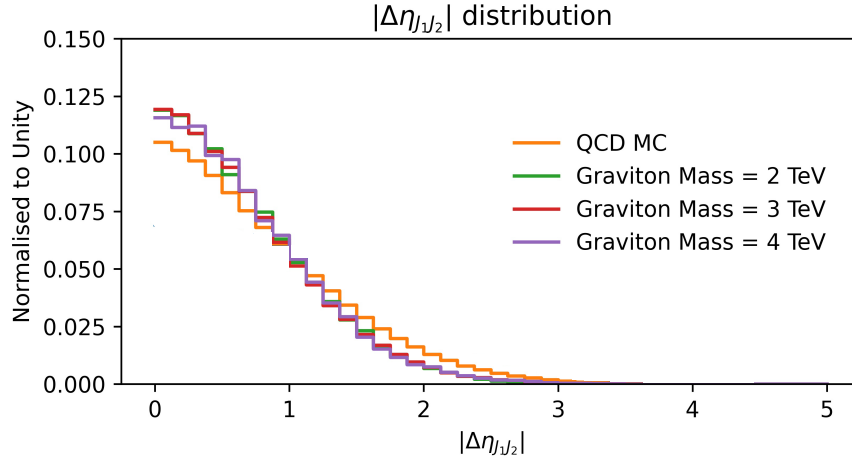


Figure 5.4: Normalised $|\Delta\eta_{J_1 J_2}^b|$ distribution for QCD multijet and three different G_{kk}^* Monte Carlo samples before the $|\Delta\eta_{J_1 J_2}^b|$ cut is applied to calorimeter jets in the Boosted analysis.

5.4.2.3 $m^{J_s}/p_T^{J_s}$ Cut (Superboosted Jets Only)

The third kinematic selection on calorimeter jets involves the ratio of jet mass to transverse momentum. This selection is applied exclusively to Superboosted jets. Although the Superboosted jet reconstruction is designed to capture extremely collimated b -hadrons from Higgs boson pair decays, optimisation is necessary to retain high signal acceptance while suppressing background events containing out-of-cone radiation. These out-of-cone radiations originate from particles produced at the interaction point that travel beyond the designated cone size employed by jet algorithms. Such particles may arise from underlying events or from heavy particle decays, such as Higgs boson decays into $b\bar{b}$.

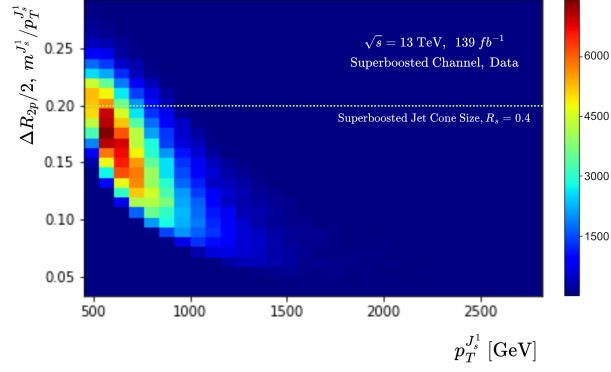
Ideally, jet algorithms are expected to capture the full set of decay products within the jet cone. However, when significant out-of-cone radiation occurs, jet energy is underestimated, leading to degraded mass reconstruction performance. To mitigate this effect, a selection is applied based on the angular separation between the two decay products from a single particle decay, ΔR_{2p} , approximated as follows:

$$\Delta R_{2p} \approx \frac{2m^P}{p_T^P} \quad (5.2)$$

where m^P and p_T^P denote the mass and transverse momentum of the parent particle undergoing a two-body decay.

Using Eq. 5.2, Superboosted jets are selected by requiring $m^{J_s}/p_T^{J_s} < 0.2$. This condition constrains ΔR_{2p} to lie within the spatial extent of the Superboosted jets, thereby suppressing events affected by significant out-of-cone radiation while maintaining high signal acceptance for very heavy resonant mass decays, as illustrated in Figure 5.5⁶.

⁶The two-dimensional histograms in Figure 5.5 serve as a reference for optimising the selection cut on $m^{J_s}/p_T^{J_s}$, which reduces the contribution from events with significant out-of-cone radiation (i.e., $m^{J_s}/p_T^{J_s} > 0.2$) while preserving high signal efficiency for large resonant mass decays. The labels "Signal Region" and "2b" refer to events that satisfy the $HH \rightarrow b\bar{b}b\bar{b}$ signal selection criteria, with a specific configuration of b -tagged trackjets in each Superboosted jet. A detailed discussion of these criteria is provided in Sections 5.4.5 and 5.4.6.



(a) Data

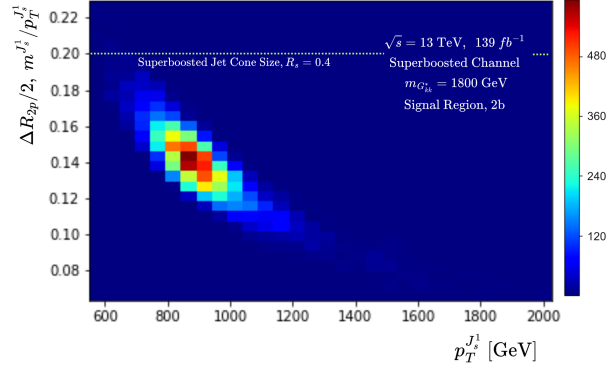
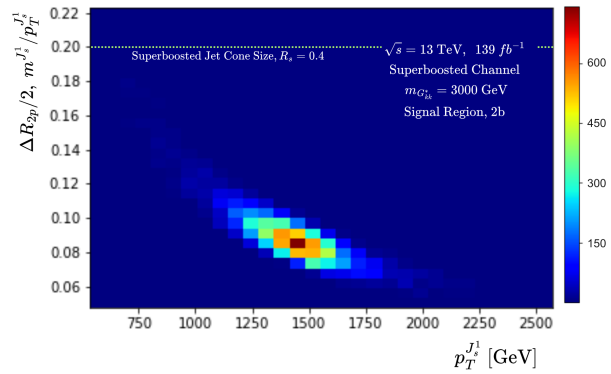
(b) Signal $m_{G_{kk}^*} = 1800$ GeV MC Sample with 2b category(c) Signal $m_{G_{kk}^*} = 3000$ GeV MC Sample with 2b category

Figure 5.5: Two-dimensional histograms of $\Delta R_{2p}/2$ versus $p_T^{J_1^1}$ in Data and Signal $m_{G_{kk}^*}$ MC Samples.

5.4.3 Collinear (Overlapping) Track Jet Veto

The third selection criterion for identifying Higgs boson pair candidates in the Boosted and Superboosted analyses is the veto on collinear (overlapping) track jets. In both analyses, track jets are reconstructed using the anti- k_t algorithm with a variable radius parameter (R), where the radius size is inversely proportional to the transverse momentum $p_T^{j_{s,b}}$ of the jet. Despite the differing sizes, soft (large R) and hard (small R) track jets can overlap when reconstructed within the same Boosted (or Superboosted) jet.

Such overlaps may result in high- p_T tracks being incorrectly assigned to softer track jets, which degrades the performance of flavour-tagging algorithms. To address this, any configuration where track jet axes pathologically overlap is vetoed, as illustrated in Fig. 5.6. The selection is implemented by requiring the angular separation between any two track jet axes to be larger than the minimum radius of the two matched track jets:

$$\Delta R(j_l^i, j_l^k) \geq \min(R_i^{j_l}, R_k^{j_l}), \quad i \neq k, \quad l = \{s, b\} \quad (5.3)$$

where $\Delta R(j_l^i, j_l^k)$ represents the angular separation in the η - ϕ plane between the two track jets, and $R_i^{j_l}$ and $R_k^{j_l}$ are the respective radii of the leading and subleading track jets contained within each Boosted or Superboosted jet.

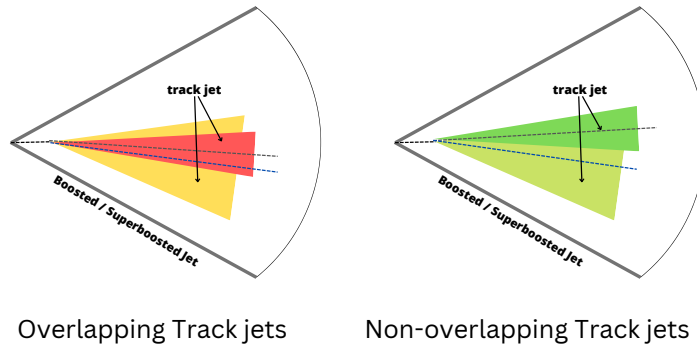


Figure 5.6: Configuration of overlapping and non-overlapping track jets inside Boosted (or Superboosted) jets. Overlapping track jets are defined when the collimated (hard) track jets intercept the axes of softer track jets.

5.4.4 Resolved Analysis Veto

The next selection criterion for identifying Higgs boson pair candidates in the Boosted and Superboosted analyses is the veto on overlapping events between the Boosted (or Superboosted) and Resolved analyses. The purpose of this veto is to enforce orthogonality between the Boosted (or Superboosted) analysis and the Resolved analysis.

The selection criteria for the Resolved analysis veto involve identifying events containing Higgs boson pair candidates reconstructed from four Resolved jets. These Resolved jets are defined as calibrated small-radius jets originating from the interaction point, each required to have a transverse momentum $p_T^{J^{res}} > 40$ GeV and pseudorapidity $|\eta^{J^{res}}| < 2.5$.

For events to satisfy the Resolved analysis veto in the Boosted (or Superboosted) analysis, two additional requirements must be fulfilled. First, at least three of the four Resolved jets must be b -tagged. Second, the events must fall within the Signal Region defined for the $HH \rightarrow b\bar{b}b\bar{b}$ decay channel in the Resolved analysis. This Signal Region is specified as an elliptical region in the plane of the two reconstructed Higgs boson candidate masses, requiring $X_{HH}^{res} < 1.6$, where X_{HH}^{res} is given by:

$$X_{HH}^{res} = \sqrt{\left(\frac{m(H_1^{res}) - 120 \text{ GeV}}{0.1 m(H_1^{res})}\right)^2 + \left(\frac{m(H_2^{res}) - 110 \text{ GeV}}{0.1 m(H_2^{res})}\right)^2} \quad (5.4)$$

Here, $m(H_1^{res})$ and $m(H_2^{res})$ denote the masses of the leading and subleading Higgs boson candidates, respectively, reconstructed from the combined four-momenta of the four Resolved jets. The values 120 GeV and 110 GeV correspond to the peaks of the two-dimensional mass distribution of Higgs boson candidate pairs obtained from Monte Carlo simulations, and are used to define the centre of the Signal Region. The requirement $X_{HH}^{res} < 1.6$ defines an elliptical region optimised to maximise the sensitivity of the Resolved analysis.

5.4.5 b -tagged Track Jet Categorisation

The fifth selection criterion for identifying Higgs boson pair candidates in both the Boosted and Superboosted analyses involves categorising events based on the number of b -tagged track jets within the Boosted (or Superboosted) jets. This categorisation aims to select events containing two highly energetic, b -tagged jets formed from merged collimated track jets.

The selection process splits events into two categories: high-tag and low-tag. The high-tag category includes events where each Boosted or Superboosted jet contains one or two b -tagged track jets. These events are used to search for Higgs boson pairs decaying into two $b\bar{b}$ pairs. In contrast, the low-tag category includes events where each Boosted (or Superboosted) jet contains one or two track jets, which may be either b -tagged or untagged. Events in the low-tag category are primarily used for background modelling.

5.4.5.1 High-tag Category

For the Boosted analysis, events in the high-tag category are divided into three subcategories:

- **2b**: Events where each Boosted jet contains only one track jet, and the track jet must be b -tagged.
- **3b**: Events where one of the two Boosted jets contains two b -tagged track jets, while the other Boosted jet contains a single b -tagged track jet.
- **4b**: Events where both Boosted jets contain two b -tagged track jets.

For the Superboosted analysis, high-tag events are split into two subcategories:

- **2b**: Events where each of the two Superboosted jets contains only one b -tagged track jet, similar to the 2b category in the Boosted analysis.
- **3b+**: Events where one of the two Superboosted jets contains either one or two b -tagged track jets, while the other Superboosted jet contains two b -tagged track jets.

When comparing the high-tag category between the two analyses, the Superboosted analysis exhibits a very limited number of events in the 4b category if the same categorisation scheme as the Boosted analysis is applied. This sparsity leads to poor background modelling and larger uncertainties in the expected limits. To address this, the 3b and 4b categories are combined into a single 3b+ category for the Superboosted analysis, thereby preserving signal acceptance and improving background modelling, ultimately enhancing the sensitivity of the analysis.

5.4.5.2 Low-tag Category

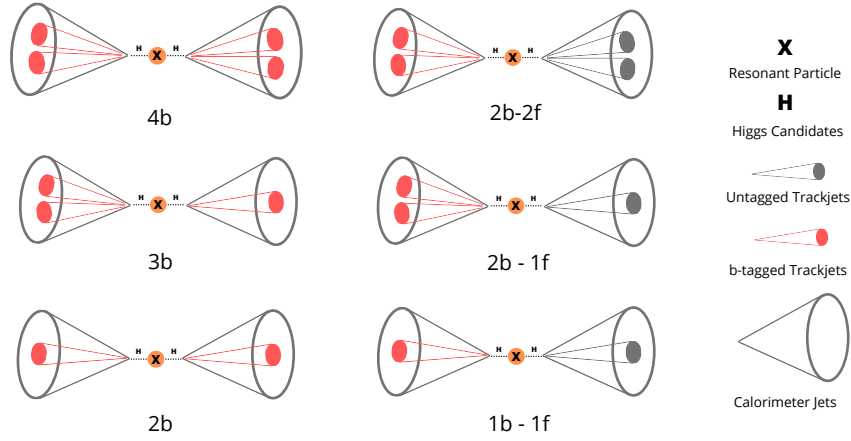
For the Boosted analysis, events in the low-tag category are separated into three subcategories:

- **1b-1f**: Events where one of the two Boosted jets contains only one b -tagged track jet, and the other Boosted jet contains only one untagged track jet.
- **2b-1f**: Events where one of the two Boosted jets contains two b -tagged track jets, and the other Boosted jet contains only one untagged track jet.
- **2b-2f**: Events where one Boosted jet contains two b -tagged track jets, and the other Boosted jet contains two untagged track jets.

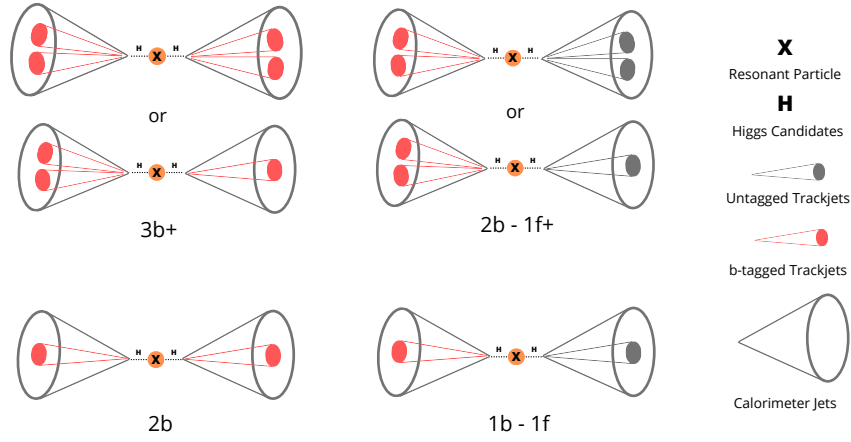
For the Superboosted analysis, events in the low-tag category are divided into two subcategories:

- **1b-1f**: Events where one of the two Superboosted jets contains only one b -tagged track jet, and the other Superboosted jet contains only one untagged track jet.
- **2b-1f+**: Events where one of the two Superboosted jets contains either one or two untagged track jets, and the other Superboosted jet contains two b -tagged track jets.

Figure 5.7 illustrates the topology of the high-tag and low-tag categories used in the Boosted and Superboosted analyses.



(a) High-tag (left) and low-tag (right) categories for the Boosted analysis.



(b) High-tag (left) and low-tag (right) categories for the Superboosted analysis.

Figure 5.7: Topology of the high-tag and low-tag categories for b -tagged track jet categorisation in the Boosted and Superboosted analyses.

5.4.6 Signal Selection for $HH \rightarrow b\bar{b}b\bar{b}$

Building upon the event selection criteria outlined in Sections 5.4.1 to 5.4.5 for identifying Higgs boson pair candidates decaying into two $b\bar{b}$ pairs, this section (5.4.6) focuses on further refining the selection within the high-tag categories. The aim is to isolate events that specifically correspond to the $HH \rightarrow b\bar{b}b\bar{b}$ signal process.

The signal selection employs an elliptical region within the two-dimensional mass plane of the reconstructed Higgs boson candidates. These masses are derived

from the combined four-momenta of the Boosted (or Superboosted) jets, their associated track jets, and any matched muons. This elliptical region, denoted as X_{HH}^i and representing the signal for the $HH \rightarrow b\bar{b}b\bar{b}$ process, defines the signal region (SR) and is expressed by the following equation:

$$X_{HH}^i = \sqrt{\left(\frac{m(H_1^i) - 124 \text{ GeV}}{0.1m(H_1^i)}\right)^2 + \left(\frac{m(H_2^i) - 115 \text{ GeV}}{0.1m(H_2^i)}\right)^2}, \quad i \in \{s, b\} \quad (5.5)$$

Here, i refers to events from either the Boosted (b) or Superboosted (s) analysis. The variables $m(H_1^i)$ and $m(H_2^i)$ denote the masses of the leading and sub-leading Higgs boson candidates, respectively, reconstructed from the combined four-momenta. The values 124 GeV and 115 GeV correspond to the two peaks observed in the two-dimensional distribution of the reconstructed Higgs boson candidate masses, based on Monte Carlo simulations.⁷ These peak values serve as the centres for defining the elliptical signal region.

Events in the high-tag categories are considered as signal candidates only if they satisfy $X_{HH}^i < 1.6$. This requirement is optimised to maximise the signal acceptance while enhancing the overall sensitivity of the analysis.

Following the signal selection, the four-momenta of the leading and sub-leading Higgs boson candidates are rescaled such that their masses are set to $m(H_1^i) = m(H_2^i) = 125.35 \text{ GeV}$. These rescaled four-momenta are subsequently used to compute the invariant mass of the combined Higgs boson candidate pair, denoted as $m(HH)_i$ (where $i \in \{s, b\}$). The $m(HH)_i$ distribution serves as the final discriminant for deriving the upper limits on the cross-section of resonant Higgs boson pair production decaying into two $b\bar{b}$ pairs. Details regarding the upper limit calculation are presented in Section 5.7. It is important to note that the data within the SR remains blinded until the expected limits, including systematic uncertainties, are established.

⁷The subleading jet J_2 has lower energy and momentum compared to the leading jet J_1 , causing the peak of the subleading Higgs boson candidate mass to shift to a lower value relative to 124 GeV.

5.4.7 Selection for Control and Validation Regions

In addition to the signal region, the Boosted and Superboosted analyses utilise the two reconstructed Higgs boson candidate masses to define regions for background modelling, as discussed in Section 5.5. Two elliptical regions are introduced: the control region (CR) and the validation region (VR).

The control region provides data for background modelling and is located in the mass plane where the background dominates the signal, while maintaining proximity to the Higgs boson mass of 125 GeV.

The validation region verifies the background model derived from the control region. Positioned between the SR and CR, it is designed to resemble the kinematics of the SR while minimising signal contamination.

The control and validation regions are defined as:

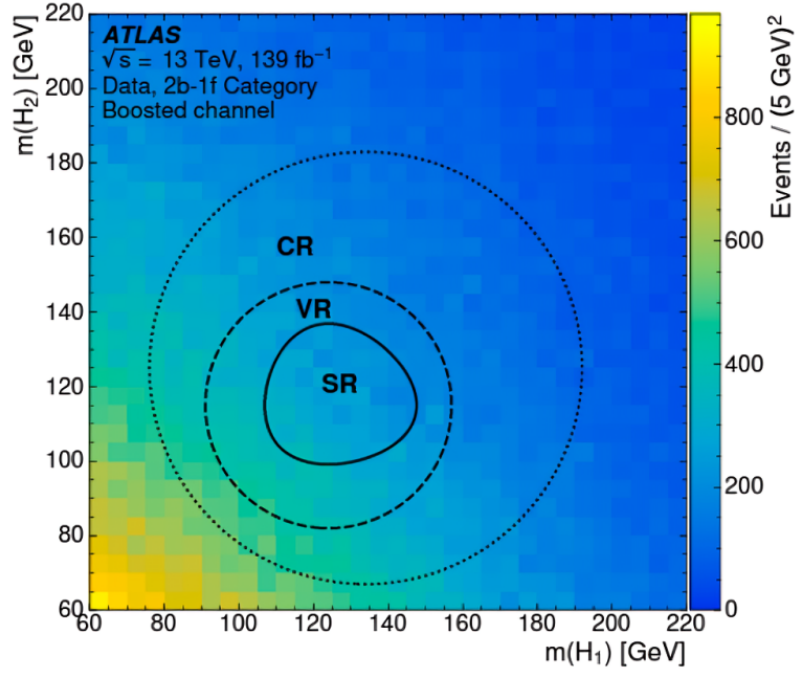
$$R_{HH_i}^{VR} = \sqrt{(m(H_1^i) - 124 \text{ GeV})^2 + (m(H_2^i) - 115 \text{ GeV})^2}, \quad i \in \{s, b\} \quad (5.6)$$

$$R_{HH_i}^{CR} = \sqrt{(m(H_1^i) - 134 \text{ GeV})^2 + (m(H_2^i) - 125 \text{ GeV})^2}, \quad i \in \{s, b\} \quad (5.7)$$

where i denotes the Boosted (b) or Superboosted (s) analyses. Events are selected in the VR if $X_{HH}^i > 1.6$ and $R_{HH_i}^{VR} < 33 \text{ GeV}$. Events are selected in the CR if $R_{HH_i}^{VR} > 33 \text{ GeV}$ and $R_{HH_i}^{CR} < 58 \text{ GeV}$.

The values 124 GeV and 115 GeV correspond to the two mass peaks observed in Monte Carlo simulations of the Higgs boson candidates. The parameters 134 GeV, 125 GeV, 33 GeV, and 58 GeV were optimised to ensure sufficient background statistics, minimal signal contamination, and kinematics closely matching the SR.

Table 5.3 and Figure 5.8 summarise the definitions of the three elliptical regions, along with the mass requirements for Higgs boson candidate pairs.



(a) Data with 2b-1f category.

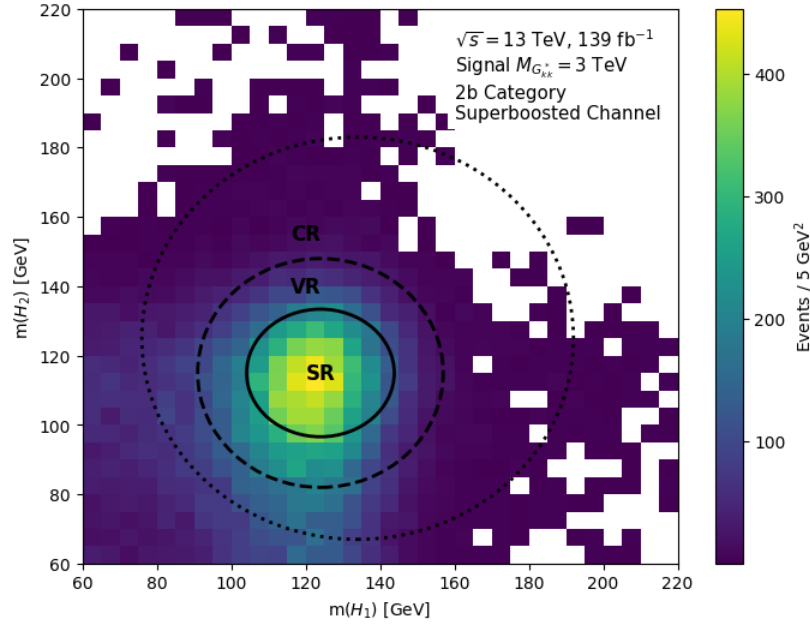
(b) Signal $M_{G_{kk}^*} = 3 \text{ TeV}$ sample with 2b category.

Figure 5.8: Contours in the $m(H_1)$ – $m(H_2)$ plane representing the Control, Validation, and Signal regions for the Boosted and Superboosted analyses [12].

Signal Region (SR)	Validation Region (VR)	Control Region (CR)
$X_{HH}^i < 1.6$	$X_{HH}^i > 1.6$ $R_{HH_i}^{VR} < 33 \text{ GeV}$	$R_{HH_i}^{VR} > 33 \text{ GeV}$ $R_{HH_i}^{CR} < 58 \text{ GeV}$

Table 5.3: Definitions of the Signal, Validation, and Control regions based on the reconstructed Higgs boson candidate masses.

5.5 Background Estimation

Following the event selection procedures outlined in Section 5.4, this section describes the background estimation strategy for the resonant Higgs boson pair production decaying into two $b\bar{b}$ pairs in the Boosted and Superboosted analyses.

The primary background processes, which mimic the $HH \rightarrow b\bar{b}b\bar{b}$ final state, are QCD multijet production and top quark pair production ($t\bar{t}$). QCD multijet events, constituting approximately 85% of the total background, arise from parton scattering processes that hadronise into jets. These jets can be misidentified as b -jets by the ATLAS detector, resulting in a final state that resembles the signal.

Top quark pair production accounts for about 14% of the total background. In $t\bar{t}$ events, each top quark decays into a bW final state. The W bosons may decay either hadronically or leptonically, with separate Monte Carlo (MC) samples used to model these decay modes. In the all-hadronic mode, both W bosons decay into light quarks, representing the dominant $t\bar{t}$ decay mode and contributing approximately 46% of the total $t\bar{t}$ background. Non-all-hadronic modes involve at least one W boson decaying leptonically, further classified into semi-leptonic and dileptonic categories. The semi-leptonic mode, where one W boson decays into leptons, constitutes about 45%, while the dileptonic mode, where both W bosons decay into leptons, represents approximately 9% of the total $t\bar{t}$ background. These $t\bar{t}$ decay modes, upon hadronisation, can produce a mixture of real and misidentified b -jets, thereby mimicking the $HH \rightarrow b\bar{b}b\bar{b}$ final state.

To account for all $t\bar{t}$ decay modes, the kinematic distributions of each all-hadronic and non-all-hadronic MC sample are rescaled to their respective theoretical cross-sections. The rescaled distributions are then summed to obtain the overall kinematics of the $t\bar{t}$ background contribution.

While ZZ production, in which a pair of Z bosons decay into two $b\bar{b}$ pairs, can also mimic the signal, its contribution is negligible (less than 1% of the total background) due to the small cross-section. Thus, it is excluded from the background estimation in the Boosted and Superboosted analyses.

The background estimation for the two dominant processes, QCD multijet and $t\bar{t}$ production, proceeds in three steps: kinematic reweighting, background modelling, and an $m(HH)$ smoothing fit.

5.5.1 Kinematic Reweighting

The first step in background estimation for the Boosted and Superboosted analyses involves kinematic reweighting of Higgs boson candidate pairs. This technique aims to align the kinematic distributions of Higgs boson candidates between the low-tag and high-tag categories.

While the low-tag category is utilised for background estimation, its kinematic distributions must resemble those in the high-tag category. However, the low-tag category contains a mixture of untagged and b -tagged Higgs boson candidates. Untagged Higgs boson candidates lack any b -tagged track jets, whereas b -tagged candidates possess at least one.

Since b -tagged and untagged candidates exhibit different kinematic shapes, these differences can introduce inaccuracies in background estimation. To address this, a reweighting procedure is applied to the untagged Higgs boson candidates in the low-tag category.

The reweighting process employs iterative cubic splines applied to several kinematic variable distributions of the untagged Higgs boson candidates. These distributions are adjusted to match those of the b -tagged candidates. The targeted variables include:

- Transverse momenta of the leading H_1^i and subleading Higgs boson candidates H_2^i containing one or two b -tagged track jets, $p_T^{j_i}(H_1^i)$ and $p_T^{j_i}(H_2^i)$,
- Transverse momenta of the leading b -tagged track jets $p_T^{j_i^1}$ and subleading b -tagged track jets $p_T^{j_i^2}$,
- Pseudorapidity of the b -tagged track jets η^{j_i} ,
- Angular separation between the leading and subleading track jets $\Delta R(j_i^1, j_i^2)$

(for Higgs boson candidates with two b -tagged and untagged track jets).

Here, i denotes whether the kinematics pertain to the Boosted (b) or Super-boosted (s) analysis. For untagged Higgs boson candidates with two track jets, one jet is randomly selected for the reweighting input.

The reweighting procedure begins by selecting all events in the control region. Next, the ratio of the distributions between b -tagged and untagged Higgs boson candidates is calculated for each kinematic variable. These ratios serve as input to the cubic splines to obtain weights as functions of the kinematic variables. The derived weights are subsequently applied to the relevant kinematic variables for each event in the data. Once the reweighted kinematic distributions are obtained, the ratio between the reweighted distribution of the untagged candidates and the original distribution of the b -tagged candidates is again used as input to the cubic splines. This iterative process produces a new weight (W_i) at each iteration, as defined in Equation (5.8).

$$W_i = W_{i-1} \times [(\Pi_j f_{ij}(x_j) - 1) \times L_i + 1] \quad (5.8)$$

Here, f_{ij} denotes the cubic spline function for the j th kinematic variable at the i th iteration. The variable x_j signifies the kinematic variable under consideration. The learning rate at the i th iteration, denoted by L_i , is defined as $L_i = 1 - 0.5^i$. The learning rate starts small to mitigate numerical instabilities from highly correlated kinematic variables in early iterations and increases gradually to accelerate convergence.

The final weight W_f for the untagged Higgs boson candidates is obtained by multiplying the weights from all iterations, as shown in Equation (5.9).

$$W_f = W_0 \times \Pi_{i=1}^{10} [(\Pi_j f_{ij}(x_j) - 1) \times L_i + 1] \quad (5.9)$$

The initial weight W_0 is set to 1 for data samples. The weights derived from each iteration are then applied to the low-tag category events in the top quark pair

($t\bar{t}$) Monte Carlo simulation samples. This approach ensures that the reweighted low-tag category Monte Carlo events reproduce the same final weight distribution as the low-tag category in data. Although reweighting $t\bar{t}$ events is not essential due to the high accuracy of the Monte Carlo modelling, applying the weights compensates for mismodelling effects in the QCD multijet normalisation scale factors calculation, as shown in Equation 5.10 in Section 5.5.2. In this calculation, the low-tag categories in $t\bar{t}$ and data appear in the same term. However, the reweighting is applied only to the low-tag category $t\bar{t}$ events, while high-tag category $t\bar{t}$ events remain unaffected.

Kinematic reweighting is applied only to the 2b, 3b, and 3b+ tag categories. For the 4b tag category, the impact of mismodelling due to b -tagging is smaller than the statistical uncertainties and can be safely neglected.

Figures 5.9 and 5.10 present the ratios of the kinematic distributions between untagged and b -tagged Higgs boson candidates across the iterations. These figures demonstrate the convergence of the spline fits after the 10th iteration.

5.5.2 Background Modelling

Following the kinematic reweighting of events in the low-tag category for both data and top quark pair $t\bar{t}$ MC samples, the next step in background estimation for the Boosted and Superboosted analyses involves modelling the QCD multijet and $t\bar{t}$ background contributions using normalisation scale factors.

5.5.2.1 Deriving Normalisation Scale Factors

Normalisation scale factors are applied to the estimated backgrounds from QCD multijet and $t\bar{t}$ production. These factors adjust the predicted background events in the high-tag category (Validation region) to match the observed data in this region. Once the background modelling is validated in the Validation region, these factors are used to predict background events in the Signal region. This ensures minimal signal contamination while closely replicating the background kinematics relevant to the $HH \rightarrow b\bar{b}b\bar{b}$ signal in the high-tag category (Signal region).

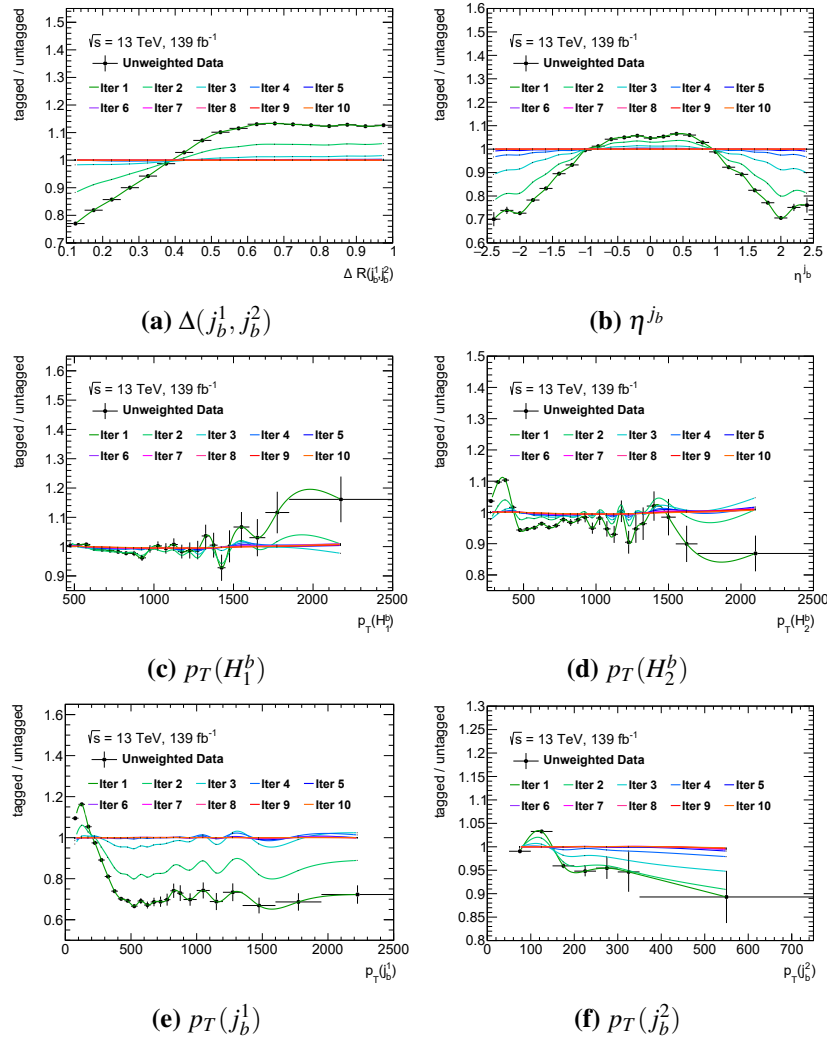


Figure 5.9: Ratio of the relevant kinematic distributions between b -tagged and untagged Higgs boson candidates in data, together with the corresponding splines after 10 iterations of kinematic reweighting in the Boosted analysis.

