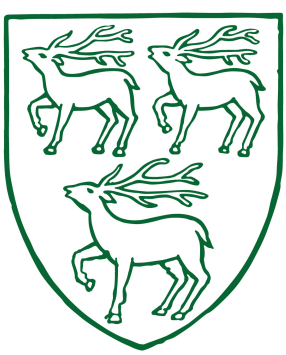
Search for New Physics through Higgs-Boson-Pair Production at the LHC and Beyond





Santiago Paredes Sáenz

Jesus College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2020

Abstract

This thesis presents a search for resonant production of highly energetic Higgs boson pairs, with each Higgs boson decaying to a b-quark pair, with the ATLAS detector. This analysis uses the complete ATLAS run 2 dataset of 139.0 fb^{-1} of proton-proton collisions at a centre-of-mass energy of $\sqrt{s} = 13$ TeV. The expanded dataset, new flavour-tagging techniques, and novel subjet reconstruction algorithms provide better sensitivity than previous studies of the process. Two beyond-the-Standard-Model resonances are chosen as benchmarks for the result: a two-Higgs-doublet model scalar S and Kaluza-Klein (KK) gravitons G^*_{kk} predicted in the bulk Randall-Sundrum (RS) model. No significant deviations from Standard Model predictions are found, and upper limits are set on the production cross section of these resonances. A phenomenology study on the analysis strategy of this search in conditions of the future High Luminosity Large Hadron Collider is presented as well. This study shows proof-of-concept improvements in sensitivity by using a deep-learning approach to separate the $hh \to 4b$ signal, and discusses the optimization of the search to constrain the Higgs boson self-coupling parameter λ_{hhh}.

Acknowledgements

The finite nature of this document, and the strict page limit, prevent me from thoroughly expressing my gratitude and acknowledging everyone who has made this thesis and my DPhil possible, but what follows is my humble attempt at it.

My supervisors Cigdem Issever, Todd Huffman, and Tony Weidberg, with their guidance, support, and enthusiasm for my research, not only taught and showed me how to be a researcher, but also kept me motivated and inspired through these years. Alan Barr, Daniela Bortoletto, Ian Shipsey, and James Frost, while not officially supervisors, also provided invaluable guidance and advice that crucially influenced my DPhil at various stages. You have all marked my life and career in the best way possible, and I feel extremely fortunate to have had you as mentors and to have had the opportunity to work together and learn from you.

As is always the case in particle physics, my research would not have been possible without collaboration and discussion with all the outstanding scientists working around me. Everyone in the Oxford Exotics group was kind enough to sit every Friday through the explanations of what I had uncovered that week, discuss it, and provide their advice and criticism, sending me on a new path for the next week. The task of the $hh \to 4b$ team was similarly tough with respect to sitting through my presentations, but with the added complications of late afternoon meetings and making progress towards getting our paper out. Working with you has been as good an experience as I could imagine, and I would like to thank you all for making it such a smooth ride.

Before getting to the many non-academic acknowledgements, it would be a flagrant omission not to thank Bill Balunas, Claudia Merlassino, Jesse Liu, and Lydia Beresford for what I've come to call 'Lamb & Flag supervision' sessions (which had recurrent appearances by other excellent scientists and friends). Together with all the physics discussions, pints, and laughs, I also got great advice on surviving through the DPhil and being a researcher, which have proven to be invaluable lessons.

I'd like to thank the whole team in Oxford's sub-department of particle physics: all the professors and academics who I learned from through lectures and discussions; the fantastic administrative team that helped me through many a bureaucratic peril with ease and kindness; and all the students and postdocs who made it so fun to be there with coffee breaks and desk-table football championships, or multi-month

long games of Diplomacy. Very special thanks to all my fellow 2016 DPhil students for all the amazing times together, including but not limited to: punting during the few warm sunny days Oxford offered, numerous Pizza Fellowship dinners, pelicans, great road trips and fun conferences together, College exchange dinners and wines & cheeses, board game sessions, table football championships, snowboarding, kayaking in lac Léman, celebrating World Cup matches and qualifiers, BBQs, housewarming parties, and just the best of times. I'd like to specially thank Beojan Stanislaus who is a member not only of this group but also the $hh \rightarrow 4b$ analys team, a great office mate and a friend through these years, not to mention a rich source of debugging advice (with the patience to deliver it).

Arriving in the UK, and adapting to Oxford was all kinds of strange for an international student from Ecuador, but being able to count on such excellent pals (some which I met the very night of my arrival!) from Jesus College, Particle Physics, Summer Schools and CERN made all the difference. All of you have made permanent marks on the fond memories of this time: Caro and the rest of our #SantiagosKitchen team figuring out this DPhil thing together; Wouter making crazy plans for trips both at College and DWB, when he wasn't busy developing peli-coin; Arj explaining cricket for the third time, and our flat's great fancy dress parties; Dan making probably the best pasta puttanesca I've ever had, right before going for a hike; Knut the awesome ski instructor and cuarenta partner; Mikkel the chef, brew-master, cuarenta nemesis, Diplomacy co-conspirator, and a fantastic friend; Priyav the determined MCR president, tennis coach, one of the two people I could probably get locked in with for months and not go crazy, and a marvelous mate; Peter, the other person I could probably get locked in with for months and not go crazy, savior of forgotten Ship street keys, terraformer, unorthodox potato chef, and an amazing amigo; and Emma the fantastic trip partner, player 2, sous-chef, and whose support, encouragement, and enthusiasm have almost effortlessly helped me through both great and tough times. Special mention to my fellow co-founders of the `S.C.A.M.` collaboration for allowing me to meet such great scientists and for fantastic collaboration meetings. It has been a major privilege to share all of this with you, and I can't wait to do all these again! You really are the best.

These last few years mark a continuation of my pursuit of curiosity and knowledge, which is a trip that began long ago, with inspiration from my grandparents, parents, and sisters. Only with their support, encouragement, and love was I able to progress, and I feel immensely fortunate to still have plenty to keep me going. These years have had many difficult times, both for me and us as a family, but with support from all of you, with our fantastic teamwork, and with the joy brought by the new members of our family, happiness managed to win out.

Institution Acknowledgements

I would like to thank the ATLAS PhD grant that made it possible for me to study at the University of Oxford, established by Fabiola Gianotti and Peter Jenni (who has followed my research and provided valuable advice throughout). I had the fortune to be supported by the sub-department of particle physics at the University of Oxford, my college at Oxford (Jesus College), and the University of Oxford in multiple areas of my DPhil. SENESCYT also provided limited partial funding.

Preface

This thesis focuses on the ATLAS experiment's search for production of highly energetic Higgs bosons which then decay each to a pair of b-quarks. The data analysis for this search, with focus on searching for heavy resonances decaying to pairs of Higgs bosons, was the main experimental physics analysis of my DPhil research. This analysis is currently in progress within ATLAS and preparing to unblind the signal region data set, and, for this reason, I separated myself from the analysis group at a late stage to produce the results in this thesis. This analysis is expected to be published this year (2020). A phenomenology study focusing on this same process, but with focus on future runs of the Large Hadron Collider (LHC) was also part of the work done for this DPhil [1].

The thesis is structured as follows. Chapter 1 gives a brief theoretical introduction to the $hh \to 4b$ process, and motivates its search. Chapter 2 describes the ATLAS detector and chapter 3 introduces particle and event reconstruction techniques, both with focus on the systems and methods relevant to the $hh \to 4b$ search. Chapters 4 and 5 provide a detailed account of the analysis, and chapter 6 focuses on further studies done for the search which are not crucial to the understanding of the search, but were important for the process and result. Chapter 7 gives a short comment on the future of di-Higgs searches and provides a brief summary of the $hh \to 4b$ phenomenology study in [1], with focus on my contributions.

Another area where I made significant contributions, but was omitted from this thesis due to it being disconnected to the main $hh \to 4b$ analysis, was the ATLAS missing transverse energy (MET) trigger system. I contributed in developing a software framework to rapidly convert incoming trigger data into quantities and plots used to evaluate the performance of the trigger. I also performed various studies during the 2017 data taking campaign to evaluate the performance of the trigger under various conditions, for example, when the LHC delivered the highest instantaneous luminosity proton-proton collisions recorded by ATLAS up to that date. These studies are documented in an ATLAS internal note [2], and are part of the recently published paper on the performance of the MET trigger during run 2 of the LHC [3].

Contents

List of Figures **xi**

List of Tables **xix**

1 Theory and Motivation **1**
1.1 Symmetries in Quantum Field Theories . . . 2
1.2 The Higgs Boson and the Higgs Sector . . . 3
1.2.1 The Higgs Field . . . 3
1.2.2 Higgs Boson Self-Coupling . . . 4
1.3 Di-Higgs Production at the LHC . . . 5
1.3.1 Standard Model Di-Higgs Production . . . 5
1.3.2 Beyond the Standard Model Di-Higgs Production . . . 7
1.3.3 Di-Higgs Decays Channels . . . 8
1.4 Current Status of Di-Higgs Searches . . . 9
1.4.1 Boosted and Resolved Regimes . . . 11
1.5 $hh \rightarrow 4b$ Searches in ATLAS . . . 12

2 The Experiment **14**
2.1 The Large Hadron Collider . . . 14
2.2 ATLAS: A Toroidal LHC ApparatuS . . . 15
2.2.1 Sub-detectors . . . 17
2.2.1.1 Inner Detector . . . 17
2.2.1.2 Calorimeters . . . 19
2.2.1.3 Muon Spectrometer . . . 23
2.2.2 Trigger . . . 23
2.2.2.1 Jet Trigger . . . 24
2.2.2.2 Study of Efficiency of 2017 and 2018 Jet Triggers . 25

3 Physics Objects and Algorithms **28**
3.1 Tracks and Vertices . . . 29
3.2 Jets . . . 30
3.2.1 Large-Radius Jets . . . 31
3.2.2 Track Jets . . . 35

3.2.2.1 Fixed-Radius Jets . . . 35
3.2.2.2 Variable-Radius Jets . . . 36
3.3 Flavour Tagging . . . 39
3.4 Muons . . . 42

4 Analysis Strategy **44**
4.1 Analysis Overview . . . 45
4.2 Data and Simulated Samples . . . 46
4.2.1 Monte Carlo Samples . . . 47
4.3 Event and Object Selection . . . 47
4.3.1 Kinematic Selections . . . 48
4.3.2 B-tag Region Selection . . . 51
4.3.3 Higgs-Mass Region Selection . . . 52
4.3.4 Vetoed Events . . . 56
4.3.4.1 Resolved Signal Region Veto . . . 56
4.3.4.2 Collinear Track Jet Veto . . . 57
4.4 Background Estimation . . . 57
4.4.1 Background Modelling in the Control Region . . . 58
4.4.2 Validation of Background Estimation . . . 59
4.5 Estimation of Systematic Uncertainties . . . 65
4.5.1 Detector & Reconstruction Systematics . . . 66
4.5.2 Background Prediction Systematics . . . 68
4.5.2.1 Background Normalization Uncertainties . . . 68
4.5.2.2 Background Shape Uncertainties . . . 72
4.5.3 Summary of Systematic Uncertainties . . . 74

5 Analysis Results **76**
5.1 Statistical Methods . . . 76
5.1.1 Likelihood . . . 77
5.1.2 Test Statistic . . . 78
5.1.3 Limit Setting Procedure . . . 79
5.1.4 Expected Limits . . . 82
5.1.4.1 Impact of Systematic Uncertainty . . . 82
5.2 Results . . . 85
5.2.1 Unblinded Signal Region Distributions . . . 86
5.2.2 Observed Limits . . . 87

6 Further Variable Radius Track Jet Studies **89**
6.1 Collinear Track-Jets Veto 90
6.1.1 Veto Motivation and Definition 90
6.1.2 Veto Cause and Impact on the Analysis 90
6.1.3 Potential fix: Increase Minimum p_{T} to Trigger Veto 93
6.1.4 Performance of Variable Radius (VR) Track Jets 96
6.2 Impact of Variable Radius (VR) Track Jets on Sensitivity 100
6.2.1 Impact of VR Specific b-tagging Training 101

7 Future Prospects **102**
7.1 Di-Higgs to $b\bar{b}b\bar{b}$ HL-LHC Phenomenology Study 104
7.1.1 Analysis Overview 105
7.1.2 Neural-Network Analysis 105
7.1.3 Feature Importance 107
7.1.4 Network correlation plots 112
7.1.5 Analysis Results 115
7.1.6 Phenomenology Study Final Remarks 116

8 Conclusions **117**

Appendices

A Full Pull and Impact of NPs **120**

B Event Selection Yields **124**

List of Figures

1.1 Diagrams depicting scalar potential $V(\phi)$ ($\mu^2 < 0$, $\lambda > 0$) with conserved (left) and broken (right) symmetry. 4
1.2 Diagrams depicting the triple and quartic Higgs couplings. 4
1.3 Di-Higgs production modes via gluon-gluon fusion. 6
1.4 Single-Higgs and double-Higgs production cross section as a function of κ_λ. The dashed line intercepts the values corresponding to the SM hypothesis ($\kappa_\lambda = 1$)[24]. 6
1.5 Higgs boson decay branching fractions. The percentages represent SM expected values from [30]. 8
1.6 Branching fraction for a pair of Higgs bosons represented by area. The numbers on each square show the approximate value (to the nearest 0.5%) of the branching fraction. The area of off-diagonal combinations are under-represented by a factor of two, but the numbers show the correct value. The percentages represent SM expected values from [30]. 9
1.7 95% CL_S limits on the production cross section of a heavy scalar (left column) and a G^*_{kk} spin-2 (right column) resonances, found by the ATLAS [31] and CMS [38] $hh \rightarrow 4b$ searches. The red lines represent the predicted cross section given a particular theoretical model. 10
1.8 Diagrams depicting the boosted and resolved topologies of a $hh \rightarrow 4b$ decay. 12

2.1 Diagram of the ATLAS detector and its sub-detectors [43]. 16
2.2 Diagram of the ATLAS coordinate system. The positive x-axis points towards the centre of the LHC, and the y-axis points upwards. . . . 17
2.3 Schematic of the sub-detectors of ATLAS representing a slice in ϕ of the plane transverse to the beam. The signatures various particles would leave in each layer are also depicted. 20
2.4 Diagram of the (a) liquid argon calorimeter and the (b) tile calorimeter of the ATLAS detector. 22

2.5 Efficiency of the `HLT_j420_a10t_lcw_jes_40smcINF_L1J100` trigger on the (a) SM and (b) 1 TeV spin 2 simulated samples. The efficiency is shown as a function of large-R jet p_T. Note the different scale in the y axis in each plot. 26

2.6 Efficiency of the `HLT_j420_a10t_lcw_jes_40smcINF_L1J100` trigger on the (a) SM and (b) 1 TeV spin 2 simulated samples. The efficiency is shown as a function of large-R jet mass. Note the different scale in the y axis in each plot. The small dip around 150 GeV in plot (b) is an artifact of the efficiency being shown as the inverse-cumulative distribution. 27

3.1 Diagram of the jet trimming process. The initial large-R jet is reclustered into smaller jets of radius R_{sub} with the k_t algorithm. A cut is then placed on the p_{T} of reclustered jets, and the resulting trimmed jet is formed from the remaining small jets [73]. 32

3.2 Jet mass resolution as a function of truth jet p_{T} (a) and jet mass distribution (b) of W/Z jets from simulated events. Plot (a) shows the jet mass resolution of calorimeter-based mass m^{calo}, track-assisted mass m^{TA}, and combined mass m^{comb}. Plot (b) shows the mass distributions for these same mass definitions before(after) calibration in dashed(solid) lines. The y axis in (a) shows half of the 68% inter-quantile range (IQnR) divided by the median of the jet mass response, which is used as an outlier-insensitive measure of the resolution [74]. 34

3.3 Sequence of calibrations and corrections applied to large-R jets. Full description in [71]. 34

3.4 Diagram depicting track jets, reconstructed with fixed and variable radius, originating from a boosted Higgs boson. The secondary vertex from the decay of the b-hadron and tracks from the resulting charged particles are shown within the track jets. Tracks shown in red are contained in both jet's cones and could be wrongly assigned. 36

3.5 Double b-labelling efficiency using `VR` subjets associated with a Higgs boson candidate (here referred to as Higgs Jet) as a function of its p_{T}. `VR` subjets are constructed with $R_{max} = 0.4$, $R_{min} = 0.02$, and various values of ρ. The efficiency of `FR` $R = 0.2$ subjets is shown as well [80]. 38

3.6 39

3.7 QCD jet rejection as a function of Higgs-boson jet efficiency for double-b-tagging using the VR sub-jet algorithm. The exclusive-kt (ExKt), center-of-mass (CoM), FR $R = 0.2$ subjets are shown as well. This plot shows efficiencies for large-R jets with mass 76 GeV $< m_{\text{jet}} <$ 146 GeV and transverse momentum 250 GeV $< p_{\text{T,jet}} <$ 400 GeV [80]. VR track jets match or exceed the performance of FR $R = 0.2$ subjets for nearly all values of Higgs-boson-jet efficiency. 39
3.8 Diagram depicting some of the parameters relevant for b-tagging. . 40
3.9 b-jet tagging efficiency vs light-flavour jet rejection for two *high-level* (MV2 and DL1) and three low-level (IP3D, SV1, and JetFitter) taggers on simulated $t\bar{t}$ events. The lower panel shows the ratio of each curve to the MV2 for comparison. Efficiency and mistag rates were computed using a sample of simulated $t\bar{t}$ events [82]. 41
3.10 The b-jet tagging efficiency (a) and light-flavour jet rejection (b) as a function of a jet's p_{T}, shown for two *high-level* (MV2 and DL1) and three *low-level* (IP3D, SV1, and JetFitter) taggers. The lower panel shows the ratio of each curve to the MV2 for comparison. Efficiency and mistag rates were computed using a sample of simulated $t\bar{t}$ events [82] 42

4.1 Diagram of the boosted $hh \to 4b$ final state. The two large-R jets from the decay of the Higgs bosons are represented by large yellow cones and the smaller purple cones within these represent the track jets associated to them. These track jets can be identified, or *tagged*, as containing a b-quark. 45
4.2 Distributions of the di-Higgs invariant mass m_{hh}, after the trigger requirement is applied, of resonant spin-2 G^*_{kk} (a) and heavy scalar (b) signal models. Data in this early stage of the selection process are shown as a rough estimate of the QCD background. 46
4.3 Diagram depicting the flow of data through the analysis selection. Size does not correspond to scale. 48
4.4 Distributions of the difference in η between the two leading Higgs boson candidates, after the trigger requirement is applied. Different signal mass resonances are shown for the spin-2 G^*_{kk} (a) and heavy scalar (b) signal models. Data in this early stage of the selection process are shown as a rough estimate of the QCD background. . . 50
4.5 Diagram of the three high-tag topologies ($4b$, $3b$, and $2b$) with the corresponding low-tag region used to estimate QCD background ($2b$-2, $2b$-1, and $1b$-1) in the lower half of the table. 52

4.6 Definitions of the Higgs boson candidate mass regions on the plane defined by the leading and subleading Higgs boson candidates masses. This plot shows data from the 1b-1 region. The regions are defined in the body of the text. 53
4.7 Efficiency of each sequential selection in the analysis for a range of masses of (a) scalar and (b) G^*_{kk} resonances. 54
4.8 Depictions of an acceptable (left) and a problematic (right) configuration of overlapping variable-radius track jets. In the right side case, the softer (purple) jet's axis falls within the hard jet (green), which can lead to a mismatch of tracks to jets when b-tagging. . . . 57
4.9 Leading Higgs boson candidate mass distribution from the background model and from the data used to normalize the QCD multijet and $t\bar{t}$ contributions in the 2, 3 and 4 b-tag analyses in the control region. The double-peak structure of the distributions is due to the removal of the validation and signal regions (see figure 4.6). 60
4.10 Di-Higgs mass in the control (left) and validation (right) for the 2b region. 61
4.11 Di-Higgs mass in the control (left) and validation (right) for the 3b region. 61
4.12 Di-Higgs mass in the control (left) and validation (right) for the 4b region. 62
4.13 QCD background estimate in the signal region (dots, these are obtained from low-tag data in the SR) for the 4b (top left), 3b (top right), and 2b (bottom) regions with the corresponding smoothing fit (red line). 62
4.14 Leading Higgs boson candidate mass distribution from the background model and from the data to validate the QCD multijet and $t\bar{t}$ background estimates in the 2, 3 and 4 b-tag analyses in the validation region. 63
4.15 Cross check of the background estimate in the signal region for the 4b (a), 3b (b), and 2b (c) regions. Low-tag and high-tag data have been replaced by QCD MC. The modelling of the SR QCD agrees with the prediction within uncertainty. 65
4.16 Expected 95% CL upper limits on the production cross section times branching ratio of the heavy scalar model accounting for only statistical uncertainty and both statistical and systematic uncertainties. 66

4.17 Comparison of the QCD model, blinded high-tag data minus $\alpha_{t\bar{t}} \times t\bar{t}$, on the left to the result of the Gaussian Process interpolation on the right. The Gaussian Process is used to assess the uncertainty from the extrapolations from control to signal regions and low-tag to high-tag regions in the background estimate. Shown are the $2b$ region (top row), $3b$ region (middle row), and the $4b$ region (bottom row). 70
4.18 Result of the smoothing fits in the $4b$ region for all functions (a) and various choices of fit range (b). The grey band shows statistical error on the nominal fit. The largest differences as compared to the nominal MJ8 function were taken as systematic uncertainties. . . . 73
4.19 Normalized m_{hh} distributions from the $2b$ region. The data points correspond to high-tag VR data, while the red line shows the SR background estimate. The ratio of the two, shown in the lower panel, is taken as a shape uncertainty. 74

5.1 Sketch of the probability density distribution of the test statistic under the B (blue) and $S + B$ (orange) hypotheses (see footnote 5). The final CL_S is given by the ratio of the two shaded areas. The limit on μ is set such that this ratio is 0.05. 80
5.2 Sketch of the plots used to show the observed and expected CL_S limits on the resonance production cross section. The solid (dashed) line traces the observed (expected) limit for each of the mass points on the x-axis, and the coloured bands show the deviation from the expected limit. For each mass point, values of cross section above the observed line are excluded. 81
5.3 Expected 95% CL upper limits on the production cross section times branching ratio of the (a) heavy scalar model and (b) RS graviton accounting for statistical uncertainties only, shown for the three b-tag regions ($4b$, $3b$, and $2b$). 83
5.4 Pulls (circle markers, bottom axis) and impacts (bars, top axis) in femtobarns on the measured cross section times branching ratio of a 1 TeV (a) heavy scalar and (b) G^*_{kk} graviton. The plot shows the 10 nuisance parameters with most impact on the expected limit from the $4b$ region. NPs related to the background estimate and jet properties were ranked as the most impactful. 84
5.5 Di-Higgs mass in the unblinded signal region in the (a) $4b$, (b) $3b$, and (c) $2b$ regions. 86
5.6 Scatter plots of the $m_{h_1} - m_{h_2}$ plane including the unblinded signal region for the (a) $4b$, (b) $3b$, and (c) $2b$ regions. 87

5.7 Observed 95% CL upper limits on the production cross section times branching ratio of the (a) heavy scalar model and (b) RS graviton accounting for both statistical and systematic uncertainties. No significant deviation from the expected limits are observed in either model. 88

6.1 Histogram showing the frequency with which pair of track jets triggered a collinear veto. The prefix $h1$ or $h2$ in the x-axis labels indicates which Higgs boson candidate the overlapping track jets are associated with, while the second pair of numbers indicate which track jets overlapped (by their order in p_{T}). 91

6.2 Distribution of the p_{T} of overlapping track jets that could cause an event to be vetoed. These plots are shown for a Standard Model non-resonant signal, and 1 TeV and 3 TeV spin-2 resonances, on each row. These plots only contain events with collinear `VR` track jets. . 92

6.3 Impact of the collinear `VR` veto on various spin-2 signal mass points. The y axis shows the fraction of events in the signal region (before any b-tagging selection) that pass the veto. Such impact is shown for various minimum track jet p_{T} values to trigger a veto. 93

6.4 Plots of the MV2c10 score for various sets of jets in a 3 TeV spin-2 signal sample. The (a) MV2c10 score for all jets and (b) the leading track jet associated to the leading Higgs boson candidate, from various samples with increasing minimum track-jet p_{T} requirement are plotted. The lower panels show the ratio of each variation to the nominal case, where the increased acceptance of b-tagged jets (a), and the small impact of varying the p_{T} threshold for vetoing (b) can be seen. 95

6.5 MV2c10 score of the jets recovered by raising the minimum track-jet p_{T} requirement from 5 GeV to 50 GeV, shown with (a) the equivalent score taken from the sample of vetoed events, and (b) with the score of jets that were never collinear. 96

6.6 Selection efficiency per signal region, for a range of spin-2 signal mass points, shown for `FR` track jets (solid lines, triangular markers) and `VR` track jets (dashed lines and circle markers). All available b-tagging working points are shown (corresponding to the number after the underscore). All analysis selections have been applied. . . 97

6.7 Efficiency of each sequential selection in the analysis for a range of mass points of G^*_{kk} resonance signal samples, made using (a) fixed-radius and (b) variable-radius track jets. 98

6.8 Angular separation ΔR between the two track jets in the leading Higgs boson candidate in a signal-free side-band region, showing the difference between Higgs boson candidates with `FR` (a) and `VR` (b) track jets. Data in this region are plotted over the stacked estimates of $t\bar{t}$ and QCD multijet backgrounds. Figure 6.8a used fixed radius and 6.8b variable radius track jets. 99
6.9 Expected 95% CL upper limits on the production cross section times branching ratio of graviton making use of fixed and variable radius track jets. The uncertainties in the limits include statistical uncertainties only and do not account for systematic uncertainties. The red line represents the expected cross section from theoretical predictions of the graviton model. This limit does *not* make use of the updated b-tagging training for `VR` track jets (see section 6.2.1). 100
6.10 Expected 95% CL upper limits on the production cross section times branching ratio of the heavy scalar model with the old (`MV2c10`) and new (`DL1r`) b-taggers. The uncertainties in the limits include statistical uncertainties only and do not account for systematic uncertainties. 101