The derivation of the normalisation factors for the Boosted and Superboosted analyses begins with the leading Higgs boson candidate mass distribution, $m(H_i^1)$, in the data and $t\bar{t}$ MC samples within the Control region (CR). For each bin in the distribution, the expected number of background events in the high-tag category is calculated by summing two terms:

- **QCD multijet term:** Represents the QCD multijet contribution to the total estimated background. This term is obtained by taking the difference between the number of reweighted events in the low-tag category from data and $t\bar{t}$ MC simulation samples.

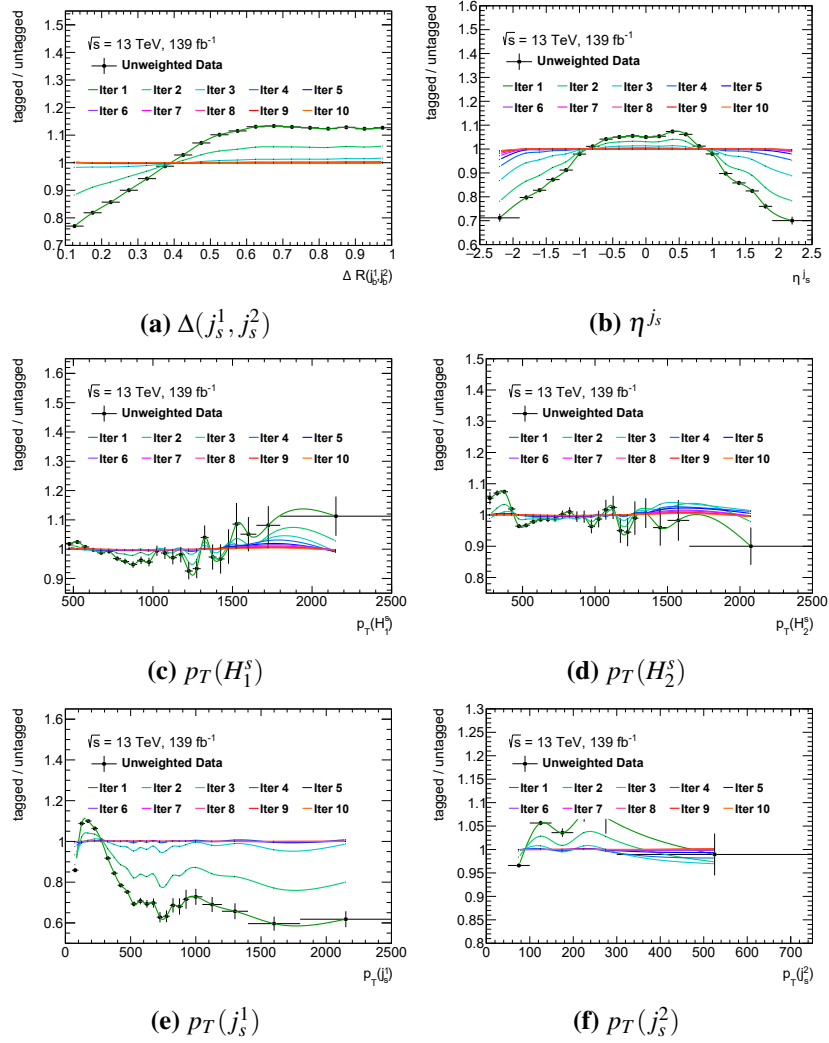


Figure 5.10: Ratio of the relevant kinematic distributions between b -tagged and untagged Higgs boson candidates in data, together with the corresponding splines after 10 iterations of kinematic reweighting in the Superboosted analysis.

- **$t\bar{t}$ term:** Represents the $t\bar{t}$ contribution to the total estimated background. This term is obtained from the number of events in the high-tag category from the $t\bar{t}$ MC simulation samples.

Both terms are scaled by their respective normalisation factors, denoted by μ_{qcd} and $\alpha_{t\bar{t}}$, as shown in Equation 5.10.

$$N_{bkg_k}^{hi} = \mu_{qcd} (N_{data_k}^{lo} - N_{t\bar{t}_k}^{lo}) + \alpha_{t\bar{t}} N_{t\bar{t}_k}^{hi} \quad (5.10)$$

Here,

- $N_{bkg_k}^{hi}$ denotes the expected total background in the k -th bin of the distribution in the high-tag category.
- $N_{data_k}^{lo}$ and $N_{t\bar{t}_k}^{lo}$ denote the total number of events in the k -th bin of the distribution in the low-tag category from data and $t\bar{t}$ MC simulation samples, respectively.
- $N_{t\bar{t}_k}^{hi}$ denotes the total number of events in the k -th bin of the distribution in the high-tag category from the $t\bar{t}$ MC simulation samples.
- μ_{qcd} is the normalisation scale factor for the QCD multijet term, accounting for differences between the low-tag and high-tag categories in $(N_{data_k}^{lo} - N_{t\bar{t}_k}^{lo})$.
- $\alpha_{t\bar{t}}$ is the normalisation scale factor for the $t\bar{t}$ term, used to estimate the number of $t\bar{t}$ events contributing to the total background.

The normalisation factors are extracted using a maximum-likelihood fit to the bins in the $m(H_i^1)$ distribution within the Control region. This approach provides the best mapping between low-tag and high-tag category events for background estimation. Once derived, the normalisation factors are applied to other kinematic variables, including $m(HH)_i$, to estimate the expected total background in each high-tag category of the Signal and Validation regions. The high-tag category in the Validation region is used to validate the background estimation obtained from the Control region.

5.5.2.2 Assumptions and Validation

The two normalisation scale factors are assumed to be applicable across the entire $m(H_1^i) - m(H_2^i)$ plane for both the Boosted (b) and Superboosted (s) analyses. This assumption is validated using a Gaussian Process Interpolator, which independently estimates the background shape in the Signal region.

A Gaussian Process Interpolator is a statistical method that models the data points in the high-tag category $m(H_1^i) - m(H_2^i)$ histogram (after subtracting $t\bar{t}$ contributions) by learning their correlations using a multivariate Gaussian distribution with a predefined covariance (kernel) function $k(x, x')$, where x and x' represent

the coordinates of two points in the $m(H_1^i) - m(H_2^i)$ plane. Once the correlations between the points are established, the Gaussian Process Interpolator predicts the expected value for each bin, producing a smooth background estimate across the entire histogram, including the blinded Signal region. Further details on the validation method are provided in Section 5.6.1.3.

The values of the normalisation scale factors derived for each high-tag category in the Boosted and Superboosted analyses are listed in Table 5.4. The observation $\mu_{qcd} \ll 1$ across all high-tag categories in Table 5.4 indicates that the number of events in the low-tag category is much larger than in the high-tag category.

	High-tag category	Boosted analysis	Superboosted analysis
μ_{qcd}	2b	0.054 ± 0.001	0.047 ± 0.002
	3b/3b+	0.120 ± 0.002	0.093 ± 0.006
	4b	0.027 ± 0.002	-
$\alpha_{t\bar{t}}$	2b	0.863 ± 0.011	0.802 ± 0.074
	3b/3b+	0.917 ± 0.234	0.786 ± 0.042
	4b	1.000 ± 0.000	-
Correlation	2b	-0.74	-0.78
	3b/3b+	-0.74	-0.65
	4b	0	-

Table 5.4: Normalisation scale factors, their corresponding uncertainties, and correlations for each high-tag category in the Boosted and Superboosted analyses.

5.5.3 $m(HH)$ Smoothing Fit

Following the derivation of the normalisation scale factors (Section 5.5.2), the final step in background estimation for QCD multijet and $t\bar{t}$ production in the Boosted and Superboosted analyses involves applying a smoothing fit to the corresponding $m(HH)_i$ distributions. This smoothing procedure aims to mitigate the effects of potential mismodelling arising from statistical fluctuations, particularly in the tails of the $m(HH)_i$ distributions. Such fluctuations can cause an underestimation or over-

estimation of sensitivity limits at high resonant masses, rendering the smoothing step crucial for both analyses.

The smoothing procedure begins by using the estimated QCD multijet and $t\bar{t}$ $m(HH)_i$ distributions obtained in Section 5.5.2 as input to the following nominal multijet (MJ) fit functions:

$$f(x) = \frac{e^{-p_0}}{x^2} (1-x)^{p_1-p_2 \ln x} \quad (\text{Boosted analysis}) \quad (5.11)$$

$$f(x) = p_0 \frac{1-x}{p_1 x} e^{-p_2 x} \quad (\text{Superboosted analysis})$$

Here, $x \equiv m(HH)_i/\sqrt{s}$ with $\sqrt{s} = 13$ TeV, and p_k are dimensionless parameters ($k = \{0, 1, 2\}$). The input distributions are normalised to unity over the chosen fit range, with bin contents treated as data points for the MJ fit functions. The fitting process determines the p_k parameters that best describe the normalised input distributions. After fitting, the distributions and their associated uncertainties are rescaled to match the expected number of events, producing the smoothed $m(HH)_i$ distributions used for background estimation. The smoothing is applied exclusively in the Validation and Signal regions.

The MJ fit function adopted for the Boosted analysis is based on its successful description of QCD multijet and $t\bar{t}$ $m(HH)_b$ distributions from previous studies. In contrast, the Boosted analysis fit function fails to accurately model the tails of the $m(HH)_s$ distribution in the Superboosted analysis, as illustrated in Figure 5.11. Consequently, a different MJ fit function is employed for the Superboosted analysis, optimised for the $m(HH)_s$ distributions of both QCD multijet and $t\bar{t}$ backgrounds.

Table 5.5 summarises the fitting ranges selected for QCD multijet and $t\bar{t}$ production in each high-tag category for the Boosted and Superboosted analyses. Fitting ranges are carefully chosen to avoid potential issues such as fitting failures at lower $m(HH)_i$ values or empty bins at higher $m(HH)_i$ values, ensuring that the

smoothing functions effectively describe the long tails of the distributions. The distributions of total background in the Validation and Signal regions for the Boosted and Superboosted analyses are shown in Figures 5.15–5.19, including $m(H_i^1)$, $m(H_i^2)$, and the smoothed $m(HH)_i$ distributions.

Background Model	Tag Category	Boosted Fit Range	Superboosted Fit Range
QCD	2b	1200–4300 GeV	1600–3600 GeV
	3b/3b+	1200–2800 GeV	1600–2900 GeV
	4b	1200–2800 GeV	-
$t\bar{t}$	2b	1200–3900 GeV	1600–3500 GeV
	3b/3b+	1200–2200 GeV	1600–2900 GeV
	4b	-	-

Table 5.5: Smoothing fit ranges for the QCD multijet and $t\bar{t}$ $m(HH)_i$ distributions in each high-tag category for the Boosted and Superboosted analyses. The smoothing procedure is applied only in the Validation and Signal regions.

5.6 Systematic Uncertainties

Following the background estimation methods described in Section 5.5, this section discusses the systematic uncertainties incorporated into the Boosted and Superboosted analyses. Although these analyses provide methodologies for deriving signal and background expectations for resonant Higgs boson pair production decaying into two $b\bar{b}$ pairs, assessing their corresponding systematic uncertainties is crucial for evaluating the accuracy of the results. These uncertainties are obtained from seven different sources:

- Background normalisation
- Background shape
- $t\bar{t}$ production modeling
- Luminosity
- Theoretical uncertainties
- b -tagging track-jet scale factors
- Jet systematics

5.6.1 Background Normalisation Uncertainty

The first component of the systematic uncertainty for the Boosted and Superboosted analyses is the background normalisation uncertainty. This uncertainty accounts for the uncertainties associated with the normalisation factors obtained from the background estimation method. Three sources contribute to this systematic uncertainty: the statistical uncertainty from the normalisation scale factors, the uncertainty from the choice of control region definitions, and the uncertainty arising from the extrapolation from low-tag to high-tag categories in the control and signal regions. The details for each component of the background normalisation uncertainty are described below.

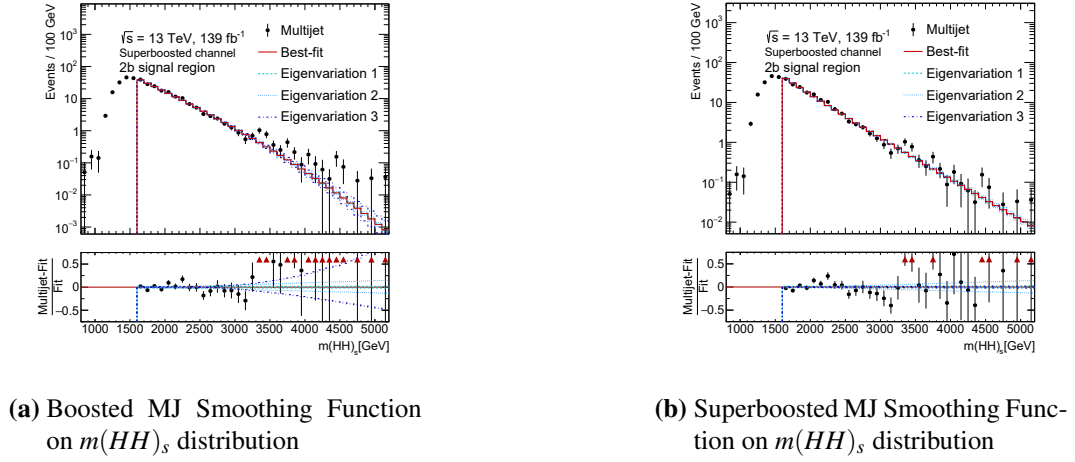


Figure 5.11: Comparison of the multijet $m(HH)_s$ distributions in the 2b split region using the Boosted and Superboosted MJ smoothing functions.

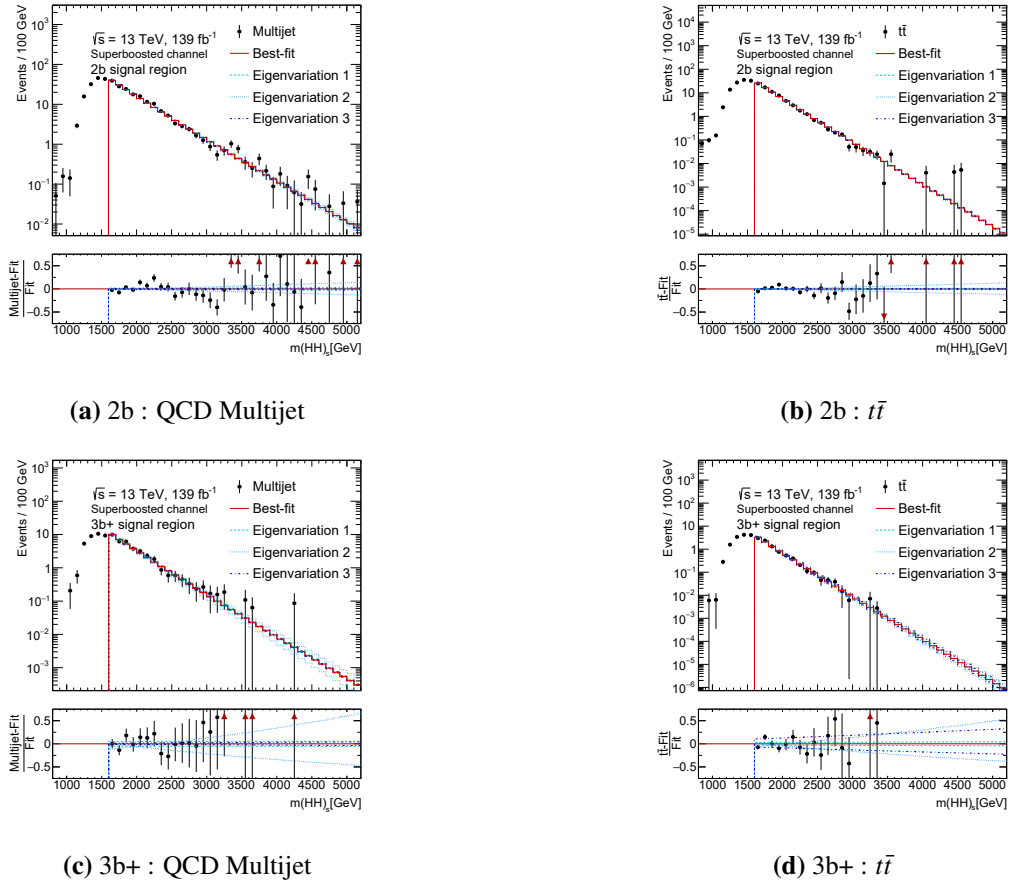


Figure 5.12: The $m(HH)_s$ distributions after applying the smoothing fit to the QCD multijet and $t\bar{t}$ backgrounds in the Signal Region of the Superboosted analysis. The eigenvariation represents the uncertainty of the smoothing, obtained from variations of the p_0 , p_1 , and p_2 parameters.

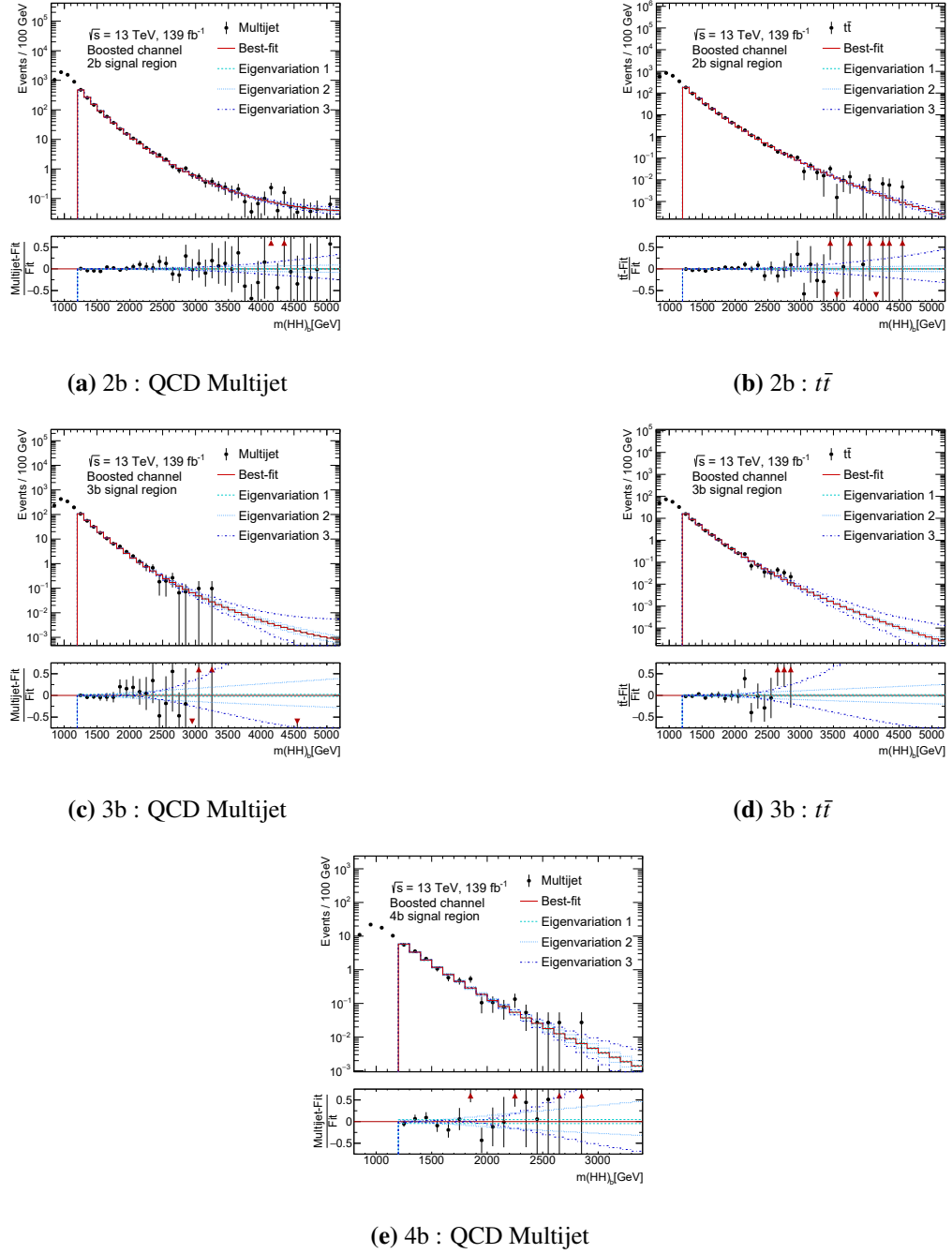


Figure 5.13: The $m(HH)_b$ distributions after applying the smoothing fit to the QCD multijet and $t\bar{t}$ backgrounds in the Signal Region of the Boosted analysis. The eigenvariation represents the uncertainty of the smoothing, derived from variations of the p_0 , p_1 , and p_2 parameters.

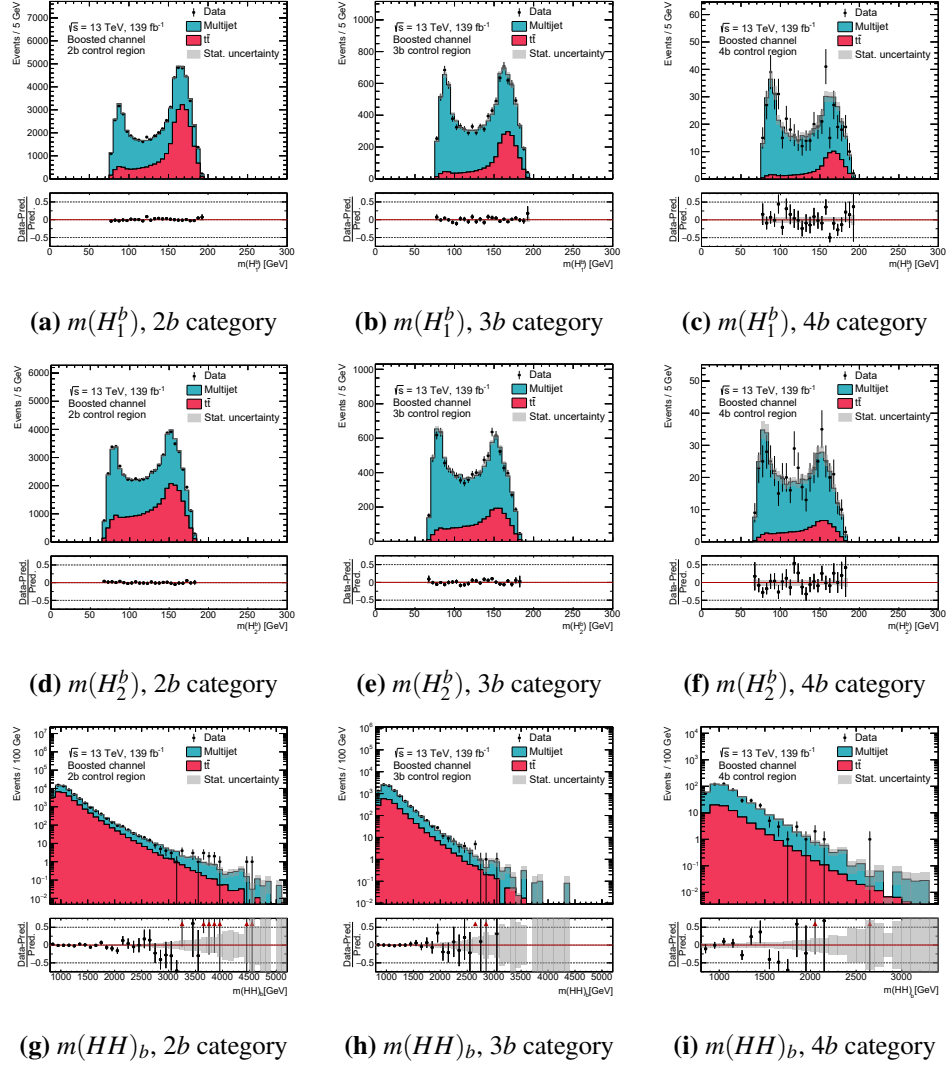


Figure 5.14: The Boosted background kinematic distributions in the Control Region for the 2b, 3b, and 4b categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic distributions are presented in Appendix D.

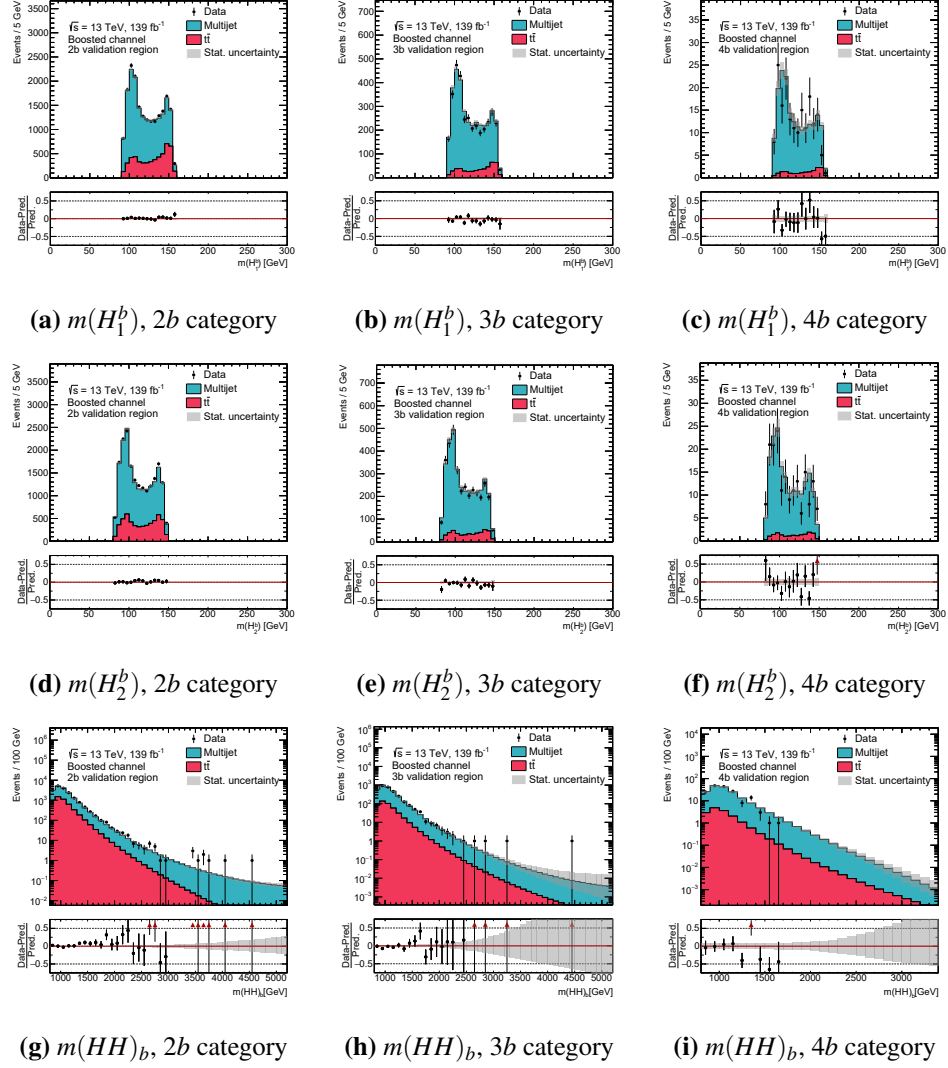


Figure 5.15: The Boosted background kinematic distributions in the Validation Region for the 2b, 3b, and 4b categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic distributions are presented in Appendix D.

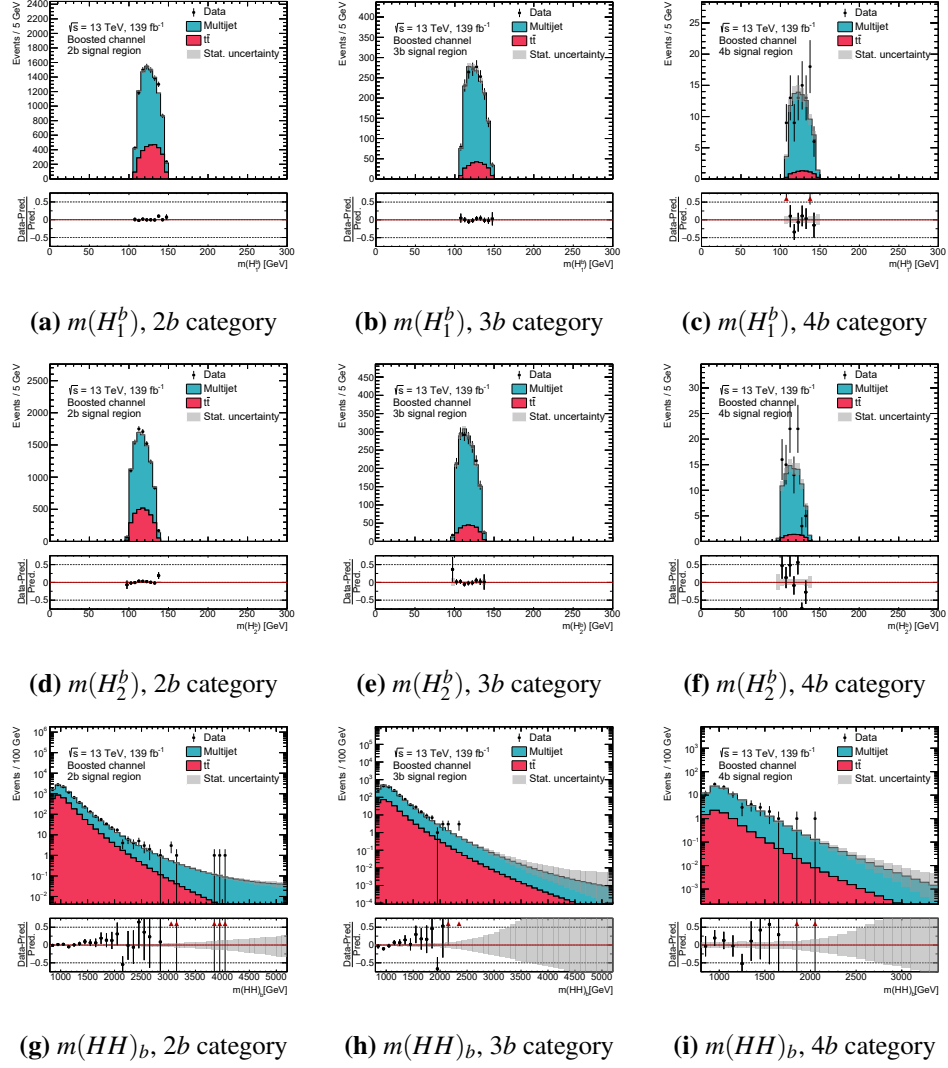


Figure 5.16: Boosted background kinematic distributions in the Signal Region for the 2b, 3b, and 4b categories. Only the leading Higgs $m(H_1^b)$, subleading Higgs $m(H_2^b)$, and $m(HH)_b$ distributions are shown. Additional kinematic variable distributions are provided in Appendix D.

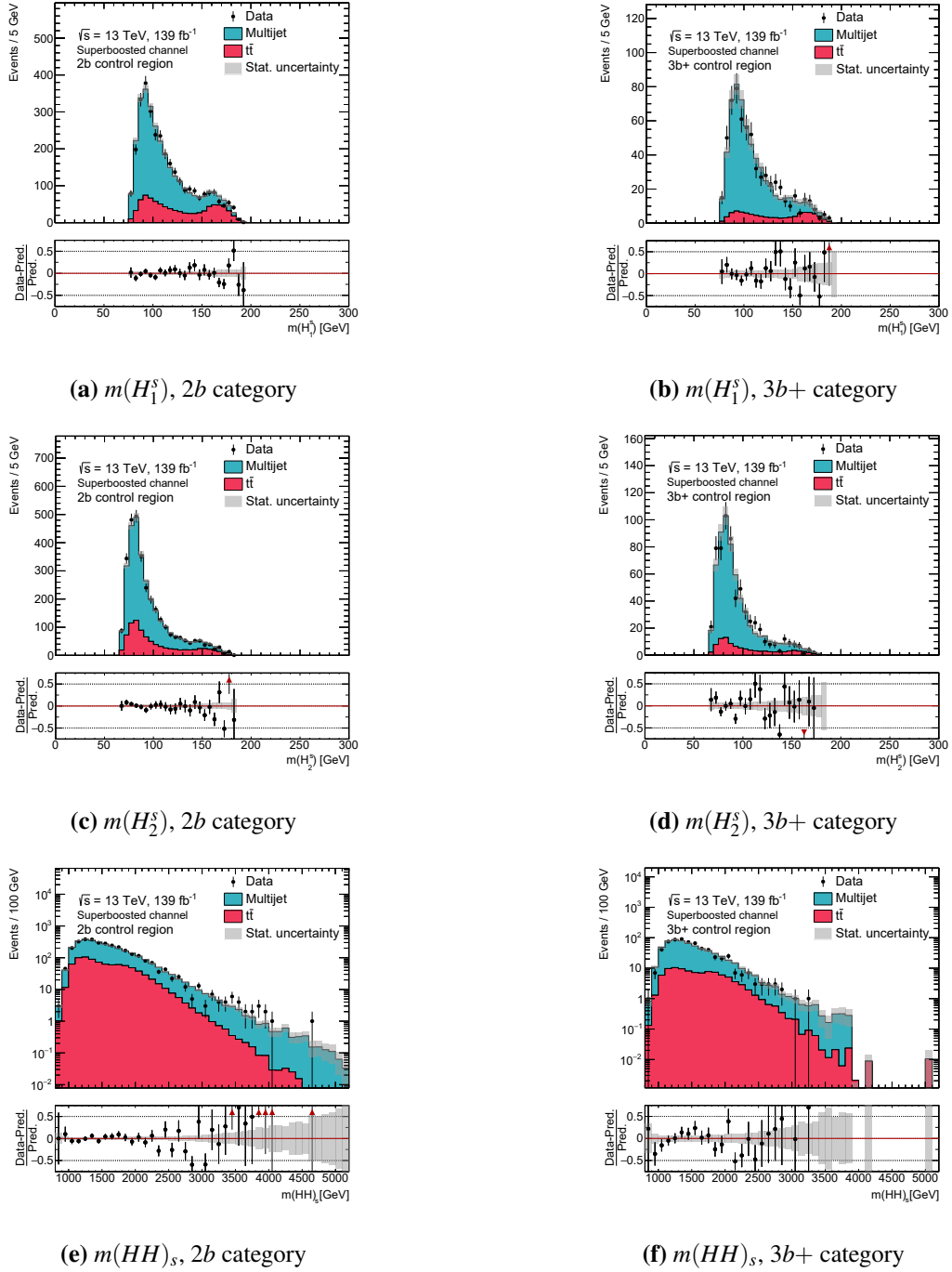


Figure 5.17: Superboosted background kinematic distributions in the Control Region for the $2b$ and $3b+$ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.

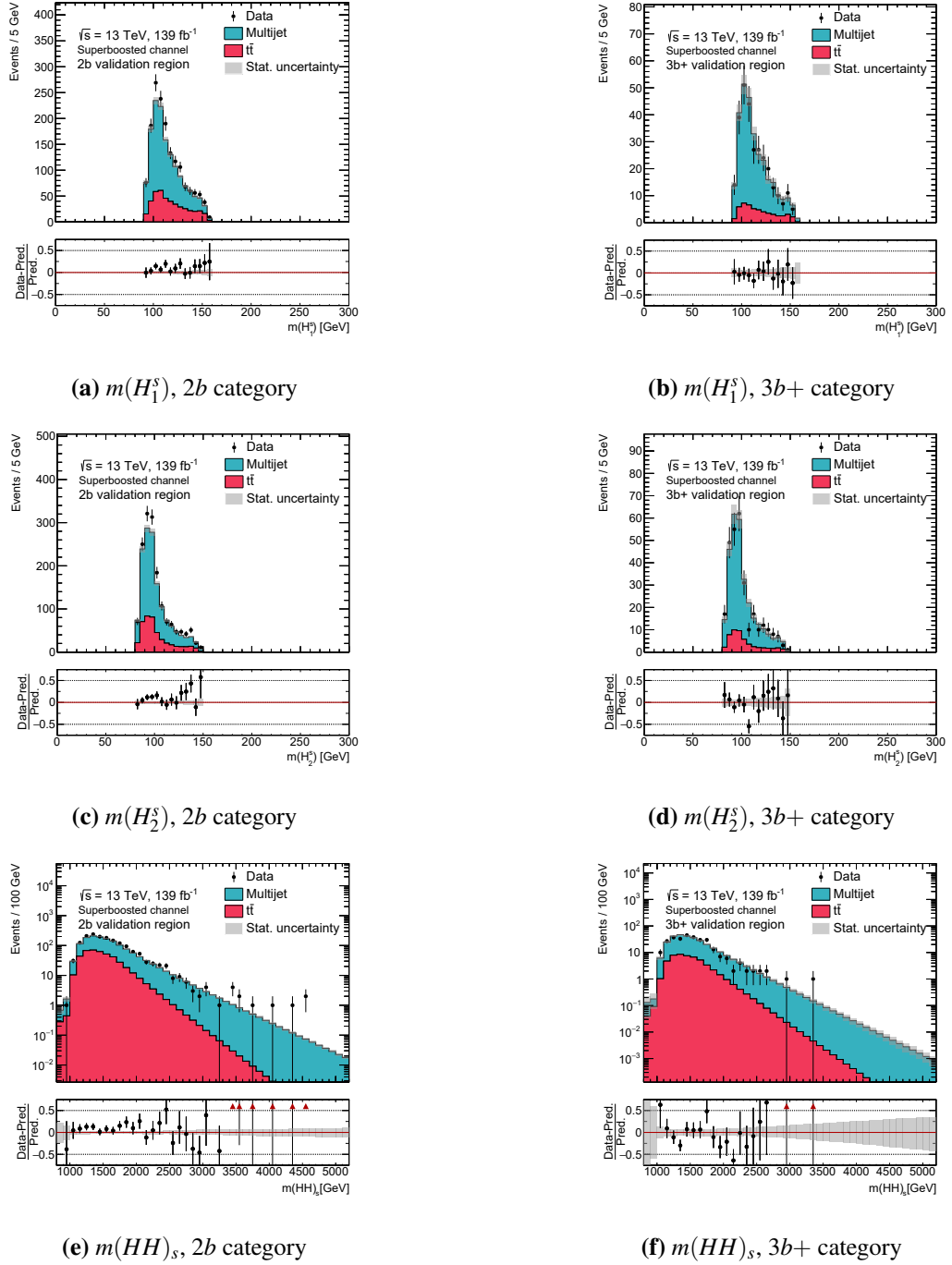


Figure 5.18: Superboosted background kinematic distributions in the Validation Region for the $2b$ and $3b+$ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.

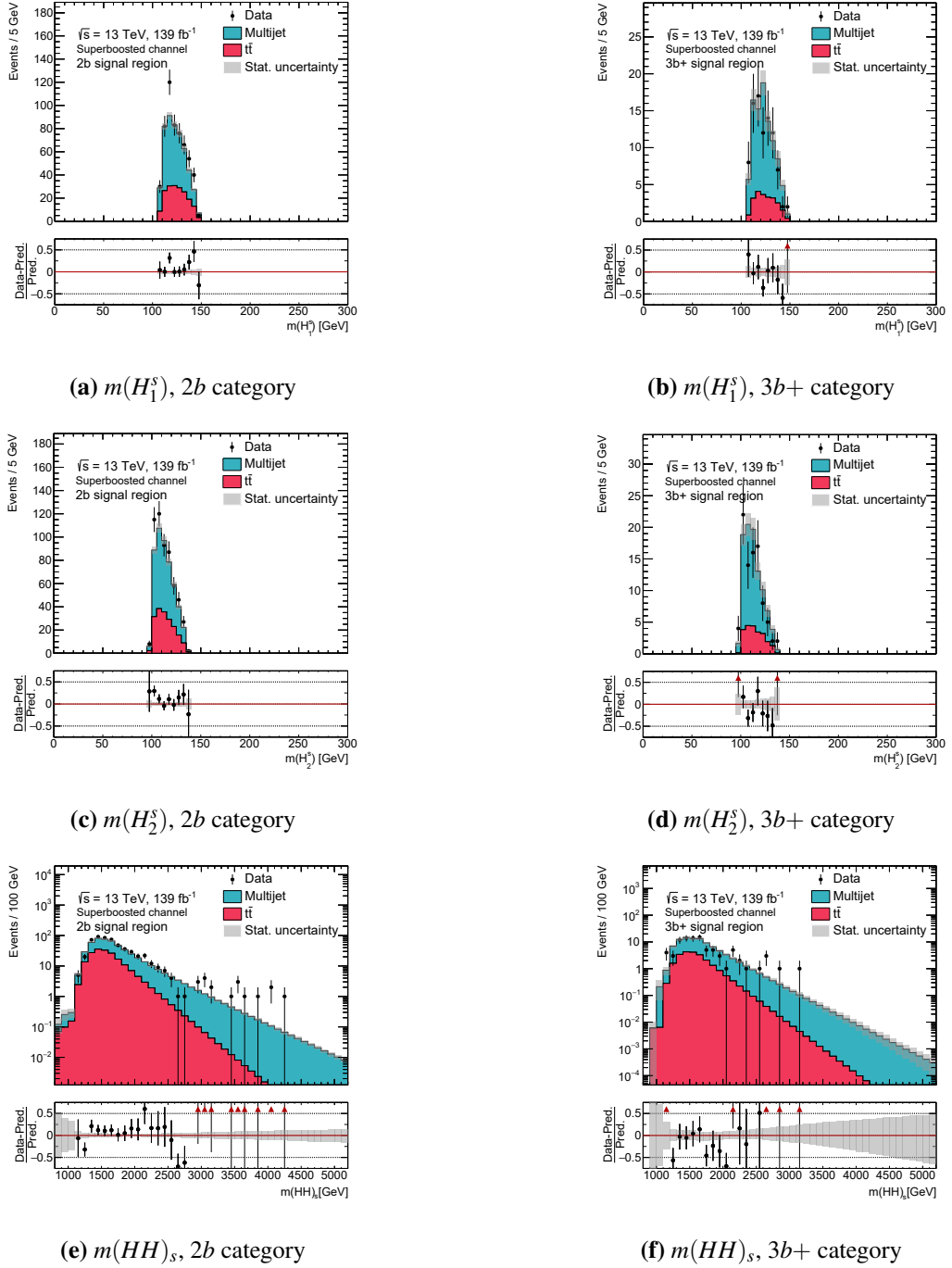


Figure 5.19: Superboosted background kinematic distributions in the Signal Region for the 2b and 3b+ categories. Only the leading Higgs $m(H_1^s)$, subleading Higgs $m(H_2^s)$, and $m(HH)_s$ distributions are shown. Additional kinematic distributions are provided in Appendix D.

5.6.1.1 Statistical Uncertainty from μ_{qcd} and $\alpha_{t\bar{t}}$

This component originates from the uncertainties in the μ_{qcd} and $\alpha_{t\bar{t}}$ normalisation scale factors. These uncertainties are derived from the covariance matrix produced by the maximum-likelihood fit used to obtain these factors, as described in equation (5.10).

5.6.1.2 Control Region Variation Systematics

This uncertainty component arises from the choice of Control Region (CR) used in the Boosted and Superboosted analyses. Background estimation relies on information from the CR in both data and $t\bar{t}$ MC simulation samples. Although the selected CR balances a high background yield with minimal signal contamination, this choice introduces an associated uncertainty. To assess this uncertainty, the impact of varying the CR size and position within the $m(H_1^i) - m(H_2^i)$ plane is evaluated.

The systematic uncertainty from CR variation is obtained by defining six different CR variations, each shifting the CR window by ± 3 GeV in the $m(H_1^i)$ and/or $m(H_2^i)$ directions:

- **Up-Up CR:** Shifts both leading and subleading Higgs candidates by -3 GeV (upward and rightward in the $m(H_1^i) - m(H_2^i)$ plane).
- **Up-Down CR:** Shifts the leading Higgs candidate by -3 GeV and the subleading candidate by $+3$ GeV (upward and leftward).
- **Down-Up CR:** Shifts the leading Higgs candidate by $+3$ GeV and the subleading candidate by -3 GeV (downward and rightward).
- **Down-Down CR:** Shifts both leading and subleading Higgs candidates by $+3$ GeV (downward and leftward).
- **Large CR:** Widens the CR band by 3 GeV in the $m(H_1^i) - m(H_2^i)$ plane by shrinking the $R_{HH_i}^{VR}$ cut and extending the $R_{HH_i}^{CR}$ cut to 3 GeV.
- **Small CR:** Narrows the CR band by 3 GeV in the $m(H_1^i) - m(H_2^i)$ plane by extending the $R_{HH_i}^{VR}$ cut and shrinking the $R_{HH_i}^{CR}$ cut to 3 GeV.

The final uncertainty from this systematic is calculated by taking the largest differences between the total expected background yield in the Signal Region derived from the nominal Control Region and from the six alternative Control Regions listed above. Table 5.6 presents the Control Region variation uncertainties for the Boosted and Superboosted analyses.

Tag Category	Control Region Variation Uncertainty (%)	
	Boosted Analysis	Superboosted Analysis
$2b$	0.88	1.00
$3b/3b+$	1.25	4.3
$4b$	6.05	–

Table 5.6: Control Region variation uncertainties for each high-tag category in the Boosted and Superboosted analyses.

5.6.1.3 Extrapolation Uncertainty

This uncertainty component accounts for potential mismodeling when extrapolating the QCD multijet background estimation across the $m(H_1^i) - m(H_2^i)$ plane. The background estimation methods in the Boosted and Superboosted analyses use events from the low-tag category in the Control Region to predict the expected QCD multijet background in the Signal Region. This procedure assumes that the normalisation scale factors μ_{qcd} remain constant throughout the $m(H_1^i) - m(H_2^i)$ plane. However, the validity of this assumption must be evaluated.

The extrapolation uncertainty is derived using Gaussian processing and events from both the low-tag and high-tag categories in data, with $t\bar{t}$ MC simulation contributions subtracted, across the $m(H_1^i) - m(H_2^i)$ plane (including the blinded Signal Region). The procedure is described as follows:

- Two separate QCD multijet distributions are generated.
- The first distribution is derived from events in the high-tag category, after subtracting $t\bar{t}$ MC simulation samples, and extrapolated using a Gaussian Process Interpolator across the $m(H_1^i) - m(H_2^i)$ plane.

- The second distribution is obtained from low-tag category data, using background normalisation factors as described in Section 5.5.2.
- The two distributions are then used to compute the expected event yields in both the Signal and Control Regions.
- A double ratio, R_{extr} , is calculated between the event yields from the two distributions in the Signal and Control Regions, using Equation (5.12). This double ratio quantifies any discrepancies due to extrapolation.

$$R_{extr} = \frac{N_{SR}^{GP}/N_{SR}^{NSF}}{N_{CR}^{GP}/N_{CR}^{NSF}} \quad (5.12)$$

Here, N_{SR}^{GP} and N_{CR}^{GP} represent the predicted numbers of QCD multijet events in the Signal and Control Regions, respectively, derived from the Gaussian Process Interpolator. N_{SR}^{NSF} and N_{CR}^{NSF} represent the predicted numbers derived using the background normalisation factors described in Section 5.5.2.

The final extrapolation uncertainty is calculated as the absolute deviation of R_{extr} from unity. This procedure is repeated for each high-tag category (ie. $2b$, $3b$, $3b+$ and $4b$). Table 5.7 summarises the extrapolation uncertainties for both the Boosted and Superboosted analyses.

Tag Category	Extrapolation Uncertainty (%)	
	Boosted Analysis	Superboosted Analysis
$2b$	1.72	4.5
$3b/3b+$	6.02	1.3
$4b$	10.94	–

Table 5.7: Extrapolation uncertainties for each high-tag category in the Boosted and Superboosted analyses.

5.6.2 Background Shape Uncertainties

The second component of the systematic uncertainty for the Boosted and Superboosted analyses arises from variations in the shape of the background $m(HH)_i$

distribution. This background shape uncertainty is crucial because $m(HH)_i$ is the primary variable used to determine cross-section limits for resonant Higgs boson pair production decaying into two $b\bar{b}$ pairs. Hence, an accurate modeling of the background distribution shape is essential for predicting the number of background events across the $m(HH)_i$ spectrum. Background shape uncertainties are divided into two components: background smoothing uncertainties and non-closure uncertainties.

5.6.2.1 Background Smoothing Uncertainties

Background smoothing uncertainties originate from variations introduced during the smoothing procedure of the background $m(HH)_i$ distribution (details in Section 5.5.3). Although the selected functional forms and fitting ranges are optimised to describe the data, residual uncertainties persist due to potential variations in these choices. To estimate their impact, the Boosted and Superboosted analyses employ alternative functional forms and fitting ranges to evaluate their effects on the smoothed background $m(HH)_i$ distributions. The procedure for extracting these uncertainties is outlined below:

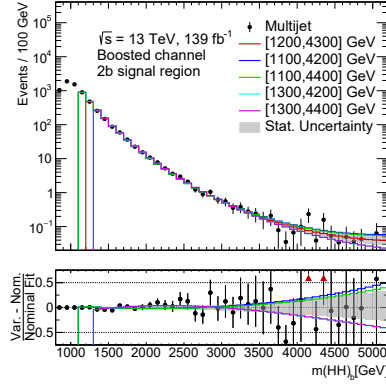
- Uncertainty from Functional Forms:
 1. The smoothing procedure is repeated using seven alternative functional forms, as listed in Table 5.8, while maintaining the same fitting range as used for the nominal form (shown in Table 5.5). This process results in seven alternative smoothed $m(HH)_i$ distributions.
 2. All eight distributions (including the nominal one) are compared, as shown in Figures 5.20–5.23. The functional forms that yield the largest upward and downward deviations from the nominal distribution are identified.
 3. The uncertainties associated with the choice of functional form are determined based on these largest deviations.

- Uncertainty from Fitting Range:

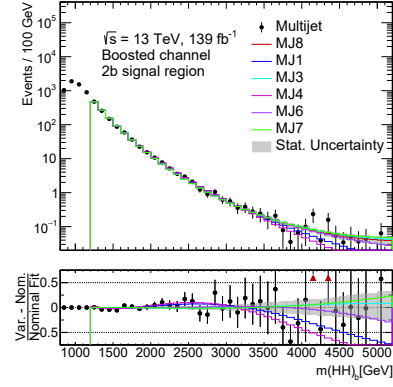
1. The smoothing procedure is repeated using the nominal functional form but with five alternative fitting ranges. This process results in five additional smoothed $m(HH)_i$ distributions.
2. All six distributions (including the nominal one) are compared, as shown in Figures 5.20–5.23. The fitting ranges that induce the largest upward and downward deviations in the $m(HH)_i$ distribution are identified.
3. The uncertainties associated with the fitting range choice are determined based on these largest deviations.

Functional Form	Boosted Analysis	Superboosted Analysis
MJ1	$f(x) = e^{-p_0}(1-x)^{p_1}x^{p_2}$	$f(x) = p_0 \frac{\cosh(p_1x)}{1+x} e^{-p_2x}$
MJ2	$f(x) = e^{-p_0}(1-x)^{p_1} e^{p_2x^2}$	$f(x) = p_0 \frac{\sqrt{1-x}}{p_1x} e^{-p_2x}$
MJ3	$f(x) = e^{-p_0}(1-x)^{p_1} x^{p_2x}$	$f(x) = p_0 \cosh(p_1x) e^{-p_2x}$
MJ4	$f(x) = e^{-p_0}(1-x)^{p_1} e^{p_2 \ln x}$	$f(x) = \frac{p_0\sqrt{1+x}}{p_1(1-x)} e^{-p_2x}$
MJ5	$f(x) = e^{-p_0}(1-x)^{p_1}(1+x)^{p_2x}$	$f(x) = \frac{p_0(1+x)}{p_1\sqrt{1-x}} e^{-p_2x}$
MJ6	$f(x) = e^{-p_0}(1-x)^{p_1}(1+x)^{p_2 \ln x}$	$f(x) = \frac{p_0(1+x)}{p_1(1-x)} e^{-p_2x}$
MJ7	$f(x) = \frac{e^{-p_0}}{x}(1-x)^{p_1-p_2 \ln x}$	$f(x) = p_0 \frac{1-\sqrt{x}}{p_1x} e^{-p_2x}$
MJ8 (Nominal)	$f(x) = \frac{e^{-p_0}}{x^2}(1-x)^{p_1-p_2 \ln x}$	$f(x) = p_0 \frac{1-x}{p_1x} e^{-p_2x}$

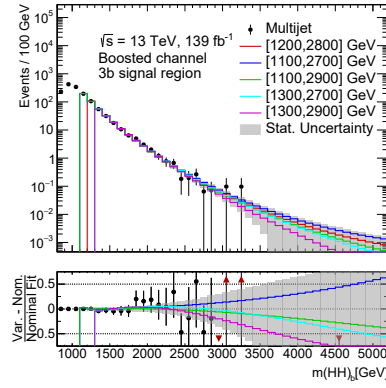
Table 5.8: Functional forms used for assessing background smoothing uncertainties.



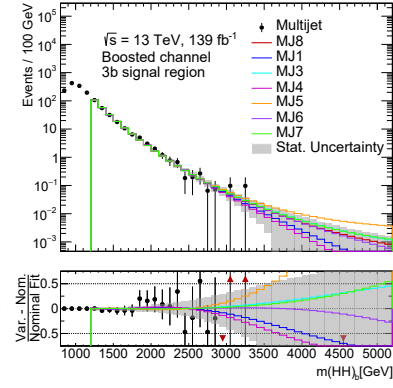
(a) Fitting range variations for 2b



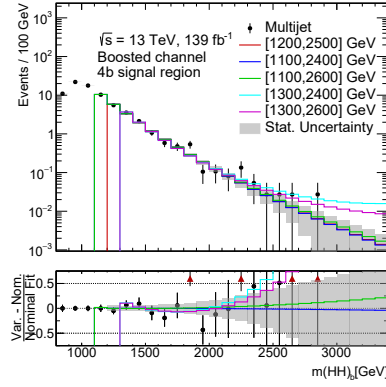
(b) Functional form variations for 2b



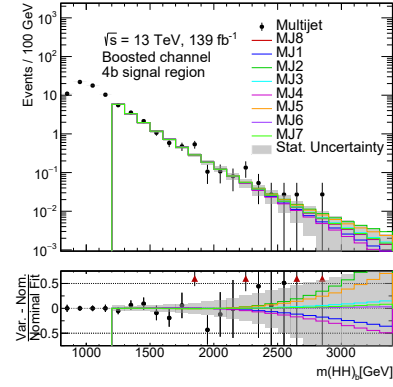
(c) Fitting range variations for 3b



(d) Functional form variations for 3b



(e) Fitting range variations for 4b



(f) Functional form variations for 4b

Figure 5.20: Smoothed QCD multijet $m(HH)_b$ distributions in the Boosted analysis, showing variations due to fitting range and functional form choices.

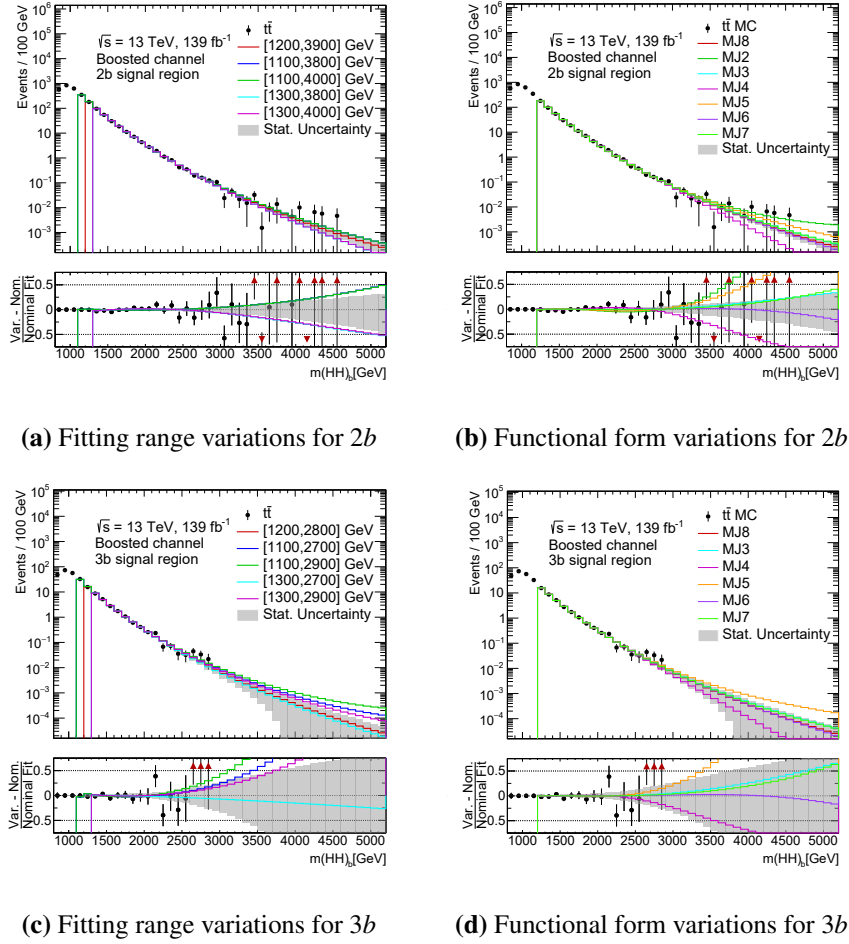


Figure 5.21: Smoothed $t\bar{t} m(HH)_b$ distributions in the Boosted analysis, showing variations due to fitting range and functional form choices.

5.6.2.2 Non-closure Uncertainties

Non-closure uncertainties are evaluated to validate the reliability of data-driven methods used in modeling the QCD multijet background. Although these methods capture the overall behavior of the QCD multijet contribution, the associated uncertainties must be assessed to quantify their accuracy. This is achieved by examining two variations: the comparison between data-driven and MC simulation-driven methods, and the comparison between the Validation and Signal regions.

Data-driven to MC Simulation-driven Comparison. The first variation involves replacing real data with fully simulated Monte Carlo (MC) samples of QCD multijet and $t\bar{t}$ processes. This approach mirrors the procedure outlined in Section 5.5.2 but

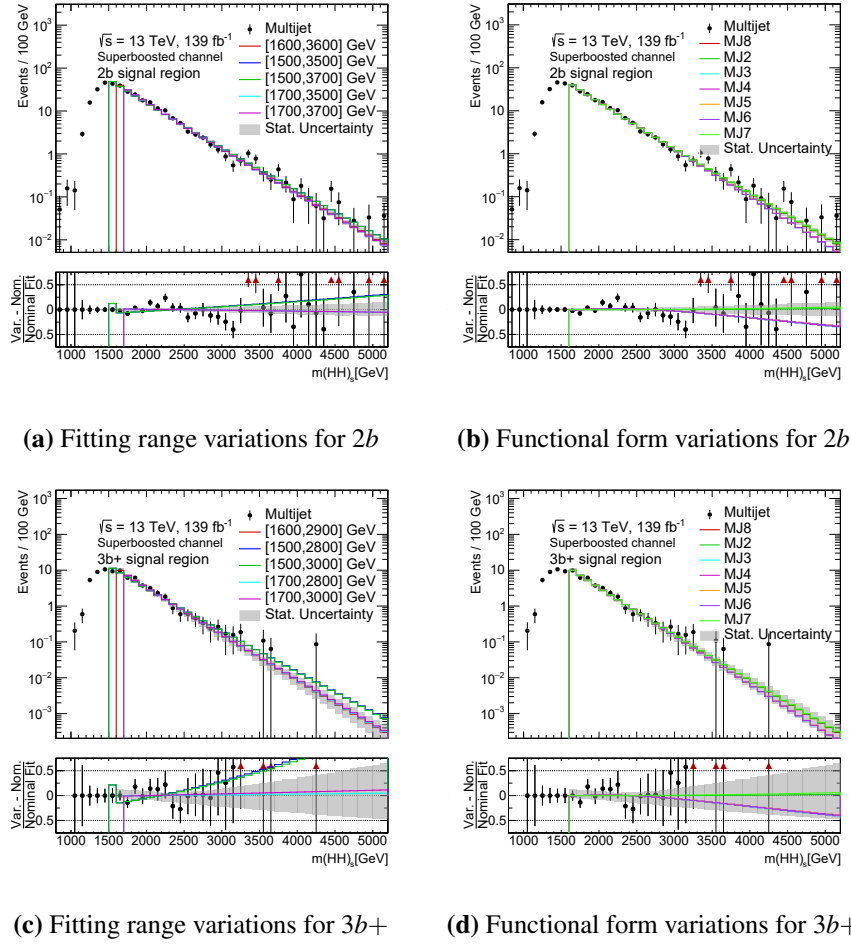


Figure 5.22: Smoothed QCD multijet $m(HH)_s$ distributions in the Superboosted analysis, showing variations due to fitting range and functional form choices.

substitutes the data with MC simulations. Notably, no smoothing is applied to the $m(HH)_i$ distribution in this comparison.

The procedure begins with extracting the kinematic distributions from QCD multijet and $t\bar{t}$ MC samples, which are then used in a reweighting procedure. Following reweighting, the combined reweighted $m(HH)_i$ distribution of QCD multijet and $t\bar{t}$, along with the reweighted $t\bar{t}$ distribution, are used in Equation 5.13 to derive the normalisation scale factors.

$$N_{bkg}^{hi,MC} = \mu_{qcd}^{MC} \left(N_{qcd+t\bar{t}}^{lo} - N_{t\bar{t}}^{lo} \right) + \alpha_{t\bar{t}} N_{t\bar{t}}^{hi} \quad (5.13)$$

Here,

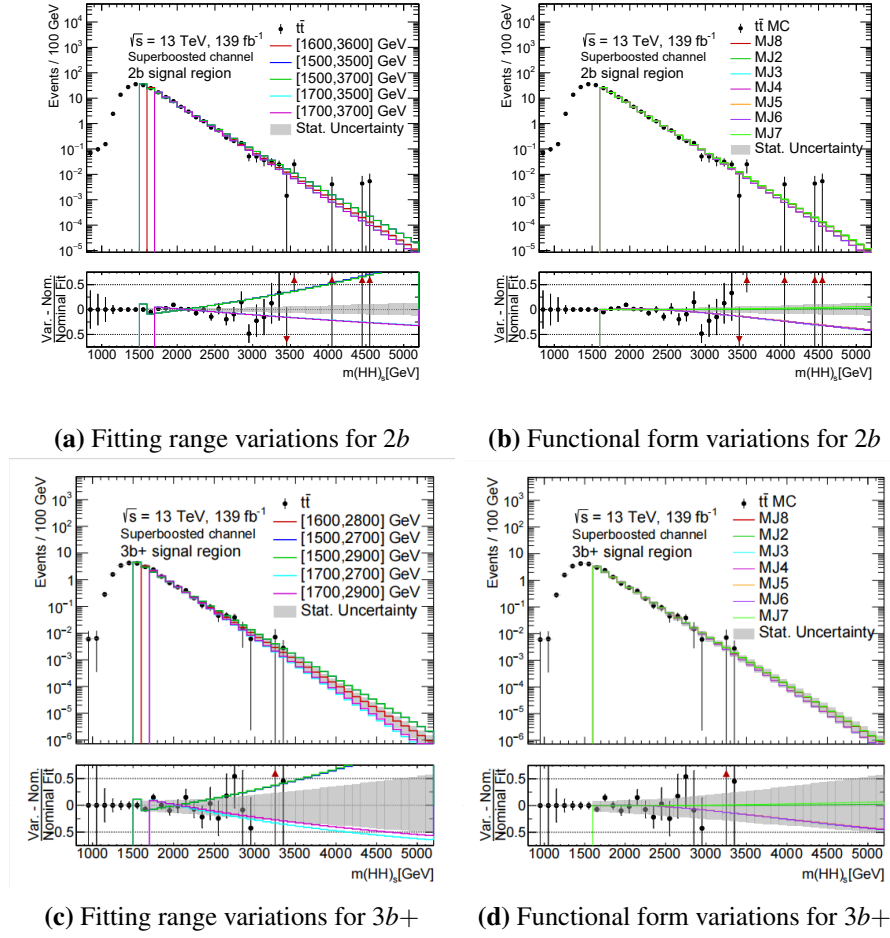


Figure 5.23: Smoothed $t\bar{t} m(HH)_s$ distributions in the Superboosted analysis, showing variations due to fitting range and functional form choices.

- $N_{bkg}^{hi,MC}$ denotes the MC-driven expected total background in the k -th bin of the high-tag category.
- $N_{qcd+t\bar{t}}^{lo}$ and $N_{t\bar{t}}^{lo}$ denote the event yields in the k -th bin of the low-tag category from the combined QCD multijet + $t\bar{t}$ and $t\bar{t}$ MC samples, respectively.
- $N_{t\bar{t}}^{hi}$ denotes the $t\bar{t}$ event yield in the k -th bin of the high-tag category from the MC sample.
- μ_{qcd}^{MC} is the normalisation factor for the QCD multijet component, capturing differences between low-tag and high-tag categories.
- $\alpha_{t\bar{t}}$ is the normalisation factor for the $t\bar{t}$ component.

After determining the normalisation scale factors, the predicted $m(HH)_i$ distribution from QCD multijet MC simulation in the combined Signal and Validation region is compared to the smoothed data-driven QCD multijet distribution in the same region. While the Signal region constrains the total background yield, the combination of Signal and Validation regions enables a more robust assessment of the QCD multijet kinematic shape due to the lower expected signal contamination.

The uncertainty is extracted by evaluating the shape differences between the MC-driven and data-driven QCD multijet $m(HH)_i$ distributions, as shown in Figures 5.24 and 5.25. These differences are smoothed using linear fits. The resulting fit functions are applied to the data-driven QCD multijet distribution to define an uncertainty envelope representing the non-closure uncertainty associated with the MC-driven estimate.

Validation to Signal Region Comparison. The second variation assesses the data-driven QCD multijet background in the Signal region relative to the data- $t\bar{t}$ distribution in the high-tag category of the Validation region. This method involves extracting the shape difference between the corresponding $m(HH)_b$ distributions. These shape differences, accompanied by their respective uncertainty bands, constitute this non-closure uncertainty variation. While the method proves effective in the Boosted analysis, it is unsuitable for the Superboosted analysis.

The differing applicability arises from the peak positions of the $m(HH)_b$ distributions in the Validation and Signal regions for the data-driven QCD multijet in the Boosted analysis. In this case, the peaks align well, as illustrated in Figure 5.26a, enabling a reliable estimation approach.

Conversely, in the Superboosted analysis, the $m(HH)_s$ distribution for the data-driven QCD multijet exhibits a minor shift toward higher $m(HH)_s$ values across the Control, Validation, and Signal regions, as shown in Figure 5.26b. A similar shift appears in the data- $t\bar{t}$ distribution for the high-tag category in the Control and Validation regions (Figure 5.26c). Consequently, the application of this method would yield large uncertainties in the Superboosted case. An alternative strategy is

therefore adopted to estimate the second non-closure systematic uncertainty.

In the Superboosted analysis, the estimation proceeds as follows:

1. The shape difference between the data-driven QCD multijet and the high-tag category of data- $t\bar{t}$ in the Validation region is determined.
2. These $m(HH)_s$ shape differences, along with their uncertainty bands, are then transformed to match the data-driven QCD multijet distribution in the Signal region.
3. The transformed shapes and their associated uncertainties represent the non-closure uncertainty in this variation.

This procedure relies on the assumption that the $m(HH)_s$ distribution shapes of the QCD multijet background derived from the low-tag category resemble those of the high-tag category of data- $t\bar{t}$ in both the Validation and Signal regions. To validate this assumption, the procedure is also applied to the Control region, providing a cross-check between the Validation and Control regions. Moreover, a comparison with the Boosted analysis confirms that this method yields uncertainty sizes similar to those obtained using the default method in that case.

Only the $2b$ category is considered for this variation in both the Boosted and Superboosted analyses, owing to the higher statistics available in the high-tag category of data- $t\bar{t}$. Figures 5.27 and 5.28 present the comparisons of $m(HH)_i$ distribution shapes between the high-tag data- $t\bar{t}$ and the data-driven QCD multijet background in the Boosted and Superboosted analyses, respectively.

5.6.3 Top-quark Pair Production Modelling Uncertainty

The third component of the systematic uncertainty in the Boosted and Superboosted analyses concerns modeling uncertainties in $t\bar{t}$ production. The objective is to assess the accuracy of the $t\bar{t}$ MC simulation samples used for background modeling, based on the specific values of the tuning parameters employed. Five key components contribute to this uncertainty: matrix element calculation, parton shower generators, renormalization and factorization scales, h_{damp} parameter, and final-state radiation

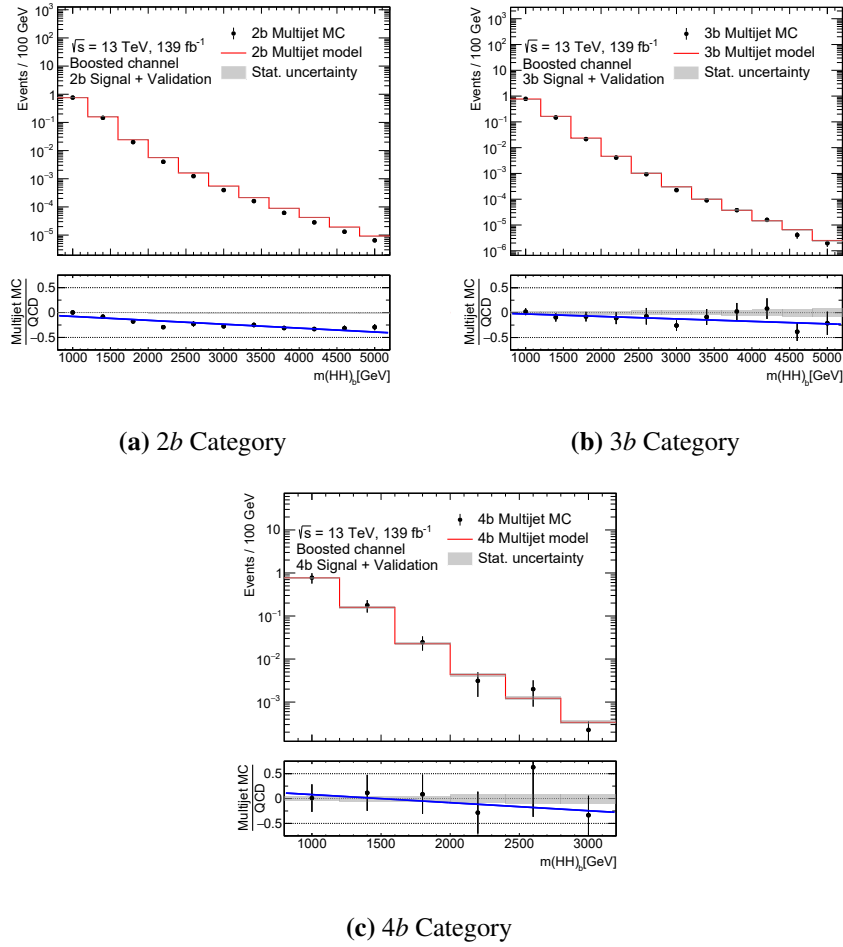


Figure 5.24: Comparison of $m(HH)_b$ distributions between MC-driven and data-driven QCD multijet models in the combined Signal and Validation regions for the Boosted analysis.