7.1 Expected value of $-2\ln\Lambda$ as a function of κ_λ (all other couplings set to their SM value) obtained in the $\kappa_\lambda = 1$ hypothesis for the single-and double-Higgs-boson decay modes[24]. 103
7.2 Architecture of the neural network. 107
7.3 Neural network score distributions $p_{\text{signal}}^{\text{DNN}}$ of benchmark signals (solid lines) and background processes (filled stacked) displayed in the legend. All event selection criteria of the *neural-network analysis* except the $p_{\text{signal}}^{\text{DNN}} > 0.75$ requirement are imposed. The neural networks are trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signals. These are displayed for (upper) resolved, (middle) intermediate and (lower) boosted regimes. The plots are normalised to $\mathcal{L} = 3000$ fb^{-1}. . . . 110
7.4 Feature importance for the networks trained on $\kappa_\lambda = 5$ signals for the (a) resolved, (b) intermediate and (c) boosted categories. The model features are ranked by their average absolute SHAP value, which is plotted on the x axis. The colour scale indicates the value of the feature on the specific event for which the SHAP value is plotted. 111
7.5 The leading Higgs boson candidate mass vs neural network scores trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal sample for the (upper) resolved, (middle) intermediate, and (lower) boosted analyses. The test samples used to make these distributions are an independent set of (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal events. 113

7.6 The leading Higgs boson candidate mass vs neural network scores trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal sample for the (upper) resolved, (middle) intermediate, and (lower) boosted analyses. The test samples used to make these distributions are an independent set of $t\bar{t}$ events. 114

7.7 Summary of 68% CL ($\chi^2 = 1$) contours in the κ_t vs κ_λ plane. These are displayed for the resolved (dark blue), intermediate (medium blue), boosted (light blue) categories and their combination (orange) for the *baseline analysis* (dashed) and *neural-network analysis* (DNN) trained on $\kappa_\lambda = 5$ (solid). A luminosity of 3000 fb^{-1} is assumed along with systematic uncertainties of 0.3%, 1% and 5% for the resolved, intermediate and boosted categories, respectively. The cross indicates the SM prediction. 115

A.1 Pulls (circle markers, bottom axis) and impacts (bars, top axis) in femtobarns on the measured cross section times branching ratio of a heavy scalar (left column) and spin-2 (right column) resonance. The plot shows the 10 nuisance parameters with most impact on the expected limit from the $3b$ (top row) and $2b$ (bottom row) regions. NPs related to the background estimate and jet properties ranked as the most impactful. 121

A.2 Pulls (circle markers, bottom axis) and impacts (bars, top axis) in fb on the measured cross section times branching ratio of a 1.5 TeV (a) scalar and (b) graviton. The axes show the same quantities as all plots in this appendix and have been omitted to enlarge the fonts on the plots. Fits for these plots were made with all b-tag regions considered. 122

A.3 Pulls (circle markers, bottom axis) and impacts (bars, top axis) in fb on the measured cross section times branching ratio of a 800 GeV G^*_{kk} graviton. NPs related to the background estimate and jet properties ranked as the most impactful. This model was a work-in-progress version used to inform the analysis and not used for any results. . 123

List of Tables

2.1 Summary of the general characteristics of the three ATLAS ID subsystems. The values cited for the resolution are representative figures of each in the barrel, as the resolution depends on factors such as each track's trajectory or the location of each module[43, 52–56]. . . 18

4.1 Large-R jet trigger names and thresholds of p_{T} and mass to pass the trigger chosen for each years' data set. 49

4.2 Overview of event selection for the boosted $hh \to 4b$ analysis. j_1 and j_2 refer to the leading and subleading large-R jets respectively (ordered in p_{T}). The b-tag regions above the double rule are the high-tag regions, while the ones below are the low-tag ones. Columns labeled $\mathrm{j}^{\text{tagged}}$ and $\mathrm{j}^{\text{no-tag}}$ refer to the number of b-tagged and *un*tagged track jets associated to a large-R jet, respectively (high-tag regions have no explicit requirements on the number of untagged jets). The definitions of Higgs-Mass regions are made exclusive (i.e. the control region does not include events in the validation region, which in turn excludes all events in the signal region). 55

4.3 Normalization factors μ_{QCD} and $\alpha_{t\bar{t}}$ derived in the CR. $\alpha_{t\bar{t}}$ in the $4b$ region was set to one. 59

4.4 Overview of systematic uncertainty sources and their impact expected limit on the production cross section 1 TeV scalar and spin-2 resonances decaying to $hh \to 4b$ in the boosted channel. The indented entries below Background Estimate correspond to the various components of this group of uncertainties. 75

7.1 Overview of event selection for the *baseline analysis* in the resolved, intermediate and boosted categories. The requirements above the upper double rule are the same as the preselection used for the *neural-network analysis* training. The requirements below the upper double rule are the signal region requirements. 106

7.2 Input variables used to train the neural network. 106

B.1 Event yield and percentage of events filtered by each step in the event selection process for data collected in 2017 and the two benchmark signal models. 125

1 Theory and Motivation

Contents

1.1 Symmetries in Quantum Field Theories 2
1.2 The Higgs Boson and the Higgs Sector 3
1.2.1 The Higgs Field 3
1.2.2 Higgs Boson Self-Coupling 4
1.3 Di-Higgs Production at the LHC 5
1.3.1 Standard Model Di-Higgs Production 5
1.3.2 Beyond the Standard Model Di-Higgs Production 7
1.3.3 Di-Higgs Decays Channels 8
1.4 Current Status of Di-Higgs Searches 9
1.4.1 Boosted and Resolved Regimes 11
1.5 $hh \to 4b$ Searches in ATLAS 12

The goal of particle physics is to understand what the most elementary building blocks of our universe are, and how they interact to form it. At one facility working towards this goal, protons are accelerated and collided at extraordinary energies with the Large Hadron Collider (LHC) at CERN (see chapter 2 for details). How protons interact during that collision, and what is produced by them, is modeled by the highly precise Standard Model of Particle Physics.

Despite its exceptional accuracy, the Standard Model does not account for some physical phenomena, notably it does not account for gravitational interactions, the existence of Dark Matter or the fact that we live in a universe of matter, without a

corresponding equal amount of anti-matter, indicating that it is an incomplete model.

1.1 Symmetries in Quantum Field Theories

Developed through most of the 20th century, the Standard Model (SM) provides an explanation for the mechanisms underlying electromagnetic, strong and weak nuclear interactions as Quantum Field Theories [4–9]. Each of these interactions emerge from local symmetries of the universe. Requiring that physics is invariant under transformations of the $U(1)$ symmetry group gives rise to electromagnetic interactions and the photon, invariance under $SU(2)_L$ [1] explain weak interactions and its mediators, and invariance under $SU(3)$ transformations brings gluons and the strong nuclear interaction into the picture. The Standard Model unifies all three into a single framework.

The SM Lagrangian that fulfills this symmetry can be compactly (omitting various sum indices for clarity) written as:

$$\mathcal{L}_{\text{SM}} = -\frac{1}{4}F^2 + i\bar{\psi}\gamma D\psi - y\phi\bar{\psi}\psi + |D\phi|^2 + \mu^2\phi^2 + \lambda\phi^4. \tag{1.1}$$

The first term $-\frac{1}{4}F^2$ captures the dynamics of gauge fields, A, that give rise to the force-mediating gauge bosons: the gluon g, the photon γ, and the weak bosons $W^{\pm}$ and Z. The tensor F is defined as function of the fields A and their derivatives, and D is the covariant derivative $D = \partial - igA$ where A are the gauge fields, and g the coupling of each of them. The second term defines the kinematics of fermions, represented by the spinors ψ, and their coupling to the gauge fields, where γ are the Dirac matrices. The remaining terms describe the Higgs sector, and its relationship to gauge bosons and fermions, and deserve a more in-depth description in the next few sections.

[1] The symmetry group of weak interactions is referred to as $SU(2)_L$ representing the weak charged current's characteristic of coupling only to left-handed particles and right-handed antiparticles.

1.2 The Higgs Boson and the Higgs Sector

As mentioned above, local gauge invariance ingeniously explains interactions in the Standard Model, but strict gauge invariance is broken by explicit mass terms in the Lagrangian. While this might not be problematic for $U(1)$ and $SU(3)$ symmetries since the electromagnetic and strong interactions are mediated by massless bosons, the massive mediators of the weak interactions violate this invariance, not to mention the mass terms of all fermions.

The Brout-Englert-Higgs mechanism [10–12] allows particles to acquire mass without explicit mass terms in the Lagrangian, thus preserving local gauge invariance.

1.2.1 The Higgs Field

A complex scalar field ϕ with non-zero vacuum expectation value v, which can couple with a massless (thus locally-invariant) gauge field B with coupling strength g, can be introduced with the potential described in the last two terms in equation 1.1

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.2}$$

where μ^2 and λ are real numbers [2] and free parameters of this potential. Expanding the field about the minima rather than zero by setting $\phi(x) = \frac{1}{\sqrt{2}}(v + h(x))$ *spontaneously breaks* its rotational $U(1)$ symmetry (figure 1.1 [3]).

After this expansion both the scalar and gauge fields acquire mass through *mass-like* terms in the Lagrangian of the form $\mathcal{L} \propto -\lambda v^2 h^2$ and $\mathcal{L} \propto -g^2 v^2 B^2$. This also gives rise to the *Higgs field* $h(x)$ and a Higgs boson of mass $m_h = \sqrt{2\lambda}v$ which has been experimentally measured [13, 14]. The Standard Model uses a similar process to give fermions masses proportional to their Yukawa couplings y to the Higgs field (expressed by the third term $-y\phi\bar{\psi}\psi$ in equation 1.1), and spontaneous

[2] Although μ^2 is a real number, spontaneous symmetry breaking only occurs for $\mu^2 < 0$.

[3] In reality the scalar potential in the SM is a complex scalar field, the x axis here would correspond to a complex plane.

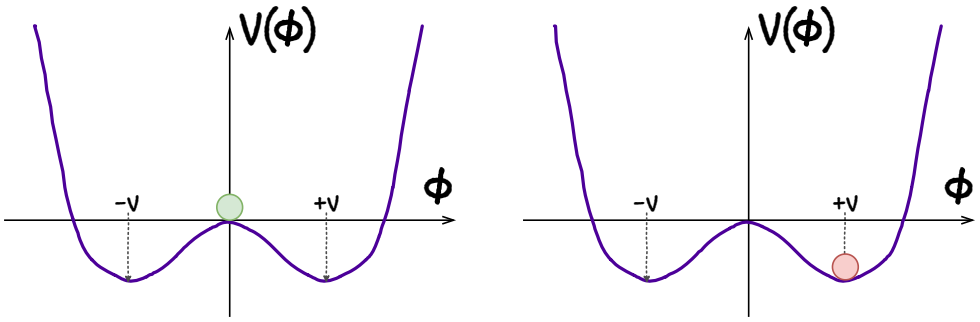


Figure 1.1: Diagrams depicting scalar potential $V(\phi)$ ($\mu^2 < 0$, $\lambda > 0$) with conserved (left) and broken (right) symmetry.

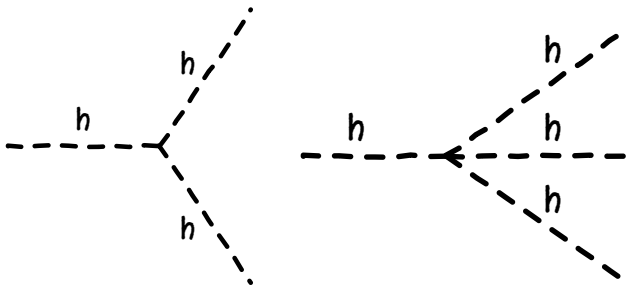
symmetry breaking allows to turn the three weak bosons massive, while keeping the photon massless (through the second term $+i\bar{\psi}\gamma D\psi$ in equation 1.1).

1.2.2 Higgs Boson Self-Coupling

Expanding the Higgs potential around v also adds interaction terms to the Lagrangian of the form

$$\mathcal{L} \propto -\lambda v h^3 - \frac{1}{4}\lambda h^4 \tag{1.3}$$

which describes the Higgs boson self-interaction through its *tri-linear self-coupling* $\lambda_{hhh} = \lambda v$ in the first term and its *quartic coupling* $\lambda_{hhhh} = \lambda_{hhh}/4v$ in the second term. These self-interaction terms allow for coupling of three or four Higgs bosons as shown in figure 1.2.



(a) Tri-linear **(b)** Quartic

Figure 1.2: Diagrams depicting the triple and quartic Higgs couplings.

Notably, the shape of the Higgs potential in figure 1.1, with μ already measured by Higgs searches, is defined by the value of λ_{hhh}. The shape of this potential

can shed light on the electroweak phase transition [4] during the evolution of the early universe [17, 18]. This phase transition of the early universe is one of the possible sources for the asymmetry between matter and antimatter that is observed in the universe [19, 20].

Higgs boson pair-production is possible at the LHC, with one of the main production channels being through the tri-linear coupling in figure 1.2a. Measuring the production cross section of *di-Higgs* events (event producing two Higgs bosons) can directly probe λ_{hhh} and the electroweak phase transition [21], making this a well-motivated area of research in collider physics.

1.3 Di-Higgs Production at the LHC

Although detecting the di-Higgs process is a new milestone for the Large Hadron Collider and its upgrades [22], the production cross section is extremely low. With a predicted gluon-gluon fusion production cross section of $\sigma_{hh} = 32.69^{+5.3\%}_{-7.7\%}\,\text{fb}$ at centre-of-mass energy of 13 TeV [23], only around 4.5×10^3 Higgs pair production events are expected [5] in the LHC run 2 dataset (integrated luminosity $139\,\text{fb}^{-1}$), which is analyzed for this work [6]. Any deviation from this small production rate could be an indication of Beyond the Standard Model (BSM) contributions. This rare process is key to completing our current understanding of the SM and cosmology, and is also highly sensitive to new physics.

1.3.1 Standard Model Di-Higgs Production

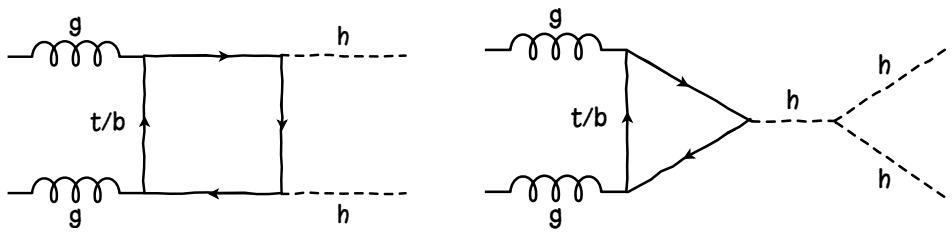
The main production channel for di-Higgs is via gluon-gluon fusion through the processes depicted in figure 1.3. The two diagrams destructively interfere and the delicate balance between them gives the expected SM cross section. Notably the

[4] Interestingly, signs of this process in the early universe may also be observable through gravitational waves produced by the transition in isolated regions, or *bubbles*, of the universe [15, 16].

[5] Triple Higgs production (figure 1.2b) is even rarer than di-Higgs production by three orders of magnitude ($\sigma^{\text{ggf}}_{hhh} < 0.1\,\text{fb}$, yielding a handful of events expected in the full run 2 dataset). Observing events producing pairs of Higgs then provides the only opportunity to directly measure the Higgs boson self-coupling in the LHC.

[6] See chapter 2 for more detail.

triangle diagram in figure 1.3b is sensitive to λ_{hhh} while the box diagram is not, due to the presence of the triple-Higgs vertex in figure 1.3b. However, due to interference between the diagrams, sensitivity to λ_{hhh} is also dependent on the coupling between the Higgs boson and the top quark as evidenced by the presence of two $t\bar{t}h$ vertices in figure 1.3a, and one in 1.3b. The resulting cross section of various single and double Higgs boson production modes as a function of κ_λ (the ratio of a value of the triple-Higgs coupling to the SM expected value) is shown in figure 1.4.



(a) Box **(b)** Triangle

Figure 1.3: Di-Higgs production modes via gluon-gluon fusion.

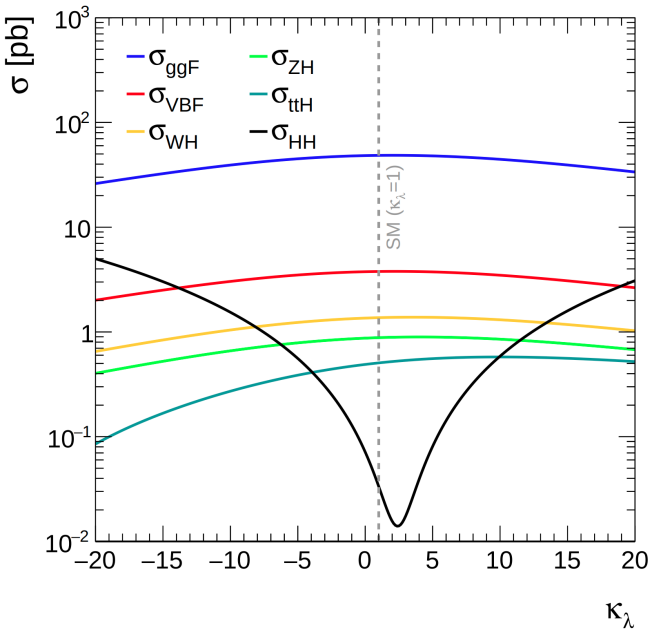


Figure 1.4: Single-Higgs and double-Higgs production cross section as a function of κ_λ. The dashed line intercepts the values corresponding to the SM hypothesis ($\kappa_\lambda = 1$)[24].

1.3.2 Beyond the Standard Model Di-Higgs Production

Heavy resonant particles from BSM physics can be produced through the triangle loop diagram in figure 1.3b (with the intermediate h replaced by the resonance), enhancing the rate at which di-Higgs events are produced at the LHC. Searching for such resonances, and placing limits on their production cross section, is the main focus of this thesis.

Two BSM models are chosen as benchmarks in the ATLAS di-Higgs search presented in this thesis: a heavy scalar S of two-Higgs-doublet models (2HDM) [25, 26], and Kaluza-Klein (KK) gravitons G^*_{kk} predicted in the bulk Randall-Sundrum (RS) [27] model, which can be produced in a similar way to figure 1.3b and subsequently decay to a pair of Higgs bosons.

Two-Higgs-doublet models are a simple extension to the SM [28]. Rather than the single doublet of scalar fields in the minimal Higgs model, these 2HDM include a second doublet. This leads to the model predicting the existence of five Higgs bosons, where the observed particle of mass $m_h = 125$ GeV would correspond to one of two neutral bosons. The other scalar boson S with similar characteristics could be lighter or heavier without any further assumptions. Varying assumptions of masses of these scalar particles, and the mixing between these, change whether the $m_h = 125$ GeV particle is lighter or heavier than the others. The heavier scalars couple to the $m_h = 125$ GeV boson, and can decay to a hh pair if kinematically allowed. The extra particles in these models can provide extra sources of CP violation that could explain the baryon asymmetry in the universe [29]. Supersymmetric models also require at least a second Higgs doublet further motivating the search for 2HDM [28].

Kaluza-Klein modes of the graviton G^*_{kk} in Randall-Sundrum (RS) models could also be produced at the LHC via gluon-gluon fusion, and then decay to a pair of Higgs bosons. RS models propose a 5th dimension exists in a separate brane from our usual 4-dimensional (3 space + 1 time) brane. All the SM particles live in the 3+1 brane and only gravitons are allowed beyond it. This leads to KK modes of the graviton appearing in the 3+1 space as massive G^*_{kk} particles. If

such excitations exist, they would couple to the Higgs boson and thus could be observed as a resonance in the di-Higgs mass spectrum.

These two models are examples of extensions to the SM that could modify the di-Higgs event production rate, and, although results in the later sections of this dissertation will be presented in terms of the resonances they predict, the search results also apply to other spin-0 and spin-2 resonances (see chapter 5). While the ATLAS analysis described in this thesis focuses on both signal models, the different spin of the resonances leads to different kinematics in the subsequent $hh \to 4b$ decay (specially in the angular distributions), leading to slightly different sensitivities for each process.

1.3.3 Di-Higgs Decays Channels

The Higgs boson decay branching fractions, depicted in figure 1.5 are predicted by the SM, and have mostly been experimentally confirmed. In general, the most common decay modes ($h \to b\bar{b}$, $h \to W^+W^-$) also suffer from the largest backgrounds in a pp collision experiment, while the more rare processes ($h \to \gamma\gamma$, $h \to \mu\mu$, $h \to ZZ^* \to 4\mu$) have unique features that distinguish them from backgrounds.

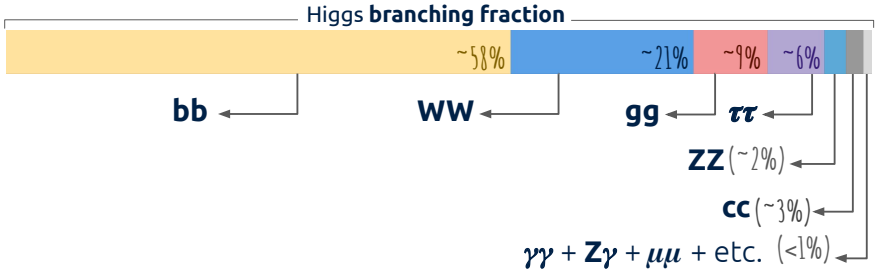


Figure 1.5: Higgs boson decay branching fractions. The percentages represent SM expected values from [30].

Different di-Higgs searches take advantage of the qualities of each Higgs boson decay. Channels incorporating one $b\bar{b}$ decay and a rarer one such as $b\bar{b}\gamma\gamma$ provide background rejection using the photons in the final state, while also having a relatively high cross section due to the $b\bar{b}$ decay. The relative trade-off in branching

fractions is depicted in figure 1.6. The $b\bar{b}b\bar{b}$ channel has the largest branching fraction with around 34% of events, but is obscured by large multijet backgrounds.

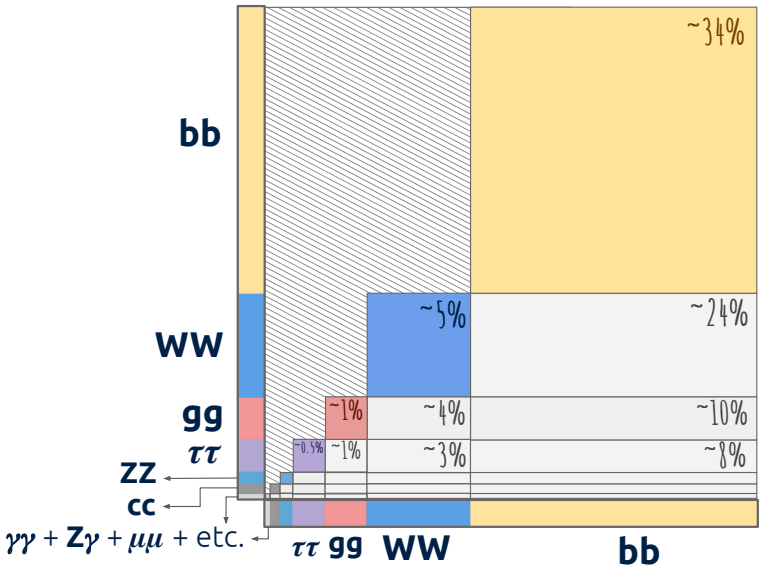


Figure 1.6: Branching fraction for a pair of Higgs bosons represented by area. The numbers on each square show the approximate value (to the nearest 0.5%) of the branching fraction. The area of off-diagonal combinations are under-represented by a factor of two, but the numbers show the correct value. The percentages represent SM expected values from [30].

This thesis focuses on the search for $hh \rightarrow 4b$, where new techniques to identify jets as coming from b-decays and new background modelling techniques promise to mitigate the complex multijet and gluon-produced b-pair backgrounds, and bring improved sensitivity.

1.4 Current Status of Di-Higgs Searches

The ATLAS and CMS experiments at CERN have searched for the Higgs pair production process in the $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$, $b\bar{b}\gamma\gamma$, $b\bar{b}W^+W^-$, $W^+W^-\gamma\gamma$ and $W^+W^-W^+W^-$ [31–37] channels with up to 36 fb^{-1} of data finding no significant deviation from SM predictions. Figure 1.7 shows the upper limit on the production cross section of G^*_{kk} and heavy scalars found by the most recent ATLAS [31] and CMS [38] $hh \rightarrow 4b$

searches. These plots show the upper limit on the production cross section [7] of a heavy scalar resonance (figure 1.7a) and a G^*_{kk} RS graviton (figure 1.7b) as a function of the resonance mass on the x-axes. The solid lines represent the maximum cross section compatible with the data observed (or the *limit*), while the dashed lines show the expected limit under the background-only hypothesis, and the green and yellow bands show the 1 and 2 standard deviation fluctuations on the expected value, respectively. The red lines in these plots represent the predicted cross section given a particular theoretical model, and any mass point where the observed limit is below the theory expectation is said to be excluded. The expected limit is a measure of the sensitivity of the search to deviations from the background-only hypothesis.

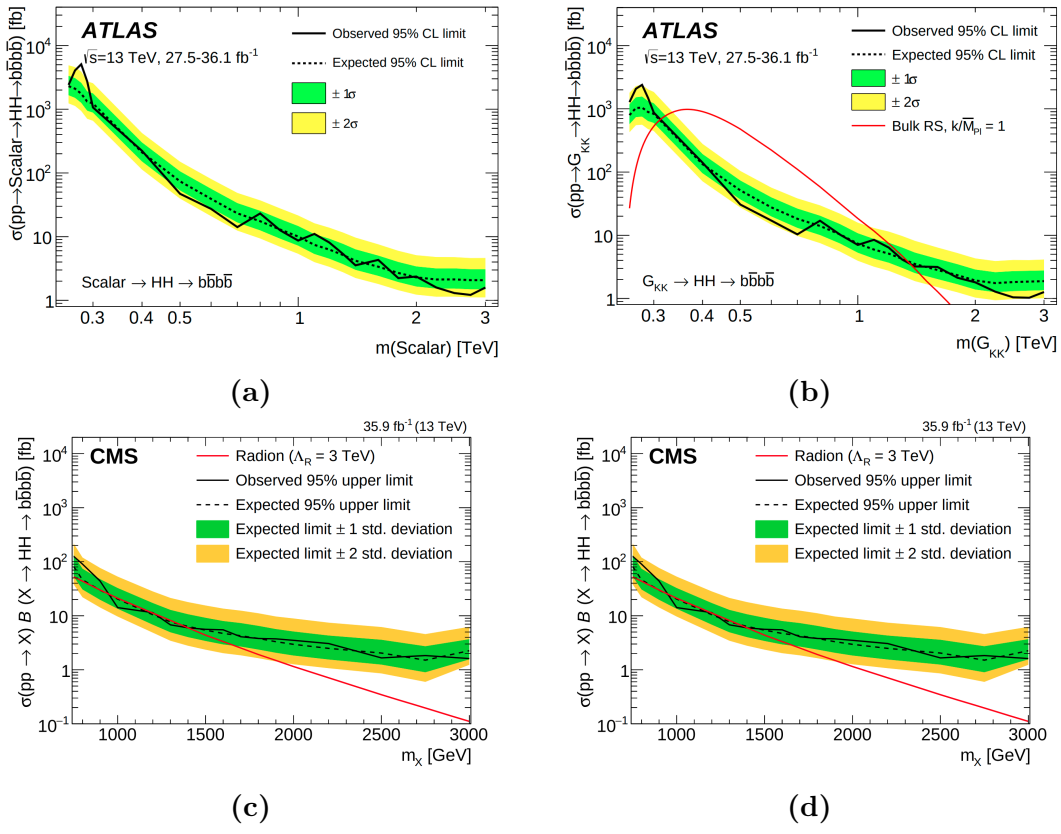


Figure 1.7: 95% CL_S limits on the production cross section of a heavy scalar (left column) and a G^*_{kk} spin-2 (right column) resonances, found by the ATLAS [31] and CMS [38] $hh \to 4b$ searches. The red lines represent the predicted cross section given a particular theoretical model.