(FSR). These components are evaluated using a combination of the nominal $t\bar{t}$ MC sample with varying tuning parameters and several alternative $t\bar{t}$ MC simulation samples, as described in Table 5.1. Comparisons are made against the nominal $t\bar{t}$ MC simulation sample with default tuning parameters in terms of the respective $m(HH)_i$ distributions.

The procedure for deriving these uncertainties is as follows:

1. Each $t\bar{t}$ MC simulation sample undergoes the procedures outlined in Sections 5.4 and 5.5 to derive the corresponding background estimation.
2. The resulting smoothed $t\bar{t}$ $m(HH)_i$ distributions are compared to the nominal

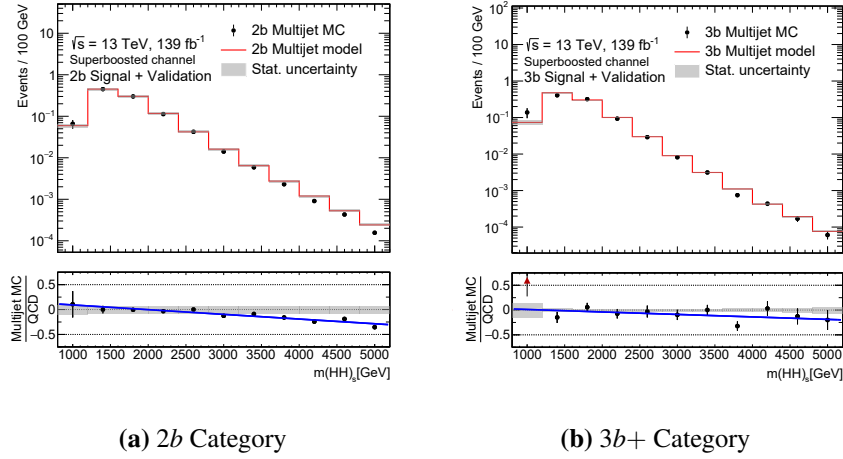
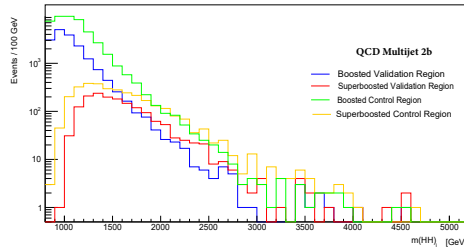
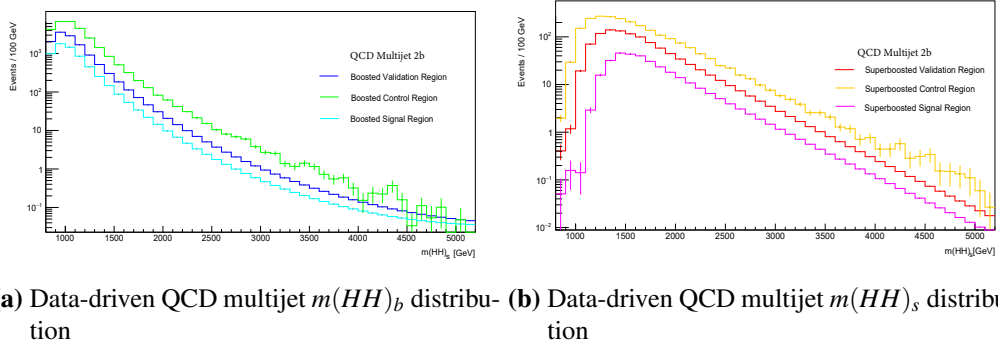


Figure 5.25: Comparison of $m(HH)_s$ distributions between MC-driven and data-driven QCD multijet models in the combined Signal and Validation regions for the Superboosted analysis.



(c) High-tag data- $t\bar{t}$ $m(HH)_i$ distribution

Figure 5.26: Comparison of 2b category $m(HH)_i$ distributions for data- $t\bar{t}$ and data-driven QCD multijet across Control, Validation, and Signal regions in the Boosted and Superboosted analyses.

$t\bar{t}$ $m(HH)_i$ distribution to identify shape differences.

3. These shape differences define the uncertainty envelope (upward and downward variations) for the $t\bar{t}$ modeling uncertainty components, applied to the

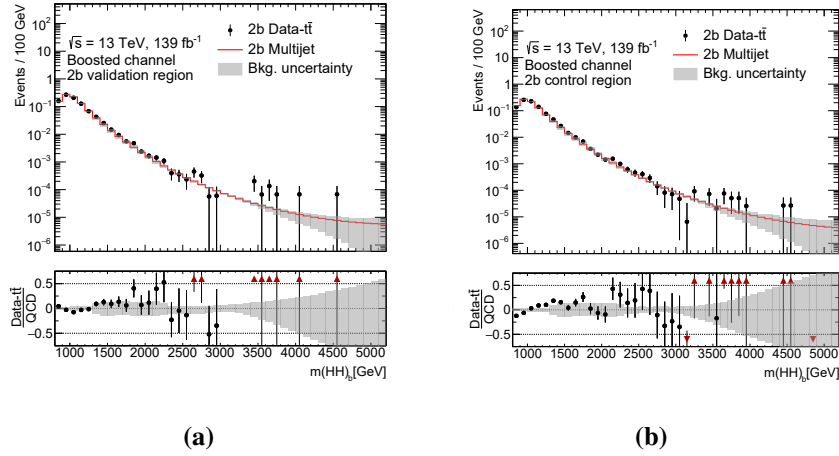


Figure 5.27: (a) Non-closure $m(HH)_b$ distribution comparison between data-driven QCD multijet derived from $1b-1f$ in the Signal region and $2b$ data- $t\bar{t}$ in the Validation region. (b) Non-closure $m(HH)_b$ comparison between QCD multijet from the Validation region and $2b$ data- $t\bar{t}$ in the Control region.

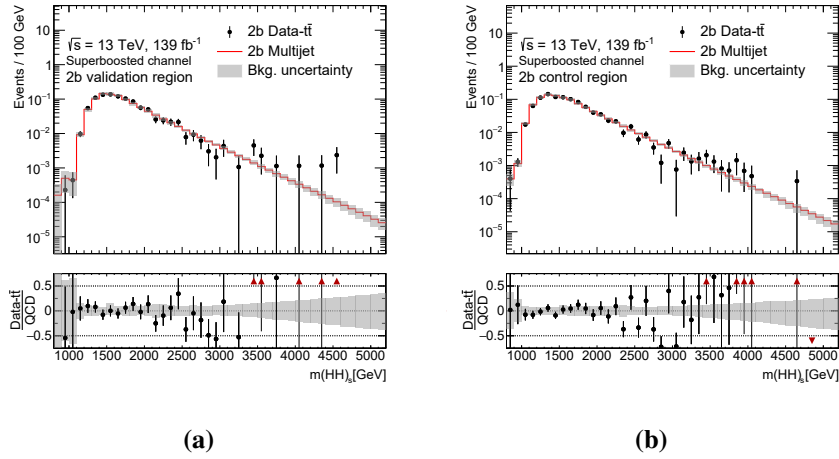


Figure 5.28: (a) Non-closure $m(HH)_s$ comparison between transformed Validation-to-Signal region $2b$ data- $t\bar{t}$ and data-driven QCD multijet derived from $1b-1f$ Signal region. (b) Non-closure $m(HH)_i$ comparison between transformed Control-to-Validation region $2b$ data- $t\bar{t}$ and data-driven QCD multijet from the Validation region.

nominal $t\bar{t}$ $m(HH)_i$ distribution.

- For single-variation comparisons, a symmetrisation method is applied to obtain the uncertainty bands for the nominal $t\bar{t}$ $m(HH)_i$ distribution corresponding to each variation.

All plots in Section 5.6.3 present unsmoothed $t\bar{t}$ distribution comparisons be-

tween the nominal and variant $t\bar{t}$ MC simulation samples. Sections 5.6.3.1 to 5.6.3.5 detail the specific components of the $t\bar{t}$ modeling uncertainty in the Boosted and Superboosted analyses.

5.6.3.1 Matrix Element

The first component of the $t\bar{t}$ production modeling uncertainty arises from differences in matrix element event generators. This uncertainty is assessed by comparing the nominal $t\bar{t}$ MC sample, generated using POWHEG, to a variant sample produced with MADGRAPH5_AMC@NLO (aMC@NLO), as detailed in Table 5.1.

The aMC@NLO sample exhibits a significant imbalance between all-hadronic and non-all-hadronic events, with the latter being considerably less populated (approximately half the number of events compared to the all-hadronic sample). When rescaling and combining the $m(HH)_i$ distributions from these two samples, this imbalance results in large deviations between the nominal $t\bar{t}$ MC sample and the aMC@NLO sample, particularly in the tails of the $m(HH)_i$ distribution (Figures 5.29 and 5.30). Consequently, the primary source of uncertainty in this component is attributed to the limited statistical power of the MC samples.

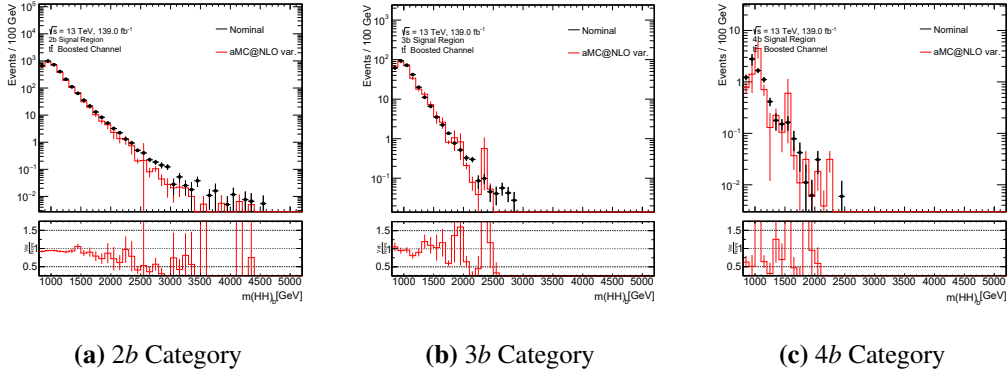


Figure 5.29: Comparison of the $t\bar{t}$ $m(HH)_b$ distributions between two different matrix element generators (Powheg and aMC@NLO) for the 2b, 3b, and 4b categories in the Boosted analysis.

5.6.3.2 Parton Shower

The second component of the $t\bar{t}$ production modeling uncertainty arises from differences in parton shower modeling. This uncertainty is evaluated by comparing the

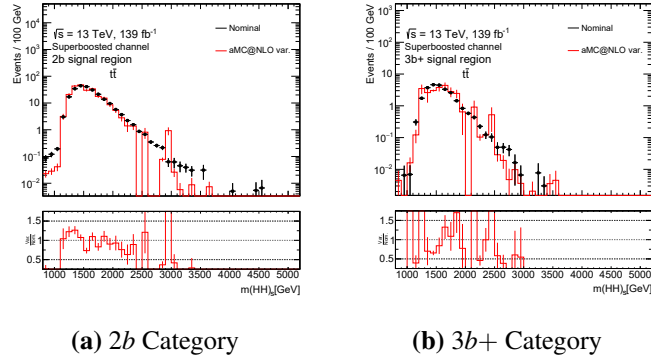


Figure 5.30: Comparison of the $t\bar{t} m(HH)_s$ distributions between two different matrix element generators (Powheg and aMC@NLO) for the $2b$ and $3b+$ categories in the Superboosted analysis.

nominal $t\bar{t}$ MC sample, which employs PYTHIA 8, to a variant sample generated with HERWIG 7, as detailed in Table 5.1. However, the variant $t\bar{t}$ sample contains significantly fewer events (approximately one-quarter) compared to the nominal sample. This statistical limitation induces fluctuations in the deviations, particularly in the tails of the $m(HH)_i$ distribution, as shown in Figures 5.31 and 5.32, and constitutes the dominant source of uncertainty for this component.

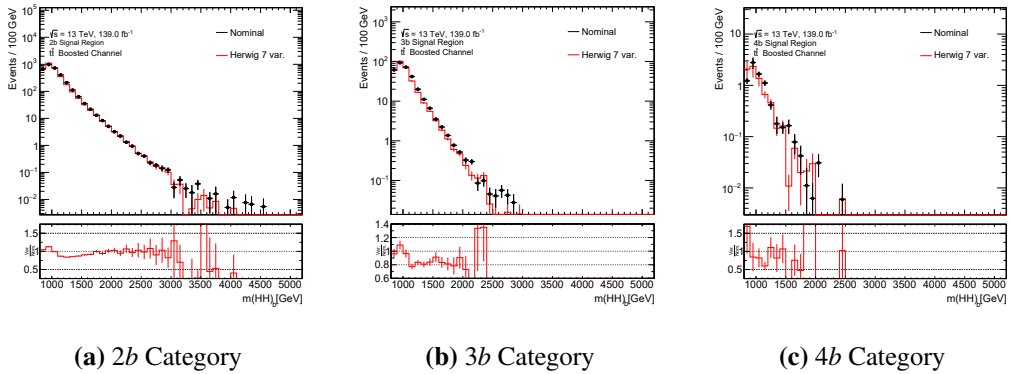


Figure 5.31: Comparison of the $t\bar{t} m(HH)_b$ distributions between two different parton shower generators in the Boosted analysis.

5.6.3.3 Parton Shower Scale

The fourth component of the $t\bar{t}$ production modeling uncertainty arises from variations in the parton shower scale. In contrast to the previous components, this uncertainty is assessed by varying the factorisation scale (μ_F) and renormalisation scale (μ_R) of the nominal $t\bar{t}$ MC sample.

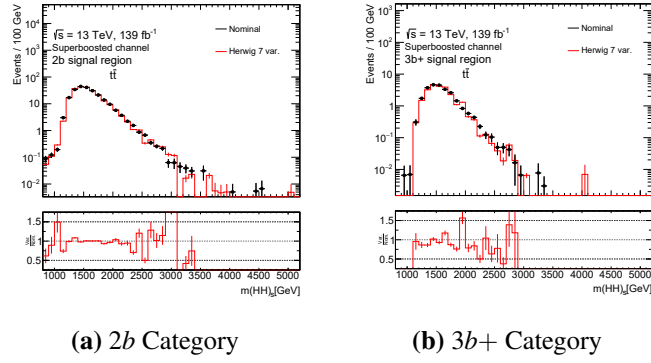


Figure 5.32: Comparison of the $t\bar{t} m(HH)_s$ distributions between two different parton shower generators in the Superboosted analysis.

The factorisation scale affects the shape of the parton distribution functions involved in the hard scattering process for NLO $t\bar{t}$ production. In parallel, the renormalisation scale adjusts the coupling constant $\alpha(\mu_R)$ in NLO $t\bar{t}$ production, mitigating large divergences in higher-order perturbative QCD calculations.

To evaluate this uncertainty, both μ_F and μ_R are varied independently from half to twice their nominal values. The resulting $m(HH)_i$ distributions are compared to the nominal distribution to quantify the impact of these scale variations. Figures 5.33–5.36 present the comparisons between the nominal and varied parton shower scales for the nominal $t\bar{t}$ samples.

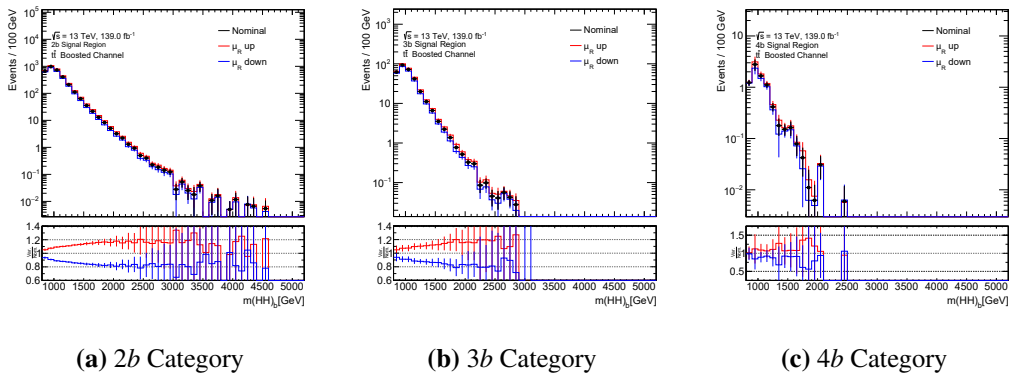


Figure 5.33: Comparison of the $t\bar{t} m(HH)_b$ distributions between different renormalisation scale (μ_R) variations in the Boosted analysis.

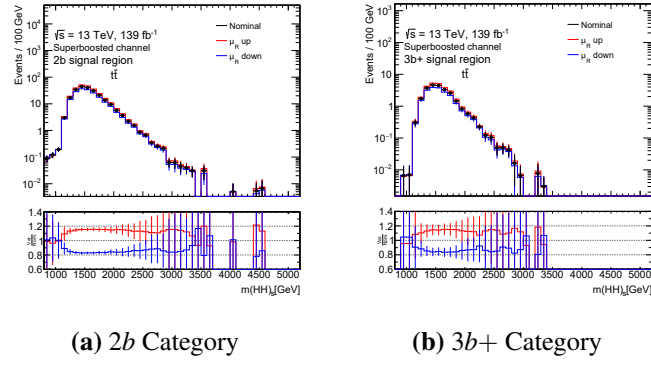


Figure 5.34: Comparison of the $t\bar{t} m(HH)_s$ distributions between different renormalisation scale (μ_R) variations in the Superboosted analysis.

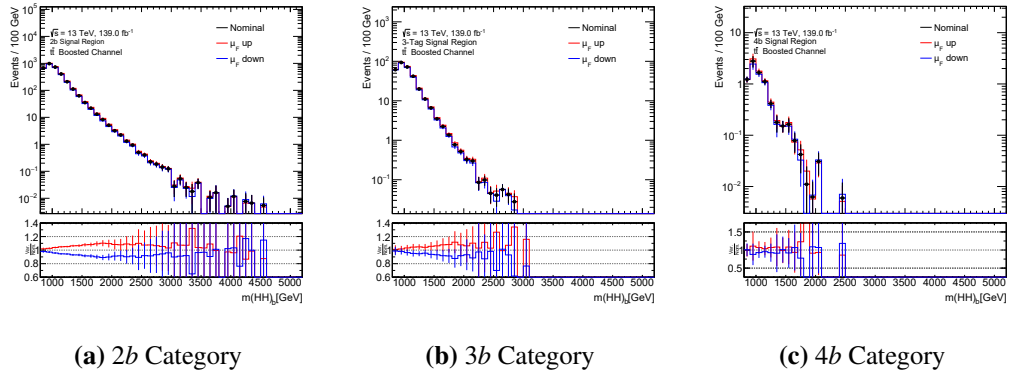


Figure 5.35: Comparison of the $t\bar{t} m(HH)_b$ distributions between different factorisation scale (μ_F) variations in the Boosted analysis.

5.6.3.4 Additional Hard Radiation (h_{damp})

Another component of the $t\bar{t}$ production modeling uncertainty originates from the h_{damp} parameter, which regulates the amount of hard radiation in $t\bar{t}$ MC samples. This uncertainty is evaluated by comparing the nominal $t\bar{t}$ MC sample to two variants with different h_{damp} values.

For the upward variation, a $t\bar{t}$ MC sample with $h_{damp} = 3m_{top}$, as specified in Table 5.1, is employed. This sample exhibits an increased amount of hard radiation compared to the nominal sample. However, an imbalance in the event distribution between all-hadronic and non-all-hadronic samples is observed, with approximately 30% fewer events in the latter. This statistical limitation results in significant deviations in the $m(HH)_i$ distribution tails (Figures 5.37 and 5.38), primarily contributing to the upward uncertainty.

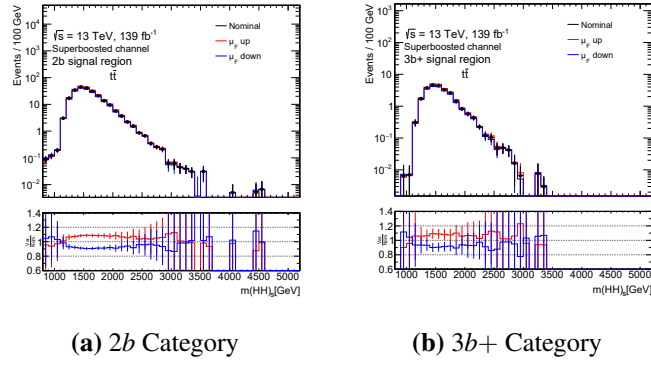


Figure 5.36: Comparison of the $t\bar{t} m(HH)_s$ distributions between different factorisation scale (μ_F) variations in the Superboosted analysis.

For the downward variation, the nominal $t\bar{t}$ MC sample with the A14 tune [91] (a set of optimised tuning parameters in the MC simulation to match data) is used, but with the Var3c down [91] variation applied and the parton shower scales (μ_F and μ_R) set to 2. The Var3c down tune reduces the Initial State Radiation (ISR) contribution in the nominal $t\bar{t}$ MC sample. Figures 5.37 and 5.38 present the $m(HH)_i$ comparison between the nominal $t\bar{t}$ sample and the two h_{damp} variations.

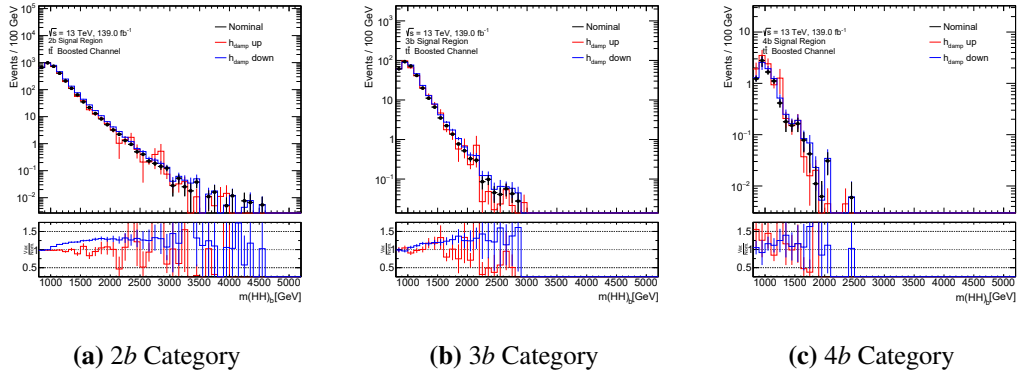


Figure 5.37: Comparison of the $t\bar{t} m(HH)_b$ distributions for two different h_{damp} parameters in the Boosted analysis.

5.6.3.5 Final State Radiation

The final component of the $t\bar{t}$ production modeling uncertainty arises from final-state radiation (FSR). This uncertainty is assessed by varying the FSR scale (μ_R^{FSR}) in the nominal $t\bar{t}$ MC sample. The FSR scale influences the amount of radiation emitted during the parton shower in the $t\bar{t}$ MC sample.



Figure 5.38: Comparison of the $t\bar{t} m(HH)_s$ distributions for two different h_{damp} parameters in the Superboosted analysis.

To evaluate this uncertainty, the μ_R^{FSR} scale is varied from half to twice the nominal value. The resulting $m(HH)_i$ distributions are compared to the nominal distribution to quantify the impact of the scale variations. Figures 5.39 and 5.40 show the comparison of $m(HH)_i$ distributions between the nominal $t\bar{t}$ MC sample and the two variations in μ_R^{FSR} .

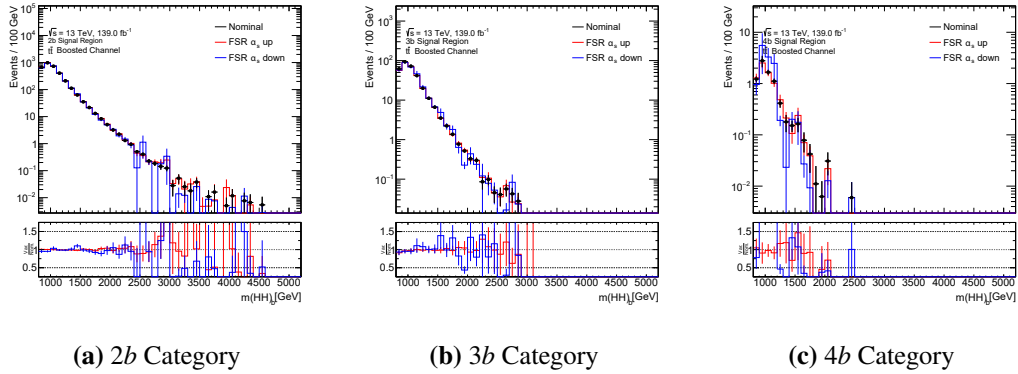


Figure 5.39: Comparison of the $t\bar{t} m(HH)_b$ distributions for final-state radiation (FSR) variations in the Boosted analysis.

5.6.4 Luminosity

The fourth component of the systematic uncertainty in the Boosted and Superboosted analyses is the luminosity uncertainty. This uncertainty, approximately 1.7%, is derived from the combined luminosity measurement for the 2015–2018 data using the LUCID-2 detector.

The obtained luminosity uncertainty is propagated to the uncertainties asso-



Figure 5.40: Comparison of the $t\bar{t} m(HH)_s$ distributions for final-state radiation (FSR) variations in the Superboosted analysis.

ciated with the background normalisation scale factors (μ_{qcd} and $\alpha_{t\bar{t}}$) derived in Section 5.5.2. It is also applied to all signal normalisation predictions, although it has a negligible impact on both analyses.

5.6.5 Theoretical Uncertainty for Signal

The fifth component of the systematic uncertainties in the Boosted and Superboosted analyses pertains to theoretical uncertainties in signal modeling. This component specifically assesses the impact of different parton shower generators on signal acceptance.

Two sets of heavy neutral scalar boson (X) MC samples are used to evaluate this uncertainty: one using HERWIG 7 (nominal) and the other using PYTHIA 8 as the parton shower generator. For the Boosted analysis, three MC samples with resonant masses of 1 TeV, 2 TeV, and 4 TeV are generated for each set. For the Superboosted analysis, two MC samples with resonant masses of 2 TeV and 4 TeV are generated for each set.

These MC samples undergo the event selection process outlined in Section 5.4 to determine their respective signal acceptances for the $4b$, $3b$, and $2b$ categories (Boosted analysis) and the $3b+$ and $2b$ categories (Superboosted analysis). The largest difference in signal acceptance between the two sets of MC samples is taken as the theoretical uncertainty for signal modeling. This uncertainty is then propagated to all signal normalisation predictions for the G_{KK}^* and X signals. The re-

sulting theoretical uncertainties are 12% for the Boosted analysis and 13% for the Superboosted analysis.

5.6.6 b -tagging Track Jet Scale Factor

The sixth component of the systematic uncertainty in the Boosted and Superboosted analyses is the b -tagging track jet scale factor. This factor corrects for differences in b -tagging efficiency between data and MC simulations. The uncertainty associated with this scale factor is assessed using the b -tagging efficiency calibration derived from dileptonic $t\bar{t}$ samples.

This uncertainty comprises four components:

- **MC Generator Modelling** [92]:

This component accounts for differences in scale factors derived from different parton shower and hadronisation generators used in $t\bar{t}$ MC samples.

- **Normalisation** [92]:

This component arises from uncertainties in the predicted cross-section of the dileptonic $t\bar{t}$ MC samples used for the b -tagging scale factor derivation.

- **Experimental** [92]:

This component incorporates detector effects and reconstructed physics objects in the $t\bar{t}$ MC sample used for the b -tagging scale factor derivation.

- **Extrapolation** [11]:

This component accounts for uncertainties in extrapolating the b -tagging scale factor to the high jet p_T regime in MC simulations.

To apply the b -tagging track jet scale factor uncertainty to the Boosted and Superboosted analyses, the following procedure is used:

1. All nominal MC samples used for signal and $t\bar{t}$ production are iteratively reproduced, incorporating all uncertainty components associated with the b -tagging scale factors.

2. For each iteration, two sets of MC samples are generated: one with an upward variation and one with a downward variation of the b -tagging scale factors.
3. These MC samples undergo the selection process outlined in Section 5.4 (for signal MC samples) and Sections 5.4 and 5.5 (for $t\bar{t}$ MC samples) to produce two $m(HH)_i$ distributions for comparison with the nominal distribution.
4. The shape differences between the nominal and varied $m(HH)_i$ distributions define the uncertainty envelope (upward and downward variations) for the b -tagging scale factor uncertainty applied to the nominal $m(HH)_i$ distribution.

5.6.7 Jet Systematics

The final component of systematic uncertainty in the Boosted and Superboosted analyses is jet systematics, which accounts for potential mismodelling in the calibration of Boosted and Superboosted jets. Four types of jet systematic uncertainties are considered: Jet Energy Scale (JES), Jet Energy Resolution (JER), Jet Mass Scale (JMS), and Jet Mass Resolution (JMR). The JES and JMS uncertainties are derived using an *in situ* method with a Category Reduction nuisance parameter scheme, which categorises all p_T -dependent JES uncertainties based on their origin.

To apply the jet systematics uncertainty, the following steps are taken:

1. All nominal MC samples for signal and $t\bar{t}$ production are iteratively reproduced, incorporating the jet systematic uncertainty components.
2. For each iteration, upward and downward variations of the MC samples are generated by modifying the kinematic variables involved in deriving the jet systematic uncertainties. For single variation systematics, only the upward variation is defined.
3. These modified MC samples undergo the selection process outlined in Section 5.4 (for signal MC samples) and Sections 5.4 and 5.5 (for $t\bar{t}$ MC samples) to produce two $m(HH)_i$ distributions for comparison with the nominal distribution.

4. The shape differences between the nominal and varied $m(HH)_i$ distributions define the uncertainty envelope (upward and downward variations) for the jet systematic uncertainty applied to the nominal $m(HH)_i$ distribution.
5. For single variation comparisons, a symmetrisation method is applied to obtain the uncertainty bands for the nominal MC samples (signal and $t\bar{t}$) $m(HH)_i$ distribution.

Further details on the derivation of jet systematic uncertainties are provided in Sections 5.6.7.1 to 5.6.7.4.

5.6.7.1 Jet Energy Scale

The JES uncertainty [77] is derived by comparing the double ratio of calibrated jet response ($R_{\text{in situ}}$) between data and MC samples as a function of jet $p_{\text{T}}^{J_i}$ and η^{J_i} . The JES uncertainty components for both the Boosted and Superboosted analyses include six main categories: η -intercalibration, Z/γ +jet topology, multijet balance, jet flavour, pileup, and Global Sequential Calibration. Each of these components is applied to the $p_{\text{T}}^{J_i}$, η^{J_i} , and E^{J_i} of the jets, where i refers to Boosted (b) or Superboosted (s) jets.

5.6.7.2 Jet Energy Resolution

The JER uncertainty is derived from the parameterised JER, $\sigma_{\text{JER}}(p_{\text{T}}^{J_i})$, in data and MC relative to $p_{\text{T}}^{J_i}$, as defined in Equation 5.14:

$$\frac{\sigma_{\text{JER}}(p_{\text{T}}^{J_i})}{p_{\text{T}}^{J_i}} = \frac{N}{p_{\text{T}}^{J_i}} \oplus \frac{S}{\sqrt{p_{\text{T}}^{J_i}}} \oplus C \quad (5.14)$$

where N is the noise term, S is the stochastic term, and C is the constant term. The modelling of JER in MC simulation involves measuring the JER using an *in situ* method, utilising momentum balance in dijet events. The JER uncertainty is obtained by smearing the $p_{\text{T}}^{J_i}$ in MC simulation to match the data. If the MC JER is smaller than the data JER, no smearing is applied.

5.6.7.3 Jet Mass Scale

The Jet Mass Scale (JMS) uncertainty [79] is derived using the R_{trk} method, which calculates the mass ratio (r_m) between the calorimeter jet mass (m^{J_i}) and the track jet mass (m^{t_i}) components as a function of the calorimeter jet transverse momentum ($p_T^{J_i}$) in data. This ratio is compared to MC samples to obtain the double ratio (R_m) as a function of $p_T^{J_i}$, providing the nominal JMS uncertainty. For Boosted jets, the combined mass of the calorimeter jet is used as the m^{J_i} component.

In addition to the nominal uncertainty, four JMS uncertainty components are considered:

- Mismodeling arising from different MC generators [79].
- Track reconstruction efficiency [93].
- Inclusion of fake tracks [94].
- Effects of weak alignment modes on tracks [95].
- Topocluster to Particle Flow mismodeling (Superboosted jets only).

Each JMS uncertainty component is applied to the mass of the calorimeter jets (m^{J_i}) in the Boosted and Superboosted analyses, where i refers to Boosted (b) or Superboosted jets (s).

5.6.7.4 Jet Mass Resolution

The Jet Mass Resolution (JMR) uncertainty is derived using information from the reconstructed mass (m^{reco}) and the truth mass (m^{truth}) of jets in MC samples [85]. The reconstructed mass corresponds to the calorimeter jet mass (m^{J_i}) or the combined mass (for Boosted jets). The nominal jet mass resolution in each bin of the truth jet transverse momentum (p_T^{truth}) is calculated as the ratio between the interquartile range of the jet response (R_m) distribution and its median value, as shown in Equation 5.15:

$$\sigma_{JMR} = \frac{R_{q84}^m - R_{q16}^m}{2 \text{Med}(R^m)}, \quad R^m = \frac{m^{reco}}{m^{truth}} \quad (5.15)$$

where σ_{JMR} is the jet mass resolution at a given p_T^{truth} bin, $\text{Med}(R_m)$ is the median value of the jet response distribution, and R_{q84}^m and R_{q16}^m represent the jet response values at the 84th and 16th percentiles, respectively.