[7] As measured by the CL_S statistical test [39], see chapter 5 for more details on the statistical analysis.

Steady progress has also been made by ATLAS and CMS to measure λ_{hhh}. With the discovery of a Higgs-like boson by the ATLAS and CMS collaborations[13][14] with mass $m_h = 125$ GeV, and with the value v which can be calculated with the gauge boson masses, a SM prediction λ_{SM} for λ_{hhh} can be made. A common way to report studies on this quantity is by calculating the ratio κ_λ[8] of a possible variation to the SM prediction:

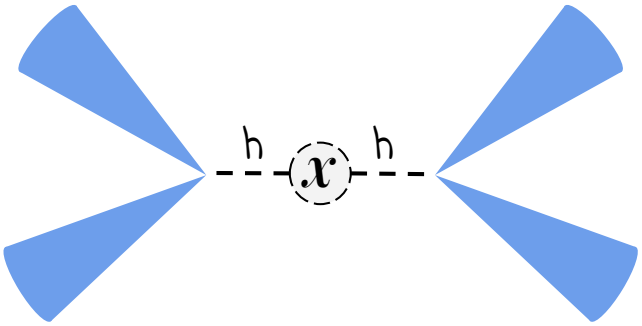
$$\kappa_\lambda = \frac{\lambda_{hhh}}{\lambda_{\text{SM}}} \tag{1.4}$$

The latest results from ATLAS constrain the Higgs self-coupling to $-2.3 < \kappa_\lambda < 10.3$ at the 95% confidence level, by combining direct constraints from di-Higgs searches with indirect constraints from single Higgs searches [24]. The most recent measurement from CMS comes from combining several di-Higgs searches and constrains the self-coupling to $-11.8 < \kappa_\lambda < 18.8$ [40].

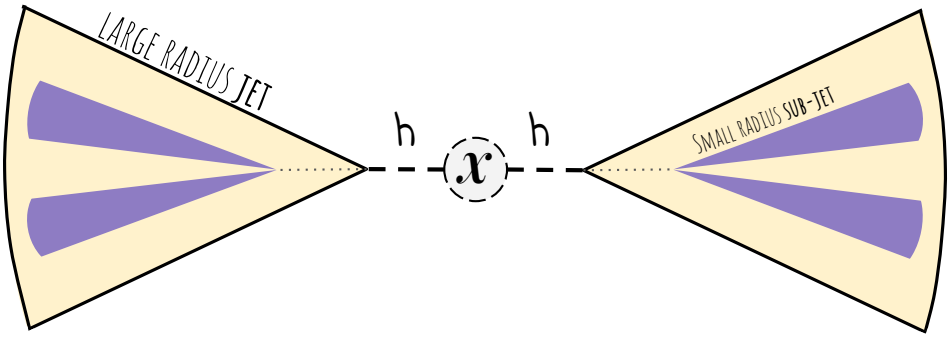
1.4.1 Boosted and Resolved Regimes

A Higgs boson produced as the decay of a massive particle can reach very high transverse momentum (p_{T}), and the signal in the detector can be substantially different for low and high p_{T}. In the case of the $hh \to 4b$ channel, low p_{T} Higgs bosons will each produce two distinct collimated showers of particles known as *jets*, each initiated by the decay of a b-hadron (which makes these b-jets) as depicted in figure 1.8a. In cases where the Higgs boson has large p_{T} (referred to as a *boosted* Higgs), the two b-hadrons can be too close to each other to be reconstructed as individual jets. These boosted decays can be more easily captured in one, larger, jet as shown in figure 1.8b. The large jets can themselves have internal structure (sub-structure) that can give clues of their origin [9]. This motivates separating the search into two regimes, one targeting events with low-p_{T} Higgs bosons by searching for four individually *resolved* b-jets and another targeting events where each high-p_{T} Higgs boson is reconstructed by a single large-radius *boosted* jet.

[8] This is often pronounced "kappa lambda".
[9] See chapter 3 for more details.



(a) Resolved di-Higgs topology



(b) Boosted di-Higgs topology

Figure 1.8: Diagrams depicting the boosted and resolved topologies of a $hh \rightarrow 4b$ decay.

Each regime has its strengths and challenges. Standard Model di-Higgs events are expected to have low p_{T} Higgs bosons, making the resolved channel ideal to search for this signal, but finding which pairs of jets need to be paired together to form a Higgs boson candidate adds a complex step to reconstruct the event. Heavy resonances are much more likely to generate boosted Higgs bosons which will be better reconstructed as a boosted jet, but recognizing properties of *sub-jets* within the large jets (also depicted in figure 1.8b) can be more complicated than in the resolved jets. Other advantage of the boosted channel is the fact that the high p_{T} and mass expected from these jets increases background rejection, where such jets are rare.

1.5 $hh \rightarrow 4b$ Searches in ATLAS

Previous searches for $hh \rightarrow 4b$ have been made in ATLAS, namely [31] which analyzed 36.1 fb^{-1} of the LHC run 2 dataset, and earlier [41] in run 1, both of which included resolved and boosted regimes. The latest result covered a resonance mass

range from 260 to 3000 GeV, and the statistically combined results of the resolved and boosted channels found no significant excess in the resonant production cross section. Only the resolved channel was used to search for non-resonant di-Higgs production, where the upper limit on the production cross section was found to be 13 times the Standard Model expectation.

The main focus of this thesis is the data analysis of the boosted regime of the search for resonant Higgs-boson-pair production, decaying via the $hh \rightarrow 4b$ channel, with the ATLAS detector using the full run 2 dataset. The search sets limits on the production cross section of two BSM benchmark models [10] , similar to the results shown in figure 1.7. This new iteration of the analysis brings several improvements apart from the added data, including new developments in b-jet identification and upgraded sub-jet algorithms.

[10]The non-resonant di-Higgs search will be presented in a separate publication, and in the past it was performed with the resolved regime only as the boosted channel is expected to have little sensitivity due to the low p_{T} Higgs bosons produced through this process.

2
The Experiment

Contents

2.1 The Large Hadron Collider 14
2.2 ATLAS: A Toroidal LHC ApparatuS 15
2.2.1 Sub-detectors 17
2.2.2 Trigger 23

The European Centre for Nuclear Physics (CERN) located in the outskirts of Geneva, Switzerland, aims to uncover the mysteries of particle physics. The largest project in the laboratory is the Large Hadron Collider (LHC), a synchrotron of 27 kilometers in circumference located underground (around 100 metres in depth) and extending into both France and Switzerland [42].

2.1 The Large Hadron Collider

The LHC accelerates two particle beams (either protons or atomic nuclei) in opposite directions around its circumference, and collides them at four points where the beams are focused and crossed. Particle detectors are located at these four points to measure collisions. These four experiments are ALICE, ATLAS, CMS and LHCb [43–46]. Two of these experiments are specialized: ALICE for heavy ion collisions

and LHCb for heavy flavour studies. The other two, ATLAS and CMS, are designed to explore a wider area of particle physics.

The first run of the LHC provided the first ever particle collisions at centre-of-mass energy of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV between 2009 and 2013 [47, 48]. After a two-year shutdown, the LHC run 2 started in 2015 and would continue until 2018. During this period ATLAS collected $139\,\text{fb}^{-1}$ of data: 3.2 fb^{-1} in 2015, 33.0 fb^{-1} in 2016, 44.3 fb^{-1} in 2017, and 58.5 fb^{-1} in 2018 [49]. The LHC proton beams were accelerated to 6.5 TeV during run 2, providing a collision centre-of-mass energy of $\sqrt{s} = 13$ TeV. The $hh \rightarrow 4b$ search presented in this thesis was conducted using proton-proton collision data gathered by the ATLAS experiment during run 2 of the LHC.

Protons in the beam are partitioned into separate segments referred to as *bunches*, typically containing around 10^{11} protons each. When two bunches collide it is likely that more than one pair of protons will interact, leading to multiple interactions per bunch crossing. Inelastic *hard scatter* collisions with high momentum transfer are considered the primary interaction (and will later be reconstructed as the primary vertex, see section 3.1), while the other, generally less energetic or diffractive interactions are called *pileup*. The average number of interactions per bunch crossing was 33.7 for the run 2 dataset [50].

2.2 ATLAS: A Toroidal LHC ApparatuS

The ATLAS detector, at 44 metres long and 25 metres high, is the largest of the LHC experiments. It is comprised of several sub-systems surrounding the proton-proton interaction point in various layers. From inside out these layers are the Inner Detector (ID, dedicated to tracking charged particle's trajectories), the calorimeter system (which destructively measures the energy of a particle that interacts with it), and the muon system (which measures a muon's momentum and tracks its trajectory) [43]. The detector is roughly cylindrical and orientated along the direction of the beam, the round section is referred to as the *barrel* and the flat ends as the *end-caps*. A schematic of ATLAS is shown in figure 2.1.

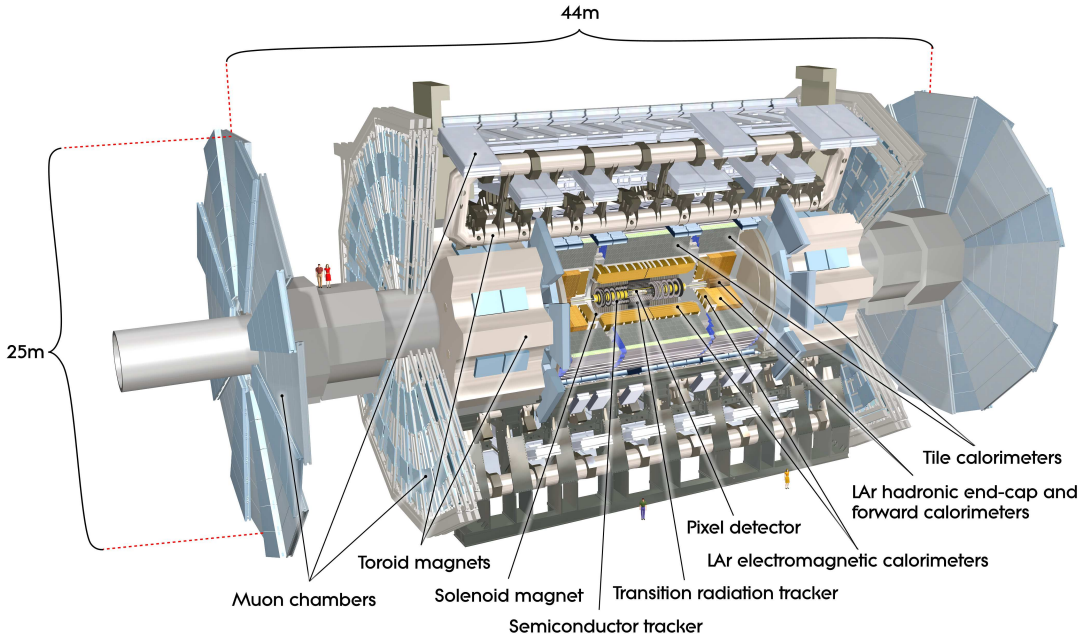


Figure 2.1: Diagram of the ATLAS detector and its sub-detectors [43].

Strong magnetic fields bend the trajectories of charged particles, and by measuring this curvature the particles momenta can be calculated. There are two systems of superconducting magnets in ATLAS for this purpose. A solenoid placed in between the ID and the calorimeters provides a 2 T field to the ID in the direction parallel to the beam [43], and three sets of magnets provide a toroidal magnetic field for the muon system, one wrapped around the barrel and two are placed at its ends.

Coordinate System

The coordinate system used at ATLAS has the z-axis pointing along the beam direction. The x-axis points towards the centre of the LHC ring, and the y axis points upwards (see figure 2.2). The positive direction of the z-axis is defined by the right hand rule.

Due to the symmetry of collisions and the detector, it is convenient to use different coordinates which are similar to spherical coordinates. Points on a plane transverse to the beam are labelled using polar coordinates: the azimuthal angle ϕ, and the distance to the beam r. The position along the direction of the beam is marked by the polar angle θ (similar to latitude on a globe). It is convention to use

the pseudorapidity $\eta = -\ln[\tan(\theta/2)]$ rather than θ since an angular separation $\Delta\eta$ is invariant under a Lorentz boost along the beam direction. It is useful to define too an arbitrary angular separation ΔR in the plane defined by η and ϕ as $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$.

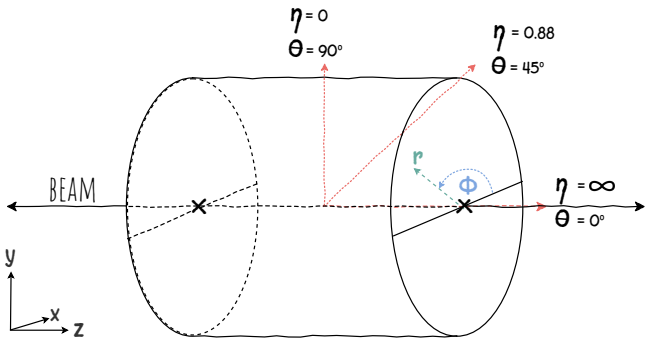


Figure 2.2: Diagram of the ATLAS coordinate system. The positive x-axis points towards the centre of the LHC, and the y-axis points upwards.

2.2.1 Sub-detectors

ATLAS's sub-systems are combinations of different detectors and technologies, exploiting various physical properties and effects to make their measurement.

2.2.1.1 Inner Detector

The ID uses two technologies to track charged particles' trajectories: silicon, and transition-radiation detectors. The main characteristics of the ATLAS detector's ID sub-systems are detailed in table 2.1. The ATLAS ID has a single-track momentum resolution of around $\sigma(1/p_{\mathrm{T}}) \approx 0.4\,\mathrm{TeV}^{-1}$ for $p_{\mathrm{T}} = 200$ GeV, and which starts to degrade in the forward regions ($|\eta| > 2$) [51].

Silicon Detectors

These detectors take advantage of the semi-conducting properties of silicon. Positively-doped silicon is embedded on a negatively-doped silicon substrate, creating a charge-carrier-depleted zone in the boundary. This depleted zone is extended by applying a reverse-bias voltage. As a charged particle travels through the depleted zone it promotes electron-hole pairs to the conduction band, which then drift in opposite

Sub-System	Radius (mm)	Size (μm)	Resolution (μm)
Pixel	$5 - 12$	50×400	10×115
IBL	25.7	50×250	10×72
SCT	$30 - 52$	$80 \times 1.26e{+}5$	17
TRT	$56 - 107$	$4 \times 1.44e{+}6$	130

Table 2.1: Summary of the general characteristics of the three ATLAS ID sub-systems. The values cited for the resolution are representative figures of each in the barrel, as the resolution depends on factors such as each track's trajectory or the location of each module[43, 52–56].

directions due to the bias voltage. The flow of charges register as a signal on the detector's readout electronics attached. Further information on particle detectors can be found in textbooks such as reference [57].

This basic effect can be implemented in a variety of sensor geometries. ATLAS uses two types of silicon sensors in the ID, namely pixel and strip modules. Pixel-detector modules have individual pixel-like cells of 50 μm by 400 μm in size [53]. These are oriented with the shorter side on the plane of the magnetic bending to maximize resolution of momentum measurement. The ID has various layers of pixels surrounding the interaction point, with four layers in the barrel and three discs covering each of the end-caps. The innermost layer of the barrel pixel detectors, the Insertable Barrel Layer (IBL) was added after run 1 of the LHC to improve the detection of B-hadron decays. This was achieved by exchanging the beam pipe with a new one equipped with a layer of pixel detectors which sit at under 3 cm from the beam [52]. The precise resolution of the pixel sensors is optimal to find tracks pointing at a vertex away from the beam that can be left by long lived particles, such as B-hadrons, decaying in the ID. The fine granularity of this detector is also key to avoid missing a particle's hit due to another particle's interaction with the sensor in conditions with high flow of particles (which cause high occupancy of sensor cells).

Silicon-strip modules work in the same way, but they house long strips of silicon (roughly 80 μm by 12 cm) rather than individual pixels [54]. Similar to pixel modules, silicon strips are oriented with their short side along the magnetic bending plane. Resolution on the long-side axis coordinate is improved by having two sensors

in each module with a small stereo angle between them. The silicon micro-strip tracker, also called semiconductor tracker (SCT) system, is formed by four layers of silicon-strip modules in the barrel region and nine layers of disks at each end-cap.

While pixel modules have superior resolution to silicon-strip ones, their cost is higher and they require more inactive material to be added to the detector for their readout [43, 53, 54]. Pixel detectors are hence used in the region closest to the beam, where the area to be covered is smaller and where the particle flux is higher. The cheaper silicon-strip technology was chosen to cover the larger outer layers, where the reduced flux allows to use their coarser granularity without occupancy problems.

Transition-Radiation Tracker

The Transition-Radiation Tracker (TRT) is made of long cylindrical *straw*-like wire chambers. Each chamber is filled with a mixture gasses and has a tungsten wire running down its length [55, 58]. As a charged particle passes through the chamber it can ionize the gas mixture causing electrons and ions to drift in opposite directions due to a voltage between the cylinder and the wire. Polymers are also placed in between the straws to generate transition radiation as a charged particle passes from one medium to another. The signal measured in the straw depends on the amount of transition radiation generated, which in turn depends on the mass of the particle. For two particles with the same momentum, the lighter one will emit more transition radiation [55, 58]. This fact is used to identify electrons in the tracker from other, heavier, charged particles commonly produced in pp collisions such as pions.

The TRT wraps around the SCT (see figure 2.3). In the barrel section the TRT has 73 planes of straws oriented parallel to the beam. In each end-cap TRT there are 80 layers of tubes oriented radially in a ring.

2.2.1.2 Calorimeters

ATLAS has two main calorimeter systems: the electromagnetic calorimeters are optimized to measure energy deposited in them by electromagnetic showers of particles, and the hadronic calorimeters are optimized to measure hadronic shower energy. It is convenient to separate the calorimeters in this way since electromagnetic

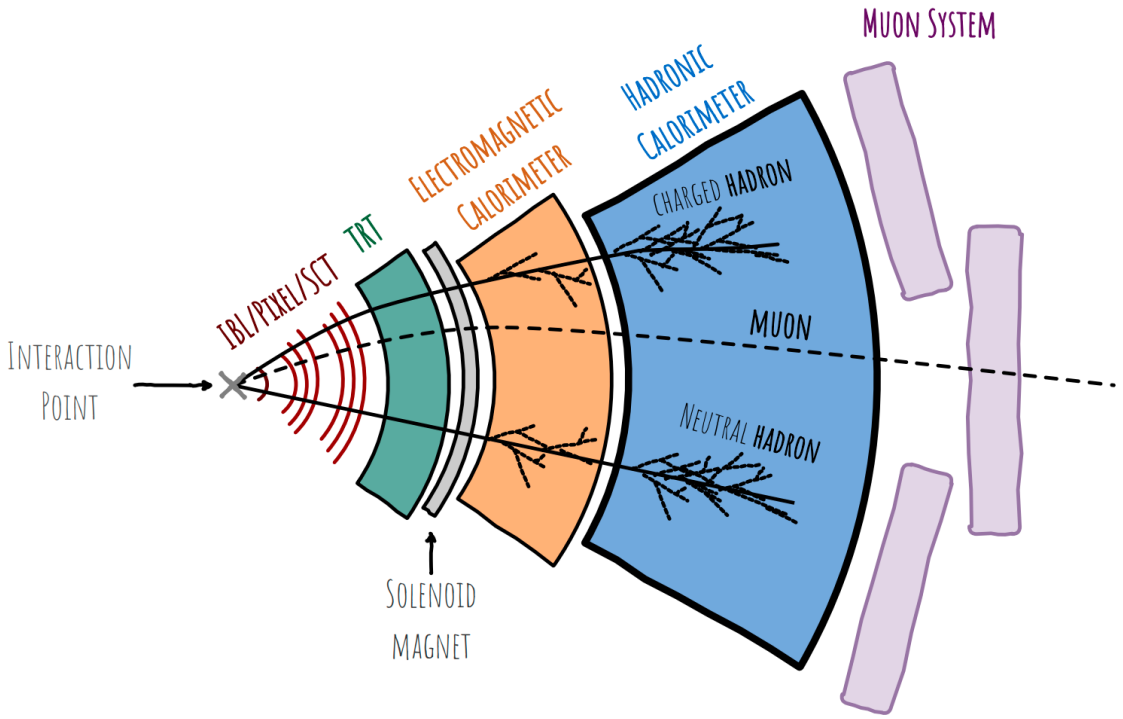


Figure 2.3: Schematic of the sub-detectors of ATLAS representing a slice in ϕ of the plane transverse to the beam. The signatures various particles would leave in each layer are also depicted.

showers tend to be smaller in volume than showers initiated by hadrons. Finer granularity is then needed in the electromagnetic calorimeter to discern features of this smaller shower, and a larger, denser, calorimeter is needed for the hadronic section to allow particles to deposit all their energy [57].

The energy resolution ($\Delta E/E$) of a calorimeter is characterized by the three types of terms in equation 2.1 (summed in quadrature): a statistical uncertainty on the number of ionizing particles gives a term proportional to $1/\sqrt{E}$, a term proportional to $1/E$ from detector and electronics effects, and a constant term due to uneven response of the calorimeter or defective modules.

$$\frac{\Delta E}{E} = \frac{a}{\sqrt{E/\text{GeV}}} \oplus \frac{b}{E} \oplus c \tag{2.1}$$

where a, b, and c are constants. In the high energy regime of the boosted $hh \rightarrow 4b$ analysis the constant term c is the most relevant. These terms are

$a_{\mathrm{EM}} = 0.1$ ($a_{\mathrm{HAD}} = 0.5$), $b_{\mathrm{EM}} = 0.4$ GeV, and $c_{\mathrm{EM}} = 0.007$ ($c_{\mathrm{HAD}} = 0.02$) for the electromagnetic (hadronic) calorimeter[1].

Overall the electromagnetic calorimeter system is 24 to 27 radiation lengths deep, and the hadronic calorimeter is around 10 hadronic interaction lengths deep, with some η-dependent variation. These allow to accurately measure the energy flow through the detector by reconstructing topological clusters of calorimeter cells (see chapter 3) [61].

Electromagnetic Calorimeter

The electromagnetic calorimeter system is formed by various liquid argon (LAr)-lead sampling calorimeters. Layers of each material are placed in alternating order. A particle traveling through the detector is likely to interact with the high-density *absorber* material, lead (Pb), and produce a shower of particles. The products of this shower can ionize the active material, LAr, leading to drifting charges that can be measured by electrodes placed between the LAr and the absorber.

The barrel electromagnetic calorimeter is divided into three concentric sections in the r direction [2]. The absorber and active layers in each section have an accordion shape (see figure 2.4a) with its waves oriented radially so as to not leave any zones without alternating material in ϕ. The end-cap LAr electromagnetic calorimeter has a similar accordion geometry but it is arranged in two wheels per end-cap [59].

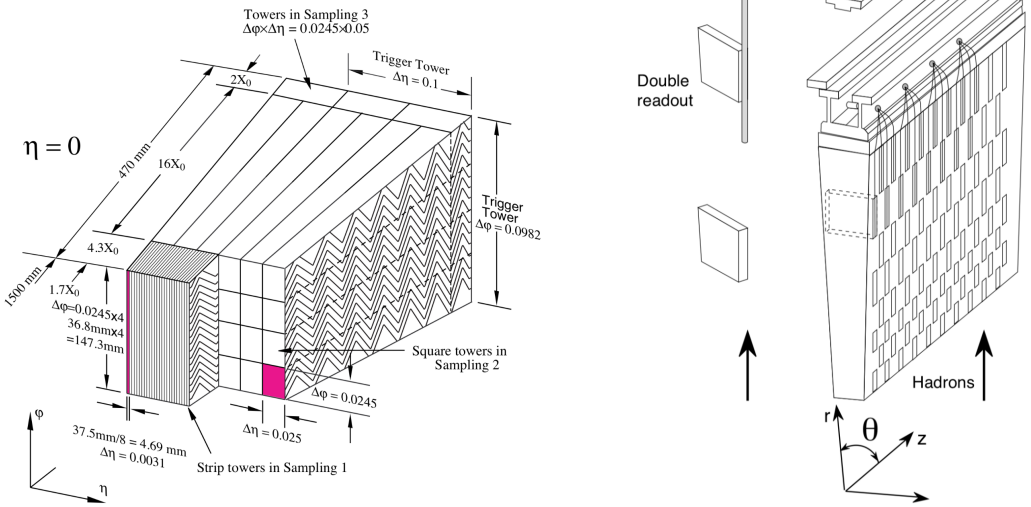
Hadronic Calorimeter

The hadronic calorimeter uses two types of sampling detector: liquid argon-copper in the end-cap region and iron-scintillating *tiles* in the barrel region. The LAr/Cu end-cap calorimeters are similar to the ones described above, with the caveat that the layers are planes rather than accordion shaped [3] as shown in figure 2.4b. The

[1] The exact value of the terms varies throughout the detector, the values presented are representative of the whole system. See references [59, 60] for details.

[2] Having some resolution in the r direction is important for a calorimeter since the 3-dimensional shape of showers initiated by hadrons differ from those initiated by photons or leptons.

[3] A separate LAr calorimeter with different geometry is placed at $3.1 < |\eta| < 4.9$ and is useful for calculating missing momentum and for forward physics, among other applications. This section, however, has no direct impact in this work.



(a) The LAr calorimeter [59]. **(b)** The tile calorimeter [60].

Figure 2.4: Diagram of the (a) liquid argon calorimeter and the (b) tile calorimeter of the ATLAS detector.

tile calorimeter also uses the dense material to maximize the interactions a high energy particle has, but in this case secondary particles from the shower will produce scintillation light when travelling through the active material, which can be gathered by wavelength-shifting fibers and read out by photomultiplier tubes [60].

Each LAr hadronic end-cap calorimeter is formed of two wheels, each divided into two segments for a total of four layers[60]. The tile barrel detector extends over the LAr hadronic calorimeter (see figure 2.1) and is divided into three sections radially [60].

A challenging issue with the hadronic calorimeter is that, due to various factors, the detector's response to electromagnetic and hadronic showers is not the same [4]. The ATLAS hadronic calorimeter uses a software-based approach to overcome this issue. This technique, other calibrations, and the algorithms used to study the showers of particles (referred to as *jets*) in the detector will be discussed in section 3.2.

[4] In other words, the calorimeter is non-compensating.

2.2.1.3 Muon Spectrometer

The muon spectrometer is the outer-most system of the ATLAS detector. Magnetic field for these detectors is provided by the toroidal superconducting magnets (the large barrel toroid in the barrel region, and two smaller toroidal magnets that fit around the end-caps).

Drift Chambers

The muon system is formed by drift tube chambers [5] , which operate similarly to the TRT straws mentioned above but differ in material choice and arrangement [62]. As a muon traverses the detector, the magnetic field will bend its trajectory. The muon's path is recorded as a sequence of hits in the drift tubes. This trajectory is used to calculate the muon's momentum.

2.2.2 Trigger

With the LHC colliding protons every 25 ns, saving data for every bunch crossing is impossible due to both information storage and recording rate limitations. The solution is to use a *trigger* system that quickly evaluates the properties of a collision searching for potentially interesting characteristics, and decides whether to record the event or not.

The ATLAS trigger system works in two stages and uses information from the calorimeter and muon systems. The first level trigger (L1) is a system of custom pieces of hardware which read the detectors, using a coarse granularity, in search for Regions of Interest (a concentration of energy deposits in the calorimeters, for example) [63]. Events passing this first filter are then analyzed by the High Level Trigger (HLT), a software-based system which can perform more complex filters such as ones requiring finer sampling of the detector or b-tagging of jets (see section 3.3). Events that pass the HLT are then saved for analysis. With the trigger system, events are recorded at an average rate of 1 kHz, down from the 40 MHz collision rate.

[5] Different technologies with faster readout are used for the muon *trigger* system. The forward region is also equipped with a layer of cathode-strip drift chambers, although this work does not use forward muons [62].

Certain interesting physics events can have an unacceptably high trigger rate to be all recorded. In these cases only a fraction of events passing a trigger are recorded, which is referred to as a *prescaled* trigger. An *unprescaled* trigger then refers to a filter for which all events that pass are recorded.

2.2.2.1 Jet Trigger

Relevant to this thesis is the ATLAS large-radius-jet trigger [64] (see chapter 3). The L1 trigger that seeds the jet trigger is the level 1 calorimeter trigger (L1Calo) [65]. The L1Calo identifies regions of interest in the calorimeters, which could potentially correspond to particles or jets. At this stage, the calorimeters are sampled in divisions called *towers* of $\Delta\eta = 0.1$ and $\Delta\phi = 0.1$ in size. A jet region of interest is defined by a local maximum of energy deposition in a region of 2×2 towers which is surrounded by a 4×4 or 8×8 ring of towers which satisfy a predefined energy threshold [6]. The jet regions of interest are then used to construct L1 jets, whose p_{T} is used to make the trigger decision.

After this first filter, events are sorted by the HLT. At the HLT level, jets are reconstructed from *topological-cell clusters* using the full set of calorimeter information and are calibrated [7] using similar processes as the ones used for the *offline* physics analyses [66, 67]. Two types of jets are reconstructed at this stage: small jets with radius $R = 0.4$, and large-radius jets of $R = 1.0$. Large-R jets can be used to capture all the radiation products of a high p_{T} Higgs boson, as is the case of the analysis presented in this thesis.

To record events for the ATLAS boosted $hh \rightarrow 4b$ search presented in this work, a single-large-R-jet trigger was used. This trigger is passed if there is at least one large-R jet with p_{T} greater than the trigger's requirement. More details on this and several other data selection criteria are presented in the following chapters 3 and 4.

[6] This threshold depends on the energy of the central towers and size of the surrounding ring, and varies for different parts of the detector to correct for uneven response [65, 66].

[7] See chapter 3 for information on topological-cell clusters and jet calibration.

2.2.2.2 Study of Efficiency of 2017 and 2018 Jet Triggers

During 2017 and 2018 the LHC delivered higher instantaneous luminosity bunch crossings than previous years. This lead to increased pileup interactions, which strains the limited computing capacity of the trigger system. The lowest unprescaled large-R jet triggers used in this boosted $hh \rightarrow 4b$ search implemented an additional requirement on jet mass [8] in order to keep the same p_T threshold from previous years, while running in higher pileup runs (see table 4.1). The extra mass requirement for these triggers was set at $m_J > 40$ GeV and $m_J > 35$ GeV for 2017 and 2018, respectively. Previous year's triggers had only a requirement on large-R p_{T}, and were fully efficient for events in the search with a leading large-R jet with $p_T > 450$ GeV.

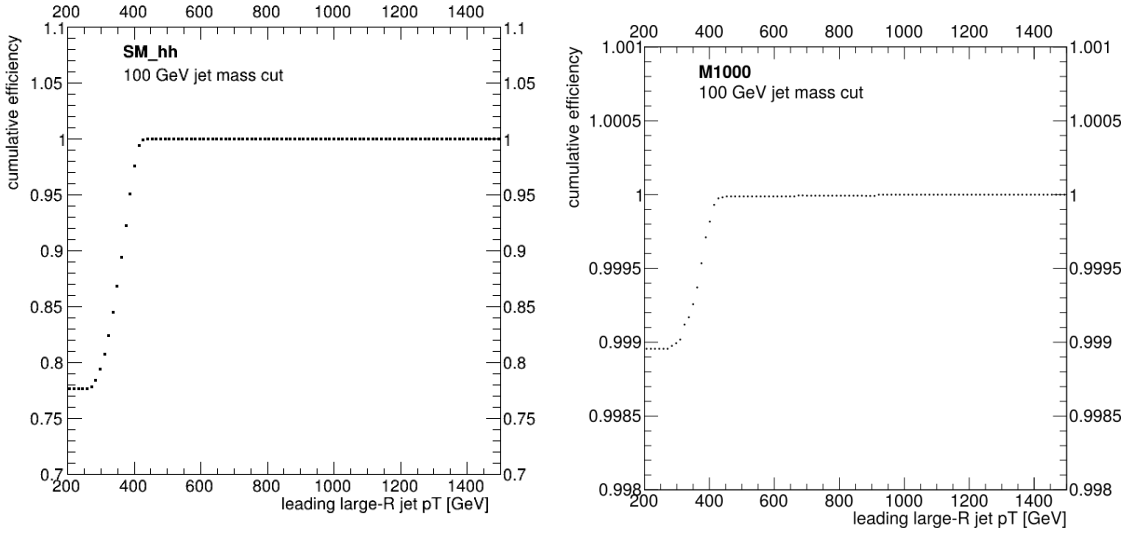
To determine the impact on analysis' signal acceptance from the extra mass requirement in the new triggers, the distributions of jet mass and p_T before and after the trigger filter were used. Two signal simulation samples were chosen to perform this test: a Standard Model $h \rightarrow hh \rightarrow 4b$ as a low-mass benchmark, and a 1 TeV RS graviton $G^*_{kk} \rightarrow hh \rightarrow 4b$ as a heavy resonance. A loose preselection similar to that of the analysis was applied, which required events to have at least two large-R jets, each with pseudorapidity $\eta < 2$ and with a pseudorapidity separation between them of $|\Delta\eta| < 1.3$ [9]. To find the efficiency of these triggers, the inverse cumulative [10] distributions of mass and transverse momentum of the leading large-R jet were calculated before and after the trigger filter, and then divided (after/before).

To find the true efficiency as a function of p_T of this combined trigger, an extra cut requiring the leading jet to have mass greater than 100 GeV was placed temporarily. The plots in figure 2.5 show the trigger cut efficiency of the 2017 trigger (with the tighter mass requirement) in both simulated samples as a function of large-R p_T. From these plots, the p_T value at which the trigger is fully efficient (425 GeV) is obtained.

[8] See chapter 3 for information on jet mass.

[9] The full analysis selection is detailed in chapter 4.

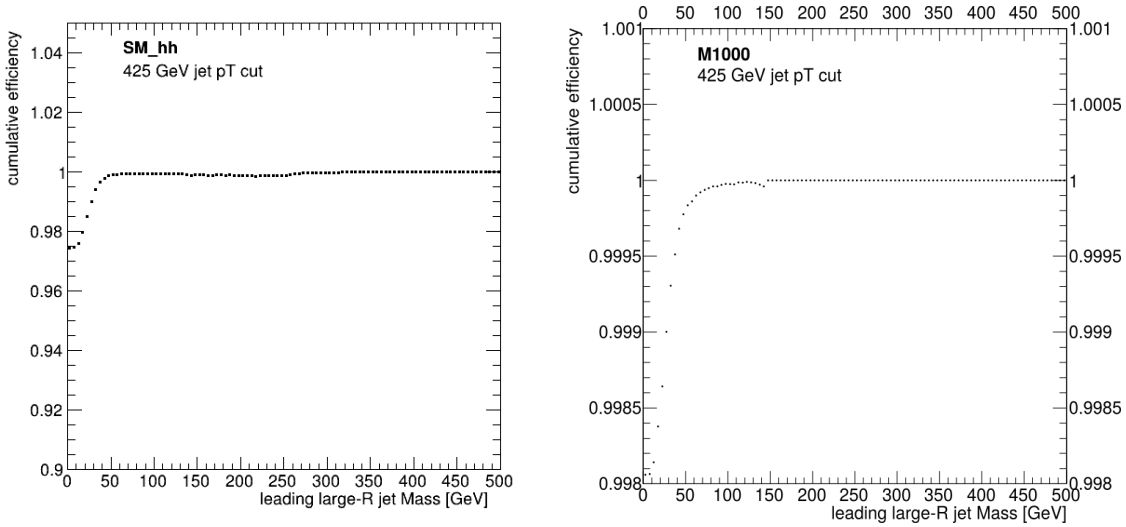
[10] The *inverse* cumulative distribution was used here since the relevant quantity for trigger efficiency is what percentage of events with p_{T} or mass of x *or higher* are accepted by the trigger, rather than the acceptance at a single value of mass or p_T.



(a) **(b)**

Figure 2.5: Efficiency of the `HLT_j420_a10t_lcw_jes_40smcINF_L1J100` trigger on the (a) SM and (b) 1 TeV spin 2 simulated samples. The efficiency is shown as a function of large-R jet p_T. Note the different scale in the y axis in each plot.

This value (425 GeV) is then used as a p_T cut to evaluate the trigger efficiency as a function of jet mass, shown in figure 2.6. These plots show that the new triggers are fully efficient at the thresholds already implemented in the analysis of large-R $p_T > 450$ GeV and mass > 50 GeV. These tests were done to evaluate the 2017 trigger, and since it proved to be fully efficient, we assume that the 2018 trigger (that uses a lower mass cut) will also be fully efficient.



(a) **(b)**

Figure 2.6: Efficiency of the `HLT_j420_a10t_lcw_jes_40smcINF_L1J100` trigger on the (a) SM and (b) 1 TeV spin 2 simulated samples. The efficiency is shown as a function of large-R jet mass. Note the different scale in the y axis in each plot. The small dip around 150 GeV in plot (b) is an artifact of the efficiency being shown as the inverse-cumulative distribution.

3

Physics Objects and Algorithms

Contents

3.1 Tracks and Vertices 29
3.2 Jets 30
3.2.1 Large-Radius Jets 31
3.2.2 Track Jets 35
3.3 Flavour Tagging 39
3.4 Muons 42

This chapter focuses on the physics objects used in the search for boosted $hh \rightarrow 4b$ and their reconstruction. Here physics objects refer to the various elementary particles, or products of their decay, that can be detected by ATLAS, for example muons, electrons, or jets of particles from hadronic decays. Reconstruction refers to how these objects are formed, and their characteristics estimated, using input from the various subsystems of the detector.

The following sections describe tracks, vertices, large-R calorimeter jets, track-jets, and muons, the physics objects used in this search. Special attention is dedicated to track-jets, since part of the work done for this thesis was to test, validate, and implement a new kind of such objects into the $hh \rightarrow 4b$ analysis. This inclusion which was a key improvement for this search (see chapter 6 which details these studies). It is worth noting, however, that ATLAS can reconstruct

many other physics objects, such as individual photons or other types of jets, but these are beyond the scope of this work.

3.1 Tracks and Vertices

As charged particles pass through the layers of the Inner Detector (ID) they deposit energy in the pixels and strips they hit, leaving a path of sequential hits on the tracker's layers. These hits are be connected to form a track that describes the trajectory of the charged particle through the ID, and measuring the curvature of a particle's track within ATLAS's magnetic field its momentum can be calculated. A track reconstruction algorithm matches the sequences of hits most likely to describe the trajectory of a particle originating from the interaction point. Requirements are set for a track to be reconstructed, such as a maximum number of missing hits on the ID layers. Tracks can be reconstructed for particles with a minimum transverse momentum (p_{T}) of 400 MeV [68]. Track reconstruction efficiency within a jet changes with the jet's p_{T} due to the change in density of particles. This is measured in data by testing if a single track's energy deposit in the tracker cells is compatible with single or multiple charged particles. The efficiency in terms of the fraction of tracks lost was found to range from 6.1% to 9.3% in the [200,1600] GeV range probed by the dedicated study in [68].

Reconstructed tracks are then used to find the vertices where protons interacted. With the luminosity conditions of the LHC during run 2 it is very likely for more than one pair of protons to interact, generating multiple vertices per bunch crossing. The vertex formed of tracks with highest $\sum p_{\mathrm{T}}^2$ is determined to be the hard scatter and this is designated as the primary vertex [69]. All the other vertices are assumed to be from collisions with lower momentum transfer and are designated as pile-up vertices. Events with a high number of pile-up interactions can be challenging for searches since particles from pile-up can interfere with the reconstruction of physics objects of interest produced by the hard scatter. The resolution of vertex finding in the transverse plane is better than 20 μm, and 30 μm along the beam

direction. The hard-scatter vertex finding efficiency depends on pileup conditions, but was found to be above 99% [69][1].

3.2 Jets

Quarks and gluons produced in pp collisions at the LHC produce a collimated shower of hadrons reconstructed as a jet in the detector. Jets can be formed from a variety of objects, including topological clusters of energy deposits of calorimeter cells (*topoclusters*[2] [61]) produced by a shower of particles, the tracks left by charged particles, *truth* particles from a simulation, and others. Jets formed from topoclusters and particle tracks are used in the search described in this thesis.

The anti-k_t algorithm

A jet clustering algorithm commonly used in collider experiments is the anti-k_t algorithm. Jets are reconstructed with a radius size determined by the parameter R in the jet clustering algorithm, measured in the η-ϕ plane, as defined in chapter 2 (see figure 2.2). This iterative jet clustering process is based on calculating the distances between all entities i and j to be clustered as $d_{ij} = \min(p_{Ti}^{-2}, p_{Tj}^{-2})\frac{\Delta R_{ij}^2}{R}$, and the distance of each to the beam B as $d_{iB} = p_{Ti}^{-2}$. The smallest of these distances is found, and if one of the d_{ij} is smallest, entities i and j are combined. The combination continues until d_{iB} is the smallest distance. At this point i is called a jet, and removed from the list of entities, and the process is resumed.

Jets clustered with the anti-k_t algorithm have the positive properties of being collinear and infrared safe[3] while also providing jets with conical shape. The jet's characteristics, such as its p_{T}, mass, or direction, are related to the properties of the particle that initiated the hadron shower [70].

[1] This efficiency assumes a large fraction of charged particles being produced (which is the case for a $hh \to 4b$ event), and is slightly lower for processes with few charged tracks.

[2] These are topological clusters of calorimeter cells in three dimensions. The cluster shape and location information can be used to apply calibrations and corrections. Calorimeter cells with large energy deposits are taken as seeds for the algorithm, and are combined with adjacent cells if they meet an energy threshold as compared to the cell noise. The clustering algorithm implicitly suppresses noise.

[3] Infrared and collinear safety refer to observables which are unchanged by soft emissions or collinear splitting of particles [70].

Jet Mass

Estimating a jet's mass is an excellent way to discriminate whether the progenitor particle was a massive top quark or Higgs boson, rather than a light hadron or a massless gluon. Nevertheless, since the definition of a jet is only given by the jet clustering algorithm that formed it, defining a jet mass is non trivial. A first approach is to estimate it by calculating invariant mass m^{calo} of the calorimeter topoclusters that formed the jet J:

$$m_{\text{calo}} = \sqrt{\left(\sum_{i \epsilon J} E_i\right)^2 - \left(\sum_{i \epsilon J} \vec{p_i}\right)^2} \tag{3.1}$$

where E_i and $\vec{p_i}$ are the energy and momentum of the i-th topocluster. These quantities are calculated assuming topoclusters are massless ($E_i = |\vec{p_i}|$) and using the spatial location of the cluster. This base definition of jet mass can be expanded by using information from other sub-detectors and calibrated depending on the jet's use case, such as the large-radius jets described in 3.2.1.

3.2.1 Large-Radius Jets

In this search, a large-radius (large-R) jet is used to reconstruct the decay products of a Higgs boson to $b\bar{b}$. The angular spread ΔR of the decay products of a particle of mass m and p_{T} varies as $\Delta R \approx 2m/p_T$. As a result the b-jets from a Higgs boson with $p_{\text{T}} = 315$ GeV will be too close together to be reconstructed as two $R = 0.4$ jets without overlap. A large-R jet could catch the contribution of both b-jets and thus could be used as a Higgs boson candidate.

Large-R jets are formed by combining topoclusters of calorimeter cells via the anti-k_t algorithm with $R = 1$ and calibrated to the hadronic scale[4]. The larger area of these jets makes them more susceptible to include unwanted contributions from sources other than the hard scatter event, such as pile-up interactions, which can degrade the energy and mass resolution. To counteract this, large-R jets go through a jet grooming process known as trimming [72], diagrammed in figure 3.1. They

[4] See the calibration section below for a brief summary of large-R jet calibrations, or the fully detailed explanation in references [61, 71].

are reclustered into subjets with radius $R_{sub} = 0.2$ using the k_t algorithm. These subjets are only kept if their p_{T} is higher than 5% of the original large-R jet's.

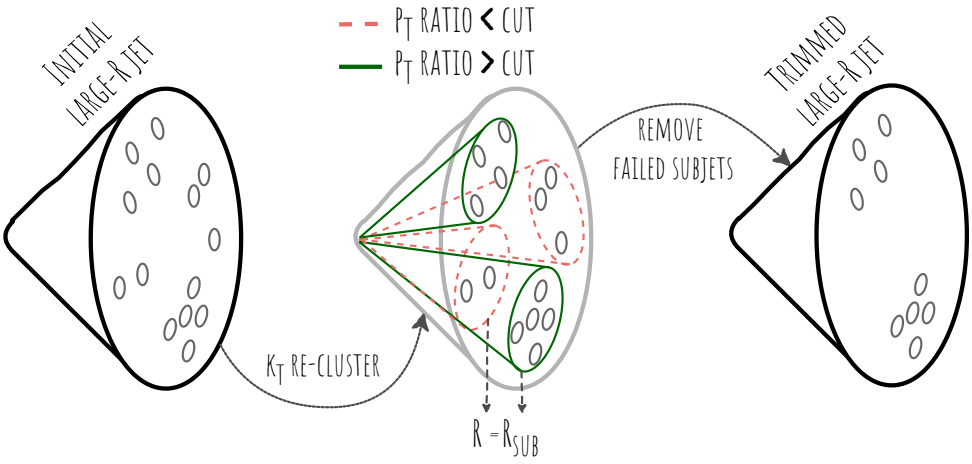


Figure 3.1: Diagram of the jet trimming process. The initial large-R jet is reclustered into smaller jets of radius R_{sub} with the k_t algorithm. A cut is then placed on the p_{T} of reclustered jets, and the resulting trimmed jet is formed from the remaining small jets [73].

Combined Jet Mass

The mass of large-R jets is then calculated with the combined mass technique, which uses both calorimeter and tracking information to improve the jet mass resolution [74]. Mass resolution of topoclusters, constructed only from calorimeter information, underperforms for highly collimated particles due to the relatively coarse spatial granularity of the calorimeters.

Information from the higher resolution tracker can be used to improve mass resolution for boosted jets. The *track-assisted* mass m^{TA} is defined as

$$m^{\mathrm{TA}} = \frac{p_T^{\mathrm{calo}}}{p_T^{\mathrm{track}}} \times m^{\mathrm{track}} \tag{3.2}$$

where p_T^{calo} is the p_{T} of the large-R jet, p_T^{track} is the p_{T} of the four-vector sum of the tracks associated with the large-R jet, and m^{track} is the invariant mass of this four-vector sum with track mass set to the mass of a pion [74].

Finally, track-assisted and calorimeter-based masses are combined to form the combined mass $m^{\mathrm{comb}} = a \times m_{\mathrm{calo}} + b \times m^{\mathrm{TA}}$ where a and b are computed to optimize

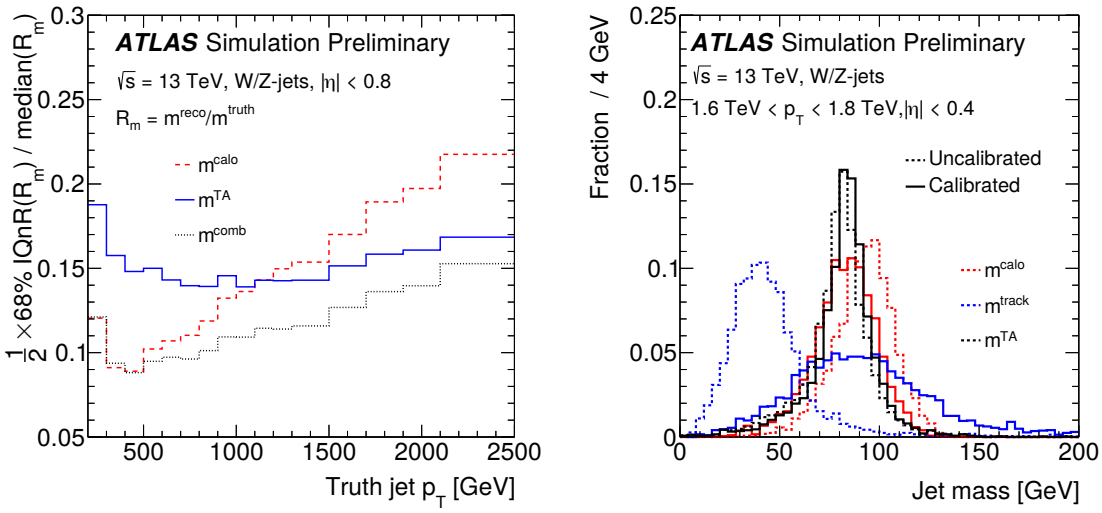
the resolution of m^{comb} [74]. This exploits the fact that the calorimeter-based mass is not explicitly used in the calculation of track-assisted mass.

Figure 3.2 shows the impact of using track information to improve jet mass measurements. The resolution of various mass calculations is shown in figure 3.2a, where the combined mass has better resolution than both the other methods throughout the jet p_{T} range. This is measured in simulated W/Z jets formed by clustering the *truth* particles with the anti-k_t algorithm to form truth-jets. This resolution is quantified by using the half of the 68% interquartile range ($\text{IQnR} \equiv 84^{\text{th}}\text{percentile} - 16^{\text{th}}\text{percentile}$) divided by the median, which corresponds to the standard deviation in the ideal Gaussian case but is robust against outliers [74]. Figure 3.2b shows the mass distribution of W/Z jets from simulated events calculated using inputs from calorimeter only (m^{calo}), tracks only (m^{track}), and the track-assisted mass (m^{TA}) which uses track information corrected with calorimeter information. The track-based mass improves dramatically by combining it with calorimeter information, making the track-assisted mass peak much sharper and reducing its width.

Large-R Jet Calibrations

Before the jets are reconstructed, calorimeter topoclusters are calibrated using the *local hadronic cell weighting* (LCW) calibration, which corrects for the non-compensation of the hadronic calorimeter, losses due to intrinsic noise suppression in the clustering algorithm, and losses due to inactive detector material near the topocluster [61].

Next, large-R jets are reconstructed and they are then trimmed as mentioned above. The energy, mass and pseudorapidity of these jets are calibrated to correct for inefficiencies due to reconstruction, inactive detector areas, pileup, or other factors. These calibrations are applied as multiplicative factors to a jet's property (p_{T}, mass, η) to correct for known mismeasurement effects [71]. These calibrations are derived in MC by comparing a reconstructed jet's properties to the properties of the truth particles in the jet [74].



(a) **(b)**

Figure 3.2: Jet mass resolution as a function of truth jet p_{T} (a) and jet mass distribution (b) of W/Z jets from simulated events. Plot (a) shows the jet mass resolution of calorimeter-based mass m^{calo}, track-assisted mass m^{TA}, and combined mass m^{comb}. Plot (b) shows the mass distributions for these same mass definitions before(after) calibration in dashed(solid) lines. The y axis in (a) shows half of the 68% inter-quantile range (IQnR) divided by the median of the jet mass response, which is used as an outlier-insensitive measure of the resolution [74].

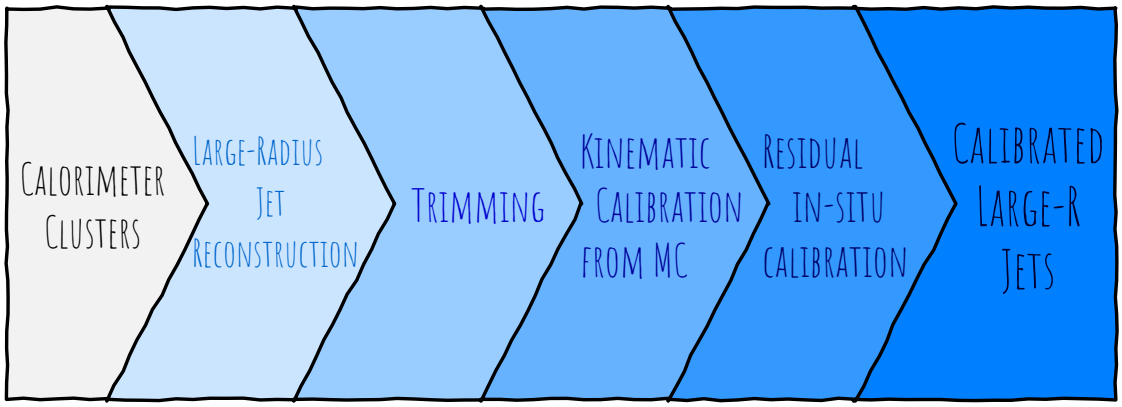


Figure 3.3: Sequence of calibrations and corrections applied to large-R jets. Full description in [71].