JMR Uncertainty for Superboosted Jets: For Superboosted jets, the JMR uncertainty is obtained by applying a 20% Gaussian smearing to the calorimeter mass (m^{Js}) as a function of the jet transverse momentum (p_T^{Js}):

$$m^{smear} = m^{Js} \times G(1, 0.66 \sigma_{JMR}^s) \quad (5.16)$$

where m^{smear} is the smeared mass of the calorimeter jets, σ_{JMR}^s is the nominal jet mass resolution for Superboosted jets, and G is a Gaussian random variable with a mean of 1 and a standard deviation of $0.66 \sigma_{JMR}^s$. This smearing is applied to all signal and nominal $t\bar{t}$ MC samples.

JMR Uncertainty for Boosted Jets: For Boosted jets, the JMR uncertainty is derived using the forward-folding method, which utilises information from the MC-corrected distribution and data from in situ calibration. This method provides improved mass resolution uncertainties compared to Gaussian smearing. The forward-folding JMR (σ_{FF}) is approximated as follows:

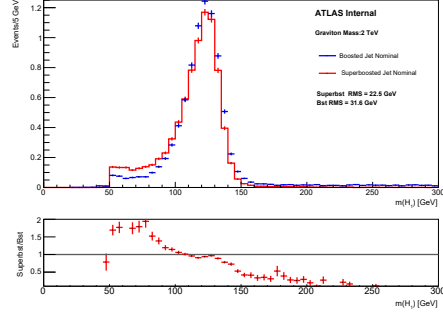
$$\sigma_{FF} = \kappa \sigma(R^m | m^{truth}, p_T^{J_b}) \quad (5.17)$$

where κ is the ratio between data-driven and MC-based JMR obtained from in situ calibration, and $\sigma(R^m | m^{truth}, p_T^{J_b})$ represents the MC-based JMR as a function of the jet mass response (R^m), the truth jet mass (m^{truth}), and the calorimeter jet transverse momentum ($p_T^{J_b}$). The forward-folding method incorporates uncertainties in κ to obtain up and down variations for the JMR, which are used as JMR uncertainties for Boosted jets, as shown in Equation 5.18:

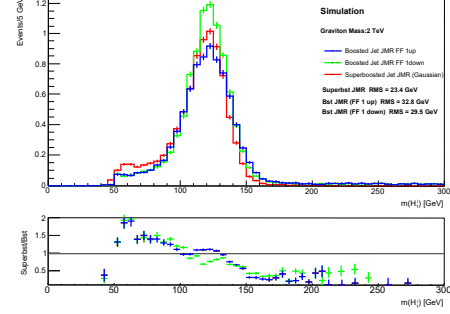
$$\sigma_{FF\pm} = (\kappa \pm \delta\kappa) \sigma(R^m | m^{truth}, p_T^{J_b}) \quad (5.18)$$

where $\pm\delta\kappa$ represents the uncertainties of κ derived from in situ calibration

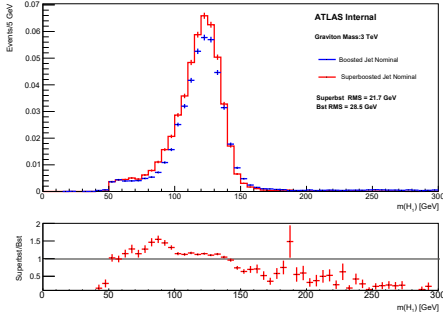
for large- R jets. These two variations are applied to the Boosted jets to obtain the systematic variations of the JMR uncertainties in the signal and nominal $t\bar{t}$ samples. Figures 5.41 and 5.42 show the comparison of the leading Higgs candidates' masses $m(H_1^i)$ in several G_{kk}^* MC samples with and without JMR uncertainties applied for both analyses.



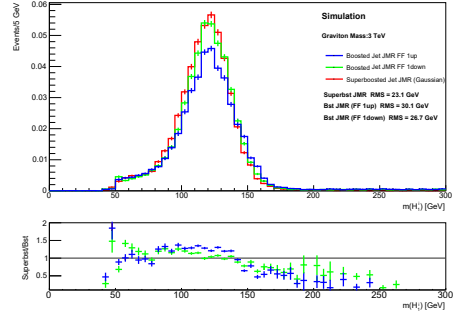
(a) $m(H_1^i)$ without JMR uncertainty
for $m_{G_{kk}^*} = 2$ TeV



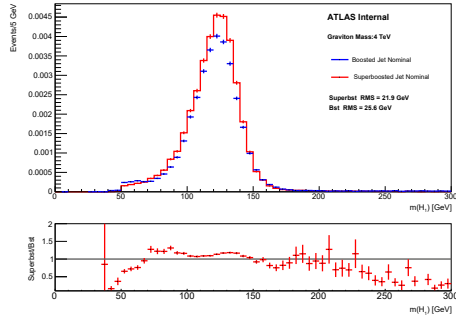
(b) $m(H_1^i)$ with JMR uncertainty
for $m_{G_{kk}^*} = 2$ TeV



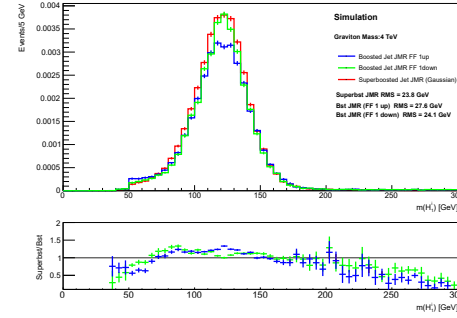
(c) $m(H_1^i)$ without JMR uncertainty
for $m_{G_{kk}^*} = 3$ TeV



(d) $m(H_1^i)$ with JMR uncertainty
for $m_{G_{kk}^*} = 3$ TeV



(e) $m(H_1^i)$ without JMR uncertainty
for $m_{G_{kk}^*} = 4$ TeV



(f) $m(H_1^i)$ with JMR uncertainty
for $m_{G_{kk}^*} = 4$ TeV

Figure 5.41: Leading Higgs candidate mass $m(H_1^i)$ distributions before and after applying JMR uncertainties at different G_{kk}^* masses. "JMR FF 1up" and "JMR FF 1down" refer to the σ_{FF+} and σ_{FF-} smearing variations from the forward-folding JMR in Boosted jets.

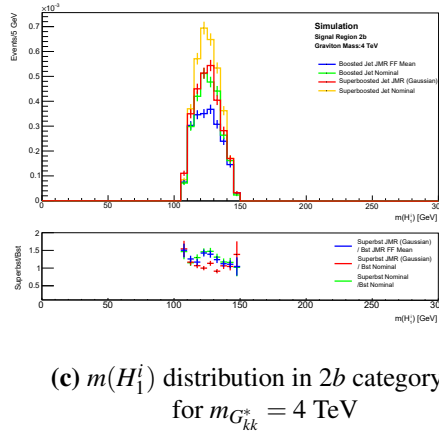
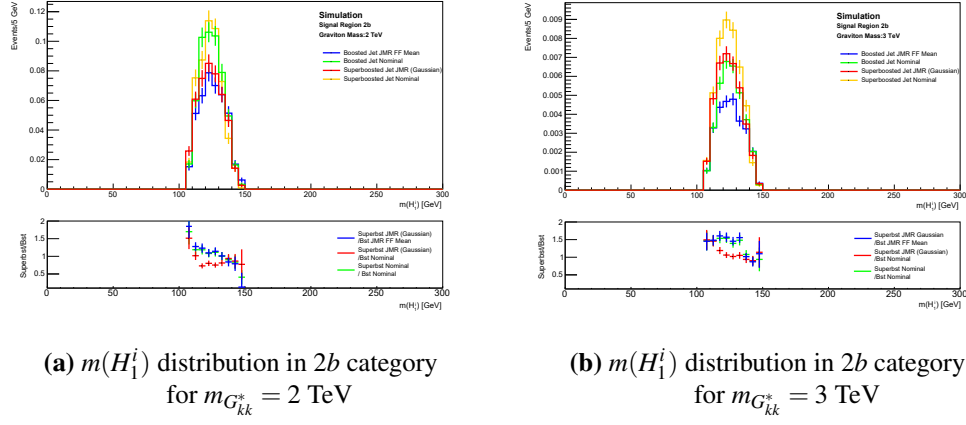


Figure 5.42: Leading Higgs candidate mass $m(H_1^i)$ distributions in the $2b$ category before and after applying JMR uncertainties at different G_{kk}^* masses. "Boosted Jet JMR FF Mean" refers to the average values of the histogram bins between the σ_{FF_+} and σ_{FF_-} variations for Boosted jets.

5.7 Statistical Analysis

Sections 5.1 to 5.6 presented the Boosted and Superboosted analysis methods employed to search for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks. Sections 5.2 to 5.4 outlined the data, Monte Carlo samples, relevant physics objects, and event selection criteria for signal and background events. Sections 5.5 and 5.6 detailed the background estimation and associated systematic uncertainties for both analyses. Building upon these sections, this section describes the statistical analysis used to derive the upper limit on the cross section for resonant Higgs boson pair production decaying into two pairs of $b\bar{b}$ quarks in both the Boosted and Superboosted analyses.

The statistical analysis for the Boosted and Superboosted analyses comprises five subsections:

- Signal interpolation for X MC samples
- Procedure for deriving the upper limit on the cross section
- Pull and impact of systematic uncertainties
- Signal event yields
- Unblinded results on the upper limits

The $m(HH)_i$ variable serves as the final discriminant for limit setting to obtain the upper limit on the cross section for $X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ production, where R denotes either an Excited Massive Kaluza-Klein Graviton (G_{kk}^*) or a Heavy Neutral Scalar Boson (X).

5.7.1 Signal Interpolation on X MC Samples

To address the large mass gaps between resonant masses in the X MC samples for $m_X > 2$ TeV, a signal interpolation procedure was employed prior to limit setting. This procedure generates interpolated signal distributions for resonant masses between existing MC simulation points without the need for additional sample generation.

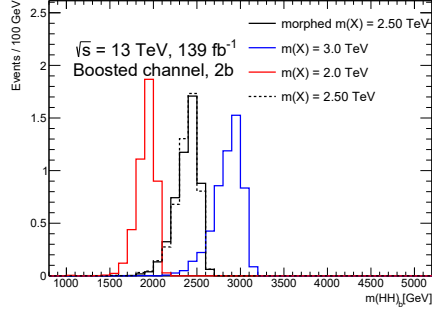
Linear moment morphing with cubic splines was used to interpolate between the $m(HH)_i$ distributions of two adjacent resonant masses. The distributions were first normalised to unity, and each bin was subsequently fitted with a cubic spline to obtain the morphed distribution for the target resonant mass. The morphed distribution was then rescaled according to the specific cross section for $X \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ at the target mass. This procedure was repeated for each systematic variation associated with the target mass.

For cross-checking and to assess the impact of interpolation, the procedure was also applied to a target mass of $m_X = 2500$ GeV, using the 2000 GeV and 3000 GeV MC samples as boundaries, as illustrated in Figures 5.43 and 5.44.

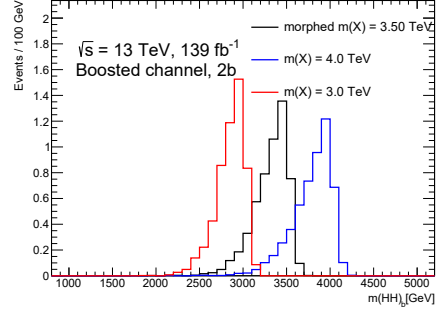
Table 5.9 summarises the target masses for signal interpolation along with their corresponding bounding resonant masses derived from MC simulation.

Target m_X for Signal Interpolation	MC Simulated Resonant Mass Bounds
2250 GeV	2000 GeV and 2500 GeV
2750 GeV	2500 GeV and 3000 GeV
3500 GeV	3000 GeV and 4000 GeV
4500 GeV	4000 GeV and 5000 GeV

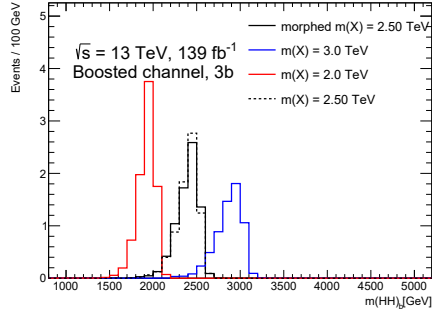
Table 5.9: Target m_X values for signal interpolation and their corresponding MC simulated resonant mass bounds.



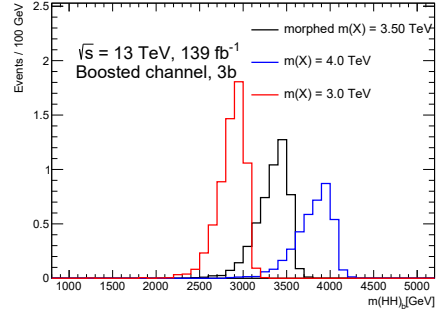
(a) Morphed signal $m(HH)_b$ distribution compared with the actual MC simulation sample for $m_X = 2.5$ TeV in the $2b$ category.



(b) Morphed signal $m(HH)_b$ distribution for $m_X = 3.5$ TeV in the $2b$ category.

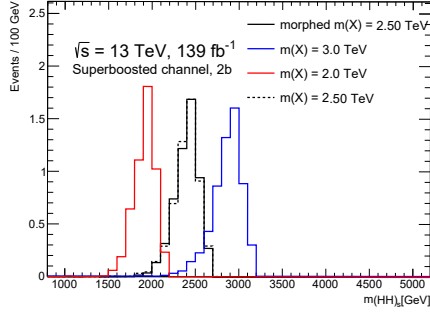


(c) Morphed signal $m(HH)_b$ distribution compared with the actual MC simulation sample for $m_X = 2.5$ TeV in the $3b$ category.

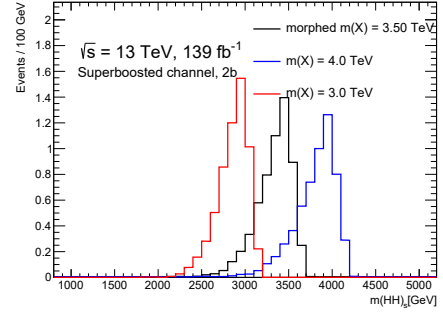


(d) Morphed signal $m(HH)_b$ distribution for $m_X = 3.5$ TeV in the $3b$ category.

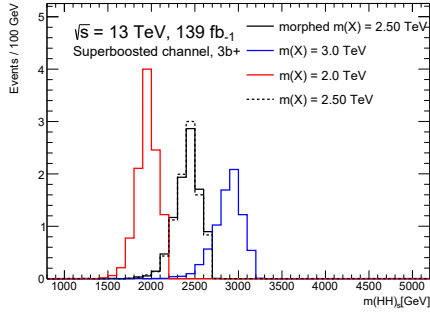
Figure 5.43: Signal interpolation $m(HH)_b$ distributions for the Boosted analysis for a Heavy Neutral Scalar Boson with $m_X = 3.5$ TeV, together with a cross-check at $m_X = 2.5$ TeV.



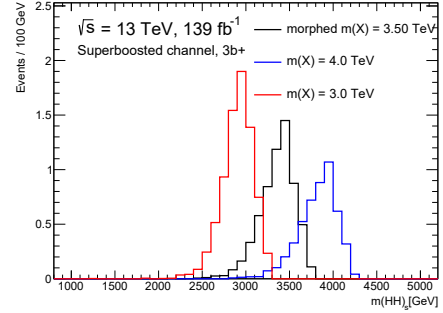
(a) Morphed signal $m(HH)_s$ distribution compared with the actual MC simulation sample for $m_X = 2.5$ TeV in the $2b$ category.



(b) Morphed signal $m(HH)_s$ distribution for $m_X = 3.5$ TeV in the $2b$ category.



(c) Morphed signal $m(HH)_s$ distribution compared with the actual MC simulation sample for $m_X = 2.5$ TeV in the $3b+$ category.



(d) Morphed signal $m(HH)_s$ distribution for $m_X = 3.5$ TeV in the $3b+$ category.

Figure 5.44: Signal interpolation $m(HH)_s$ distributions for the Superboosted analysis for a Heavy Neutral Scalar Boson with $m_X = 3.5$ TeV, together with a cross-check at $m_X = 2.5$ TeV.

5.7.2 Limit Setting Procedure

The limit-setting procedure for the Boosted and Superboosted analyses consists of four steps: building the profile likelihood, testing the hypothesis and calculating confidence levels, deriving expected limits, and combining limits from tag categories.

5.7.2.1 Building the Profile Likelihood

The first step in the limit-setting procedure involves constructing the profile likelihood function for the signal and background $m(HH)_i$ histograms. The profile likelihood function, L , is defined as the product of Poisson distributions for each bin in the histogram (Equation 5.19) [96]:

$$L(\mu, \theta) = \prod_{k=1}^N \frac{(\mu s_k + b_k)^{n_k}}{n_k!} e^{-(\mu s_k + b_k)} C(\theta) \quad (5.19)$$

where:

- μ is the signal normalization;
- n_k is the number of observed data entries in bin k ;
- N is the total number of histogram bins;
- $\mu s_k + b_k$ is the expected total number of signal and background events in bin k ;
- θ represents the set of nuisance parameters derived from the $m(HH)_i$ distributions and associated systematic uncertainties;
- $C(\theta)$ represents the product of constraint terms for the nuisance parameters.

For the background-only case, μ is set to zero. Nuisance parameters are incorporated as Gaussian and log-normal constraint terms in Equation 5.19. Shape variations are treated as Gaussian constraints, while normalization uncertainties are treated as log-normal constraints. Each systematic uncertainty is correlated with a

Systematic Uncertainty	Uncertainty Type	Constraint Term
Luminosity	Normalisation	Log-Normal
Signal Theoretical Uncertainty	Normalisation	Log-Normal
Background Normalisation	Normalisation	Log-Normal
Background Shape	Shape, Normalisation	Gaussian, Log-Normal
$t\bar{t}$ Production Modelling	Shape, Normalisation	Gaussian, Log-Normal
b -tagging Scale Factor	Shape, Normalisation	Gaussian, Log-Normal
Jet Systematics	Shape, Normalisation	Gaussian, Log-Normal

Table 5.10: Summary of systematic uncertainties, uncertainty types, and corresponding constraint terms in the profile likelihood.

single nuisance parameter. Table 5.10 summarizes the systematic uncertainty categories and their corresponding constraint types.

5.7.2.2 Testing the Hypothesis

The second step involves testing the signal-plus-background hypothesis against the background-only hypothesis. This is achieved using a one-sided negative log-likelihood ratio test statistic, $\tilde{q}_\mu$, defined as follows (Equation 5.20):

$$\tilde{q}_\mu = \begin{cases} -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(0, \hat{\theta}(0))} & \text{if } \hat{\mu} < 0 \\ -2 \ln \frac{L(\mu, \hat{\theta}(\mu))}{L(\hat{\mu}, \hat{\theta}(\mu))} & \text{if } 0 \leq \hat{\mu} < \mu \\ 0 & \text{if } \mu < \hat{\mu} \end{cases} \quad (5.20)$$

where:

- $\hat{\mu}$ is the maximum likelihood estimate of the signal strength parameter μ ;
- $\hat{\theta}$ is the maximum likelihood estimate of the nuisance parameters θ ;

- $\hat{\hat{\theta}}(\mu)$ is the maximum likelihood estimate of the nuisance parameters for a fixed signal strength μ ;
- $L(\mu, \hat{\hat{\theta}}(\mu))$ is the constrained likelihood for a given signal strength μ ;
- $L(\hat{\mu}, \hat{\theta}(\mu))$ is the unconstrained maximum likelihood.

The choice of test statistic is based on the requirement that an upward fluctuation of the signal does not constitute evidence against the signal hypothesis (for the case $\mu < \hat{\mu}$), and that a downward fluctuation of the background does not constitute evidence against the background hypothesis (for the case $\hat{\mu} < 0$).

5.7.2.3 Finding μ from Confidence Levels

Once the test statistic $\tilde{q}_\mu$ is defined, it is used to calculate the 95% confidence level (CL_s) for excluding a given resonant mass of the signal. The CL_s method [97] ensures that the derived value of μ is statistically significant for setting the upper limit. The CL_s value is calculated using Equation 5.21:

$$CL_s = \frac{P_{s+b}(\tilde{q}_\mu \leq q_{\text{obs}})}{P_b(\tilde{q}_\mu \leq q_{\text{obs}})} \quad (5.21)$$

where:

- $P_{s+b}(\tilde{q}_\mu \leq q_{\text{obs}})$ denotes the probability of the signal-plus-background model producing a test statistic value less than or equal to the observed value q_{obs} ,
- $P_b(\tilde{q}_\mu \leq q_{\text{obs}})$ denotes the probability of the background-only model producing a test statistic value less than or equal to the observed value q_{obs} .

To obtain the 95% confidence level, the signal strength parameter μ is varied for each resonant mass point until the CL_s value reaches 0.05. The corresponding μ value is then converted into a cross-section limit, $\sigma(X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b})$.

For the calculation of the probabilities in Equation 5.21, two methods are employed: the asymptotic approximation [98] and the Toy Monte Carlo simulation.

Asymptotic Approximation: For large datasets, the asymptotic approximation, based on Wilks' theorem and the Wald approximation, provides an efficient estimation of the CL_s value. Wilks' theorem states that the distribution of the negative log-likelihood ratio test statistic, $f(\tilde{q}_\mu|\mu, \theta)$, asymptotically follows a non-central chi-squared (χ^2) distribution with one degree of freedom for large sample sizes. The Wald approximation yields an approximate expression for the negative log-likelihood ratio test statistic, $\tilde{q}_\mu$, as shown in Equation 5.22:

$$\tilde{q}_\mu \approx \frac{(\mu - \hat{\mu})^2}{\sigma_{\hat{\mu}}^2} + O(1/\sqrt{n}) \quad (5.22)$$

where:

- $\sigma_{\hat{\mu}}$ denotes the standard deviation of the maximum likelihood estimate $\hat{\mu}$,
- $O(1/\sqrt{n})$ represents the error term due to finite sample size (n).

This approximation is valid for profile likelihoods constructed from large datasets, where the contribution of the error term $O(1/\sqrt{n})$ becomes negligible.

Toy Monte Carlo Simulation: For cases with low event counts, the Toy Monte Carlo simulation method is employed. This approach generates a large number of pseudo-experiments to empirically determine the distribution of the test statistic. Approximately 2500 pseudo-experiments are generated for each resonant mass point. The pseudo-data are sampled from Poisson distributions with expected values derived from fixed input values of the signal strength parameter, μ . The resulting distribution of the negative log-likelihood ratio test statistic, $f(\tilde{q}_\mu|\mu, \theta)$, is then used to calculate the CL_s value for a given μ . This procedure is repeated for different μ values to obtain the cross-section limit, $\sigma(X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b})$. Although offering improved accuracy, this method requires a substantial amount of computational resources.

The choice of method depends on the resonant mass. For resonant masses above 3 TeV, Toy Monte Carlo simulation is used due to its superior accuracy in low-statistics regions. For resonant masses below 3 TeV, the asymptotic approximation

provides sufficient accuracy. The asymptotic approximation yields reliable values for $\tilde{q}_\mu$ and a faithful shape for the distribution $f(\tilde{q}_\mu|\mu, \theta)$ only when the sample size n is sufficiently large. In cases with low event counts, significant deviations of $\tilde{q}_\mu$ from the true values can occur, and the true distribution $f(\tilde{q}_\mu|\mu, \theta)$ can differ markedly from the ideal χ^2 distribution. Such deviations lead to poor estimations of $\tilde{q}_\mu$ and μ at high $m(HH)_i$ values, potentially causing an underestimation of the upper limit approximation for the cross-section at large resonant masses.

5.7.2.4 Expected Limits

Prior to unblinding the observed data, expected limits on the cross section are determined to assess the sensitivity of the Boosted and Superboosted analyses. Lower expected cross-section limits indicate greater sensitivity to signal detection.

Expected limits are calculated using an Asimov dataset, where the observed data in each bin of Equation 5.19 are replaced by the expected background contribution, that is, $n_k = b_k$. The profile likelihood is then constructed based on this dataset, and the corresponding $\tilde{q}_\mu$ and μ values for a CL_s of 0.05 are determined for each point of the resonant mass. These values are subsequently converted into expected limit cross sections, $\sigma(X_R \rightarrow HH \rightarrow b\bar{b}b\bar{b})$.

For resonant masses below 3 TeV, the asymptotic approximation is used to calculate the CL_s value, whereas for masses above 3 TeV, the Toy Monte Carlo simulation method is employed.

5.7.2.5 Combining Limits from Tag Categories

Building upon the limit-setting procedures outlined in Sections 5.7.1.1 to 5.7.1.3, this section discusses the combination of limits from different tag categories in both the Boosted and Superboosted analyses. Each category is combined into a single limit within specific resonant mass ranges.

- **Boosted Analysis:**

- The $3b$ and $4b$ categories are combined into a single limit for the resonant mass range between 900 GeV and 3 TeV.

- The $4b$ category is excluded from the combined limit for the resonant mass range above 3 TeV due to reduced sensitivity beyond this point.
- The $3b$ and $2b$ categories are combined into a single limit for the resonant mass range between 2 TeV and 5 TeV.

- **Superboosted Analysis:**

- The $3b/3b+$ category is used for the resonant mass range between 1600 GeV and 5 TeV.
- The $2b$ category is combined with the $3b/3b+$ category for the resonant mass range between 2 TeV and 5 TeV.

Table 5.11 summarises the tag categories and resonant mass ranges used for calculating the combined expected limits.

Resonant Mass Range	Tag category for combined limit	
	Boosted Analysis	Superboosted Analysis
900 GeV–2 TeV	$4b, 3b$	–
1600 GeV–2 TeV	–	$3b+$ only
2 TeV–3 TeV	$4b, 3b, 2b$	$2b, 3b+$
3.5 TeV–5 TeV	$3b, 2b$	$3b+, 2b$

Table 5.11: Tag categories for each resonant mass range used in the combined limit calculation for the Boosted and Superboosted analyses.

5.7.3 Pull and Impact

5.7.3.1 Pull

A pull quantifies the deviation of a nuisance parameter from its nominal value during the maximum likelihood estimation procedure. It is defined as the difference between the maximum likelihood estimate of the nuisance parameter ($\hat{\theta}$) and the expected value (θ_0), divided by the standard deviation (σ_θ), as shown in Equation 5.23:

$$\text{pull}(\theta) = \frac{\hat{\theta} - \theta_0}{\sigma_\theta} \quad (5.23)$$

In a well-behaved statistical model, pulls for nuisance parameters are generally contained within one standard deviation. Nuisance parameters exhibiting larger pulls may indicate significant mismodeling within the analysis, potentially leading to notable deviations in the limit calculation and requiring further investigation.

5.7.3.2 Impact

The impact of a nuisance parameter measures the variation in the maximum likelihood estimate of the signal strength parameter ($\hat{\mu}$) when the nuisance parameter is shifted. This evaluation identifies nuisance parameters contributing significantly to the total systematic uncertainty. For the Boosted and Superboosted analyses, the impact is determined by fixing each nuisance parameter to $\theta_0 \pm \sigma_\theta$ and re-evaluating $\hat{\mu}$. The resulting change in $\hat{\mu}$ relative to the nominal value quantifies the impact, as given in Equation 5.24:

$$\text{impact}(\theta) = \hat{\mu}_{\theta_0 \pm \sigma_\theta} - \hat{\mu} \quad (5.24)$$

A large impact value indicates a substantial contribution of the corresponding nuisance parameter to the overall systematic uncertainty. Figures 5.45-5.47 present the constraint plots, displaying pulls (circles and black lines, lower scale) and corresponding impacts for the Boosted and Superboosted analyses. Table 5.12 summarises the abbreviations used for the nuisance parameters shown in these plots.

In Figure 5.48, the $m(X) = 5$ TeV signal in the Superboosted analysis shows a pronounced asymmetric pattern, with nuisance parameters predominantly shifted in one direction. This asymmetry arises primarily from the effect of background smoothing systematics in low-statistics categories such as 3b+, as illustrated in Figure 5.22. Similar features are also present in the Boosted analysis (Figure 5.46). An additional source of asymmetry in the Superboosted channel stems from Jet Mass Scale (JMS) uncertainties, which become increasingly relevant at high transverse momenta and large invariant masses—particularly for narrow-width signals like the heavy neutral scalar boson (X).

A comparison between the 5 TeV X and the G_{kk}^* signals in Figure 5.49 demonstrates that the X signal exhibits a more substantial one-sided deviation under $\pm 1\sigma$ systematic shifts. In contrast, the G_{kk}^* signal maintains a more symmetric response. This distinction is attributed to the broader natural width of the G_{kk}^* resonance,

which reduces sensitivity to edge-induced smearing effects. The behavior of JMS uncertainties across the relevant kinematic phase space is shown in Figure B.1, further highlighting the enhanced sensitivity of narrow-width X signals to such effects in the Superboosted regime.

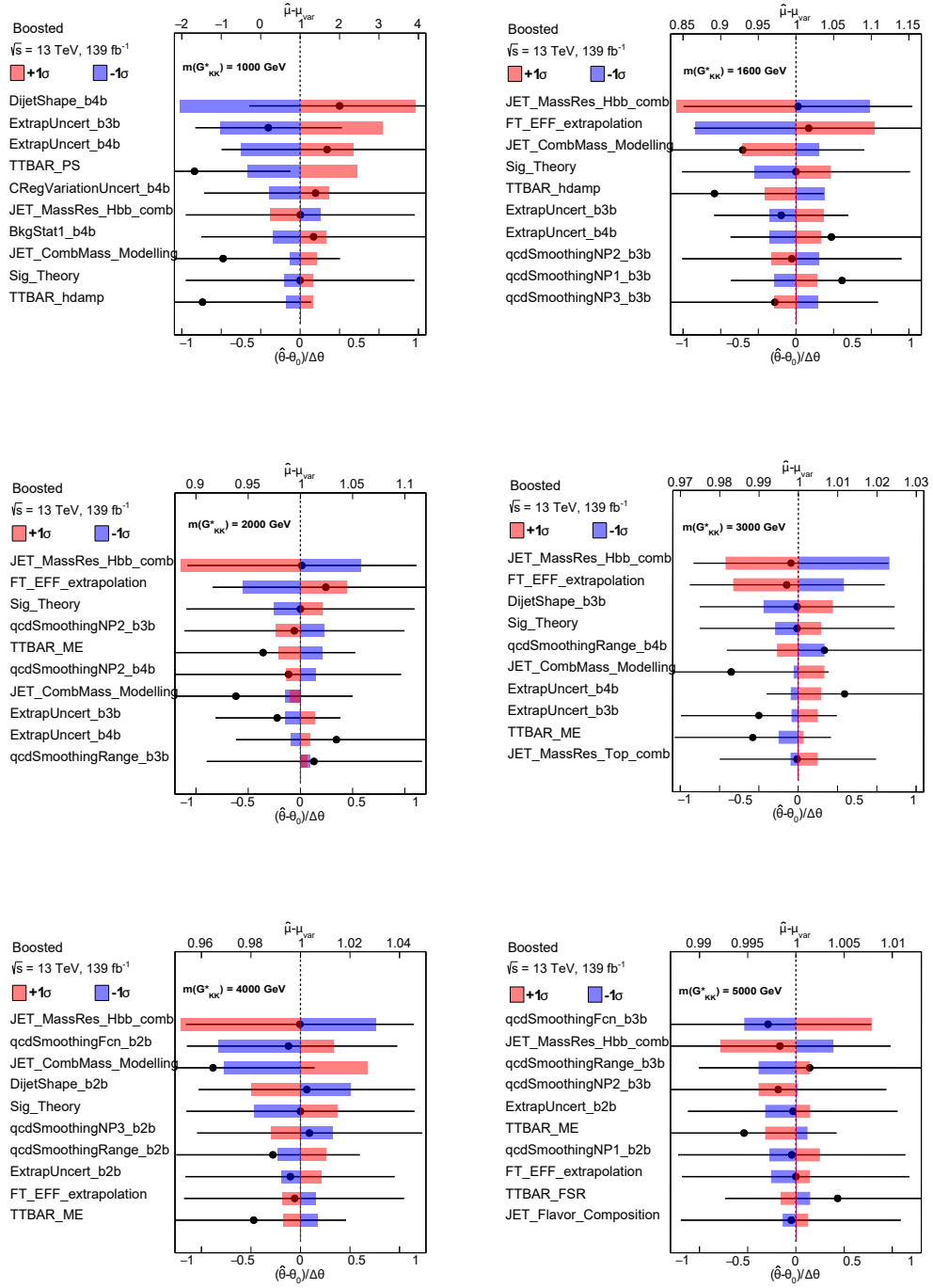


Figure 5.45: Pulls (circles and black lines with bottom scale) and impacts (blue and red bars with top scale) for various G_{kk}^* masses in the Boosted analysis.

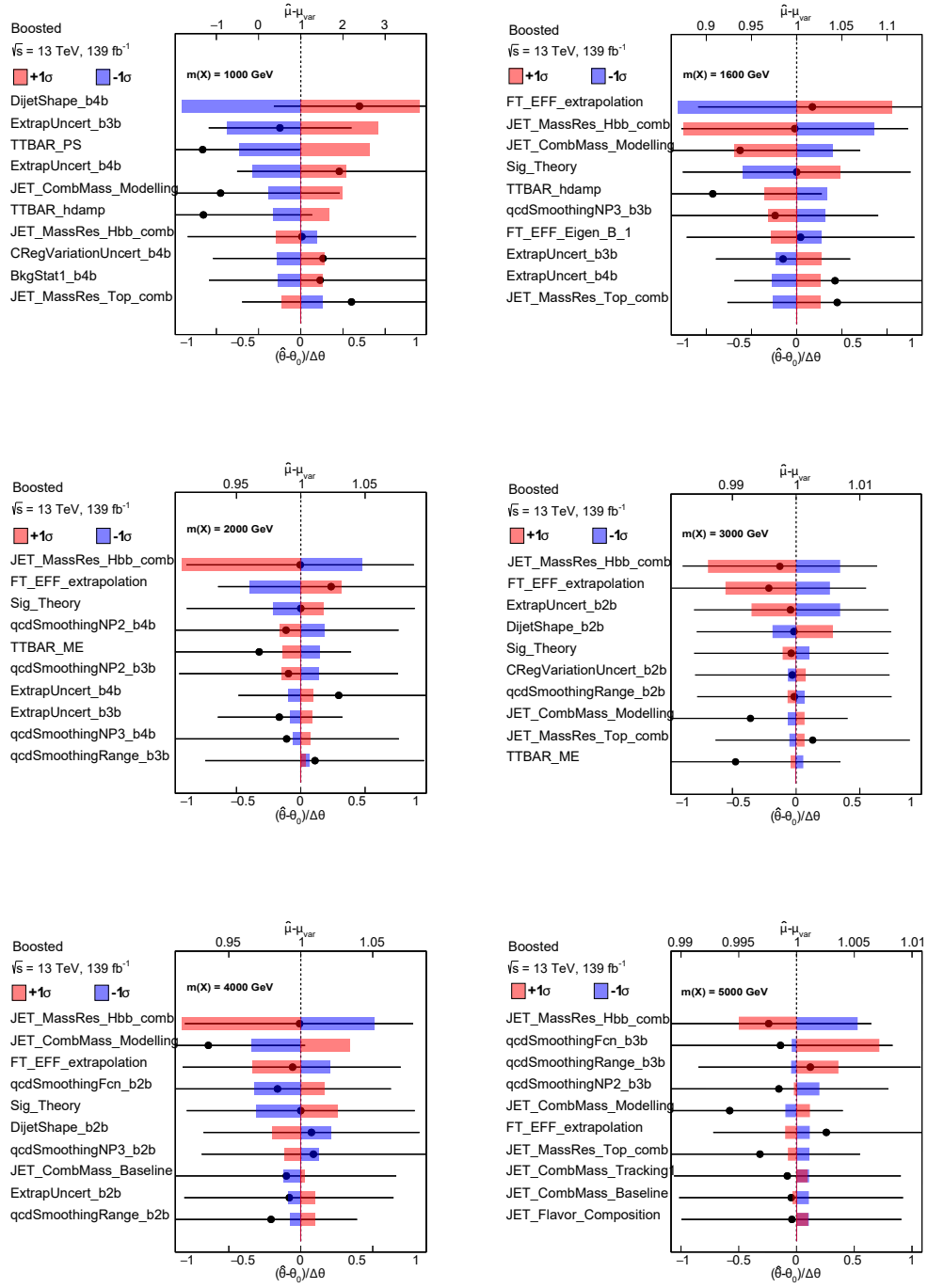


Figure 5.46: Pulls (circles and black lines with bottom scale) and impacts (blue and red bars with top scale) for various X masses in the Boosted analysis.