Extra *in-situ* calibrations are applied to data only to reduce the uncertainty on the jet's energy and mass. *In-situ* calibrations are derived from data using events with large-R jets recoiling against an already well-calibrated object, such as a photon or a reconstructed Z-boson. Unaccounted for inefficiencies in the jet reconstruction or the detector response can be corrected by balancing the jet with

the other well-measured object [75]. The sequence of calibrations applied to the large-R jet used in this analysis are summarized in the diagram in figure 3.3.

3.2.2 Track Jets

Clusters of tracks made using the anti-k_t algorithm, called track jets, are associated with a large-R jet with the ghost-association method[5]. Using jets formed by tracks, rather than calorimeter cells, provides a way to obtain an independent b-tag (see section 3.3) that can be associated with any calorimeter-based jet, such as groomed large-R jets which are optimized for highly boosted topologies [77].

Large-R jets can be reconstructed with small cone sizes to allow identification of collimated b-jets within the large-R jet as two separate objects, which contributes to a better Higgs boson candidate reconstruction. Tracks around the axis of a track jet can be passed through a b-tagging algorithm [78] (see section 3.3) to determine if they come from b-quarks. The track jets associated with a large-R jet are then used to infer if the large-R jet could be from a Higgs boson. Track jets are also required to have $|\eta| < 2.5$ since the jet must be within the coverage of the ID tracker (see chapter 2).

Two types of track jets, fixed-radius (`FR`) and variable-radius (`VR`), are used in this study. `FR` track jets are used as a benchmark since this was the object used in a previous search for this process [31], and `VR` track jets are explored as a new option to overcome the shortcomings of the older technique.

3.2.2.1 Fixed-Radius Jets

Fixed-radius track jets are made by clustering tracks with the anti-k_t algorithm with the distance parameter $R = 0.2$. Of the track jets associated with a large-R jet only the ones composed of two or more tracks and with $p_{\mathrm{T}} > 10$ GeV are kept in order to suppress light flavoured jets in favour of b-jets.

[5] Ghost-association consists of assigning a negligible p_{T} value to a track jet and then including it within the objects that get clustered into a large-R jet. With such a small p_{T} the track jet does not change the clustering of the large-R jet, but this way each track jet gets associated exclusively to one jet [76].

The fixed size of this type of track jets sets a threshold on their capacity to reconstruct highly collimated jets individually. As the p_{T} of a Higgs boson gets larger, the angular separation between the b-jets coming from its decay will decrease, to a point where they start overlapping. In this scenario, depicted in the left half of figure 3.4, some tracks could be clustered into the wrong track jet, which both greatly reduces the efficiency of b-tagging and risks wrongly b-tagging a jet not coming from a b-quark. Ultimately this effect limits sensitivity of the search for cases with highly boosted Higgs bosons.

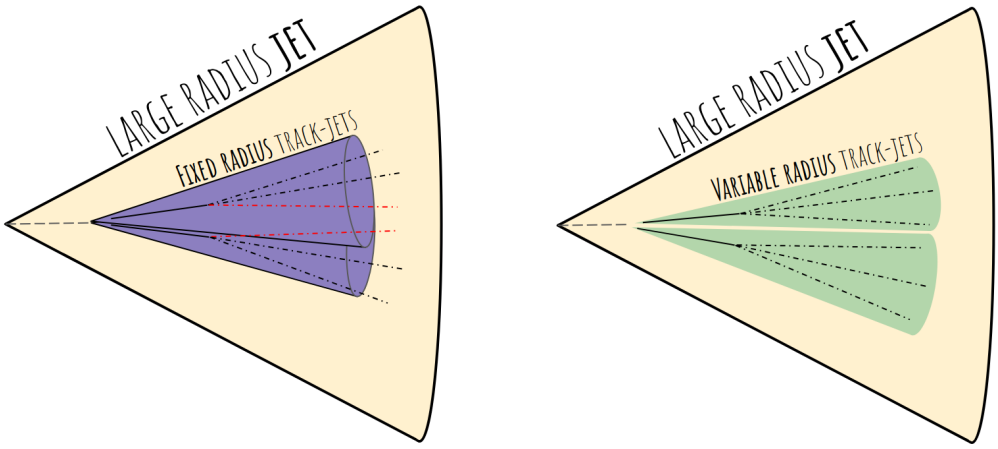


Figure 3.4: Diagram depicting track jets, reconstructed with fixed and variable radius, originating from a boosted Higgs boson. The secondary vertex from the decay of the b-hadron and tracks from the resulting charged particles are shown within the track jets. Tracks shown in red are contained in both jet's cones and could be wrongly assigned.

3.2.2.2 Variable-Radius Jets

One way to mitigate the issue of overlapping track jets tested in this study is to allow their radii to decrease as their p_{T} increases[6]. The jet algorithm, introduced in reference [79], is combined with anti-k_t which allows the radius R_{eff} of a VR jet of transverse momentum p_{T} to scale as

$$R_{eff} = \frac{\rho}{p_T} \tag{3.3}$$

[6] The boosted $hh \rightarrow 4b$ analysis was one of the first adopters of these objects, and one of the prime use cases. I was responsible for implementing them in the analysis, testing their performance, and working in conjunction with the Flavour Tagging group to investigate their behaviour. Findings of these studies are detailed in chapter 6.

where ρ is a parameter capturing how fast the radius reduces with p_{T}. VR track jets in this search have this parameter set to $\rho = 30$ GeV. The size thresholds $R_{max} = 0.4$ and $R_{min} = 0.02$ are also specified in the VR algorithm to prevent jets from growing too large, or shrinking below the tracker resolution.

The chosen values of R_{max}, R_{min} and ρ are determined by scanning a range in each variable and finding the combination that gives the best result over the larger Higgs boson p_{T} range[7]. All optimization and tests of VR track jets (detailed in reference [80]) were performed on a simulated sample of RS gravitons G^*_{kk} decaying to Higgs bosons, which subsequently decay to pairs of b-quarks. To gauge the efficiency of the VR jet algorithm, the *double subjet b-labelling* efficiency $\epsilon_{\mathrm{truth}}^{\text{Double subjet b-label}}$ is defined as

$$\epsilon_{\mathrm{truth}}^{\text{Double subjet b-label}} = \frac{N(\text{Double subjet b-label} \mid \text{Higgs jet})}{N(\text{Higgs jet})} \tag{3.4}$$

where Higgs jet refers to a large-R jet which has a truth Higgs boson and two truth b-hadrons within its area, and a subjet is b-labelled if its axis is within $\Delta R < 0.3$ of a truth b-hadron[8]. A large-R jet will only get the double subjet b-label if the two truth b-hadrons are matched to its two highest p_{T} subjets [80].

Figure 3.5 shows the double subjet b-labelling efficiency of the VR jet algorithm as a function of Higgs jet p_{T}, for various values of the ρ parameter. The same efficiency measure is shown for FR jets with $R = 0.2$. As the Higgs boson p_{T} increases, the curve corresponding to FR jets drops sharply at the point where the subjets would start merging, while the others, corresponding to VR jets, stay comparatively flat. VR subjets can accommodate for highly collimated b-jets better than FR as shown in figure 3.4. Avoiding the subjet overlap is critical since this will ensure the b-tagging algorithm gets the correct tracks as input for each subjet.

While figure 3.5 shows that b-labelling efficiency at high p_{T} could be improved substantially by using VR track-jets, it is also important to ensure the capacity to reject non-Higgs-boson jets is also improved, or at least maintained. Figure

[7] See figure 3.5 for an example of how varying ρ affects the algorithm's performance.

[8] If more than one subjet's axis is within this ΔR range, the label will be given to the closest one.

3.7 shows the QCD jet rejection (defined as the inverse of the QCD jet efficiency) as a function of Higgs-boson jet efficiency for double-b-tagging[9]using the VR sub-jet algorithm. QCD jet rejection is matched or improved by VR track jets as compared to FR ($R = 0.2$) for the vast majority of values of Higgs-boson-jet efficiency, especially above 0.7.

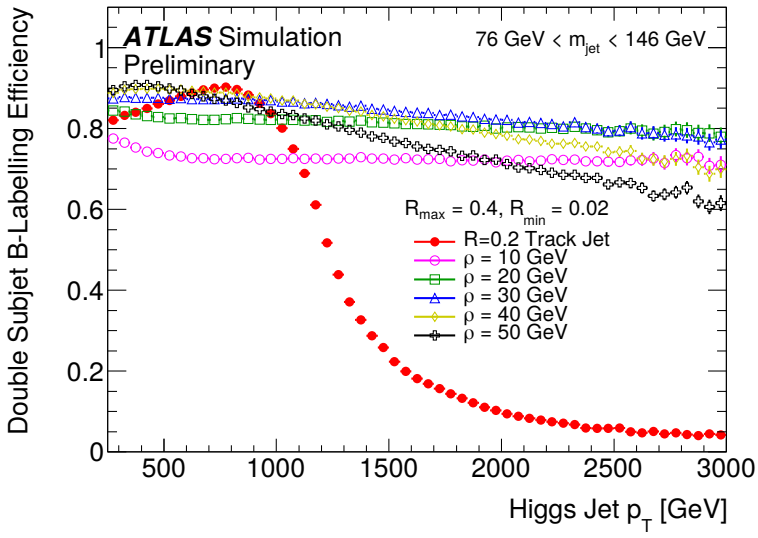


Figure 3.5: Double b-labelling efficiency using VR subjets associated with a Higgs boson candidate (here referred to as Higgs Jet) as a function of its p_{T}. VR subjets are constructed with $R_{max} = 0.4$, $R_{min} = 0.02$, and various values of ρ. The efficiency of FR $R = 0.2$ subjets is shown as well [80].

[9] In this plot, double-b-*tagging* differs from the double subjet b-*labelling* in figure 3.5. Here subjets are tagged using the MV2c10 b-tagging algorithm, rather than truth-labelling subjets by matching them to simulated particles.

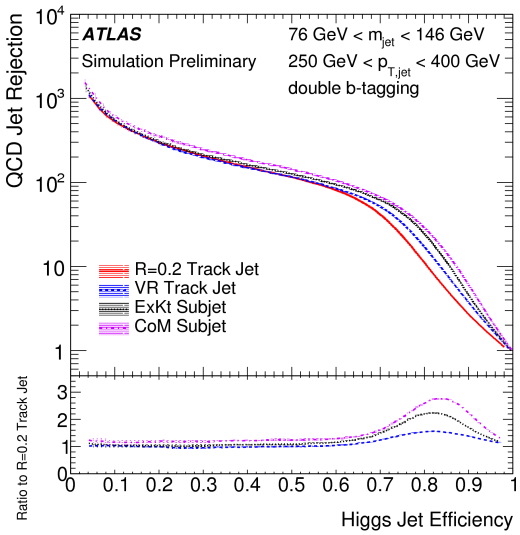


Figure 3.6

Figure 3.7: QCD jet rejection as a function of Higgs-boson jet efficiency for double-b-tagging using the VR sub-jet algorithm. The exclusive-kt (ExKt), center-of-mass (CoM), FR $R = 0.2$ subjets are shown as well. This plot shows efficiencies for large-R jets with mass 76 GeV $< m_{\text{jet}} <$ 146 GeV and transverse momentum 250 GeV $< p_{\text{T,jet}} <$ 400 GeV [80]. VR track jets match or exceed the performance of FR $R = 0.2$ subjets for nearly all values of Higgs-boson-jet efficiency.

3.3 Flavour Tagging

Identifying whether a jet originated from a b decay (*b-jet*), as opposed to a lighter quark (*light-jet*), can be a powerful tool for searches for new particles or for precision measurements of the Standard Model.

Flavour-tagging algorithms use the longer lifetime of b-hadrons to separate b-jets from light jets during the b-tagging process. This extended lifetime allows the b-hadron to travel from the primary vertex into the ID before decaying[10], which results in a set of tracks pointing to a secondary vertex displaced from the primary vertex. Information such as the impact parameter of the tracks associated with a jet, or the presence of a secondary vertex reconstructed from these tracks, is used in to predict the likelihood that the jet originated from a b-decay. A diagram showing these parameters relevant for b-tagging is shown in figure 3.8. These algorithms not only use kinematic information directly, but they exploit kinematic dependences

[10] A b-hadron with $p_{\text{T}} = 50$ GeV already has a mean flight path around of 5 mm.

of the relevant parameters to improve their efficiency.

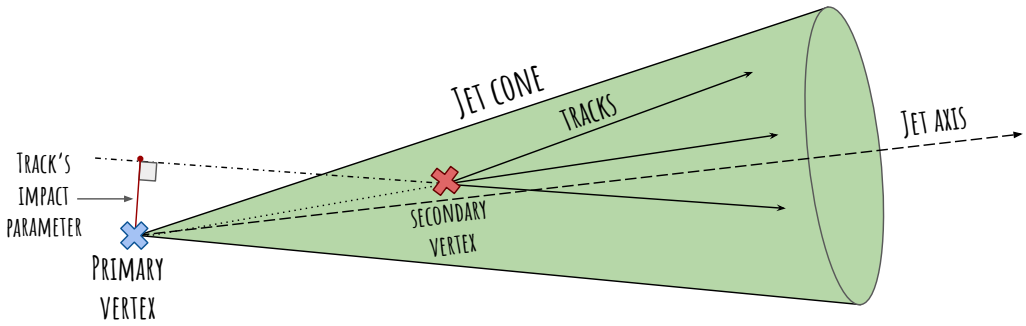


Figure 3.8: Diagram depicting some of the parameters relevant for b-tagging.

Various working points are defined for the b-tagging algorithm, each with different false positive (light-jets identified as b-jets) and false negative (b-jets identified as light-jets) rates. A tighter working point corresponds to a lower false positive rate, but it will also yield more false negatives than a looser working point. In this study the 77% working point is used, meaning that it is expected to b-tag 77% of the true b-jets. Other working points are set at 60, 70, and 85% b-jet efficiency.

Two b-tagging algorithms were used during the course of this study: the `MV2c10` and `DL1r` algorithms [81]. These two take similar basic inputs and combine them in different ways to reach the final b-tagging decision. A boosted decision tree is used to calculate the final discriminating variable of the `MV2c10`, while a recursive neural network does this job in the case of `DL1r`.

Figure 3.9 shows the b-tagging efficiency versus the light-flavour jet rejection rate (defined as the inverse of the false positive rate) on the y-axis for various flavour tagging algorithms on a simulated sample of $t\bar{t}$ events[11]. The first two taggers are the ones used in this study, `MV2c10` and `DL1r` , which are referred to as *high-level*

[11] The studies detailed in reference [82] were conducted on a sample of $t\bar{t}$ events and b-jets were reconstructed as hadronic jets built from topoclusters of radius $R = 0.4$. Optimization and performance measurements are done in the same way for the track jets used in the boosted $hh \to 4b$ search.

taggers because they use the output of the other three *low-level* algorithms on the list as inputs[12]. The low-level tagger `IP2D` uses the transverse impact parameter to create a discriminant, while `IP3D` also uses the longitudinal impact parameter. The `SV1` algorithm tries to find only secondary vertex with the provided tracks to calculate the tagging score, while the more complex `JetFitter` algorithm tries to reconstruct the full hadron decay chain.

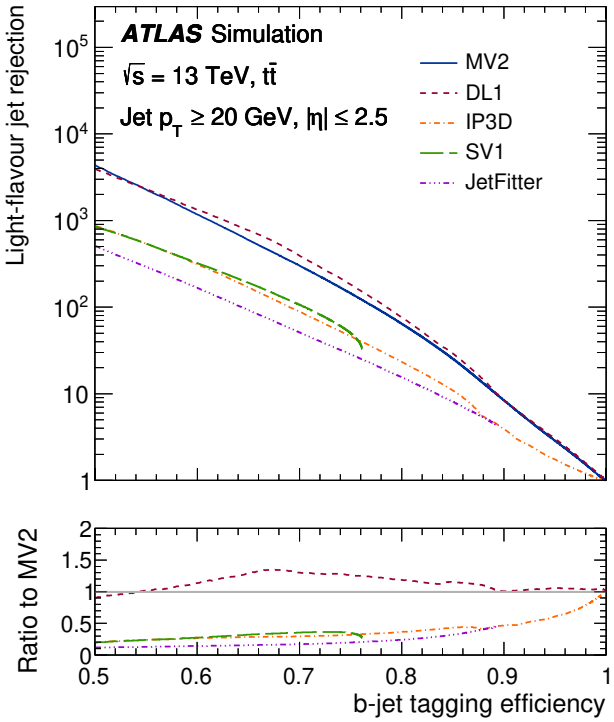


Figure 3.9: *b*-jet tagging efficiency vs light-flavour jet rejection for two *high-level* (`MV2` and `DL1`) and three low-level (`IP3D`, `SV1`, and `JetFitter`) taggers on simulated $t\bar{t}$ events. The lower panel shows the ratio of each curve to the `MV2` for comparison. Efficiency and mistag rates were computed using a sample of simulated $t\bar{t}$ events [82].

Since *b*-tagging efficiency and mistag rates are calculated using simulated events, a calibration is applied to data to correct for mismodelling of the inputs to *b*-tagging algorithms in simulated events. As described in reference [82], a data sample enriched in *b*-jets is used to derive the efficiencies in data ϵ^{data}[13] and a p_{T} dependent

[12] This plot shows the performance of the `DL1` tagger, a precursor of `DL1r` which shares most of its characteristics, the main difference being the architecture of the neural network used to construct the discriminant variable. `MV2c10` was renamed to `MV2`.

[13] This sample is obtained by selecting $t\bar{t}$ events and exploiting the near 100% branching fraction of $t \to bW$.

scale factor defined as $SF = \epsilon^{\text{data}}/\epsilon^{\text{MC}}$ is calculated (a similar *anti-tag* scale factor is derived for jets failing b-tagging). These multiplicative scale factors can then be applied as necessary in each event to add per-jet corrections.

A feature of b-tagging algorithms relevant to this study is their suboptimal performance for high p_{T} jets. High p_{T} jets tend to have more tracks associated to them leading to an increased number of fake secondary vertices being reconstructed, which degrades the b-tagging efficiency [81]. This effect is evidenced in the plots shown in figure 3.10. Both light-jet rejection and b-tagging efficiency drop as the jet's p_{T} increases.

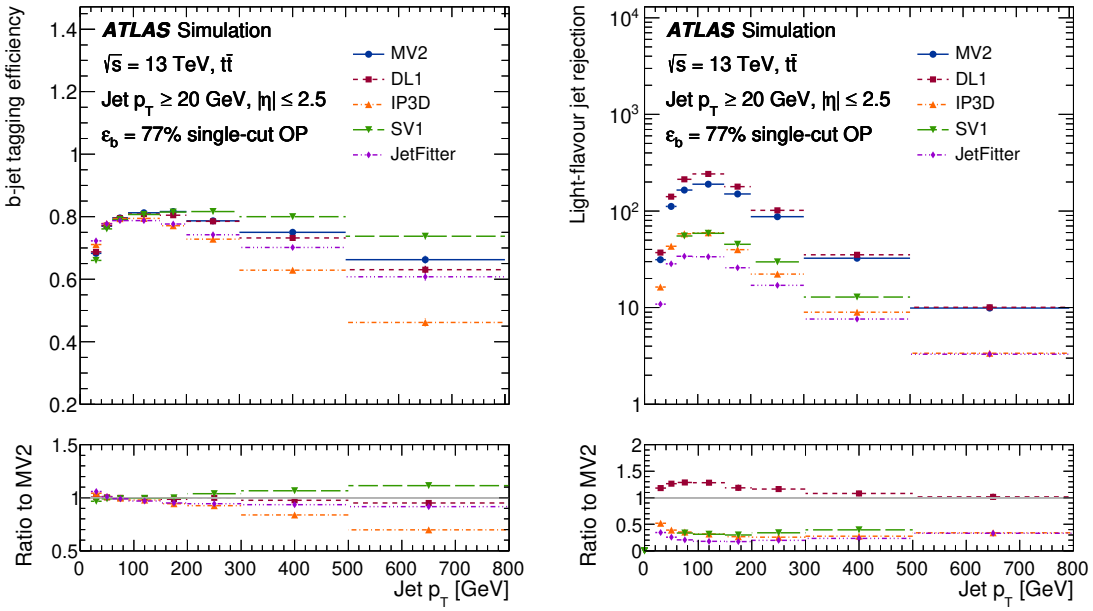


Figure 3.10: The b-jet tagging efficiency (a) and light-flavour jet rejection (b) as a function of a jet's p_{T}, shown for two *high-level* (`MV2` and `DL1`) and three *low-level* (`IP3D`, `SV1`, and `JetFitter`) taggers. The lower panel shows the ratio of each curve to the `MV2` for comparison. Efficiency and mistag rates were computed using a sample of simulated $t\bar{t}$ events [82]

3.4 Muons

Muons are reconstructed using information from both the Muon Spectrometer (MS) and ID. Muon candidates are independently reconstructed from hits in the ID and

MS, and matched to form a final muon candidate. Muon momentum is calculated from the curvature of the muon's track under ATLAS's magnetic field [83].

Muons are used in this study only to correct the large-R jets for potential energy losses from semi-leptonic b-decays. If a muon is spatially matched to one of the track jets associated with a large-R jet, and if this track jet is b-tagged, the muon's four-vector is added to the large-R jet's, and the mass of the latter is recalculated.

4
Analysis Strategy

Contents

4.1 Analysis Overview 45
4.2 Data and Simulated Samples 46
4.2.1 Monte Carlo Samples 47
4.3 Event and Object Selection 47
4.3.1 Kinematic Selections 48
4.3.2 B-tag Region Selection 51
4.3.3 Higgs-Mass Region Selection 52
4.3.4 Vetoed Events 56
4.4 Background Estimation 57
4.4.1 Background Modelling in the Control Region 58
4.4.2 Validation of Background Estimation 59
4.5 Estimation of Systematic Uncertainties 65
4.5.1 Detector & Reconstruction Systematics 66
4.5.2 Background Prediction Systematics 68
4.5.3 Summary of Systematic Uncertainties 74

This chapter details the strategy of the data analysis performed to search for boosted $hh \to 4b$ events using the ATLAS run 2 dataset. Section 4.1 gives a brief rundown of the analysis as an introduction. The next few sections (4.2-4.4) provide a thorough overview of the analysis, Chapter 5 presents the final results of this search, and chapter 6 gives a more extensive description of specific contributions in areas relevant to the $hh \to 4b$ analysis performed as part of the work of this thesis.

4.1 Analysis Overview

A pair of Higgs bosons at rest decaying to four b-quarks is expected to produce four clearly separated jets. As the transverse momentum p_{T} of the Higgs boson increases, its decay products collimate closer together. Due to this effect Higgs bosons with high p_{T} (boosted) and those with low p_{T} have a different signature, and can be studied separately (as discussed in section 1.4.1).

This analysis focuses on the search for pairs of boosted Higgs bosons decaying to 4 b-quarks. A schematic of the signal events of this search is shown in figure 4.1, with two large-radius jets which each have two track-jets associated to them. These track-jets have been tagged to indicate whether they are likely the product of a b-decay or not, and can be used to separate signal events from backgrounds which are less likely to contain b-jets. Additional signal regions with fewer b-tagged jets (which are not depicted in the graphic in figure 4.1, but later on in figure 4.5) are also defined for this analysis in order to accept events with highly boosted Higgs bosons where the b-jets are too closely collimated to be reconstructed as two separate jets.

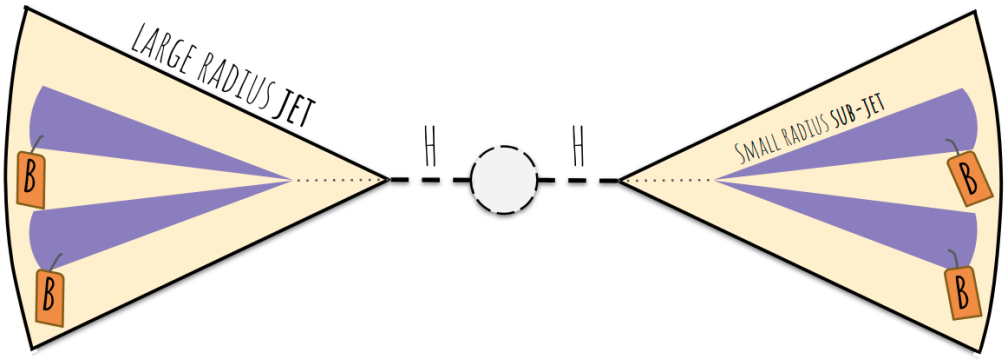
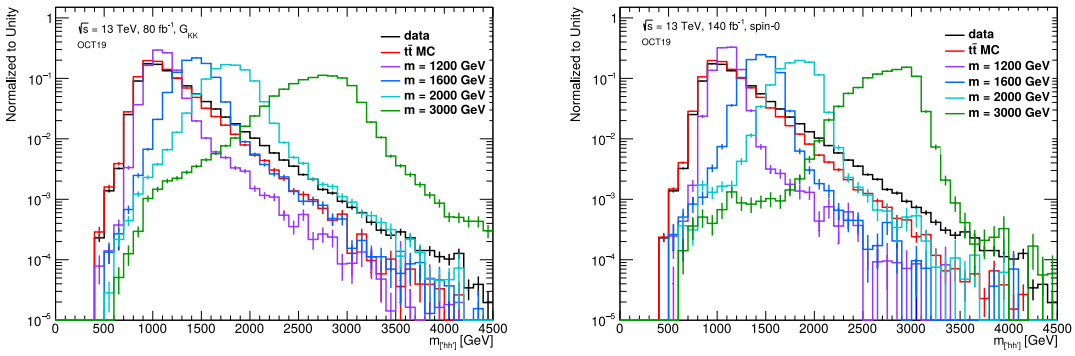


Figure 4.1: Diagram of the boosted $hh \rightarrow 4b$ final state. The two large-R jets from the decay of the Higgs bosons are represented by large yellow cones and the smaller purple cones within these represent the track jets associated to them. These track jets can be identified, or *tagged*, as containing a b-quark.

The largest background processes with a similar final state are QCD multijet and $t\bar{t}$, with QCD accounting for up to 95% of the background (these backgrounds, and the method to estimate them, are discussed in section 4.4). Various kinematic selections are applied to the data to separate signal from background (section

4.3), and the remaining background events are estimated using a mostly data-driven technique: the shape of the distribution of the highest-transverse-momentum jet's mass is obtained for both background types, and the two are independently scaled up or down until their addition best agrees to the experimental data in an auxiliary signal-free region.

The di-Higgs system's invariant mass m_{hh} is chosen as the final discriminant for this search. This distribution, shown in 4.2, captures the mass of the various resonances which would present a peak structure rather than the smoothly falling spectrum expected from SM background processes. Histograms of this observable in a signal region defined in the plane outlined by the masses of the two large-R jets are used to perform the statistical analysis detailed in section 5.1 to set upper limits of the production cross section of heavy scalar and spin-2 resonances, and test the agreement of the data with the background-only hypothesis.