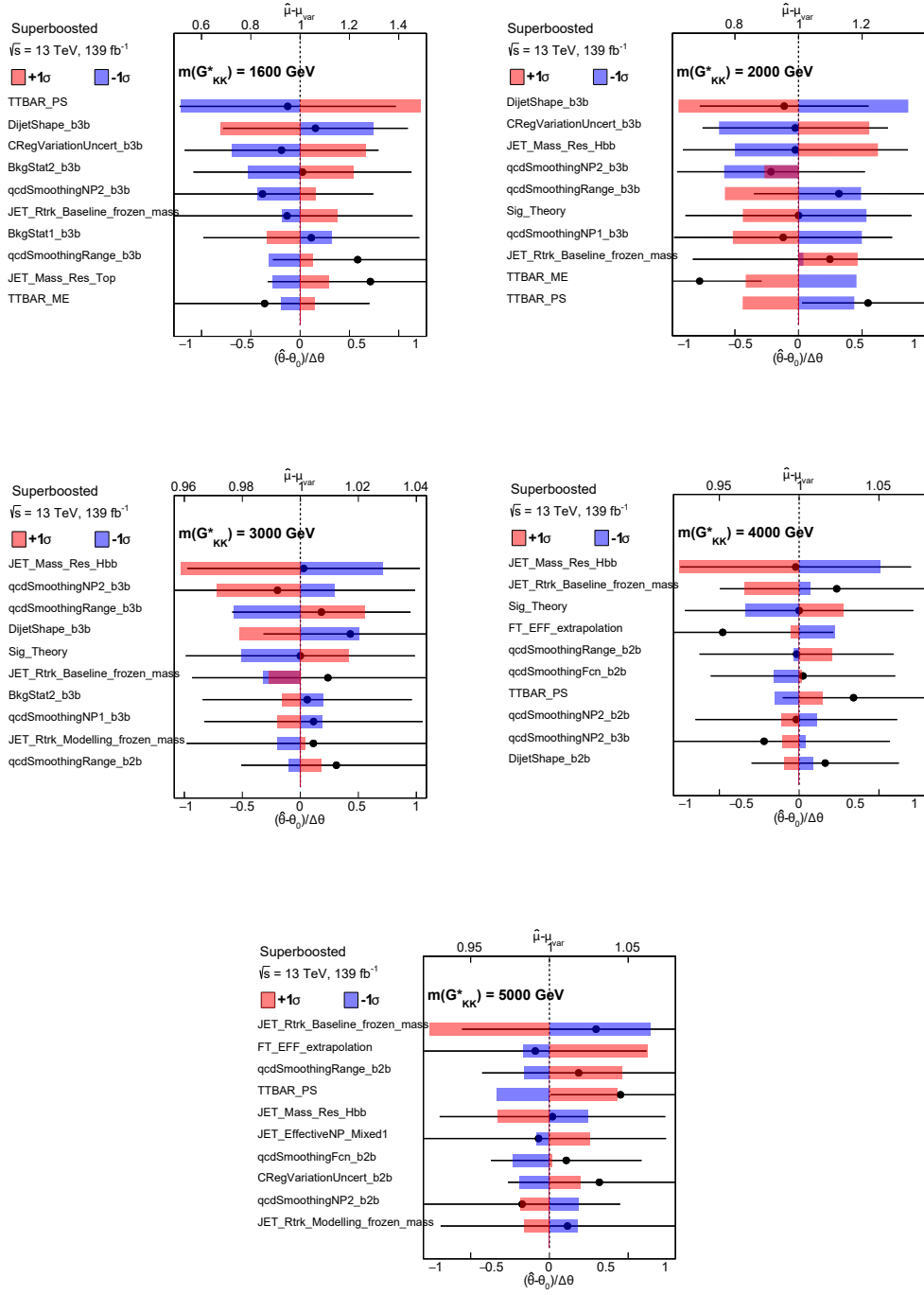


Figure 5.47: Pulls (circles and black lines with bottom scale) and impacts for different G_{kk}^* masses in the Superboosted analysis.

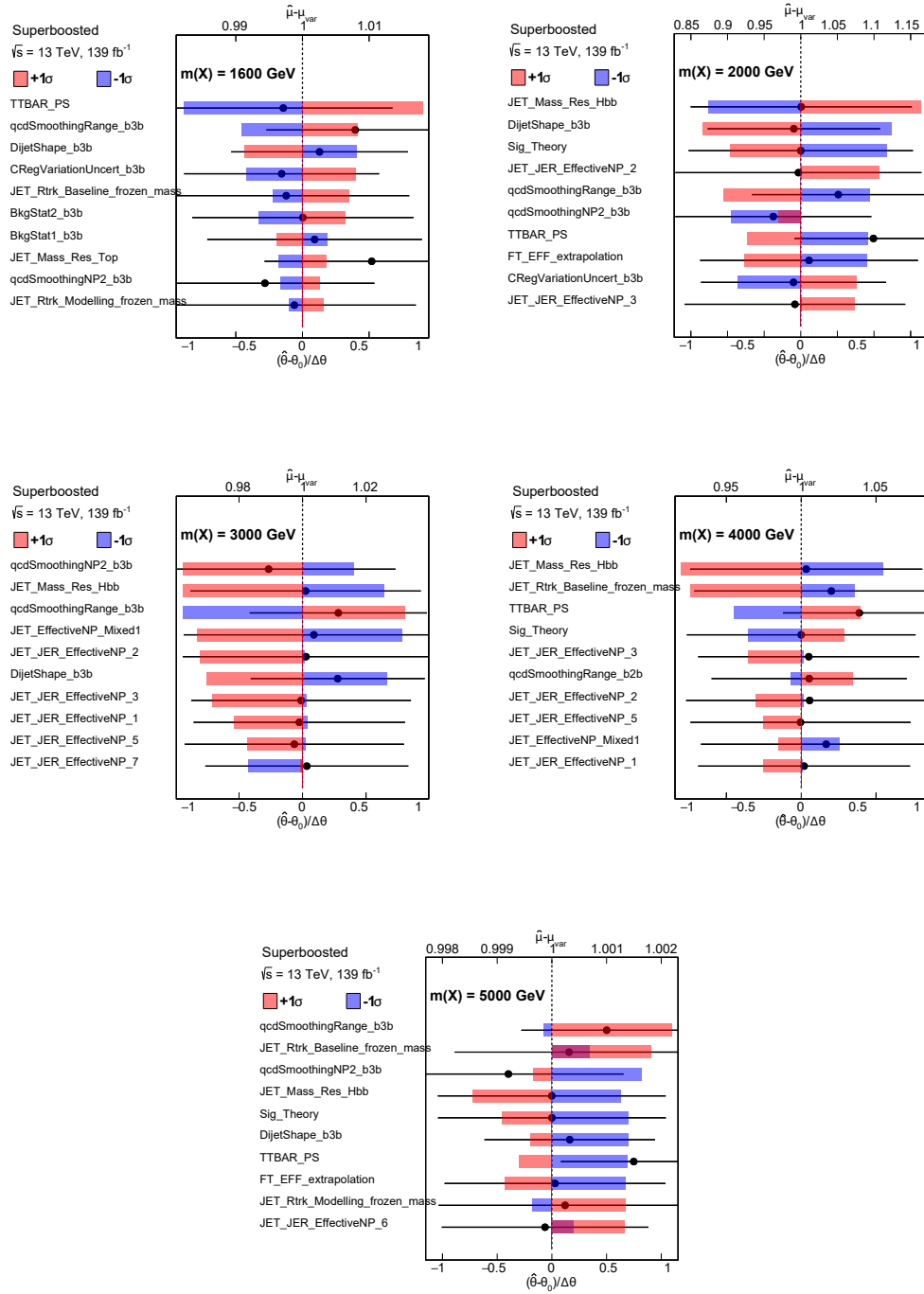


Figure 5.48: Pulls (circles and black lines with bottom scale) and impacts for different X masses in the Superboosted analysis.

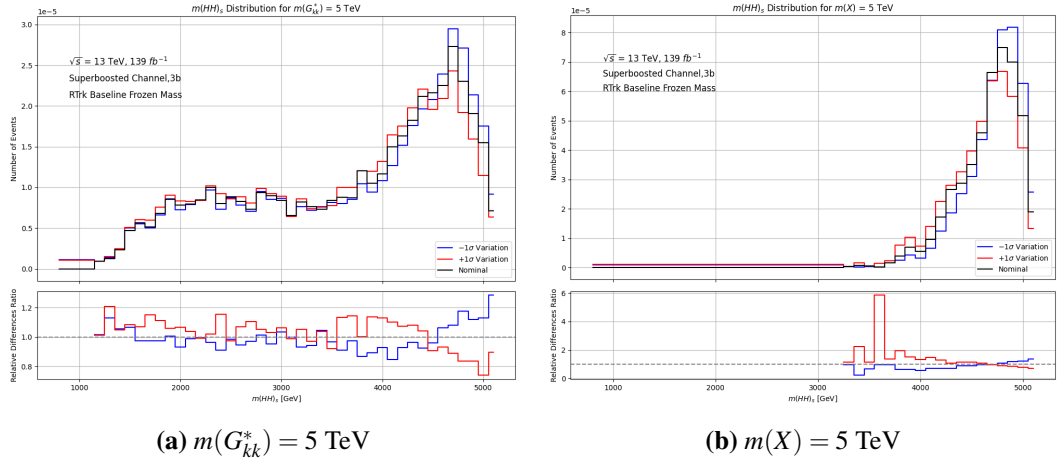


Figure 5.49: Comparison of signal $m(HH)_s$ distributions under $\pm 1\sigma$ variations of a Jet Mass Scale (JMS) uncertainty component for the 5 TeV G_{kk}^* and X MC samples. The X exhibits stronger asymmetric distortion due to its narrower natural width, whereas the broader G_{kk}^* resonance displays a more balanced shift.

Abbreviation	Nuisance Parameter Name / High-tag Category
b4b	4b category
b3b	3b or 3b+ category
b2b	2b category
Sig_Theory	Theoretical uncertainty for signal
BkgStat	Statistical uncertainty on background modelling
DijetShape	Data-driven to MC simulation-driven comparison
ExtrapUncert	Extrapolation uncertainty
CRegVariationUncert	Control region variation
qcdSmoothingNP	Background smoothing uncertainty (dimensionless parameters p_k)
qcdSmoothingRange	Background smoothing uncertainty — fitting range
qcdSmoothingFcn	Background smoothing uncertainty — functional forms
TTBAR_ME	$t\bar{t}$ modelling — Matrix element
TTBAR_PS	$t\bar{t}$ modelling — Parton shower
TTBAR_hdamp	$t\bar{t}$ modelling — h_{damp} parameter
TTBAR_FSR	$t\bar{t}$ modelling — Final state radiation
JET	Jet energy scale uncertainty components
JET_MassRes_Hbb	Jet mass resolution — $H \rightarrow b\bar{b}$ modelling
JET_MassRes_Top	Jet mass resolution — $t\bar{t}$ modelling
JET_RTrk	Jet mass scale uncertainty components (Superboosted jets)
JET_JER	Jet energy resolution uncertainty components (Superboosted jets)
JET_CombMass	Jet mass scale uncertainty components (Boosted jets)
FT_EFF	b -tagging scale factor calibration uncertainty components

Table 5.12: Summary of abbreviations for nuisance parameter names used in the pull and impact plots for the Boosted and Superboosted analyses.

5.7.3.3 Relative Impact of Groups of Systematic Uncertainties

The relative impact of each category of systematic uncertainties on the analysis sensitivity is evaluated by calculating the expected limits while fixing all components of a single group of nuisance parameters to their best-fit values. Six groups of nuisance parameters, representing one or two categories of systematic uncertainties (outlined in Table 5.10), are considered. The expected limits obtained with fixed nuisance parameters are compared to the expected limits derived without fixing any parameters. This comparison highlights the contribution of each systematic uncertainty category to the overall analysis sensitivity. Tables 5.13–5.16 summarize the impact of each category of systematic uncertainties for the Boosted and Superboosted analyses.

Uncertainty category	Relative Impact (%)					
	1000	1600	2000	3000	4000	5000
JMR	4.36	7.12	15.78	9.11	7.8	6.45
JES+JMS	0.05	1.23	0.83	4.41	1.7	0.51
B-Tagging Calibration	0.62	2.09	2.90	0.50	2.11	6.37
Theory ($t\bar{t}$ background)	5.59	0.49	0.56	0.50	2.42	0.18
Theory (Signal)	1.75	1.4	1.30	1.37	2.1	1.07
Background $m(HH)$ model	20.8	1.1	0.69	1.02	1.0	1.02
All systematic uncertainties	34.91	13.4	21.22	14.61	15.6	14.9

Table 5.13: Relative impact on expected limits for each systematic uncertainty in the Heavy Neutral Scalar boson (X) model in the Boosted analysis.

Uncertainty category	Relative Impact (%)					
	1000	1600	2000	3000	4000	5000
JMR	5.72	8.87	15.66	8.30	6.51	6.78
JES+JMS	0.047	2.30	1.62	5.50	1.31	0.68
B-Tagging Calibration	0.79	1.73	2.96	0.66	6.88	5.59
Theory ($t\bar{t}$ background)	5.34	0.91	1.08	0.77	0.03	0.45
Theory (Signal)	1.74	1.46	1.51	1.45	0.48	1.13
Background $m(HH)$ model	31.88	1.81	1.30	1.57	1.86	1.96
All systematic uncertainties	46.84	16.58	22.03	15.28	16.37	15.44

Table 5.14: Relative impact on expected limits for each systematic uncertainty in the Excited Kaluza-Klein Graviton (G_{kk}^*) model in the Boosted analysis.

Uncertainty category	Relative Impact (%)				
	1600	2000	3000	4000	5000
JMR	4.52	2.91	3.92	5.44	4.08
JES+JMS+JER	0.52	0.75	0.29	0.48	0.24
B-Tagging Calibration	1.02	2.02	0.62	2.09	4.88
Theory ($t\bar{t}$ background)	9.61	0.42	0.16	0.03	0.3
Theory (Signal)	1.78	1.46	1.27	1.33	0.96
Background $m(HH)$ model	8.90	4.43	2.99	1.29	0.27
All systematic uncertainties	24.91	11.52	8.94	10.76	11.54

Table 5.15: Relative impact on expected limits for each systematic uncertainty in the Heavy Neutral Scalar boson (X) model in the Superboosted analysis.

Jet Mass Resolution (JMR) and b -tagging scale factor uncertainties have the most significant impact at higher resonant masses for both analyses. The Superboosted analysis demonstrates slightly lower JMR uncertainties in these regions due to superior jet mass resolution performance at higher resonant masses. Consequently, signal loss from smearing is reduced in Superboosted jets compared to the forward-folding method employed in the Boosted analysis (Figures 5.41 and 5.42).

The dominant source of b -tagging scale factor uncertainty arises from the large extrapolation uncertainty associated with high- p_T jets containing b -hadrons, as determined from MC simulation. This is primarily due to the limited number of $t\bar{t}$

Uncertainty category	Relative Impact (%)				
	1600	2000	3000	4000	5000
JMR	4.24	3.15	5.15	6.17	4.07
JES+JMS+JER	0.35	0.48	0.49	0.33	0.4
B-Tagging Calibration	1.12	2.31	0.87	1.00	3.87
Theory ($t\bar{t}$ background)	9.16	1.02	0.23	0.16	0.16
Theory (Signal)	1.81	1.56	1.73	1.37	1.08
Background $m(HH)$ model	11.78	6.89	4.73	1.46	0.79
All systematic uncertainties	26.71	14.46	12.43	10.44	10.52

Table 5.16: Relative impact on expected limits for each systematic uncertainty in the Excited Kaluza-Klein Graviton (G_{kk}^*) model in the Superboosted analysis.

events decaying into high- p_T b -hadrons in the data (Figure 4.4) [10][99].

Background $m(HH)_i$ modeling uncertainties contribute significantly to the overall systematic uncertainty at low resonant masses. This is largely attributed to mismodeling between MC-driven and data-driven estimates in tag regions with limited statistical power, such as the $4b$ category in the Boosted analysis and the $3b+$ category in the Superboosted analysis. In addition, background normalization uncertainties, including extrapolations from Gaussian processing (Boosted analysis) and variations in control regions (Superboosted analysis), further contribute to the overall uncertainty. These components are visible in the constraint plots shown in Figures 5.45–5.48.

5.7.4 Signal Region Event Yields

Event yields are calculated by integrating the bins within the central region of the $m(HH)_i$ signal histogram until approximately 90% of the total events are encompassed for each signal distribution. The selected bins are subsequently applied to the data and background histograms to determine their respective event yields. Each signal histogram is normalized to 1 fb. Tables 5.17–5.20 present the event yields for background, signal, and data at various masses of G_{kk}^* and X .

Mass of X (GeV)	Tag region	Event Yields		
		Data	Background	Signal
1000	3b	1076	1109.3 ± 28.0	11.2 ± 2.0
	4b	62	58.4 ± 5.8	3.9 ± 0.7
2000	2b	77	77.8 ± 3.5	2.7 ± 0.9
	3b	14	11.5 ± 0.8	5.2 ± 1.6
	4b	2	0.8 ± 0.1	2.5 ± 1.1
	2b	7	5.8 ± 0.5	2.0 ± 0.8
3000	3b	0	0.6 ± 0.2	2.4 ± 0.8
	4b	0	0.1 ± 0.1	0.7 ± 0.4
	2b	3	1.1 ± 0.2	4.9 ± 1.9
4000	3b	0	0.1 ± 0.1	3.5 ± 1.0
	2b	0	0.7 ± 0.2	2.5 ± 1.1
5000	3b	0	0.01 ± 0.1	1.1 ± 0.3
	2b	0	0.7 ± 0.2	2.5 ± 1.1

Table 5.17: Event yields for the selection of individual masses of the Heavy Neutral Scalar Boson X in the Boosted analysis.

Mass of G_{kk}^* (GeV)	Tag region	Event Yields		
		Data	Background	Signal
1000	3b	1282	1298.2 ± 29.4	12.0 ± 2.1
	4b	72	71.0 ± 6.9	4.1 ± 0.8
2000	2b	223	221.0 ± 8.4	3.8 ± 1.2
	3b	38	33.5 ± 1.7	7.1 ± 2.1
	4b	3	2.2 ± 0.3	3.2 ± 1.3
	2b	19	16.5 ± 1.2	2.4 ± 0.9
3000	3b	3	2.8 ± 0.4	2.8 ± 0.8
	4b	1	0.6 ± 0.2	0.8 ± 0.4
	2b	10	7.5 ± 0.7	8.6 ± 3.4
4000	3b	6	4.1 ± 0.6	6.5 ± 1.7
	2b	18	13.6 ± 1.1	3.5 ± 1.4
5000	3b	41	35.5 ± 1.9	2.1 ± 0.5

Table 5.18: Event yields for the selection of individual masses of the Excited Kaluza-Klein Graviton G_{kk}^* in the Boosted analysis.

Mass of X (GeV)	Tag region	Event Yields		
		Data	Background	Signal
2000	2b	108	102.0 ± 6.2	8.8 ± 1.1
	3b+	14	15.9 ± 1.9	8.8 ± 1.1
3000	2b	11	14.0 ± 1.2	4.1 ± 0.5
	3b+	5	1.7 ± 0.4	4.2 ± 0.5
4000	2b	8	2.3 ± 0.3	3.6 ± 0.4
	3b+	0	0.2 ± 0.1	3.6 ± 0.4
5000	2b	1	0.5 ± 0.1	4.2 ± 0.5
	3b+	0	0.03 ± 0.1	4.2 ± 0.5

Table 5.19: Event yields for the selection of individual masses of the Heavy Neutral Scalar Boson X in the Superboosted analysis.

Mass of G_{kk}^* (GeV)	Tag region	Event Yields		
		Data	Background	Signal
2000	2b	177	174.2 ± 10.3	13.0 ± 1.5
	3b+	22	27.6 ± 3.3	13.0 ± 1.5
3000	2b	22	27.7 ± 2.2	5.4 ± 0.6
	3b+	7	4.6 ± 0.8	5.4 ± 0.6
4000	2b	18	11.1 ± 1.1	4.3 ± 0.5
	3b+	9	6.5 ± 1.1	4.3 ± 0.5
5000	2b	20	18.4 ± 1.7	4.3 ± 0.5
	3b+	18	16.3 ± 2.0	4.1 ± 0.4

Table 5.20: Event yields for the selection of individual masses of the Excited Kaluza-Klein Graviton G_{kk}^* in the Superboosted analysis.

5.7.5 Observed Limits and Resolved+Boosted Combination

5.7.5.1 Calculating Local Significance

After unblinding the results, the observed limit is evaluated to determine whether the background-only hypothesis can be rejected in favor of the alternative (signal) hypothesis. This is achieved by calculating the local p-value using Equation 5.25:

$$p_0 = 1 - \Phi(\sqrt{q_0}), \quad (5.25)$$

where Φ denotes the Gaussian cumulative distribution function and q_0 represents the test statistic from Equation 5.20 under the background-only hypothesis ($\mu = 0$). The local significance, Z_0 , is then obtained from the inverse cumulative distribution function Φ^{-1} , as shown in Equation 5.28:

$$Z_0 = \Phi^{-1}(1 - p_0). \quad (5.26)$$

If the p-value is less than 10^{-7} , the background-only hypothesis is rejected in favor of the alternative hypothesis, corresponding to a 5σ deviation from the background-only expectation. Conversely, if the p-value is close to 0.5 (typically corresponding to Z_0 within 2σ), the alternative hypothesis is rejected, and the background-only hypothesis is retained. For the Boosted and Superboosted analyses, the alternative hypothesis corresponds to the presence of Excited Kaluza-Klein Graviton (G_{kk}^*) and Heavy Neutral Scalar Boson (X) particles.

5.7.5.2 Look Elsewhere Effect and Global Significance

If the local significance at a mass point is sufficiently large, additional statistical tests must be performed to account for the Look Elsewhere Effect (LEE) [100], in order to obtain the global p-value. The LEE corrects the local significance for multiple searches conducted within the specified resonant mass range, ensuring that a large upward fluctuation at a particular mass point is not attributed to chance. The global p-value, p_{global} , is calculated using the Gross-Vitells method, as shown in Equation 5.27:

$$p_{\text{global}} = N_{\text{up}} e^{-\frac{1}{2}(Z_0^2 - Z_{\text{ref}}^2)} + p_0, \quad (5.27)$$

where:

- N_{up} is the number of upward crossings above the reference value Z_{ref} within the search range,
- Z_0 is the local significance,
- p_0 is the local p-value.

The global significance, Z_{global} , is obtained from the inverse cumulative distribution function Φ^{-1} , as shown in Equation 5.28:

$$Z_{\text{global}} = \Phi^{-1}(1 - p_{\text{global}}). \quad (5.28)$$

5.7.5.3 Observed Limits

Figures 5.50 and 5.51 show the observed (unblinded) limits, including systematic uncertainties, for the Boosted and Superboosted analyses. No significant excess is observed in any model in either analysis.

The largest deviation in the Boosted analysis occurs around $m_R = 4$ TeV, with deviations of 1.41σ for the G_{kk}^* model and 1.8σ for the X model. However, the global significance in the Boosted analysis reduces to approximately 0σ for the G_{kk}^* model and 0.25σ for the X model.

In the Superboosted analysis, the largest deviation also occurs at $m_R = 4$ TeV, with values of 1.95σ for the G_{kk}^* model and 1.56σ for the X model. Nevertheless, the global significance reduces to approximately 0.64σ for the G_{kk}^* model and 0.01σ for the X model, indicating that the deviations are not statistically significant. Figure 5.52 shows the observed local p-values, p_0 , for all signal masses in both the Boosted and Superboosted analyses.

5.7.5.4 Resolved and Boosted Analyses Combined Upper Limits

For the current $HH \rightarrow b\bar{b}b\bar{b}$ analysis, the upper limits from the Resolved and Boosted analyses are combined to obtain a single upper limit on the production

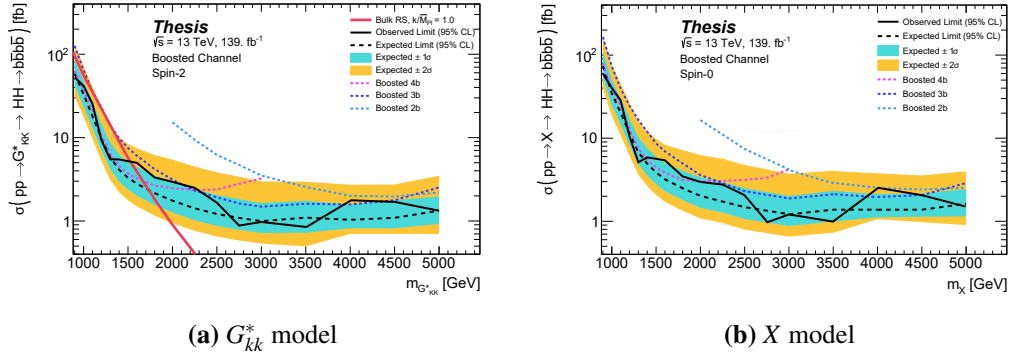


Figure 5.50: Observed limits for the G_{kk}^* and X models in the Boosted analysis with systematic uncertainties.

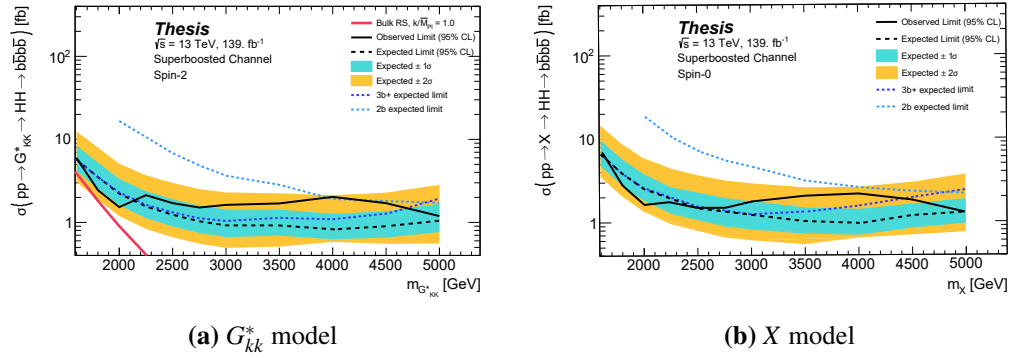


Figure 5.51: Observed limits for the G_{kk}^* and X models in the Superboosted analysis with systematic uncertainties.

cross section. This combination includes the G_{kk}^* and X models over the mass range of 251 GeV to 5 TeV. All uncertainties between the Resolved and Boosted analyses are treated as uncorrelated, except for the signal theory and luminosity uncertainties, which are considered correlated.

The combined results reveal no significant excess, with the largest deviation observed at a mass point of 1100 GeV. The local significance of this deviation is 2.3σ for the X model and 2.5σ for the G_{kk}^* model. The dominant contribution arises from the Resolved analysis, with corresponding deviations of 2.25σ and 2.31σ for the X and G_{kk}^* models, respectively (see Figure 5.54). However, the global significance at 1100 GeV for the combined observed limit is 0.4σ for the X model and 0.5σ for the G_{kk}^* model [13], indicating that the observed deviations are not statistically significant. Figure 5.53 presents the combined limits from the Resolved

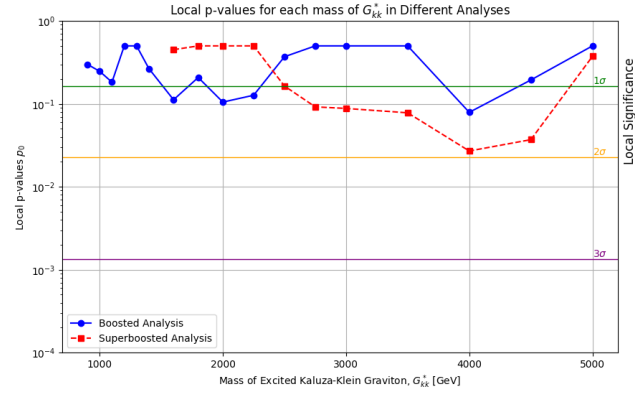
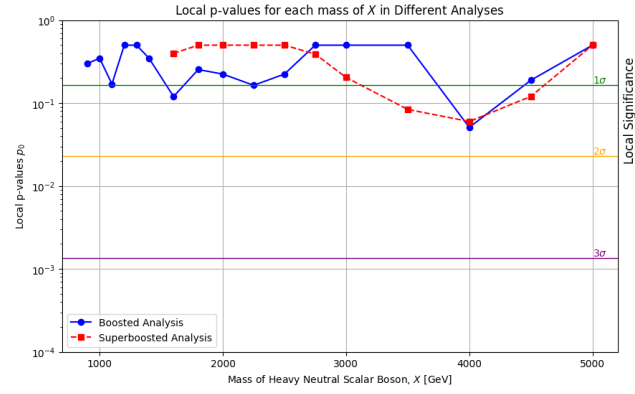
(a) Kaluza-Klein Graviton (G_{kk}^*) model(b) Heavy Neutral Scalar Boson (X) model

Figure 5.52: Observed local p -values for the Excited Kaluza-Klein Graviton (G_{kk}^*) and Heavy Neutral Scalar Boson (X) models in the Boosted and Superboosted analyses.

and Boosted analyses.

Due to time constraints, the combination of the Superboosted analysis with the Resolved and Boosted analyses has not been performed in this thesis. However, a full combination of all three analyses is recommended for future work.

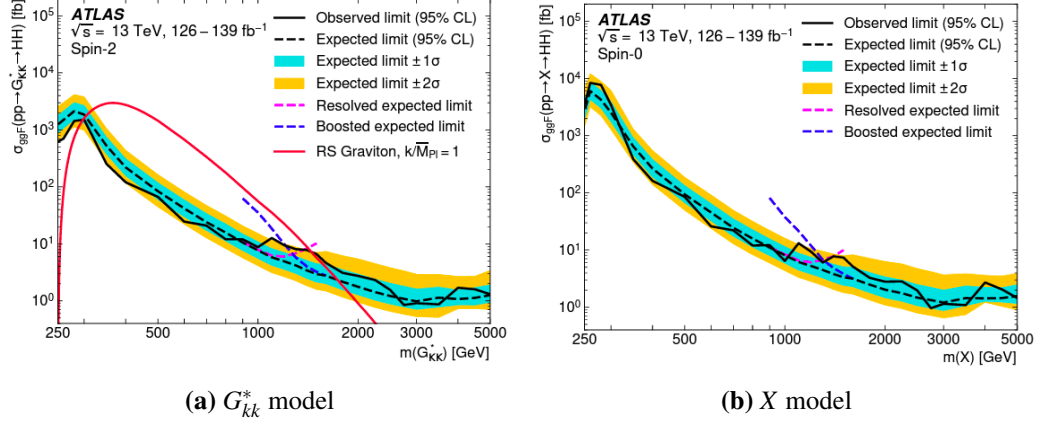


Figure 5.53: Expected and observed 95% CL upper limits on the production cross section for resonant HH decays into two $b\bar{b}$ pairs, from the combined Boosted and Resolved analyses [13].

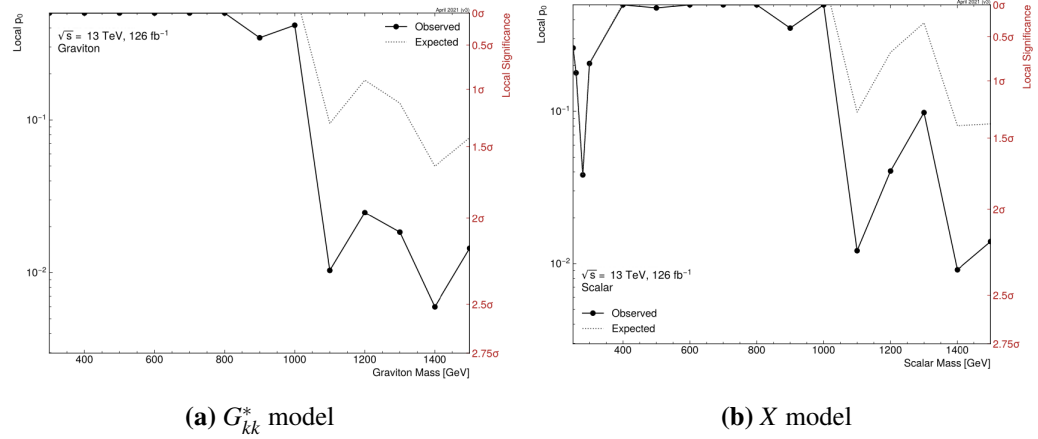


Figure 5.54: Local p -values as a function of the resonance mass for the G_{kk}^* and X models in the Resolved analysis.

Chapter 6

Conclusions

A search for resonant Higgs boson pair production via the process $pp \rightarrow HH \rightarrow b\bar{b}b\bar{b}$ has been performed using 139 fb^{-1} of Run-2 data collected by the ATLAS detector. The analysis was conducted using both the Boosted and Superboosted strategies. No significant signal excess is observed in either analysis for the G_{kk}^* and X models. Similarly, the combination of the Resolved and Boosted analyses for the G_{kk}^* and X models does not yield any significant excess.