(a) **(b)**

Figure 4.2: Distributions of the di-Higgs invariant mass m_{hh}, after the trigger requirement is applied, of resonant spin-2 G^*_{kk} (a) and heavy scalar (b) signal models. Data in this early stage of the selection process are shown as a rough estimate of the QCD background.

4.2 Data and Simulated Samples

The boosted $hh \rightarrow 4b$ search detailed in this work uses $139\,\text{fb}^{-1}$ of data at $\sqrt{s} = 13$ TeV collected by ATLAS, corresponding to the full run 2 dataset. The background estimation is mostly data driven, only using Monte-Carlo (MC)

simulation in a minor way. MC simulation is used to test the heavy scalar S and RS graviton G^*_{kk} signal benchmarks.

4.2.1 Monte Carlo Samples

A full simulation of the ATLAS detector using the `GEANT4` toolkit [84] to simulate interactions of particles with the detectors, was used for all the simulated samples. A $t\bar{t}$ MC sample is used to model the shape of this background to the $hh \to 4b$ search (see chapter 4 for more detail). This process was simulated at next to leading order (NLO) in α_s with the POWHEG method [85], using Pythia 8 [86] for hadronization and showering, with EvtGen for heavy flavour decays [87]. The nominal PDF is NNPDF 2.3 leading order (LO) [88] with the A14 tune set of underlying event parameters [89].

The heavy scalar process $S \to hh \to b\bar{b}b\bar{b}$ was simulated at LO in α_s using MadGraph [90] [1]. Hadronization and parton showering were done with Herwig 7 [91] with EvtGen, and the nominal PDF is NNPDF 2.3 LO. Masses from 900 GeV up to 5 TeV were analyzed for this boosted analysis[2]. The heavy scalar is assumed to have a width smaller than the detector resolution and no other new particles are taken into account in the simulation.

The spin-2 RS graviton resonance $G^*_{kk} \to hh \to b\bar{b}b\bar{b}$ was also simulated at LO in α_s using MadGraph, with hadronization and parton showering done using Pythia 8, with EvtGen for heavy flavour decays. The nominal PDF is NNPDF 2.3 LO. Resonance masses ranging from 800 GeV to 3 TeV are considered.

4.3 Event and Object Selection

Events for the boosted $hh \to 4b$ analysis are first filtered via a set of kinematic selections, described below. Next, events are then assigned to various *regions* depending on the large-R jet mass or number of b-tagged track-jets. Finally, certain events are vetoed from the analysis to allow the boosted and resolved channels of

[1] See chapter 1 for a brief description of the benchmark signal models.
[2] Lighter masses are covered by a resolved counterpart to this analysis.

the $hh \rightarrow 4b$ search to use fully independent sets of events (allowing their results to be statistically combined without double-counting), and to ensure b-tagging algorithms behave correctly in all events.

4.3.1 Kinematic Selections

Figure 4.3 shows an overview of how data are filtered, starting with the kinematic selections described in this section. Events are then divided into the various b-tag and Higgs-mass regions used in the analysis yielding the final subsets of data represented by the rightmost elements in the graph.

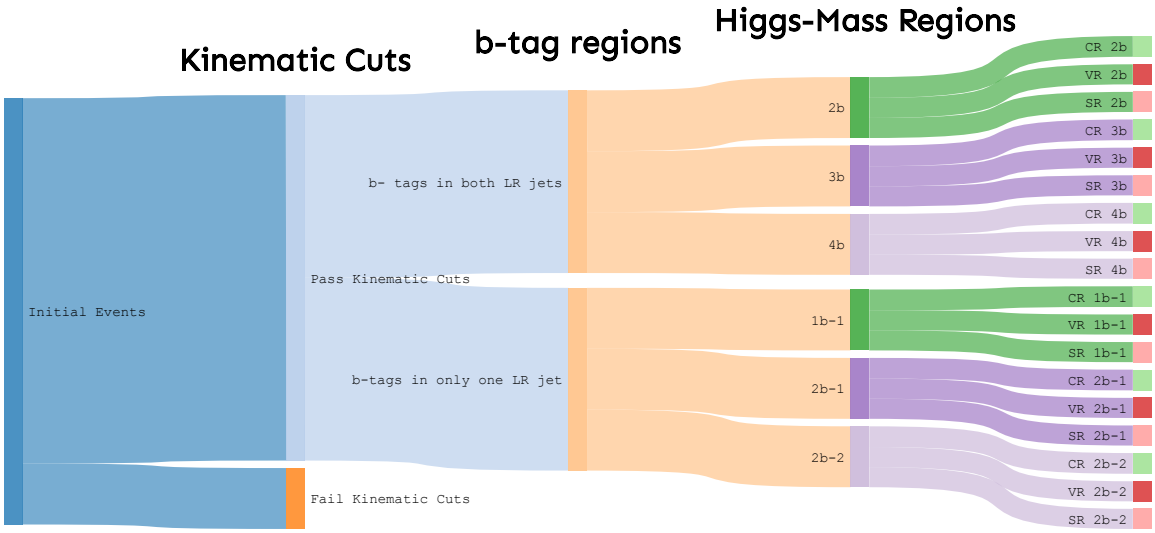


Figure 4.3: Diagram depicting the flow of data through the analysis selection. Size does not correspond to scale.

Trigger

To begin the selection of events with a boosted $hh \rightarrow 4b$ final state (figure 4.1), the *lowest unprescaled*[3] single large-R jet trigger was used. This trigger was slightly different for each year's data due to the different luminosity and pileup conditions in each year. Table 4.1 summarizes the trigger names[4] and large-R jet requirements

[3] Certain interesting physics events can have an unacceptably high trigger rate to be all recorded. In these cases only a fraction of events passing a trigger are recorded, which is referred to as a *pre-scaled* trigger. An *unprescaled* trigger then refers to a filter for which all events that pass are recorded.

[4] Trigger names contain the HLT trigger threshold and object calibrations as well as the L1 trigger that seeded it. See section 2.2.2 for more details.

Table 4.1: Large-R jet trigger names and thresholds of p_{T} and mass to pass the trigger chosen for each years' data set.

Year	Trigger Name	Large-R Jet p_{T} [GeV]	Large-R Jet Mass [GeV]
2015	`HLT_j360_a10_lcw_sub_L1J100`	> 360	—
2016	`HLT_j420_a10_lcw_L1J100`	> 420	—
2017	`HLT_j420_a10t_lcw_jes_40smcINF_L1J100`	> 420	> 40
2018	`HLT_j420_a10t_lcw_jes_35smcINF_L1J100`	> 420	> 35

to pass the trigger per year. During 2015 and 2016 triggers were designed to find events with large-R jet p_{T} of at least 360 GeV and 420 GeV respectively[5], as these p_{T} thresholds were the lowest that allowed to keep the rate within the available bandwidth. In order to keep the p_{T} threshold at 420 GeV during the increased luminosity conditions in 2017 and 2018, a large-R jet mass requirement was added to this trigger at 40 GeV and 35 GeV respectively. The effect on acceptance of events due to the addition of the large-R jet mass requirement on the 2017 and 2018 triggers was studied as part of the work for this thesis finding no impact (this study is described separately in section 2.2.2.2).

Large-R Jet Kinematic Selection

Events passing the trigger are then required to have at least two reconstructed large-R jets. If the event has more than two, only the two jets with highest p_{T} are considered. The *leading* large-R jet in p_{T} is required to have $p_{\mathrm{T}} > 450$ GeV to ensure the trigger is fully efficient for all events examined. The p_{T} threshold for the *subleading* large-R jet is set to 250 GeV, as a Higgs-jet with this momentum (or higher) is expected to be contained, on average, within a $R = 1$ jet [92] as per the rule of thumb:

$$R \approx \frac{2m}{p_{\mathrm{T}}}. \tag{4.1}$$

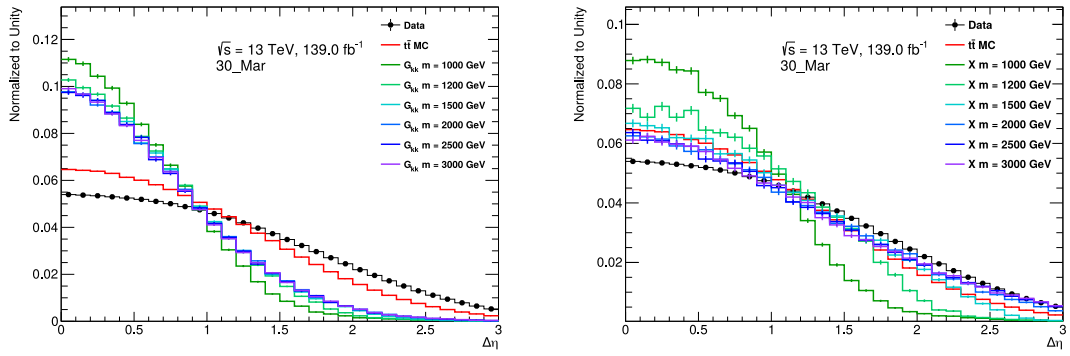
Large-R jets are also required to have $|\eta| < 2$ to ensure tracker acceptance, and mass > 50 GeV to reject jets from QCD multijet background.

Next, events are sorted using information from track jets that have been ghost-associated to each large-R jet (see section 3.2.2). Only events where both the

[5] These p_{T} values refer to *HLT-trigger-level* reconstruction.

leading and the subleading large-R jets have at least one track jet associated are kept[6]. Large-R jets that pass the requirements up to this stage are referred to as *Higgs boson candidates*.

To further reduce the QCD multijet background, a requirement is placed on the pseudorapidity separation between the Higgs boson candidates of $|\Delta\eta| < 1.3$. Heavy resonances are produced via s-channel processes, while background processes can also be produced via t or u-channel. Jets from a heavy resonance are then likely to have a small $|\Delta\eta|$, while background processes' jets, such as those in QCD multi-jet events, will have a more uniform distribution in $|\Delta\eta|$. The separation in $|\Delta\eta|$ between signal and background can be seen in figure 4.4, where the observable is plotted for various resonance masses, as well as the MC $t\bar{t}$ background sample. In these two plots events are only filtered by the trigger, and data at this stage are shown as an estimate of the QCD background shape. The optimal value for the $|\Delta\eta|$ requirement was found to be different for the two signal benchmark models (with the G^*_{kk} signal having clearer separation for all mass points), and the chosen value of $|\Delta\eta| < 1.3$ was close to optimal for both searches. Events failing any of the above requirements are removed from the dataset.



(a) **(b)**

Figure 4.4: Distributions of the difference in η between the two leading Higgs boson candidates, after the trigger requirement is applied. Different signal mass resonances are shown for the spin-2 G^*_{kk} (a) and heavy scalar (b) signal models. Data in this early stage of the selection process are shown as a rough estimate of the QCD background.

[6] While the signal topology requires at least one b-tagged track jet per large-R jet, events with large-R jets containing *untagged* track jets are used to model the background. See section 4.4.

Track Jet Kinematic Selection

In order to be considered for this analysis, track jets must have $|\eta| < 2.5$ which assures they are within the coverage of the tracker. Jets must also be formed of at least two tracks and have $p_{\mathrm{T}} > 5$ GeV (these conditions assure that the jets can be passed through the b-tagging algorithm). Track jets failing these requirements are not considered for the analysis. The remaining ones are ghost-associated to large-R jets, and only up to two jets per large-R jet are considered for b-tagging (the two with highest p_{T}).

4.3.2 B-tag Region Selection

After the kinematic selection is performed, several subsets (or *regions*) of data are defined. Large-R jets originating from $h \to b\bar{b}$ are expected to have two associated energetic track jets, corresponding to each b-hadron (b-jets). Identifying these b-jets can be a powerful discriminant for the $hh \to 4b$ signal.

Events are separated into three categories, or *regions*, depending on the b-tagging state of the track jets associated to the Higgs boson candidates. In the $4b$ region, each large-R jet is required to contain exactly two b-tagged track jets. Due to the relatively low b-tagging efficiency (exponentiated to the fourth power by requesting four b-jets) in the boosted regime, as well as the fact that the two b-quarks from a Higgs boson decay can be closely collimated, most of the signal events do not actually have two b-tagged track jets associated with each large-R jet and thus fall outside the $4b$ region. In order to recover some of these signal events, the $3b$ and $2b$ regions are also defined. The $3b$ region requires that one large-R jet be associated with exactly two b-tagged track jets, and the other large-R jet be associated with exactly one b-tagged track jet. Events with each large-R jet associated with exactly one b-tagged track jet are assigned to the $2b$ region.

The b-tag categories described above are defined to accept signal events and thus require at least one b-tagged track jet per Higgs boson candidate. These categories are referred to as *high-tag* regions. Events with similar topology to each high-tag region, but which contain b-tags in only one of the large-R jets, are used

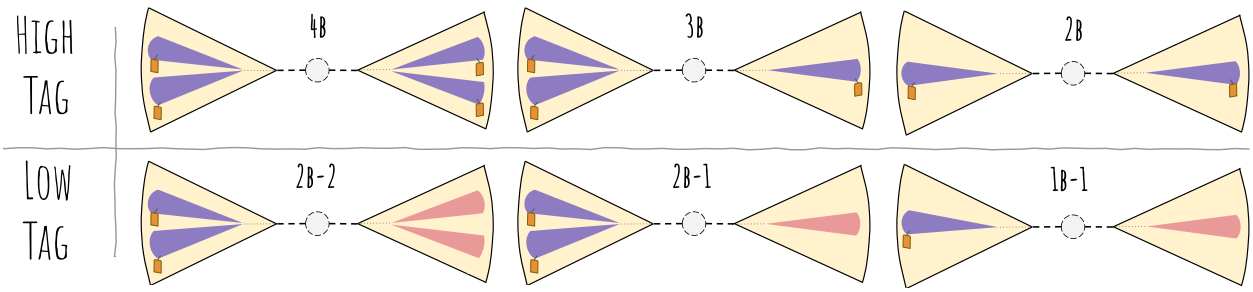


Figure 4.5: Diagram of the three high-tag topologies (4b, 3b, and 2b) with the corresponding low-tag region used to estimate QCD background (2b-2, 2b-1, and 1b-1) in the lower half of the table.

to estimate the QCD multijet background. These three categories are collectively named *low-tag* regions, and they are used only to estimate the contribution of QCD multijet processes to the background. A diagram depicting the three high-tag (and their corresponding low-tag) regions is shown in figure 4.5. The QCD multijet background in the 4b region is estimated using events that have two b-tagged track jets in one large-R jet and two untagged track jets in the other large-R jet (the 2b-2 region). Similarly, the QCD multijet background for the 3b(2b) region is estimated using events with two tagged(one) track jet(s) in one large-R jet and one untagged track jet in the other large-R jet named 2b-1(1b-1) region.

4.3.3 Higgs-Mass Region Selection

All b-tag regions are further divided into a *Control Region* (CR), a *Validation Region* (VR[7]), and a *Signal Region* (SR) according to the mass of the Higgs boson candidates in the event. The distribution the di-Higgs invariant mass m_{hh} The CR and VR are defined in order to produce and validate the background estimation. Events in these regions are assumed to have similar kinematic properties to the signal region, but negligible amount of signal, which makes them good candidates to model the background expected in the signal region. The key assumption of the background estimation strategy is that the proportion of the two components

[7]The acronym VR is unfortunately duplicated: it is used for the validation region in this analysis and for variable radius track-jets within ATLAS. In this thesis I have used different fonts for each acronym, VR for validation region and VR for variable radius, in an attempt to avoid further confusion.

of the background ($t\bar{t}$ and QCD multijet) is roughly constant throughout the m_{h_1}– m_{h_2} plane. This proportion is derived in the CR, tested in the VR, and finally used to estimate the background in the SR (see section 4.4 for details on the background estimation process).

Figure 4.6 shows a scatter plot of the large-R jet mass for each large R jet. The leading jet mass is associated with the first Higgs boson candidate (m_{h1}) while the sub-leading jet mass is associated with the second Higgs boson candidate (m_{h2}). It is in this plane that the control, validation and signal regions are defined.

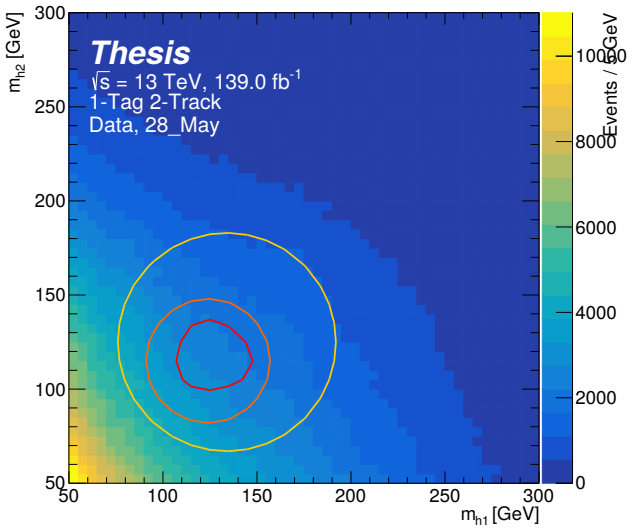


Figure 4.6: Definitions of the Higgs boson candidate mass regions on the plane defined by the leading and subleading Higgs boson candidates masses. This plot shows data from the 1b-1 region. The regions are defined in the body of the text.

The regions are defined by the following boundaries in the m_{h_1}– m_{h_2} plane:

$$X_{hh} \equiv \sqrt{\left(\frac{m_{h1} - 124\,\text{GeV}}{0.1 m_{h1}}\right)^2 + \left(\frac{m_{h2} - 115\,\text{GeV}}{0.1 m_{h2}}\right)^2} \tag{4.2a}$$

$$R_{hh}^{VR} \equiv \sqrt{(m_{h1} - 124\,\text{GeV})^2 + (m_{h2} - 115\,\text{GeV})^2} \tag{4.2b}$$

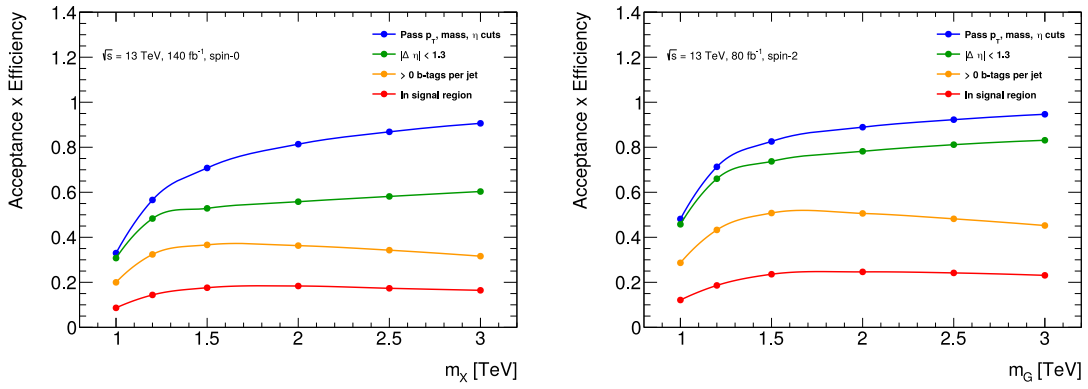
$$R_{hh}^{CR} \equiv \sqrt{(m_{h1} - 134\,\text{GeV})^2 + (m_{h2} - 125\,\text{GeV})^2}. \tag{4.2c}$$

The SR is inside the red X_{hh} boundary, the VR is the ring between R_{hh}^{VR} in orange and the X_{hh} boundary, and the CR is the outer ring between the R_{hh}^{CR} in yellow and R_{hh}^{VR}. The VR is defined by $R_{hh}^{VR} < 33$ GeV, and the CR by $R_{hh}^{CR} <$

58 GeV. The centre of the CR contour is moved to higher masses than the SR and VR to avoid the low-mass peak in the background distribution, while also encompassing a number of events sufficient to estimate and validate the background model in the corresponding regions.

The X_{hh} variable is designed to represent the distance of an event from the di-Higgs peak in the m_{h_1}– m_{h_2} plane, and the signal region is chosen to have X_{hh} < 1.6. This value was chosen since it contains $> 90\%$ of the di-Higgs peak in G^*_{kk} simulated events.The centre of this region is chosen below the Higgs boson mass to account for jet constituents falling outside the jet cone (specially relevant for the subleading Higgs boson candidate), neutrinos from leptonic b decays, and inactive areas of the detector. The denominator in X_{hh} represents the resolution of the reconstructed Higgs boson candidates masses.

Figure 4.7 shows the efficiency of the analysis selection on the two signal models. The plots show the union of all b-tag regions (as the selection requires at least one b-tagged track jet in each large-R jet). The largest difference in the selection efficiency between the models come from filters that depend on angular distributions, which vary between the spin-0 and spin-2 samples, such as the $|\Delta\eta|$ or b-tagging selections.



(a) **(b)**

Figure 4.7: Efficiency of each sequential selection in the analysis for a range of masses of (a) scalar and (b) G^*_{kk} resonances.

Selection Stage	Requirement			
Kinematic Selection				
N(large-R jets)	≥ 2			
Large-R jet mass	$m_{\mathrm{j}_1}, m_{\mathrm{j}_2} \geq 50$ GeV			
Large-R jet p_{T}	$p_{\mathrm{T}}^{\mathrm{j}_1} \geq 450$ GeV			
	$p_{\mathrm{T}}^{\mathrm{j}_2} \geq 250$ GeV			
Large-R jet $\lvert\eta\rvert$	≤ 2			
N(track jets in large-R jets)	≥ 1 each			
$\Delta\lvert\eta\rvert$ between large-R jets	$\Delta\eta_{\mathrm{j}_1\mathrm{j}_2} < 1.3$			
b-tag Region Selection	b-tag Status			
	$\mathrm{j}_1^{\text{tagged}}$	$\mathrm{j}_2^{\text{tagged}}$	$\mathrm{j}_1^{\text{no-tag}}$	$\mathrm{j}_2^{\text{no-tag}}$
4b	2	2	-	-
3b	1	2	-	-
	2	1	-	-
2b	1	1	-	-
2b-2	2	0	0	2
	0	2	2	0
2b-1	2	0	0	1
	0	2	1	0
1b-1	1	0	0	1
	0	1	1	0
Higgs-Mass Region Selection				
Control Region	$R_{hh}^{CR} < 58$ GeV			
Validation Region	$R_{hh}^{VR} < 33$ GeV			
Signal Region	$X_{hh} < 1.6$			

Table 4.2: Overview of event selection for the boosted $hh \to 4b$ analysis. j_1 and j_2 refer to the leading and subleading large-R jets respectively (ordered in p_{T}). The b-tag regions above the double rule are the high-tag regions, while the ones below are the low-tag ones. Columns labeled $\mathrm{j}^{\text{tagged}}$ and $\mathrm{j}^{\text{no-tag}}$ refer to the number of b-tagged and *un*tagged track jets associated to a large-R jet, respectively (high-tag regions have no explicit requirements on the number of untagged jets). The definitions of Higgs-Mass regions are made exclusive (i.e. the control region does not include events in the validation region, which in turn excludes all events in the signal region).

4.3.4 Vetoed Events

Table 4.2 summarizes the event selection criteria of this analysis. After this selection, events go through one last filter to veto events in the resolved counterpart of this analysis, and events with certain track-jet configurations. The event yield after each step in the selection process of part of the data, and one mass point of each signal benchmark, can be found in appendix B.

4.3.4.1 Resolved Signal Region Veto

In order to ensure that the resolved and boosted channels (see section 1.4.1) are mutually exclusive, events used in the signal, control or validation regions of one analysis are vetoed in the other one. Events are vetoed in the following order:

- In the boosted analysis, any event passing the resolved signal region selection (including requiring 4 b-tagged resolved jets[8]), is vetoed from being used anywhere in the boosted analysis.

- In the resolved analysis, events in the control or validation regions which pass the signal, control or validation region requirements of the boosted analysis (including 2, 3 or 4 b-tagged variable radius track jets), are vetoed from the resolved analysis.

Resolved signal region events are thus given the highest priority, and no resolved signal events are removed from the resolved analysis, and boosted signal, validation, and control region events take priority over non-signal resolved events. This decision was made on the basis that the resolved analysis has more sensitivity in its high mass region (low mass region of the boosted analysis), and also has substantially more events in its control and validation regions than boosted, whose equivalent regions are statistically limited.

[8] Resolved jets are $R = 0.4$ topocluster anti-k_t jets corrected with information from tracks matched spatially to calorimeter clusters.

4.3.4.2 Collinear Track Jet Veto

While VR track jets have the benefit of adjusting their size as their p_{T} increases, this also makes it possible for one high-p_{T} jet to be fully contained within the cone of a jet with lower transverse momentum. Some of these cases can cause tracks to be assigned to the wrong track jet (see figure 4.8), which could produce unpredictable results if attempting to b-tag these jets. This unpredictable behaviour can happen because key factors for b-tagging, such as track impact parameter or finding of a secondary vertex, heavily rely on having the correct tracks. Events are then vetoed if a high p_{T} track jet contains the axis of a lower p_{T} one within its cone. A detailed discussion of this effect, the veto to avoid it, and the effect of the veto on the analysis has its dedicated section 6.1.2.

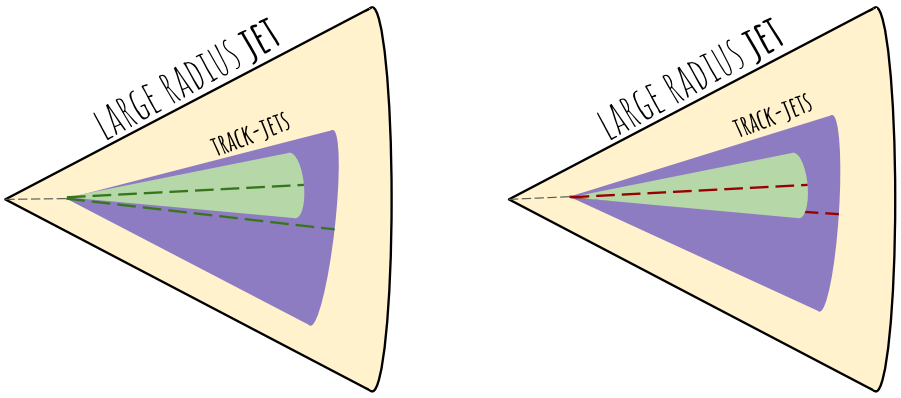


Figure 4.8: Depictions of an acceptable (left) and a problematic (right) configuration of overlapping variable-radius track jets. In the right side case, the softer (purple) jet's axis falls within the hard jet (green), which can lead to a mismatch of tracks to jets when b-tagging.

4.4 Background Estimation

The main background comes from multijet events from QCD interactions, and it accounts for 80 to 95% of background events, depending on the b-tag region. A second

smaller source of background is from $t\bar{t}$ events decaying hadronically, which accounts for virtually all the remaining background. Contributions from other processes such as single Higgs boson production in association with a top quark pair, $ZZ \to b\bar{b}b\bar{b}$ or $Z+$ jets were found to be below the percent level in studies by the analysis team and in previous work [31], and thus are not accounted for in this analysis.

4.4.1 Background Modelling in the Control Region

The $t\bar{t}$ MC simulated data are passed through the same event selection and divided into the same regions as the data to estimate the $t\bar{t}$ contribution in the CR. The contribution of QCD multijet background is estimated using a data-driven method. As mentioned in section 4.3.2, low-tag (1b-1, 2b-1, and 2b-2 regions) data are used to model the shape of QCD background in the corresponding high-tag region (2b, 3b, and 4b). The low-tag regions in data are used only to estimate the shape of the QCD multijet background. The high-tag $t\bar{t}$ contribution (taken from the simulated data) is subtracted from the low-tag data (after normalizing to account for the yields in low- and high-tag data) to capture the shape of QCD multijet only.

In the control region, for each b-tag region, the leading Higgs boson candidate mass distribution shapes from these two background samples are fit simultaneously until the best match to the corresponding high-tag distribution is found [9]. The fit independently varies the normalization of each background (labelled μ_{QCD} and $\alpha_{t\bar{t}}$ for the QCD multijet and $t\bar{t}$ backgrounds respectively) to find the overall normalization factors that maximize the likelihood of the estimated background given the high-tag data.

For each bin, the estimated number of background events $N_{\text{estimated}}^{\text{high-tag}}$ is

$$N_{\text{estimated}}^{\text{high-tag}} = \mu_{QCD}(N_{\text{data}}^{low-tag} - N_{t\bar{t}}^{low-tag}) + \alpha_{t\bar{t}} N_{t\bar{t}}^{\text{high-tag}} \tag{4.3}$$

The distributions after this *binned maximum-likelihood fit* is completed are shown in figure 4.9, and the resulting low-to-high tag correction factors μ_{QCD} and $\alpha_{t\bar{t}}$ are shown in table 4.3 for each of the b-tag regions. The values of μ_{QCD} and $\alpha_{t\bar{t}}$ are the

[9] Other observables' distributions were tested, finding no significant difference in the normalization factors μ_{QCD} and $\alpha_{t\bar{t}}$ found using the leading Higgs boson candidate mass.

Table 4.3: Normalization factors μ_{QCD} and $\alpha_{t\bar{t}}$ derived in the CR. $\alpha_{t\bar{t}}$ in the $4b$ region was set to one.

b-tag region	μ_{QCD}	$\alpha_{t\bar{t}}$
$4b$	0.0055 ± 0.00029	1.00
$3b$	0.096 ± 0.0018	1.16 ± 0.058
$2b$	0.055 ± 0.00054	1.19 ± 0.015

factors by which the QCD multijet background estimate (obtained from low-tag data) and the $t\bar{t}$ background estimate (obtained from simulation) were scaled to best match the high-tag data in that region. From figure 4.9 it is noted that the low statistics of the $4b$ region can not give a meaningful constraint on the $t\bar{t}$ background, so $\alpha_{t\bar{t}}$ is set to one for this b-tag region. Due to low statistics of simulated $t\bar{t}$ events in the $4b$ region, the $t\bar{t}$ distribution in the $3b$ region is used instead.

Figures 4.10-4.12 show the di-Higgs system's invariant mass m_{hh} for the $2b$, $3b$, and $4b$ regions, which will be the final discriminant for the analysis (see statistical analysis section 5.1) obtained from background estimation and observed data in the control and validation regions.

To reduce the effect of statistical fluctuations at high m_{hh} in the boosted analysis, the multijet distribution is fitted to a *smoothing* functional form in a series of studies performed by the analysis group. The QCD background smoothing in the SR is shown in figure 4.13 for all b-tag regions [10].

4.4.2 Validation of Background Estimation

The background model derived in the CR is then validated in the VR. The assumption of this background estimation method is that the correction factors μ_{QCD} and $\alpha_{t\bar{t}}$ are approximately constant throughout the m_{h_1}– m_{h_2} plane. This assumption is validated by applying these same correction factors to the VR data and verifying the agreement is similar to the agreement seen in the CR. Figure 4.14 shows the same distribution used to normalize the backgrounds (figure 4.9), the

[10]The types of functions used are commonly referred to as *di-jet* functions as they have been used successfully to fit di-jet spectra in past analyses. See references [93] and [94] for details.

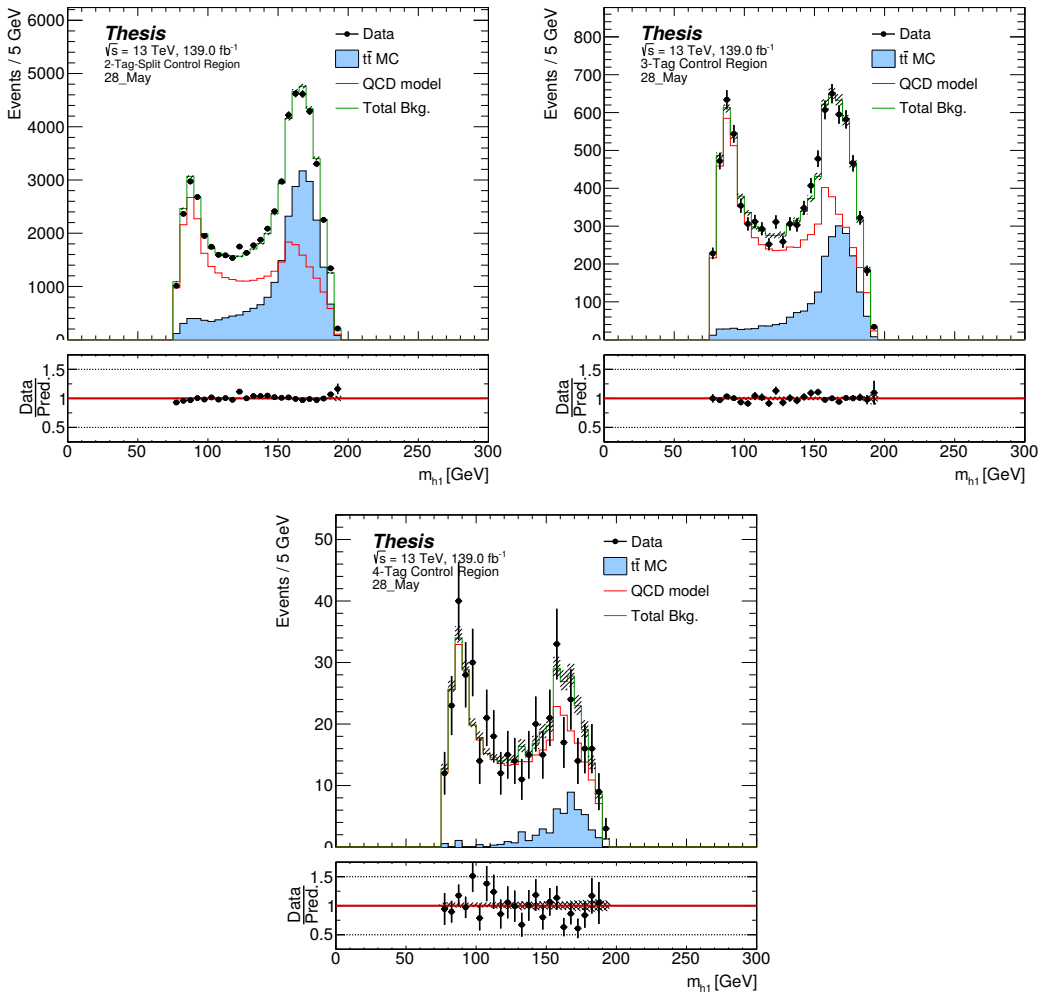
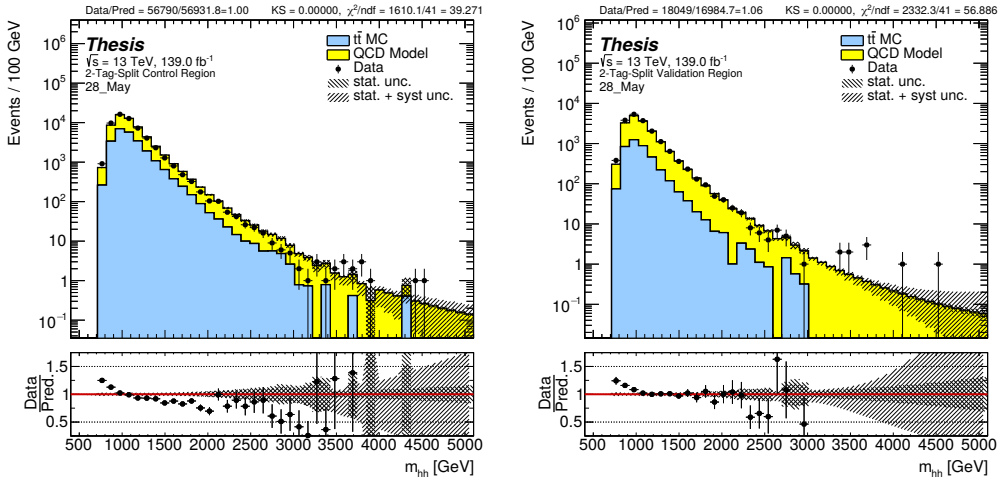


Figure 4.9: Leading Higgs boson candidate mass distribution from the background model and from the data used to normalize the QCD multijet and $t\bar{t}$ contributions in the 2, 3 and 4 b-tag analyses in the control region. The double-peak structure of the distributions is due to the removal of the validation and signal regions (see figure 4.6).

leading Higgs boson candidate's mass, in the validation region. Agreement between data and prediction is comparable between the control and validation regions.

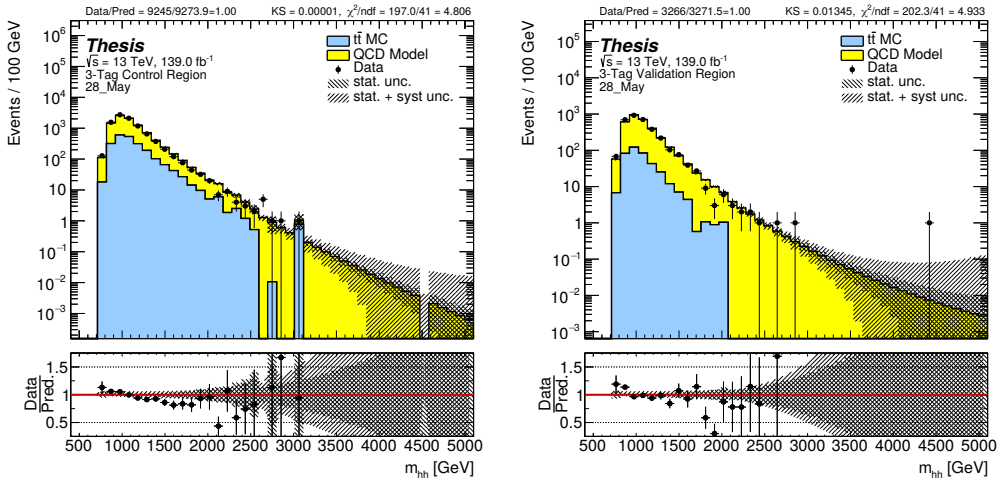
Closure Tests in Validation Region

Plots of this validation are shown in figures 4.10- 4.12 and show similar agreement of background estimation to data between the control and validation regions. Any statistically significant disagreements in this test as evidenced by the bottom panels, such as figure 4.10b, are included as an additional uncertainty on the background estimation (see section 4.5.2).



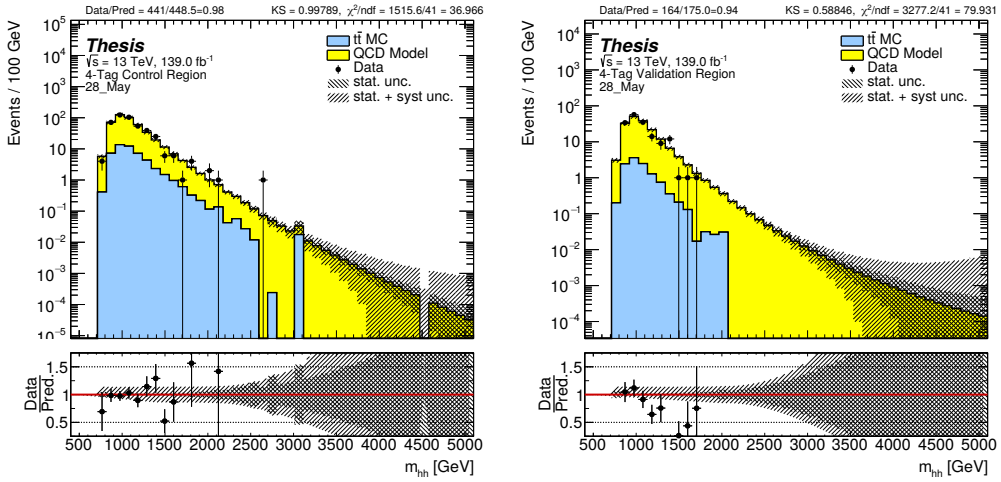
(a) **(b)**

Figure 4.10: Di-Higgs mass in the control (left) and validation (right) for the $2b$ region.



(a) **(b)**

Figure 4.11: Di-Higgs mass in the control (left) and validation (right) for the $3b$ region.



(a) (b)

Figure 4.12: Di-Higgs mass in the control (left) and validation (right) for the 4b region.

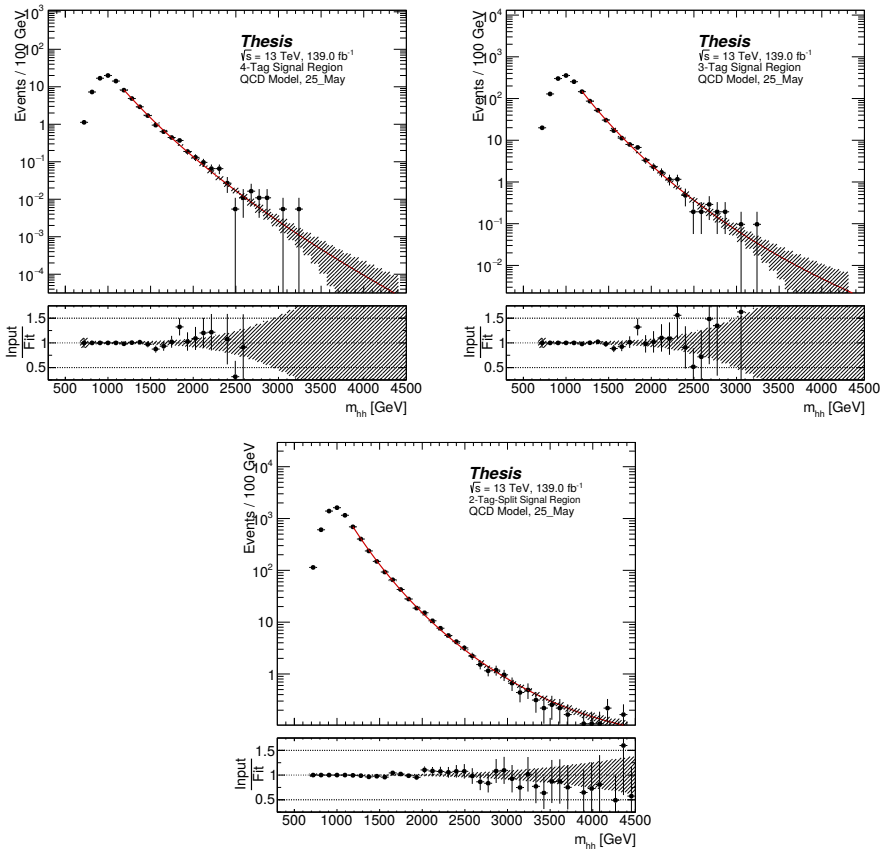


Figure 4.13: QCD background estimate in the signal region (dots, these are obtained from low-tag data in the SR) for the 4b (top left), 3b (top right), and 2b (bottom) regions with the corresponding smoothing fit (red line).

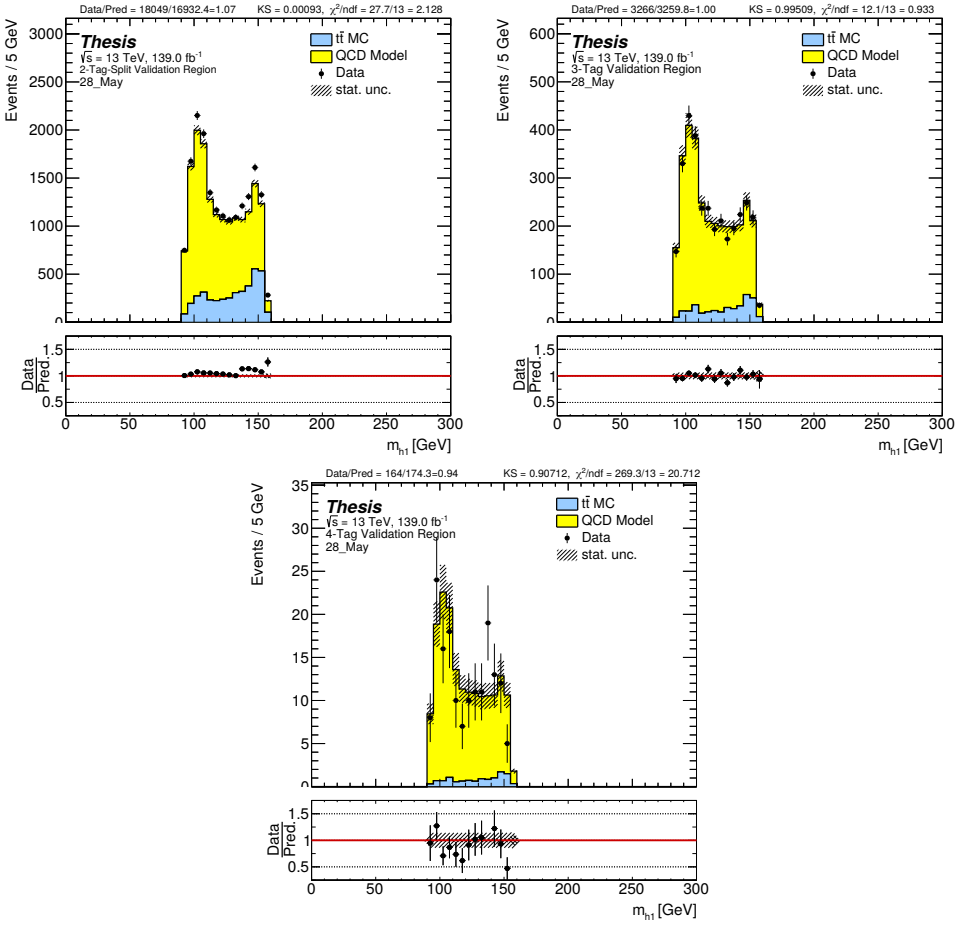


Figure 4.14: Leading Higgs boson candidate mass distribution from the background model and from the data to validate the QCD multijet and $t\bar{t}$ background estimates in the 2, 3 and 4 b-tag analyses in the validation region.

Background Estimate Rehearsal Using Simulated QCD Events

While obtaining enough Monte Carlo simulated events to accurately model the QCD multijet background in the $hh \rightarrow 4b$ analysis is not possible, such a sample can offer a valid cross-check of the background estimation method. A dataset of simulated di-jet events was used in place of experimental data for this validation. These datasets were passed through the same filters and selections. The same background estimation process was then re-run substituting the high-tag data with a combination of di-jet and $t\bar{t}$ events, allowing the background estimation in the signal region to be validated without unblinding the experimental data.

The comparison of the background estimation prediction to the signal region data is shown in figure 4.15, and shows that modelling of the SR QCD Monte Carlo data is adequate within the uncertainties for the $2b$ and $3b$ regions. Although the $2b$ plot shows some under-prediction by the model, this is expected to be well within the considered uncertainties. Note that the uncertainty bands in this figure are statistical only [11]. The statistical uncertainty in the $4b$ region is too wide to meaningfully test the modelling.

[11] The study by the analysis group of this cross check including systematic uncertainties is ongoing.

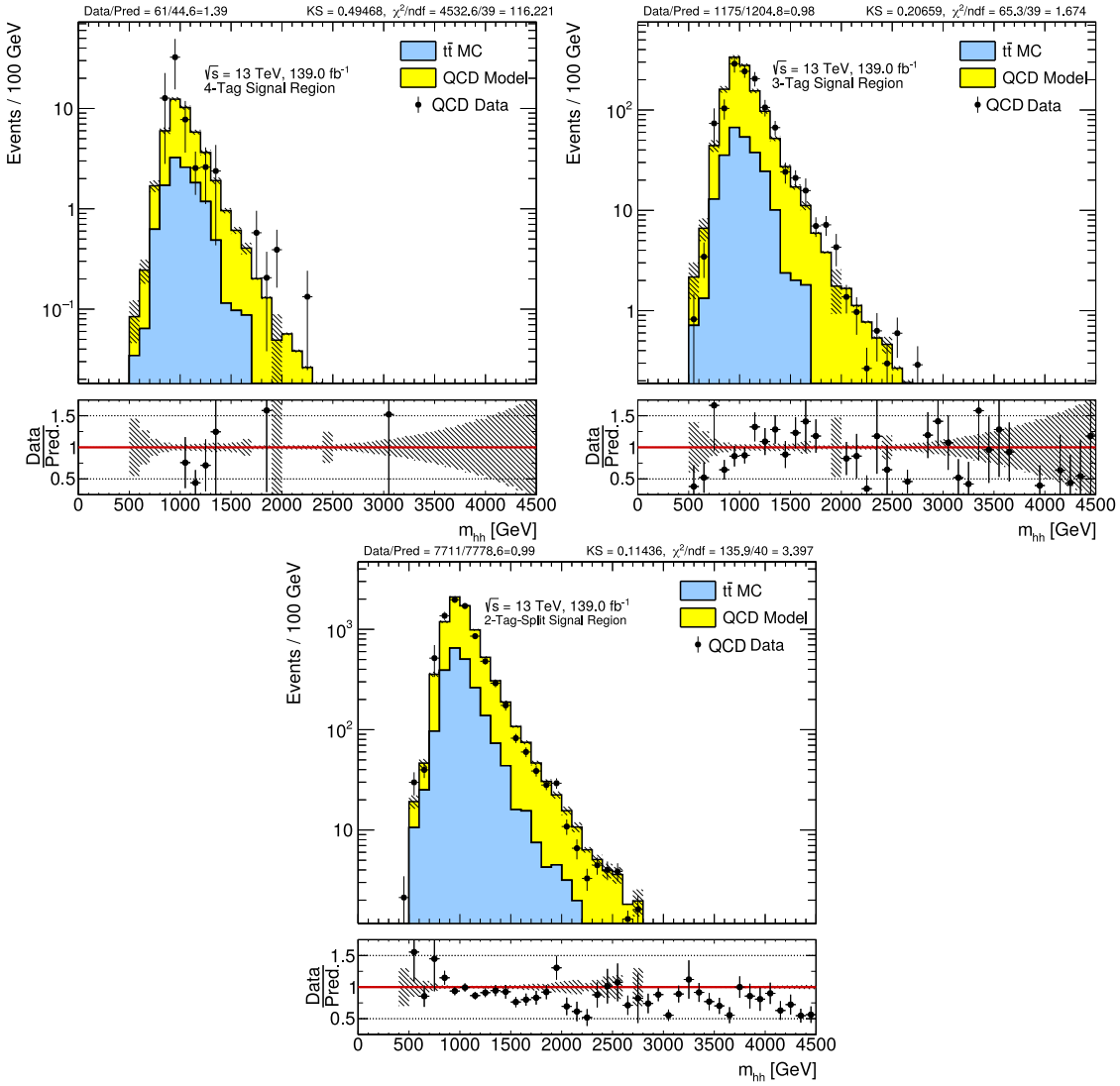


Figure 4.15: Cross check of the background estimate in the signal region for the 4b (a), 3b (b), and 2b (c) regions. Low-tag and high-tag data have been replaced by QCD MC. The modelling of the SR QCD agrees with the prediction within uncertainty.

4.5 Estimation of Systematic Uncertainties

Due to the low cross section of boosted di-Higgs, relatively few events are expected in this analysis. Therefore the statistical uncertainty on the data is dominant over the systematic uncertainties, as evidenced in figure 4.16. This plot shows the expected limit on the production cross section of a heavy scalar expected to be set by this search when accounting for statistical uncertainty only, and also for both statistical and systematic uncertainties [12]. The degradation of the limit by adding

[12] See section 5.1.4 for more detail.

systematic uncertainties falls within one standard deviation in the measurement accounting for all uncertainties. This highlights the degree to which statistical uncertainties dominate over systematic effects. It is also worth noting that this low-mass range, where significant impact is seen, is where the resolved counterpart to this analysis is expected to have equal or better sensitivity, and where the resolved event veto has its biggest impact (section 4.3.4).

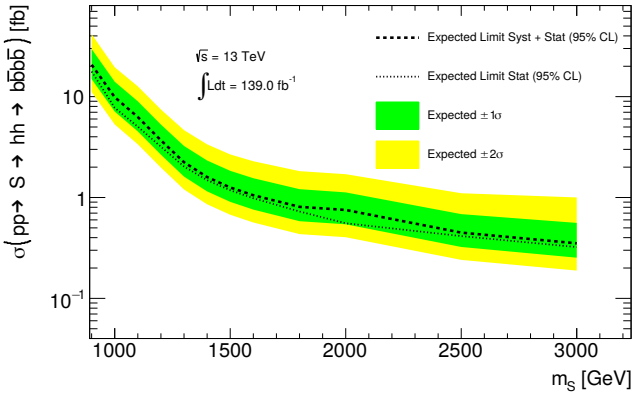


Figure 4.16: Expected 95% CL upper limits on the production cross section times branching ratio of the heavy scalar model accounting for only statistical uncertainty and both statistical and systematic uncertainties.

The most important systematic uncertainties considered for the analysis are mentioned in the following sections, with special emphasis on the uncertainties that have the most impact on the analysis sensitivity (jet-related and background-estimation-related systematics). The impact of systematics on the analysis is further discussed in section 5.1.4.1, and a brief summary of the impact of each group of systematics is presented in section 4.5.3.

4.5.1 Detector & Reconstruction Systematics

Systematic uncertainties arising from detector and reconstruction effects were propagated through the analysis using standard recommendations from dedicated groups within ATLAS. These are applied to all MC simulation samples.