Appendix A

Ghost Association vs ΔR Track Jet Matching Comparison for Boosted Jets

This section presents a comparison study between Ghost Association (GA) and ΔR -based matching of track jets to calorimeter jets in the context of Boosted jets. In the current analysis, Boosted jets employ the GA technique to match track jets to calorimeter jets. However, this method has not been developed for small- R jets used in the Superboosted regime within ATLAS. Consequently, the ΔR matching method is adopted as an alternative for Superboosted jets.

To validate the applicability of ΔR -based matching for Superboosted jets, a cross-check is performed by comparing GA and $\Delta R = 1.0$ matching on Boosted jets. The results, shown in Figures A.1 to A.4, indicate that both methods yield very similar distributions in the reconstructed masses $m(HH)$, $m(H_1)$, and $m(H_2)$, demonstrating the viability of the ΔR method.

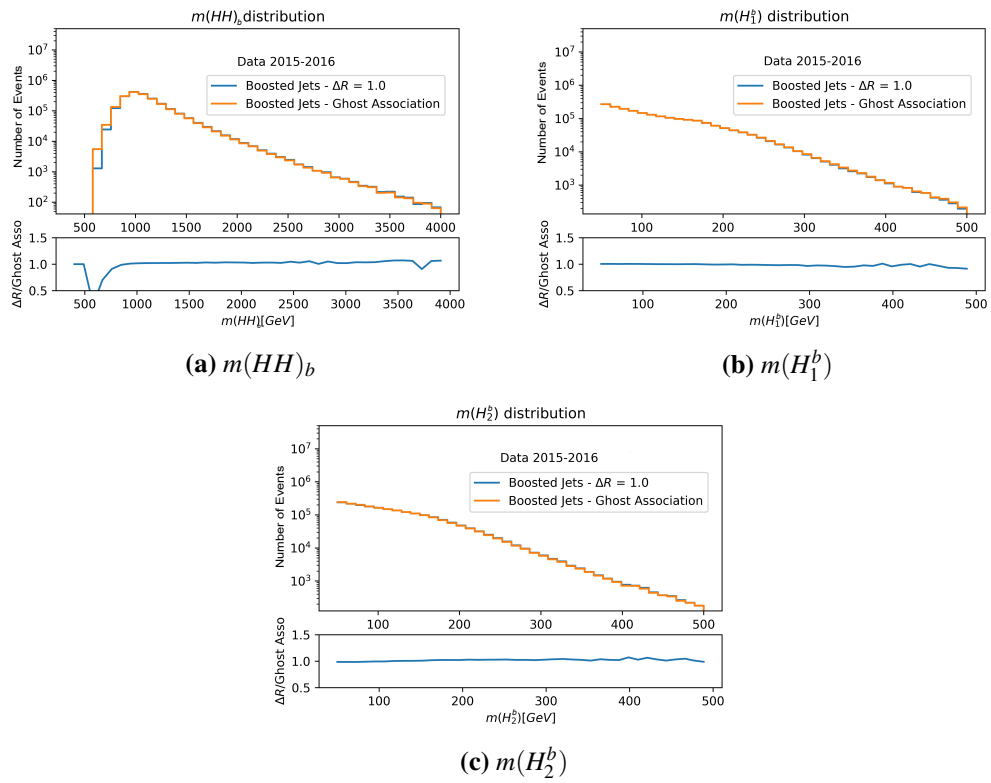


Figure A.1: Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ track jet matching in data without a Higgs mass cut.

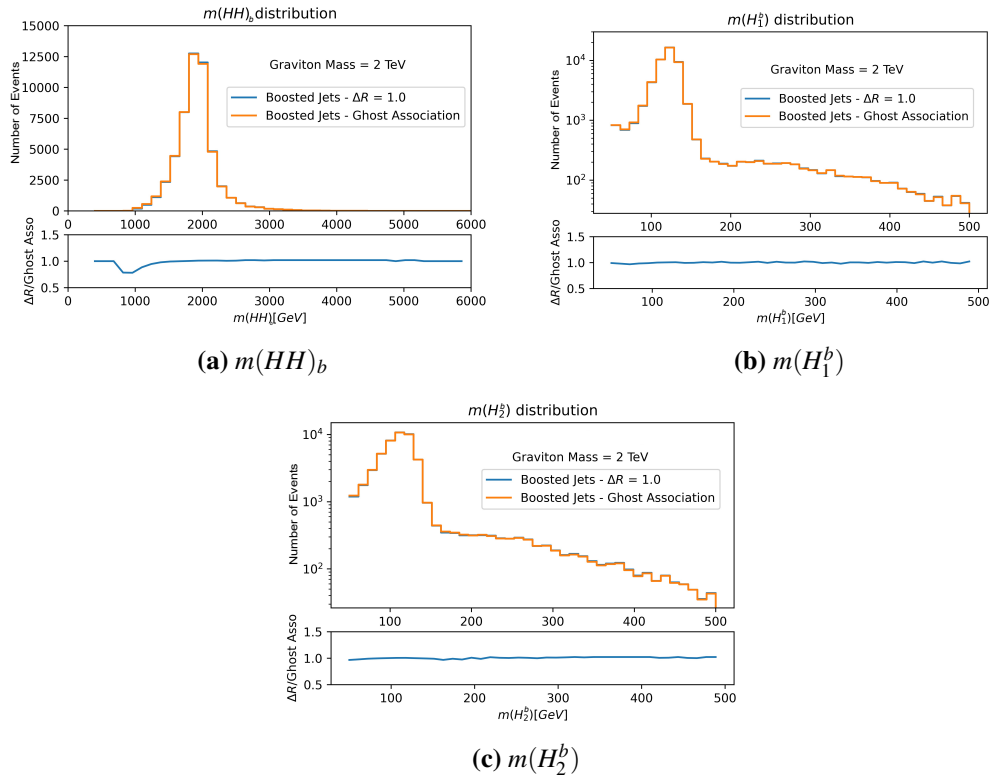


Figure A.2: Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 2$ TeV graviton samples without a Higgs mass cut.

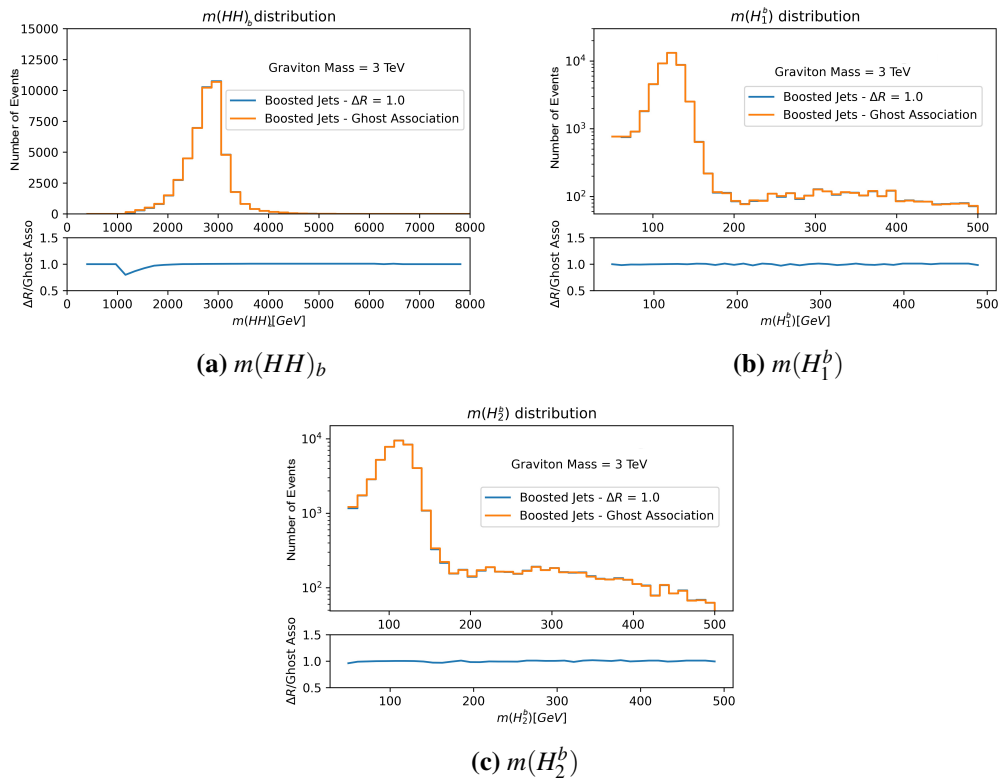


Figure A.3: Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 3$ TeV graviton samples without a Higgs mass cut.

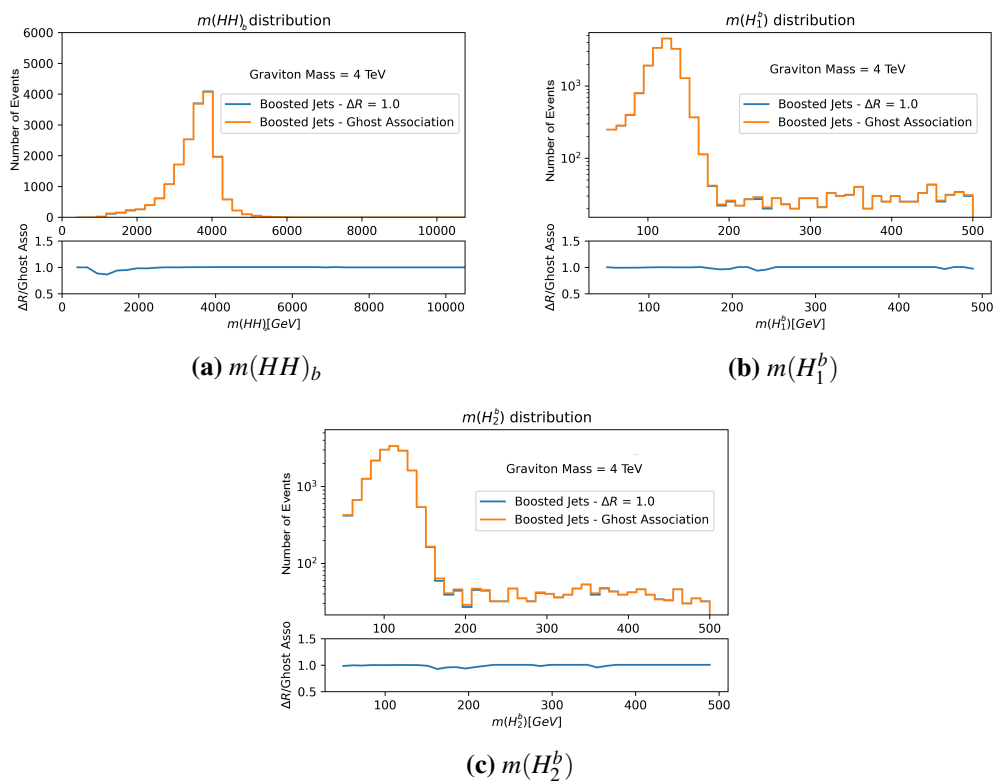


Figure A.4: Comparison of $m(HH)_b$, $m(H_1^b)$, and $m(H_2^b)$ between Ghost Association and $\Delta R = 1.0$ matching for $m_{G_{kk}^*} = 4$ TeV samples without a Higgs mass cut.

Appendix B

Jet Mass Scale (JMS) Uncertainties for Small- R Jets

This section describes the performance of small- R jets across different calorimeter jet m/p_T and p_T regions. Figure B.1 presents the JMS uncertainties for Topocluster jets, derived using the R_{trk} method. The current implementation of the R_{trk} approach assumes no correlation between the track and calorimeter components of Topocluster jets. In contrast, the Particle Flow algorithm integrates tracking information directly into the calorimeter jet clustering process, resulting in differences in reconstructed jet mass and momentum compared to Topocluster jets.

To account for these differences, an additional uncertainty is introduced based on the ratio between Topocluster and Particle Flow jets. Figure B.2 shows the corresponding JMS uncertainties arising from this comparison for small- R jets. For the JMS systematic uncertainties used in the Superboosted analysis, the contributions from Figures B.1 and B.2 are combined in quadrature.

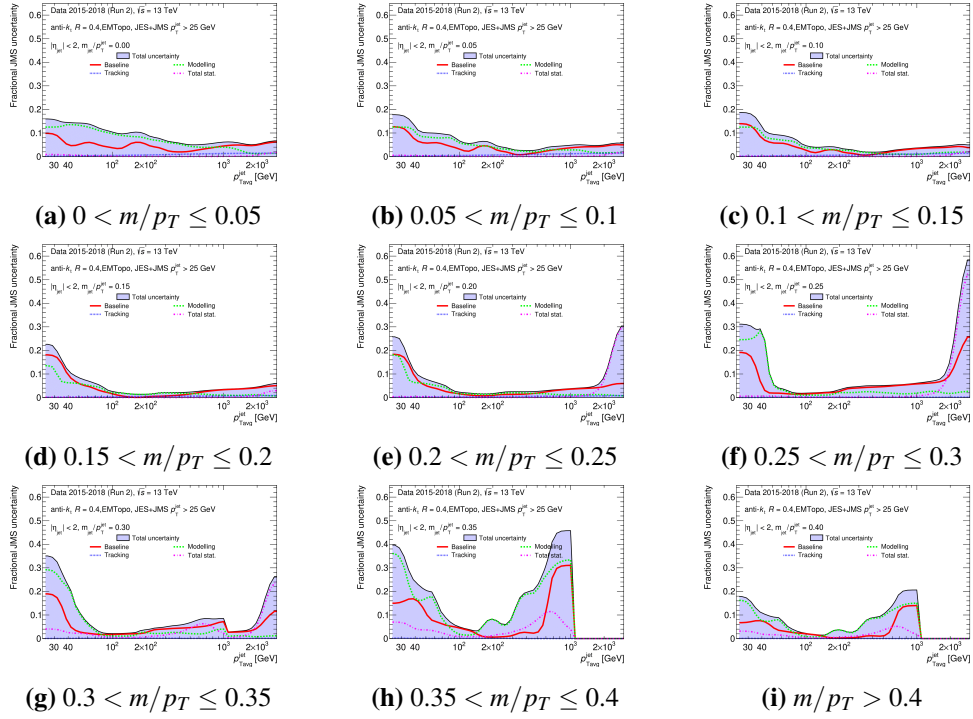


Figure B.1: JMS uncertainties for small- R jets across different jet m/p_T intervals, derived using the R_{trk} method.

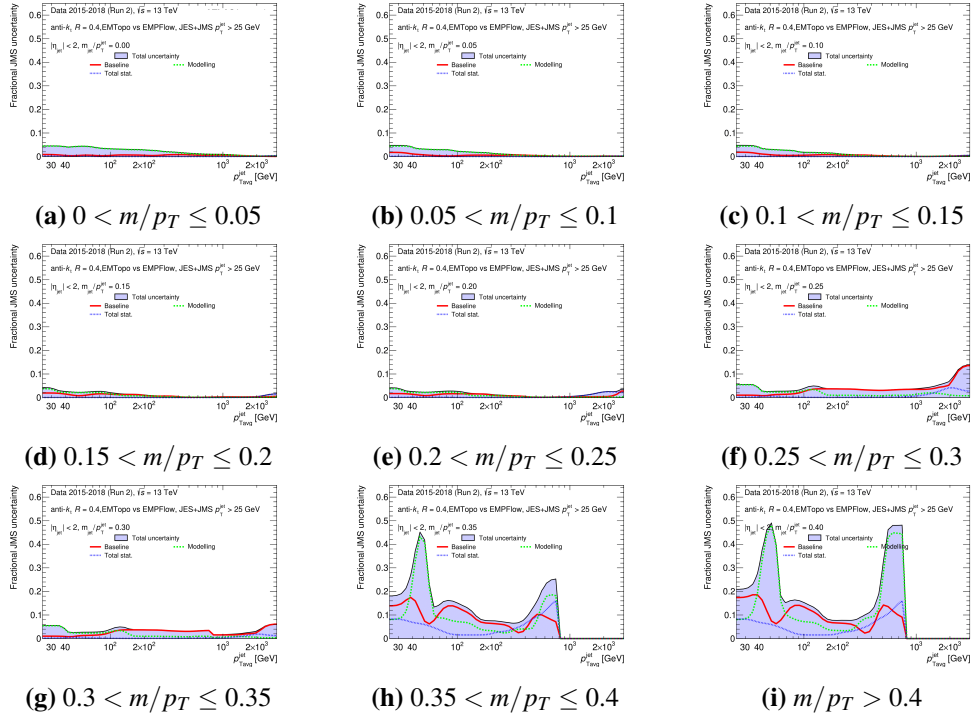


Figure B.2: JMS uncertainties for small- R jets due to differences between Topocluster and Particle Flow algorithms, shown for various m/p_T ranges.

Appendix C

Signal Efficiency and Higgs Boson Candidate Mass Resolution Comparison Between Boosted and Superboosted Analyses

This section presents a preliminary study comparing the signal efficiencies and leading Higgs boson candidate mass resolutions between the Boosted and Superboosted analyses. The motivation for this study is to assess the feasibility of the Superboosted strategy at higher resonant masses.

Signal efficiency for each analysis is defined as the number of events passing the Signal Region (including all high tag categories) divided by the total number of events before event selection in the X Monte Carlo (MC) samples. Subsequently, the distributions of leading Higgs boson candidate masses and transverse momenta are compared between the two analyses, using the same number of MC events for each.

Figure C.1 displays the signal efficiency as a function of the resonant mass for both Boosted and Superboosted analyses. At larger resonant masses ($m_{X_R} \geq 1.8$ TeV), the Superboosted analysis retains more events than the Boosted approach. This improvement is primarily due to the broader Higgs boson candidate mass resolution observed in the Boosted analysis at very high p_T , which leads to a loss in

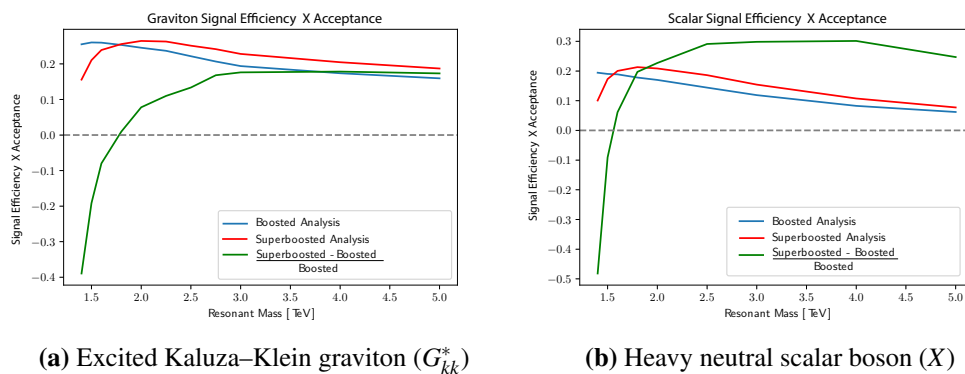


Figure C.1: Signal efficiency comparison between Boosted and Superboosted analyses in the G_{kk}^* and X MC samples. The signal is defined as the number of events passing all high tag categories in the Signal Region.

signal acceptance. In contrast, the Superboosted analysis achieves a narrower mass resolution at high p_T values, as illustrated in Figure C.2, resulting in improved signal retention. These observations suggest that the Superboosted strategy offers advantages for probing highly collimated Higgs boson decays, particularly in the high-mass regime.

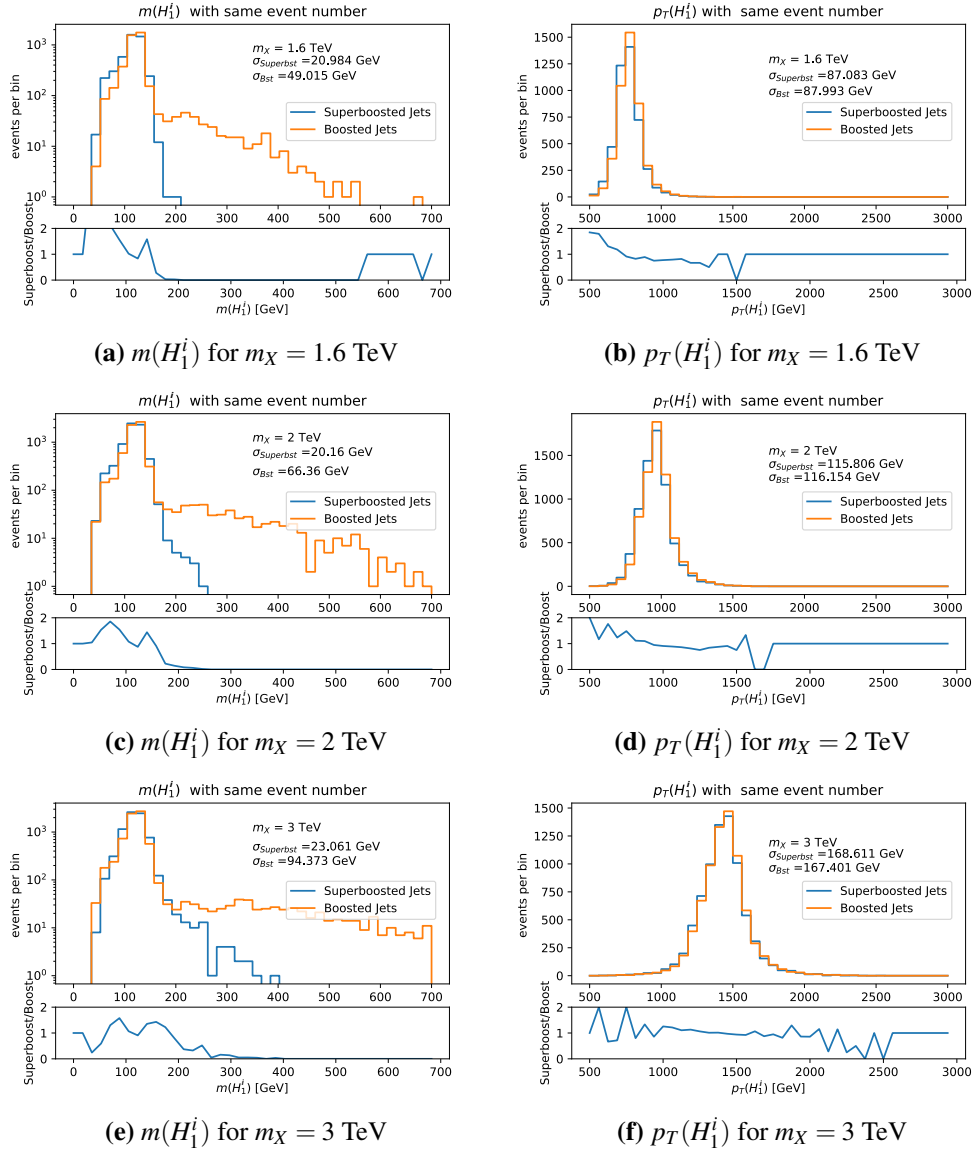


Figure C.2: Comparison of leading Higgs boson candidate mass $m(H_1^i)$ and transverse momentum $p_T(H_1^i)$ between the Boosted and Superboosted analyses for three different m_X values.

Appendix D

Boosted and Superboosted Background kinematics plots

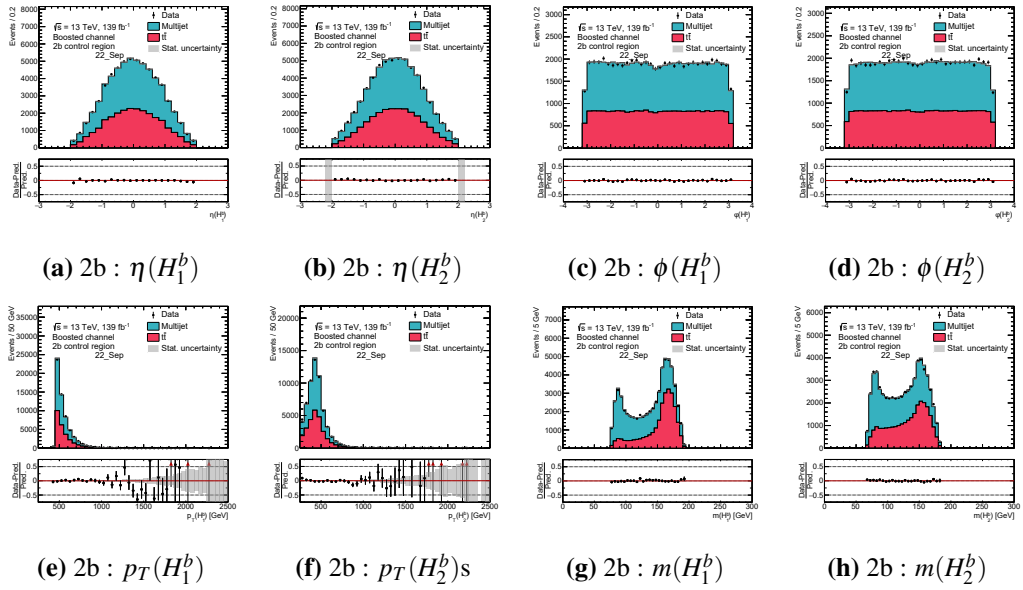


Figure D.1: The Background kinematics in 2b Control Region Boosted Analysis

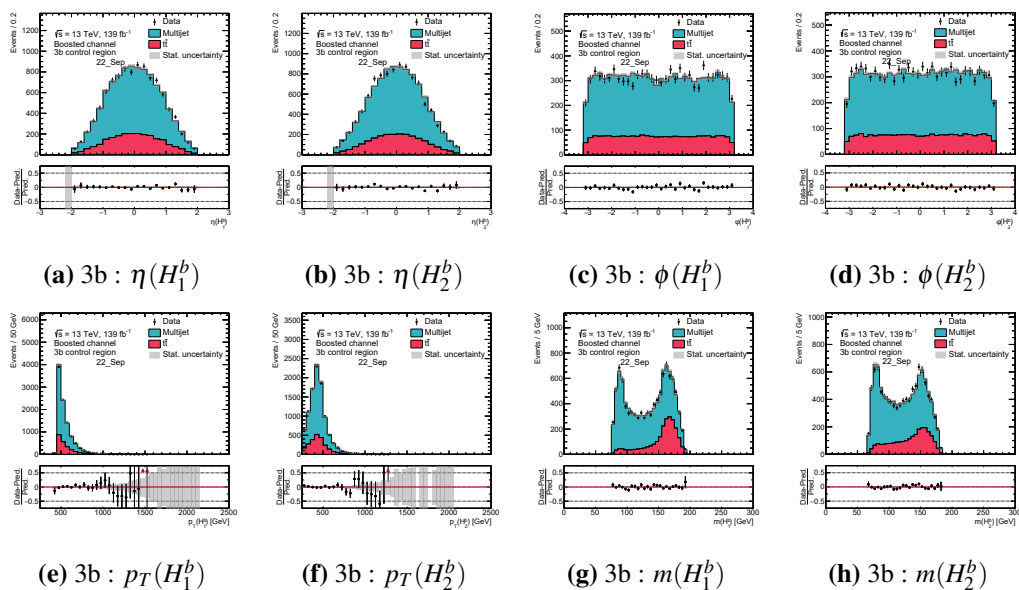


Figure D.2: The Background kinematics in 3b Control Region in Boosted Analysis

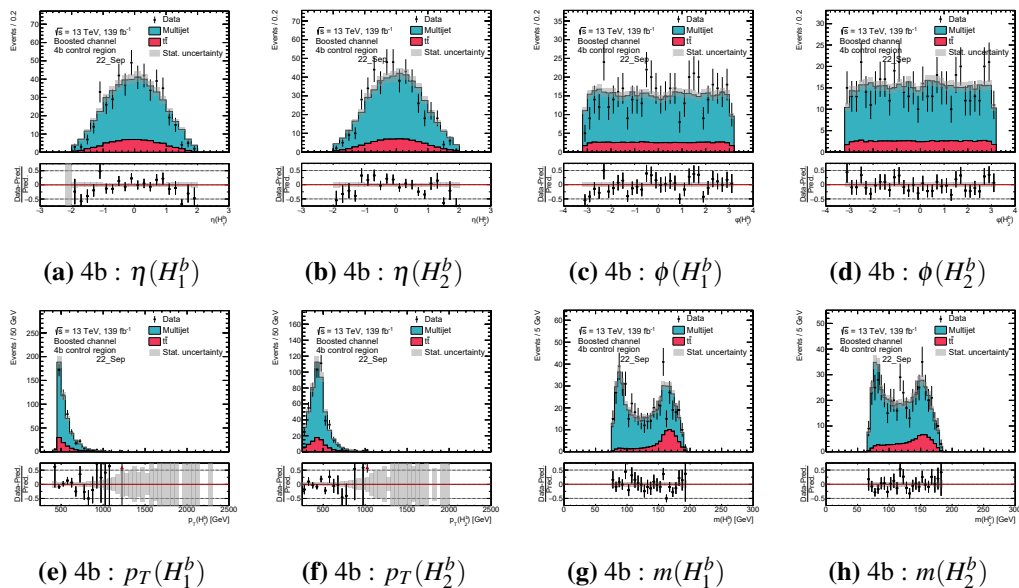


Figure D.3: The Background kinematics in 4b Control Region Boosted Analysis

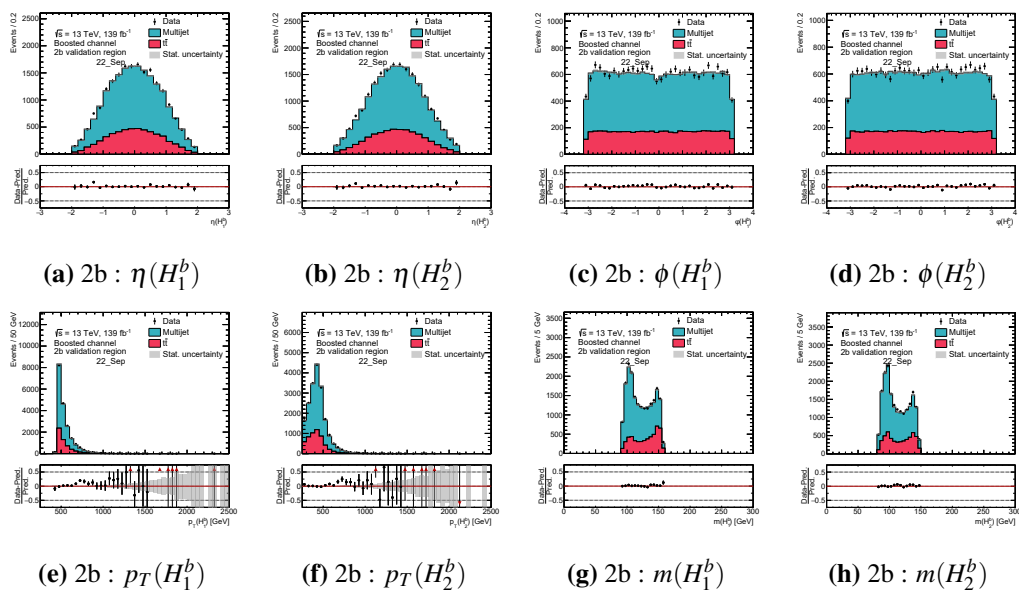


Figure D.4: The Background kinematics in $2b$ Validation Region Boosted Analysis

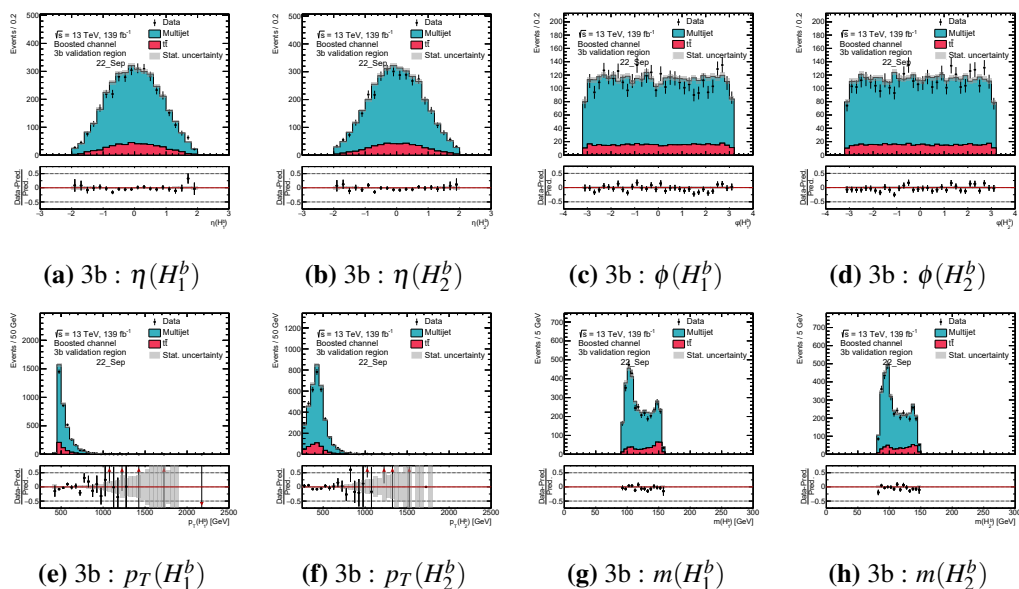


Figure D.5: The Background kinematics in $3b$ Validation Region in Boosted Analysis

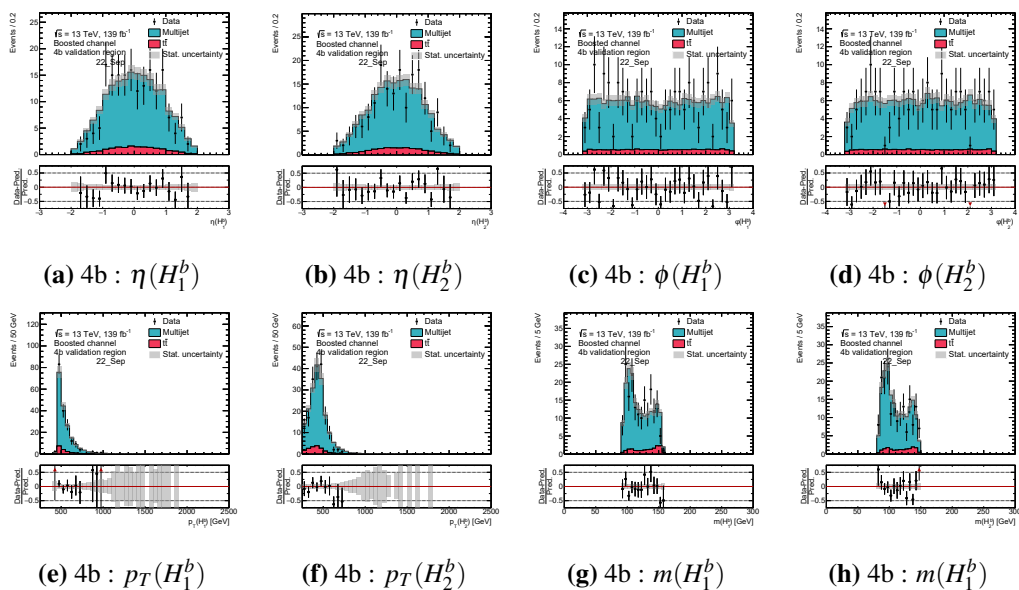


Figure D.6: The Background kinematics in 4b Validation Region Boosted Analysis

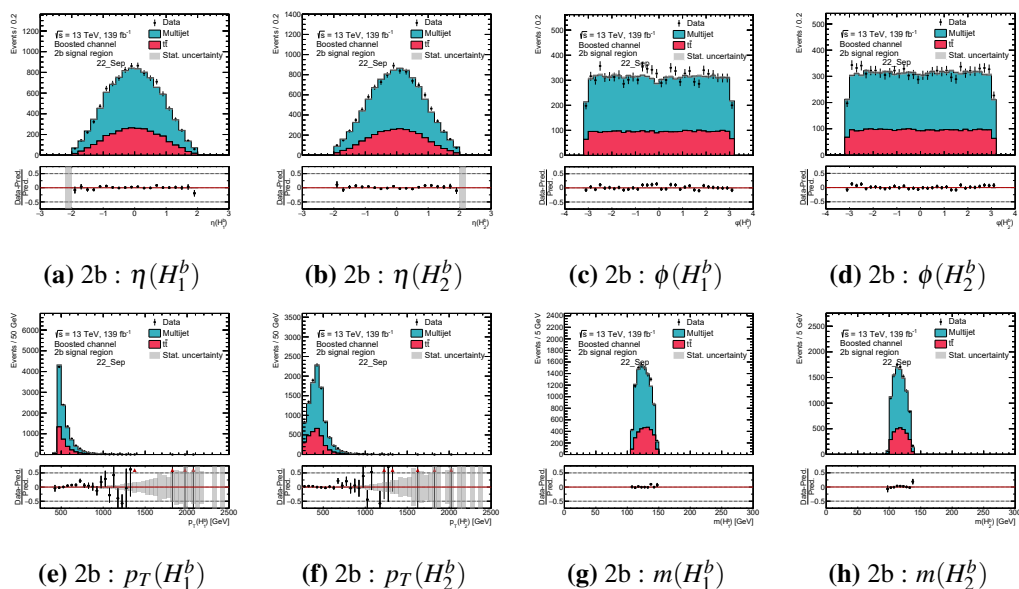


Figure D.7: The Background kinematics in 2b Signal Region Boosted Analysis

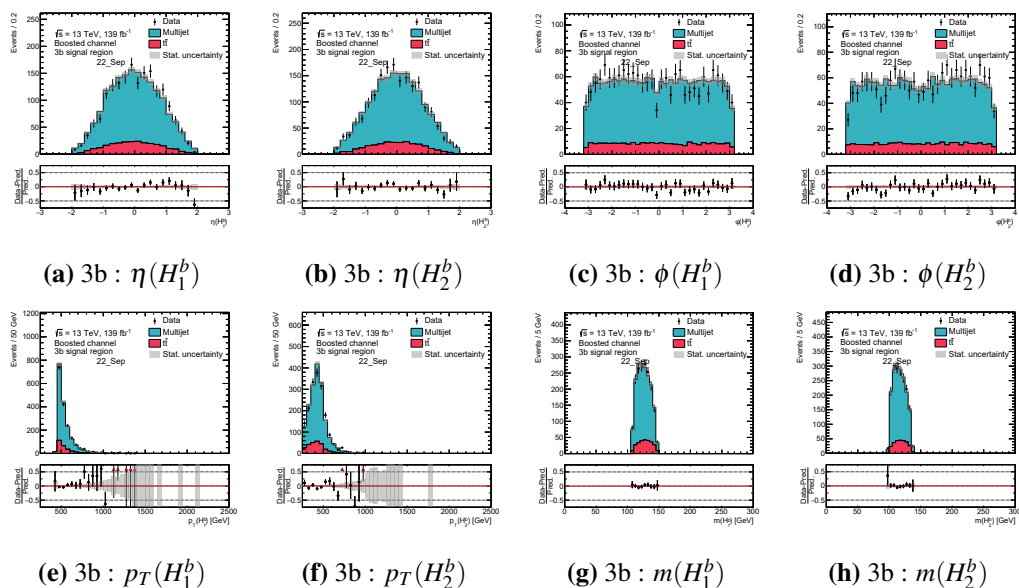


Figure D.8: The Background kinematics in $3b$ Signal Region in Boosted Analysis

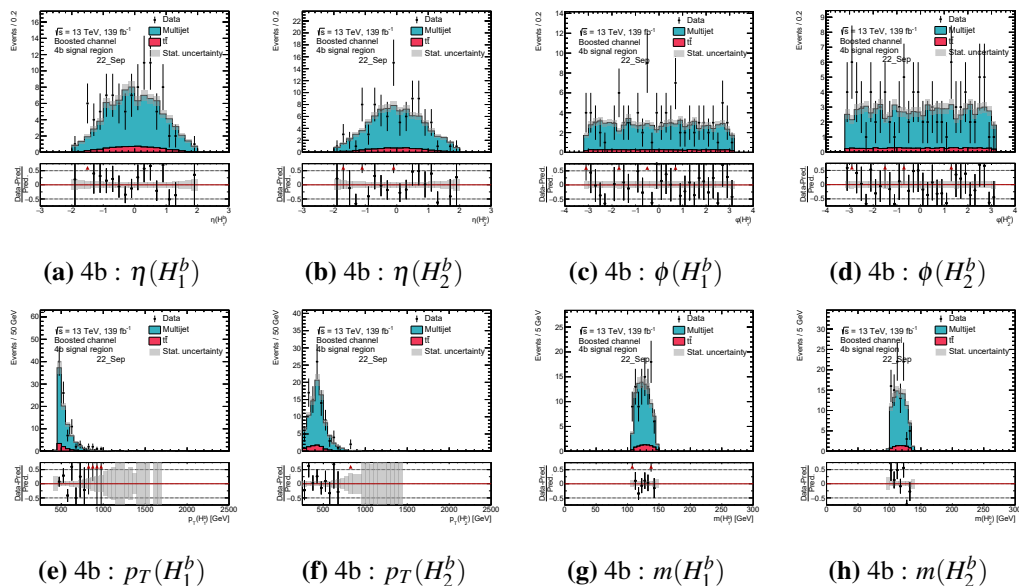


Figure D.9: The Background kinematics in $4b$ Signal Region Boosted Analysis

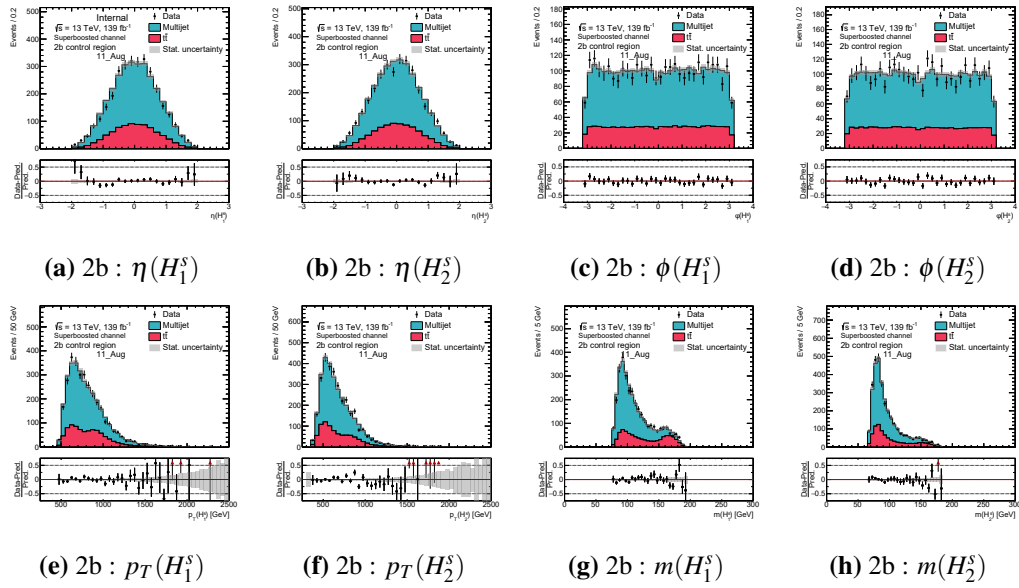


Figure D.10: The Background kinematics in 2b Control Region Superboosted Analysis

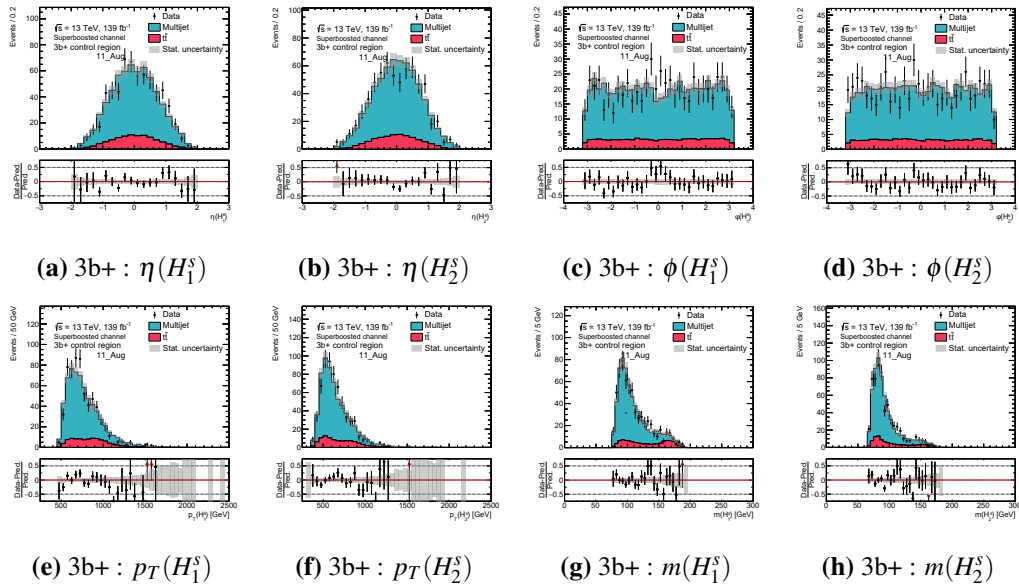


Figure D.11: The Background kinematics in 3b+ Control Region in Superboosted Analysis

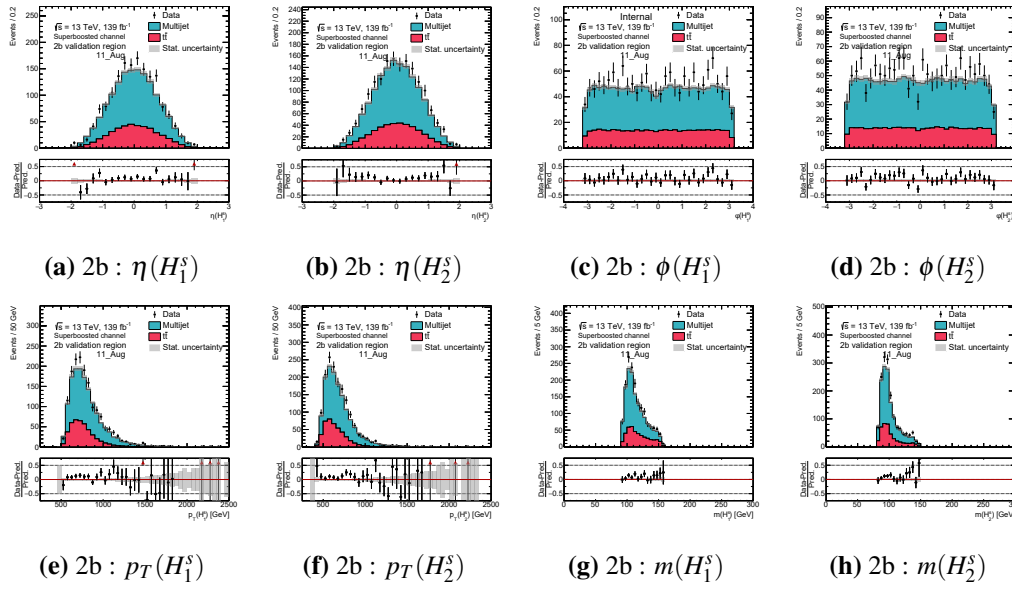


Figure D.12: The Background kinematics in 2b Validation Region Superboosted Analysis

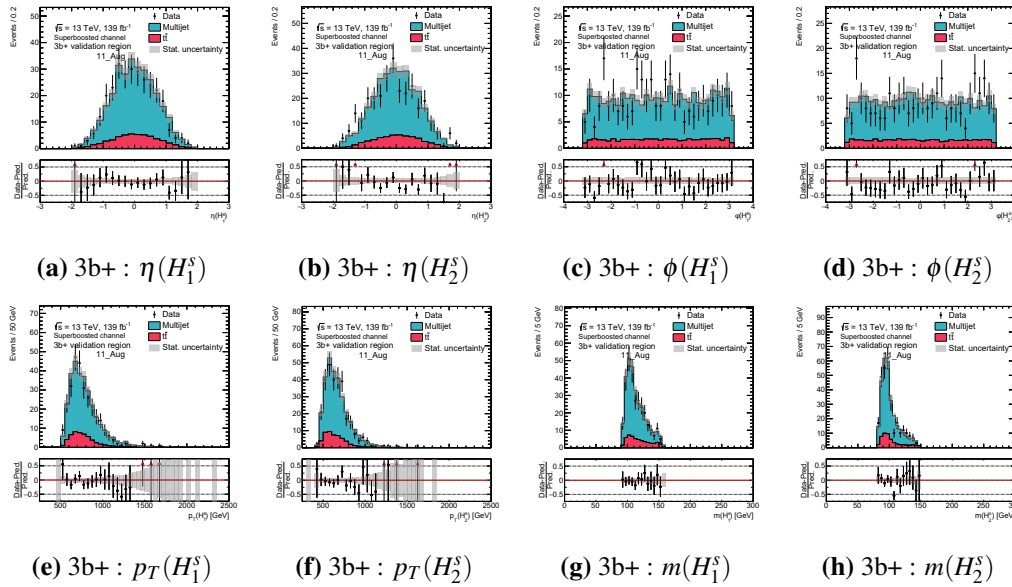


Figure D.13: The Background kinematics in 3b+ Validation Region in Superboosted Analysis

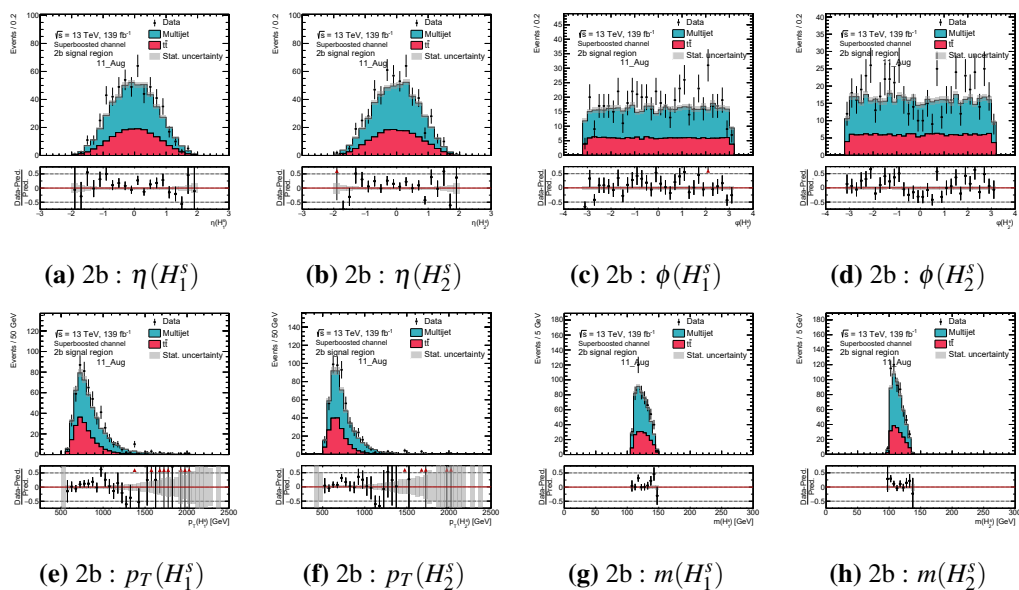


Figure D.14: The Background kinematics in 2b Signal Region Superboosted Analysis

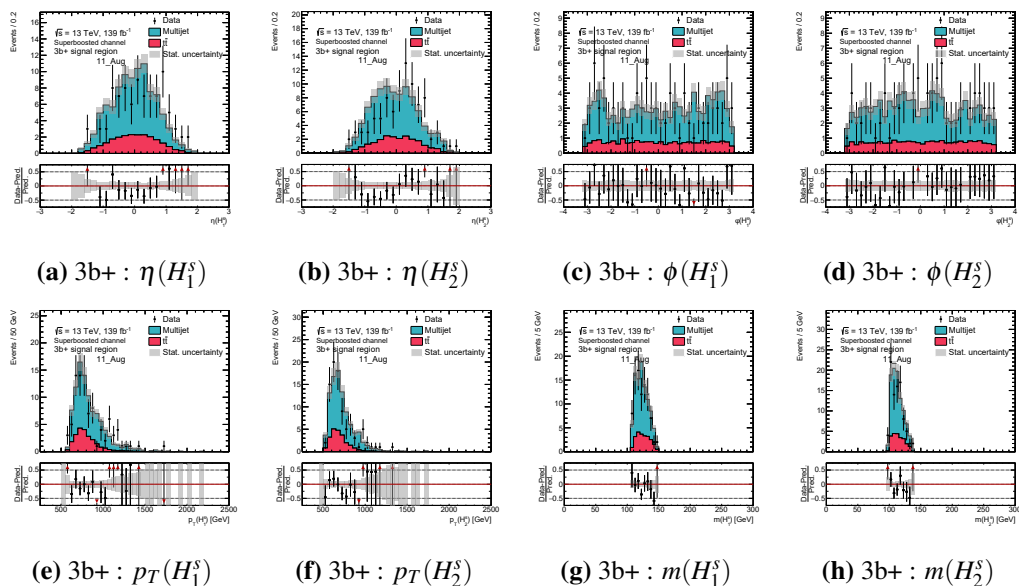


Figure D.15: The Background kinematics in 3b+ Signal Region in Superboosted Analysis

Bibliography

- [1] C. Grojean. Higgs Physics. In *8th CERN–Latin-American School of High-Energy Physics*, pages 143–158, 2016. <https://doi.org/10.5170/CERN-2016-005.143>.
- [2] Jason Veatch. Searches for resonant scalar boson pair production using run 2 lhq proton-proton collision data. *Symmetry*, 14(2), 2022.
- [3] Esma Mobs. The CERN accelerator complex in 2019. Complexe des accélérateurs du CERN en 2019. 2019. General Photo <https://cds.cern.ch/record/2684277>.
- [4] Joao Pequeno. Computer generated image of the whole ATLAS detector. <https://cds.cern.ch/record/1095924>, 2008.
- [5] Joao Pequeno. Computer generated image of the ATLAS inner detector. <http://cds.cern.ch/record/1095926>, 2008.
- [6] Joao Pequeno. Computer Generated image of the ATLAS calorimeter. <https://cds.cern.ch/record/1095927>, 2008.
- [7] Joao Pequeno. Computer generated image of the ATLAS Muons subsystem. <https://cds.cern.ch/record/1095929>, 2008.
- [8] The ATLAS Collaboration. Operation of the ATLAS trigger system in Run 2. *JINST*, 15(10):P10004, 2020. <https://doi.org/10.1088/1748-0221/15/10/P10004>.

- [9] ATLAS Collaboration. Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS. 6 2017.
- [10] Abbott Aad, G. et al. ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J. C*, 79(11):970, 2019. <https://doi.org/10.1140/epjc/s10052-019-7450-8>.
- [11] ATLAS Collaboration. Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta in the ATLAS experiment. 2021. <https://cds.cern.ch/record/2753444>.
- [12] ATLAS Collaboration. Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 105(9):092002, 2022.
- [13] ATLAS Collaboration. Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev. D*, 105(9):092002, 2022. <https://link.aps.org/doi/10.1103/PhysRevD.105.092002>.
- [14] K. A. Olive et al. Review of Particle Physics. *Chin. Phys. C*, 38:090001, 2014. <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [15] M. Tanabashi et al. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.
- [16] G. Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012. <https://doi.org/10.1016/j.physletb.2012.08.020>.

- [17] S. Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012. <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [18] Steven Weinberg. A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266, Nov 1967. <https://link.aps.org/doi/10.1103/PhysRevLett.19.1264>.
- [19] Abdus Salam. Weak and Electromagnetic Interactions. *Conf. Proc. C*, 680519:367–377, 1968. <https://doi.org/10.1007/BF02726525>.
- [20] F. Englert and R. Brout. Broken Symmetry and the Mass of Gauge Vector Mesons. *Phys. Rev. Lett.*, 13:321–323, 1964. <https://doi.org/10.1103/PhysRevLett.13.321>.
- [21] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964. <https://link.aps.org/doi/10.1103/PhysRevLett.13.508>.
- [22] Peter W. Higgs. Spontaneous symmetry breakdown without massless bosons. *Phys. Rev.*, 145:1156–1163, May 1966. <https://link.aps.org/doi/10.1103/PhysRev.145.1156>.
- [23] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys.*, 22:579–588, 1961. [https://doi.org/10.1016/0029-5582\(61\)90469-2](https://doi.org/10.1016/0029-5582(61)90469-2).
- [24] T. W. B. Kibble. Symmetry breaking in non-abelian gauge theories. *Phys. Rev.*, 155:1554–1561, Mar 1967. <https://link.aps.org/doi/10.1103/PhysRev.155.1554>.
- [25] M. Aker et al. Direct neutrino-mass measurement with sub-electronvolt sensitivity. *Nature Phys.*, 18(2):160–166, 2022. <https://doi.org/10.1038/s41567-021-01463-1>.

- [26] Peter Skands. Introduction to QCD. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pages 341–420, 2013. <https://doi.org/10.1016/j.physrep.2012.02.002>.
- [27] Richard Keith Ellis, William James Stirling, and Bryan R Webber. *QCD and collider physics*. Cambridge monographs on particle physics, nuclear physics, and cosmology. Cambridge University Press, Cambridge, 2003. Photography by S. Vascotto.
- [28] David J. Gross and Frank Wilczek. Ultraviolet behavior of non-abelian gauge theories. *Phys. Rev. Lett.*, 30:1343–1346, Jun 1973. <https://link.aps.org/doi/10.1103/PhysRevLett.30.1343>.
- [29] H. David Politzer. Reliable perturbative results for strong interactions? *Phys. Rev. Lett.*, 30:1346–1349, Jun 1973. <https://link.aps.org/doi/10.1103/PhysRevLett.30.1346>.
- [30] C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957.
- [31] J R Andersen et al. Handbook of LHC Higgs Cross Sections: 3. Higgs Properties. 7 2013. <https://doi.org/10.48550/arXiv.1307.1347>.
- [32] ATLAS Collaboration. Observation of $h \rightarrow b\bar{b}$ decays and vh production with the atlas detector. *Physics Letters B*, 786:59–86, 2018.
- [33] ATLAS Collaboration. Observation and measurement of higgs boson decays to WW^* with the atlas detector. *Phys. Rev. D*, 92:012006, Jul 2015.
- [34] CMS Collaboration. Observation of the higgs boson decay to a pair of leptons with the cms detector. *Physics Letters B*, 779:283–316, 2018.
- [35] CMS Collaboration. Search for higgs boson and observation of z boson through their decay into a charm quark-antiquark pair in boosted topologies

- in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Rev. Lett.*, 131:041801, Jul 2023.
- [36] ATLAS Collaboration. Measurement of the higgs boson mass in the $h \rightarrow zz \rightarrow 4l$ and $h \rightarrow \gamma\gamma$ channels with $\sqrt{s} = 13$ TeV pp collisions using the atlas detector. *Physics Letters B*, 784:345–366, 2018.
- [37] ATLAS and CMS Collaboration. Evidence for the higgs boson decay to a z boson and a photon at the lhc. *Physical Review Letters*, 132(2), January 2024.
- [38] CMS Collaboration. Evidence for higgs boson decay to a pair of muons. *Journal of High Energy Physics*, 2021(1), January 2021.
- [39] Isabella Masina and Mariano Quiros. On the veltman condition, the hierarchy problem, and high-scale supersymmetry. *Physical Review D*, 88(9), November 2013.
- [40] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, and João P. Silva. Theory and phenomenology of two-higgs-doublet models. *Physics Reports*, 516(1-2):1–102, jul 2012. <https://doi.org/10.1016/j.physrep.2012.02.002>.
- [41] Csaba Csáki. The minimal supersymmetric standard model. *Modern Physics Letters A*, 11(08):599–613, 1996.
- [42] Vasiliki A. Mitsou. Searches for Supersymmetry with the ATLAS and CMS detectors. *PoS*, CORFU2022:074, 2023.
- [43] Luciano Maiani. The gim mechanism: origin, predictions and recent uses, 2013.
- [44] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak interactions with lepton-hadron symmetry. *Phys. Rev. D*, 2:1285–1292, Oct 1970.

- [45] Yi-Lun Chung, Kingman Cheung, and Shih-Chieh Hsu. Sensitivity of two-higgs-doublet models on higgs-pair production via $b\bar{b}b\bar{b}$ final state. *Phys. Rev. D*, 106:095015, Nov 2022.
- [46] Lisa Randall and Raman Sundrum. Large mass hierarchy from a small extra dimension. *Phys. Rev. Lett.*, 83:3370–3373, Oct 1999. <https://link.aps.org/doi/10.1103/PhysRevLett.83.3370>.
- [47] Abdel Pérez-Lorenzana. An introduction to the brane world, 2004. <https://arxiv.org/abs/hep-ph/0406279>.
- [48] Abdus Salam and J Strathdee. On kaluza-klein theory. *Annals of Physics*, 141(2):316–352, 1982.
- [49] Liam Fitzpatrick, Jared Kaplan, Lisa Randall, and Lian-Tao Wang. Searching for the kaluza-klein graviton in bulk rs models. *Journal of High Energy Physics*, 2007(09):013–013, September 2007.
- [50] Arnaud Ferrari. Searches for Higgs boson pair production with ATLAS. In *13th Conference on the Intersections of Particle and Nuclear Physics*, 9 2018. <https://doi.org/10.48550/arXiv.1809.08870>.
- [51] ATLAS Collaboration. Search for pair production of boosted Higgs bosons via vector-boson fusion production in the $b\bar{b}b\bar{b}$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical report, CERN, Geneva, 2024. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2024-003>.
- [52] Roberto Contino, Margherita Ghezzi, Mauro Moretti, Giuliano Panico, Fulvio Piccinini, and Andrea Wulzer. Anomalous Couplings in Double Higgs Production. *JHEP*, 08:154, 2012. [https://doi.org/10.1007/JHEP08\(2012\)154](https://doi.org/10.1007/JHEP08(2012)154).

- [53] G. Aad et al. Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector. *Phys. Lett. B*, 800:135103, 2020. <https://doi.org/10.1016/j.physletb.2019.135103>.
- [54] Morad Aaboud et al. Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s}=13$ TeV. *JHEP*, 05:124, 2019. [https://doi.org/10.1007/JHEP05\(2019\)124](https://doi.org/10.1007/JHEP05(2019)124).
- [55] Albert M Sirunyan et al. Search for Higgs boson pair production in events with two bottom quarks and two tau leptons in proton–proton collisions at $\sqrt{s}=13$ TeV. *Phys. Lett. B*, 778:101–127, 2018. <https://doi.org/10.1016/j.physletb.2018.01.001>.
- [56] G Aad et al. Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s}=13$ TeV with the ATLAS detector. *Phys. Rev. D*, 106:052001, 2022. <https://doi.org/10.1103/PhysRevD.106.052001>.
- [57] Morad Aaboud et al. Search for Higgs boson pair production in the $\gamma\gamma WW^*$ channel using pp collision data recorded at $\sqrt{s}=13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 78(12):1007, 2018. <https://doi.org/10.1140/epjc/s10052-018-6457-x>.
- [58] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. <https://cds.cern.ch/record/1129811>.
- [59] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008. <https://cds.cern.ch/record/1129810>.
- [60] The ALICE Collaboration. The ALICE experiment at the CERN LHC. *JINST*, 3:S08002, 2008. <https://cds.cern.ch/record/1129812>.

- [61] Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC. 6 2019. <https://cds.cern.ch/record/2677054>.
- [62] Expected performance of the ATLAS detector at the High-Luminosity LHC. 2019. <https://cds.cern.ch/record/2655304>.
- [63] Joseph William Carter. Luminosity measurement with the ATLAS experiment at the LHC. *PoS, ICHEP2020*:726, 2021. <http://cds.cern.ch/record/2753050>.
- [64] Carla Sbarra. The LUCID Detector for LHC Run-2. *Universe*, 5(1):11, 2019. <https://doi.org/10.3390/universe5010011>.
- [65] S. Agostinelli et al. GEANT4 – a simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003. [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- [66] G. Aad et al. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70:823–874, 2010. <https://doi.org/10.1140/epjc/s10052-010-1429-9>.
- [67] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867, 2008. <https://doi.org/10.1016/j.cpc.2008.01.036>.
- [68] X. Artru and G. Mennessier. String model and multiproduction. *Nucl. Phys. B*, 70:93–115, 1974. [https://doi.org/10.1016/0550-3213\(74\)90360-5](https://doi.org/10.1016/0550-3213(74)90360-5).
- [69] Johannes Bellm et al. Herwig 7.0/Herwig++ 3.0 release note. *Eur. Phys. J. C*, 76(4):196, 2016. <https://doi.org/10.1140/epjc/s10052-016-4018-8>.

- [70] Richard D. Field and Stephen Wolfram. A QCD Model for e^+e^- Annihilation. *Nucl. Phys. B*, 213:65–84, 1983. [https://doi.org/10.1016/0550-3213\(83\)90175-X](https://doi.org/10.1016/0550-3213(83)90175-X).
- [71] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. *JHEP*, 11:070, 2007. <https://doi.org/10.1088/1126-6708/2007/11/070>.
- [72] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [73] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008. <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [74] G. Aad et al. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J. C*, 77:490, 2017. <https://doi.org/10.1140/epjc/s10052-017-5004-5>.
- [75] M. Aaboud et al. Jet reconstruction and performance using particle flow with the ATLAS Detector. *Eur. Phys. J. C*, 77(7):466, 2017. <https://doi.org/10.1140/epjc/s10052-017-5031-2>.
- [76] ATLAS Collaboration. Jet mass reconstruction with the ATLAS Detector in early Run 2 data. 7 2016. <https://cds.cern.ch/record/2200211>.
- [77] G. Aad et al. Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Eur. Phys. J. C*, 81(8):689, 2021. <https://doi.org/10.1140/epjc/s10052-021-09402-3>.

- [78] G. Aad et al. Optimisation of large-radius jet reconstruction for the ATLAS detector in 13 TeV proton–proton collisions. *Eur. Phys. J. C*, 81(4):334, 2021. <https://doi.org/10.1140/epjc/s10052-021-09054-3>.
- [79] M. Aaboud et al. In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector. *Eur. Phys. J. C*, 79(2):135, 2019. <https://doi.org/10.1140/epjc/s10052-019-6632-8>.
- [80] Matteo Cacciari and Gavin P. Salam. Pileup subtraction using jet areas. *Phys. Lett. B*, 659:119–126, 2008. <https://doi.org/10.1016/j.physletb.2007.09.077>.
- [81] ATLAS Collaboration. Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment. 3 2017. <https://cds.cern.ch/record/2256687>.
- [82] ATLAS Collaboration. Secondary vertex finding for jet flavour identification with the ATLAS detector. 6 2017. <https://cds.cern.ch/record/2270366>.
- [83] ATLAS Collaboration. Topological b -hadron decay reconstruction and identification of b -jets with the JetFitter package in the ATLAS experiment at the LHC. 2018. <https://cds.cern.ch/record/2645405>.
- [84] ATLAS Collaboration. Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. 7 2017. <https://cds.cern.ch/record/2273281>.
- [85] ATLAS Collaboration. Jet Mass Resolutions in ATLAS using Run 2 Monte Carlo Simulation. Technical report, CERN, Geneva, 2018. <https://cds.cern.ch/record/2631339>.

- [86] Richard Ball et al. Parton distributions with LHC data. *Nucl. Phys. B*, 867:244–289, 2013. <https://doi.org/10.1016/j.nuclphysb.2012.10.003>.
- [87] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016. <https://doi.org/10.1088/0954-3899/43/2/023001>.
- [88] L. A. Harland-Lang, A. D. Martin, P. Motylinski, and R. S. Thorne. Parton distributions in the LHC era: MMHT 2014 PDFs. *Eur. Phys. J. C*, 75(5):204, 2015.
- [89] Tomáš Ježo, Jonas M. Lindert, Niccolo Moretti, and Stefano Pozzorini. New NLOPS predictions for $t\bar{t} + b$ -jet production at the LHC. *Eur. Phys. J. C*, 78(6):502, 2018. <https://doi.org/10.1140/epjc/s10052-018-5956-0>.
- [90] Giulia Negro. Top Quark Modelling and Tuning in ATLAS and CMS. In *14th International Workshop on Top Quark Physics*, 1 2022. <https://doi.org/10.48550/arXiv.2201.03517>.
- [91] ATLAS Collaboration. ATLAS Pythia 8 tunes to 7 TeV data. Technical report, CERN, Geneva, 2014. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2014-021>.
- [92] Morad Aaboud et al. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018.
- [93] ATLAS Collaboration. Track Reconstruction Performance of the ATLAS Inner Detector at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2015-018, CERN, Geneva, Jul 2015. <https://cds.cern.ch/record/2037683>.

- [94] M. Aaboud et al. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J. C*, 77(10):673, 2017. <https://doi.org/10.1140/epjc/s10052-017-5225-7>.
- [95] ATLAS Collaboration. Studies of radial distortions of the ATLAS Inner Detector. Technical Report ATL-PHYS-PUB-2018-003, CERN, Geneva, Mar 2018. <https://cds.cern.ch/record/2309785>.
- [96] Robert D. Cousins, James T. Linnemann, and Jordan Tucker. Evaluation of three methods for calculating statistical significance when incorporating a systematic uncertainty into a test of the background-only hypothesis for a Poisson process. *Nucl. Instrum. Meth. A*, 595(2):480, 2008. <https://doi.org/10.1016/j.nima.2008.07.086>.
- [97] Alexander L. Read. Presentation of search results: The CL(s) technique. *J. Phys. G*, 28:2693–2704, 2002. [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313).
- [98] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. <https://doi.org/10.1140/epjc/s10052-011-1554-0>.
- [99] ATLAS Collaboration. Simulation-based extrapolation of b-tagging calibrations towards high transverse momenta in the ATLAS experiment. 2021. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-003>.
- [100] Eilam Gross and Ofer Vitells. Trial factors for the look elsewhere effect in high energy physics. *Eur. Phys. J. C*, 70:525–530, 2010. <https://doi.org/10.1140/epjc/s10052-010-1470-8>.