Luminosity Uncertainty

The uncertainty on the measurement of the integrated luminosity is 1.7 % and is measured by the specialized detector LUCID-2 [95].

Large-R Jet Uncertainties

Throughout the multiple calibrations and corrections to jet energy and mass described in section 3.2.1, various sources of uncertainty are introduced. These arise due to factors such as the statistical uncertainty on the datasets used, event selection choices, assumptions on the jet flavour composition, choices made in the MC simulations, among others. These are evaluated individually, often by changing the choices made and re-calibrating jets, and setting differences in the jets' p_{T} and η as uncertainties.

Considering every source of uncertainty on the jet energy scale from each of the steps of calibration yields a large number of systematic uncertainties. For physics analyses, instead of using the extensive list of uncertainties, a decomposition of the full set is made that preserves the correlation between the uncertainties while reducing the number of necessary *nuisance parameters* [13] to a computationally-feasible dimension [96]. The jet energy resolution is determined separately from momentum-balance as explained in section 3.2.1. The uncertainties on the jet mass scale and resolution are calculated in a similar way [75]. An extensive account of jet uncertainty calculations can be found in references [73–75].

Alternative versions of all MC datasets are made by varying one uncertainty up or down at a time. This is done by applying a multiplicative factor dependent on the estimate of the uncertainty to each of the jets used in the analysis. This results in up/down variations of all signal and $t\bar{t}$ samples which will later be considered in the statistical analysis (section 5.1).

[13] See section 5.1.1 for definition.

Flavour Tagging Uncertainties

Flavour tagging efficiency measurements and calibrations are obtained from dedicated $t\bar{t}$ data samples with high purity, and $t\bar{t}$ MC (see section 3.3). These are then prone to uncertainties from MC modelling (MC statistics and flavour composition, for example) in the samples used, and the purity of the $t\bar{t}$ data samples.

Similarly to large-R jet uncertainties, these go through a decomposition to reduce the number of uncertainties while conserving the correlations [96]. Each uncertainty is then added as an event weight scale factor that is applied for each track jet inspected by the b-tagging algorithm [80–82]. The analysis is then repeated for each variation, yielding one set of histograms per variation that will later be part of the final statistical analysis (section 5.1).

4.5.2 Background Prediction Systematics

The uncertainty in the background estimation procedure is parametrized as two types of uncertainty: the normalization of the two main components of the background ($t\bar{t}$ and QCD multijet), and the uncertainty on the shape of the m_{hh} spectrum. The most impactful systematic uncertainties for this search proved to be ones arising from the background prediction, as described in section 5.1.4.1.

4.5.2.1 Background Normalization Uncertainties

Three elements in the background normalization introduce systematic uncertainty. These are the extrapolations from the CR to the SR and from the low to high tag regions, the limited statistics in the CR where normalization factors are derived, and the size and placement of the control region itself[14].

Low-tag to high-tag & CR to SR extrapolations

Normalization uncertainties are assigned due to the assumptions that the low-tag regions model the QCD multijet background [15], and that the composition

[14] Part of this effect is already covered by the extrapolation uncertainty but it is included as an extra precaution to ensure the effect is accounted for. The impact of the CR-related uncertainties is small when compared to the extrapolation uncertainty as discussed in section 4.5.3 below.

[15] There is also a *shape* uncertainty associated to this assumption, described in section 4.5.2.2 below.

of the background (as described by the normalization factors μ_{QCD} and $\alpha_{t\bar{t}}$) is uniform across the m_{h_1}– m_{h_2} plane.

These uncertainties are evaluated using a Gaussian process interpolation [97, 98] across the blinded signal region in the m_{h_1}– m_{h_2} plane, to obtain an independent estimate of the background in the SR. The m_{h_1}– m_{h_2} plane for each of the b-tag regions are shown in figure 4.17 next to their corresponding interpolations. This interpolation assumes smoothness in the background distribution, represented using low-tag data in this case. The $t\bar{t}$ shape from MC is is subtracted from the low-tag data to capture only the QCD multijet shape before the interpolation.

The yield estimate in the signal region from this interpolation is added to the expected $t\bar{t}$ yield from MC to give an expected yield of background events in the signal region $N_{\mathrm{SR_{exp}}}^{\text{high-tag}}$, which is then used to calculate the double ratio:

$$R_{\text{extr}} \equiv \frac{N_{\mathrm{SR_{exp}}}^{\text{high-tag}}/N_{\mathrm{SR}}^{\text{low-tag}}}{N_{\mathrm{CR}}^{\text{high-tag}}/N_{\mathrm{CR}}^{\text{low-tag}}}. \tag{4.4}$$

R_{extr} quantifies the consistency in the yields of the background estimates across both extrapolations. $N_{\mathrm{SR}}^{\text{low-tag}}$ and $N_{\mathrm{CR}}^{\text{low-tag}}$ are the yields obtained from similar Gaussian Process fits to the low-tag region, and $N_{\mathrm{CR}}^{\text{high-tag}}$ and $N_{\mathrm{SR_{exp}}}^{\text{high-tag}}$ are calculated from the interpolation shown in figure 4.17. The uncertainty is then estimated by $|R_{\text{extr}} - 1|$, with values of 2.34%, 5.79%, and 13.9% in the $2b$, $3b$, and $4b$ regions respectively.

Limited CR statistics

The limited statistics in the control region used to determine μ_{QCD} and $\alpha_{t\bar{t}}$ introduce a statistical uncertainty on the values of the normalization factors. The size of this uncertainty is evaluated as yield differences in the signal region from up/down variations of μ_{QCD} and $\alpha_{t\bar{t}}$ taken from the CR fit uncertainty These uncertainties are 0.73%, 1.63%, and 0.03% for the $2b$, $3b$, and $4b$ regions respectively[16] (see table 4.3).

[16] This is the fit shown in figure 4.9.

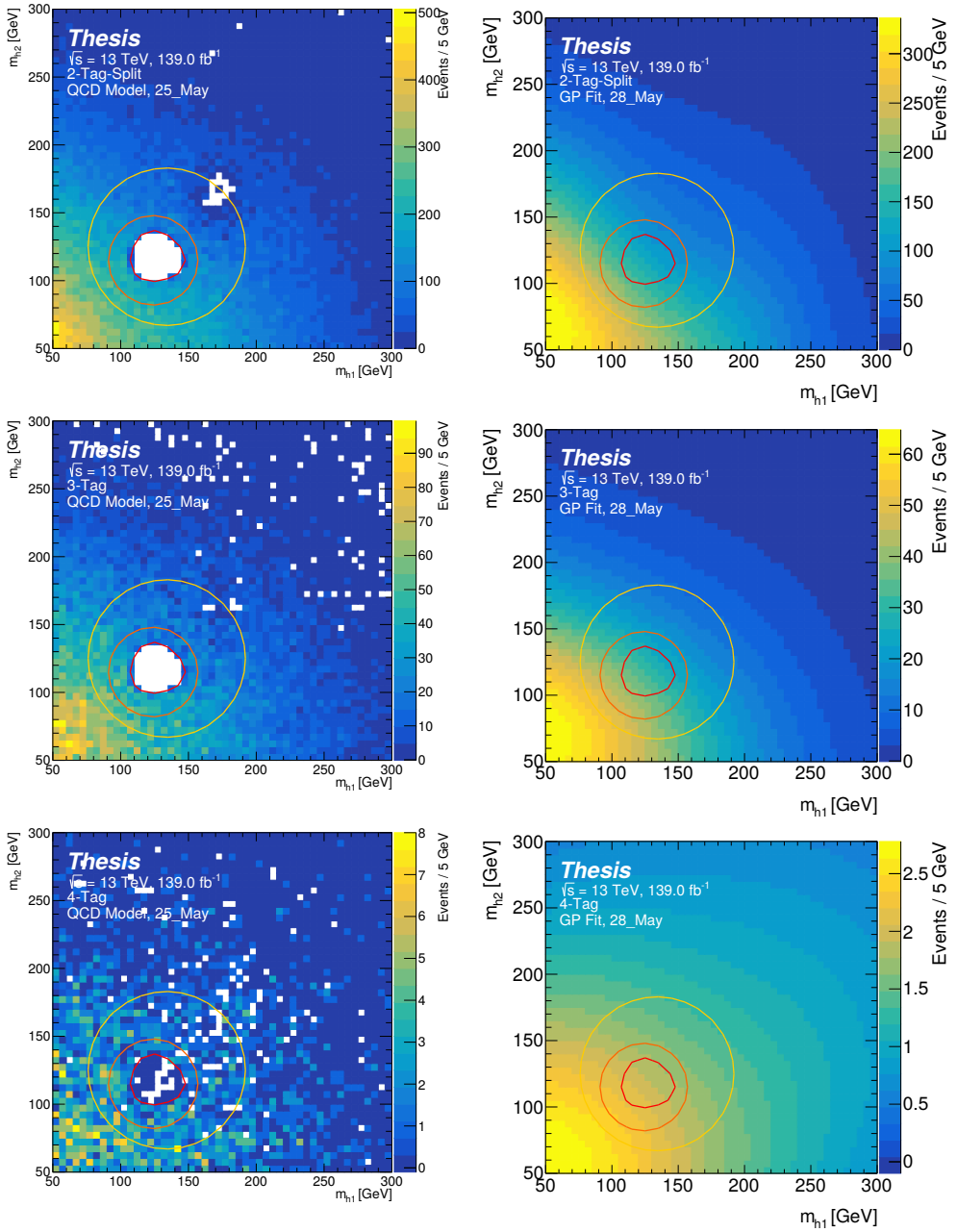


Figure 4.17: Comparison of the QCD model, blinded high-tag data minus $\alpha_{t\bar{t}} \times t\bar{t}$, on the left to the result of the Gaussian Process interpolation on the right. The Gaussian Process is used to assess the uncertainty from the extrapolations from control to signal regions and low-tag to high-tag regions in the background estimate. Shown are the $2b$ region (top row), $3b$ region (middle row), and the $4b$ region (bottom row).

Size and Placing of the CR

The exact size and placement of the control region has an impact on the values of the normalization factors μ_{QCD} and $\alpha_{t\bar{t}}$, which in turn have an impact on the final background estimation. To account for the impact of these choices, an additional

uncertainty is assigned to the choice of CR. This uncertainty is evaluated by creating a set of alternative control regions with varying size and position, re-running the background estimation using each one, and comparing the difference in the yields of the final background estimates in the SR. The size of this uncertainty is taken as the largest difference in yield between any of the variations and the nominal background estimate [17]. Six alternative control regions (essentially six changes to the yellow and orange circles in figure 4.6) were made:

1. **Up-up control region** The centre of the circles that define the control region were moved up and to the right by 3 GeV in both leading and subleading Higgs boson candidate mass.

2. **Up-down control region** The centre of the circles that define the control region were moved right by 3 GeV in the m_{h_1} axis and down by 3 GeV in the m_{h_2} axis.

3. **Down-up control region** The centre of the circles that define the control region were moved left by 3 GeV in the m_{h_1} axis and up by 3 GeV in the m_{h_2} axis.

4. **Down-down control region** The centre of the circles that define the control region were moved left by 3 GeV in the m_{h_1} axis and down by 3 GeV in the m_{h_2} axis.

5. **Large control region** The R_{hh}^{VR} requirement is decreased by 3 GeV and the R_{hh}^{CR} increased by 3 GeV, shrinking the inner radius and expanding the outer boundary of the control region.

6. **Small control region** The R_{hh}^{VR} requirement is increased by 3 GeV and the R_{hh}^{CR} decreased by 3 GeV, shrinking the outer radius and expanding the inner boundary of the control region.

[17] This uncertainty is replaced by the statistical uncertainty on the nominal yield itself if it is larger that the yield difference. This also applies to other uncertainties involving yield differences.

While these changes also modify the validation region, they leave the blinded signal region unchanged. The variations are done separately for each of the b-tag regions $4b$, $3b$, and $2b$, with final uncertainties of 6.0%, 1.6%, and 0.93%, respectively.

4.5.2.2 Background Shape Uncertainties

As previously mentioned, the assumption that the low-tag regions provide a model for the QCD background in the high-tag regions introduces uncertainty. Relevant to this analysis, the inaccuracy of this assumption would lead to mismodelling in the m_{hh} distribution. Other parts of the analysis, such as the smoothing fit applied to the background estimate, can add to the shape uncertainty. The uncertainty on the shape of the m_{hh} spectrum is evaluated in two ways.

Uncertainty on the m_{hh} smoothing procedure

The different choices of fitting function and fit range for the smoothing process produce slightly different results. Uncertainties on these choices are derived from using alternative functions and changing the range of the fit, as shown in figure 4.18 (see footnote 10). The fits with the largest difference compared to the nominal one were taken as a systematic uncertainty on the shape of the m_{hh}. The envelope defined by the two most different fits in each direction (larger or smaller) are taken as up and down variations of this systematic uncertainty [18]. These uncertainties correspond to the `MJ2` and `MJ4`[19] curves in the lower panel of figure 4.18 (left) and the fits with ranges [1236,2802] GeV and [1131,2698] GeV in the lower panel of figure 4.18 (right). This was studied and implemented by the analysis group.

Shape Non-Closure Uncertainty

Any residual mismodelling in the m_{hh} spectrum was assigned a systematic uncertainty by comparing the background prediction in the signal region obtained from the nominal method to the shape of the high-tag data in the validation region. The shape uncertainty was taken as the ratio of the unit-normalized SR background

[18] If *all* variations are smaller or larger than the nominal, then a symmetric envelope is formed from the largest deviation.

[19] See footnote 10.

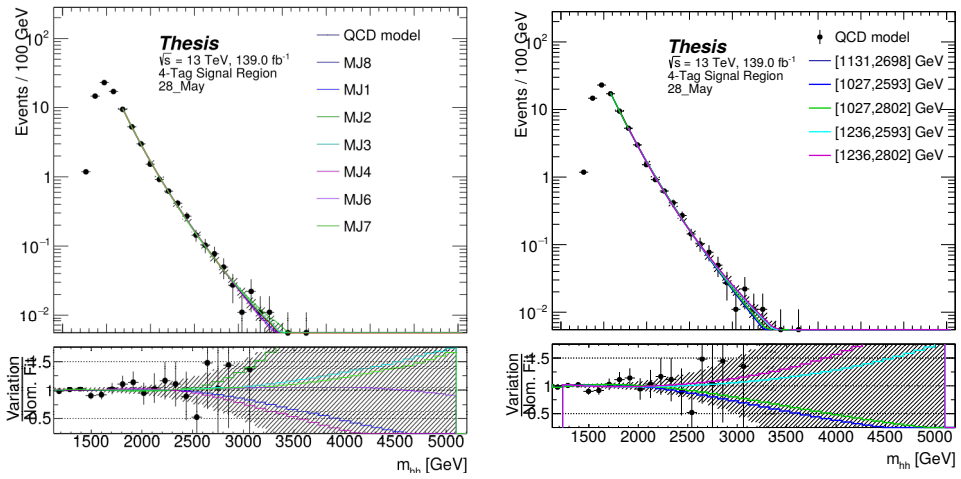


Figure 4.18: Result of the smoothing fits in the 4b region for all functions (a) and various choices of fit range (b). The grey band shows statistical error on the nominal fit. The largest differences as compared to the nominal MJ8 function were taken as systematic uncertainties.

prediction and VR high-tag data as shown in figure 4.19 for the 2b region (with the ratio shown in the bottom panel being the shape uncertainty). Notably, the background model seems to underestimate the yield in the low m_{hh} region in the 2b region which can be observed in figure 4.10b. This uncertainty was only relevant for the 3b and 2b regions, since the statistical uncertainty on the 4b background prediction was larger than any mismodelling observed. It is also worth noting that while some plots shown in this chapter include systematic uncertainty bands, these *do not* include the shape non-closure uncertainty as it was derived after the process where all the other uncertainties are plotted [20].

[20]This comment is only relevant to plots in the 2b region.

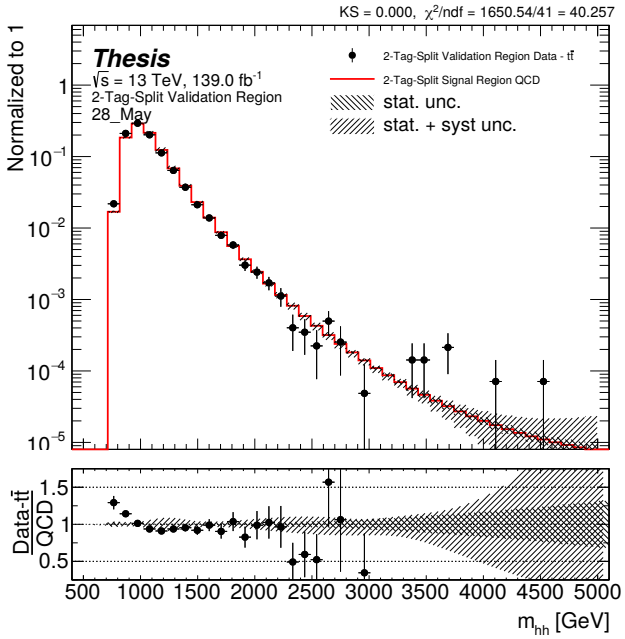


Figure 4.19: Normalized m_{hh} distributions from the $2b$ region. The data points correspond to high-tag VR data, while the red line shows the SR background estimate. The ratio of the two, shown in the lower panel, is taken as a shape uncertainty.

4.5.3 Summary of Systematic Uncertainties

Table 4.4 shows the impact of each group of systematics on the expected limit on the two benchmark models. The impact is shown for the 1 TeV mass resonance [21] of each model. To calculate the impact of each group of systematics, the expected limit was recalculated taking only one group of systematic uncertainties (b-tagging, jet, and background estimation systematics) into account at a time, and this limit was compared to the nominal expected limit accounting for statistical uncertainty only. The dominant systematic uncertainties are the ones associated with the background estimation procedure with an impact of 22.5% and 44.3% on the 1 TeV scalar and graviton expected limits respectively, particularly the uncertainty on the extrapolation from low- to high-tag, and control to signal regions which had an impact of 21.3% and 41.4%.

[21] This mass point was chosen since systematic uncertainties have most effect in the low-mass end of the m_{hh} distribution, where the statistical uncertainty is smaller than in the high-mass end of the spectrum.

Source of Uncertainty	Impact on Expected Limit (%)	
	$m_S = 1$ TeV	$m_{G^*_{kk}} = 1$ TeV
B-tagging	0.25	0.36
Jet	3.93	4.18
Background Estimate	22.50	44.34
Extrapolation	21.28	41.41
CR Definition	6.24	10.23
CR Statistical	3.17	4.61
m_{hh} Smoothing	0.03	0.14
Shape non-closure	< 0.01	0.06

Table 4.4: Overview of systematic uncertainty sources and their impact expected limit on the production cross section 1 TeV scalar and spin-2 resonances decaying to $hh \to 4b$ in the boosted channel. The indented entries below Background Estimate correspond to the various components of this group of uncertainties.

5
Analysis Results

Contents

5.1 Statistical Methods 76
5.1.1 Likelihood 77
5.1.2 Test Statistic 78
5.1.3 Limit Setting Procedure 79
5.1.4 Expected Limits 82
5.2 Results 85
5.2.1 Unblinded Signal Region Distributions 86
5.2.2 Observed Limits 87

5.1 Statistical Methods

The statistical analysis, implemented using the `RooFit` framework [99], in this search is similar to techniques used in [31], and the results shown in this section are obtained with the Heavy Scalar and graviton G^*_{kk} benchmark models. The di-Higgs invariant mass m_{hh} in the Signal Region (SR) was chosen as the discriminating variable to evaluate the upper limit on the production cross section for the heavy scalar S and graviton G^*_{kk} predicted by these models.

In broad terms, the statistical analysis of the data proceeds as follows. After passing the data, and the signal and $t\bar{t}$ simulated samples through the event selection

and background estimation processes detailed in section 4.4, histograms of m_{hh} in the SR of all b-tag-regions ($4b$, $3b$, and $2b$) are obtained for observed data, signal and the estimated background. The observed data are then compared to the corresponding estimated background histogram, and to the signal *plus* the estimated background. This comparison determines if the observed data agree with the background estimation B only, or if it is more consistent with the signal plus background $(S + B)$ and, if so, how significant this improved agreement is.

In reality, the agreement of the data to the various hypotheses is evaluated using a binned maximum-likelihood fit. The *likelihood* is the probability to measure the obtained data under the assumption of a hypothesis, $(S + B)$ or B in the case of this search. During the fit, the signal histogram is scaled up or down by the *signal strength* parameter μ, that tests *how large* of a signal, added on top of the estimated background, gives the best compatibility to the data [1]. Important details of the calculation of the test statistic and cross section upper limits are explained below, and although the statistical process to obtain them is rather complex, these values form the fundamental observations of this boosted $hh \rightarrow 4b$ search.

5.1.1 Likelihood

The likelihood of an event distribution, given the observed data, for a signal strength μ accounting for the nuisance parameters (NPs) in the vector $\vec{\theta}$ is defined as:

$$L\left(\text{data}|\mu,\vec{\theta}\right) = \prod_i \mathcal{P}\left(y_i|B_i(\vec{\theta}) + \mu S_i(\vec{\theta})\right) \prod_j \mathcal{G}\left(\theta_j\right) \tag{5.1}$$

where i runs over all bins in the $4b$, $3b$, and $2b$ histograms and j over all sources of systematic uncertainty. y_i represents the yield of data in bin i of the m_{hh} distribution, and S_i and B_i denote the expected signal and background yields for that bin. $\mathcal{G}\left(\theta_j\right)$ represents the Gaussian of the nuisance parameter j [2], and

[1] For example, $\mu = 0$ corresponds to no signal, i.e. the background only hypothesis B. A more detailed description of the fit is presented in the following sections.

[2] Each source of systematic uncertainty is modeled by a NP with a Gaussian probability density function of mean 0 and standard deviation 1. Their nominal effect on S and B is estimated in auxiliary measurements in signal-free regions (see section 4.5).

$\mathcal{P}(y|y_{\text{exp}})$ denotes the Poisson probability density function of observing a yield y when y_{exp} are expected under some parameters μ and $\vec{\theta}$.

The likelihood can be calculated for various values of the parameters μ and $\vec{\theta}$, and the best fit parameters (denoted $\hat{\mu}$ and $\hat{\vec{\theta}}$) are the ones that maximize the likelihood. This function, and its fit to data, can then be used not only for final measurements of the search, but to test the impact of individual nuisance parameters, evaluate expected limits to asses the searches sensitivity, among other uses.

5.1.2 Test Statistic

The figure of merit, or *test statistic*, for the final analysis of the data in this $hh \to 4b$ search is a one-sided negative profile log-likelihood [3] ratio [100] $\tilde{q_\mu}$ (based on the definition of likelihood L above, the arrow over $\vec{\theta}$ has been omitted):

$$\tilde{q_\mu} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0 \\ -2\ln\frac{L(\mu,\hat{\hat{\theta}}(\mu))}{L(\hat{\mu},\hat{\theta})} & 0 \leq \hat{\mu} < \mu \\ 0 & \mu < \hat{\mu} \end{cases} \tag{5.2}$$

where μ is the tested signal normalization, $\hat{\mu}$ is value of μ that maximizes the likelihood given the data, θ is the set of nuisance parameters, $\hat{\theta}$ is the value of θ that maximizes the likelihood, and $\hat{\hat{\theta}}$ is the value of θ that maximizes the likelihood given a fixed μ. L is the profile likelihood computed with the given parameters of μ and θ. $L(\hat{\mu},\hat{\theta})$ is the *unconstrained* likelihood, meaning that μ and all nuisance parameters are allowed to float during the fit. $L(\mu,\hat{\hat{\theta}})$ is the *constrained* likelihood found by setting μ to a particular value, but letting θ float.

When testing a simple hypothesis the ratio of the constrained and unconstrained likelihood constitutes the most *statistically powerful* variable to reject one hypothesis in favour of another one [101], which motivates the ratios of likelihoods in the definition of the test statistic. $\tilde{q_\mu}$ tests the compatibility of the data with the $B + \mu S$ hypotheses, where smaller $\tilde{q_\mu}$ corresponds to better compatibility. One

[3] The negative of logarithm of the likelihood is used for ease of computation, since minima in this function correspond exactly to maxima in L.

can then minimize $\tilde{q_\mu}$ varying or fixing the model parameters, such as μ or specific NPs, to test various features of the model.

The chosen test statistic ensures that a downward fluctuation of the background is not evidence against the background hypothesis. If there is a down fluctuation of the background, the best fit could correspond to an unphysical, *negative*, signal strength $\hat{\mu} < 0$. In this case the chosen test statistic simply sets the best-fit signal strength $\hat{\mu}$ in the profile likelihood to zero.

This definition of $\tilde{q_\mu}$ also assures that an upwards fluctuation of the signal does not serve as evidence against a signal hypothesis. This is a desirable characteristic when placing upper limits, considering that if the best-fit signal strength is larger than the one being tested (i.e. $\mu < \hat{\mu}$), this should not be evidence against that tested μ [4]. If this happens the test statistic is set to zero, indicating the maximum compatibility with that signal hypothesis.

5.1.3 Limit Setting Procedure

The test statistic $\tilde{q_\mu}$ is then used to calculate the 95% CL_{S} [39] confidence level statistic for exclusion of a given signal mass, a standard procedure for searches for exotic particles. The CL_{S} method takes into account not only the probability to reject the $S + B$ hypothesis were it to be true (p-value of the measurement), but also weights it by the probability of accepting the background only hypothesis B if this one was correct (also known as the *power* of the measurement):

$$\mathrm{CL_s} = \frac{P_{\mathrm{s+b}}(q_{\mathrm{obs}})}{P_{\mathrm{b}}(q_{\mathrm{obs}})} \tag{5.3}$$

where $P_{\mathrm{s+b}}(q_{\mathrm{obs}})$ is the probability of the signal+background model to produce equal or greater *in*compatibility to the data than observed, and $P_{\mathrm{b}}(q_{\mathrm{obs}})$ is the probability of the background only model to produce equal or better agreement to the data than observed. Here q_{obs} is the value of the test statistic calculated for a specific value of μ being probed. Note that while $P_{\mathrm{s+b}}$ corresponds to the p-value

[4] If we did take this as evidence against μ, this could lead to rejecting a particular value of μ because we observe *too large* of a signal in the data!

of the $S + B$ hypothesis, $P_{\mathrm{b}}(q_{\mathrm{obs}})$ corresponds to 1 - p-value of the background only hypothesis. A sketch depicting the distribution of the test statistic under the B (in blue) and $S + B$ (in orange) hypotheses, as well as the relevant probabilities, is shown in figure 5.1 [5]. The shaded areas under each curve and delimited by q_{obs} give the values to calculate CL_{S}.

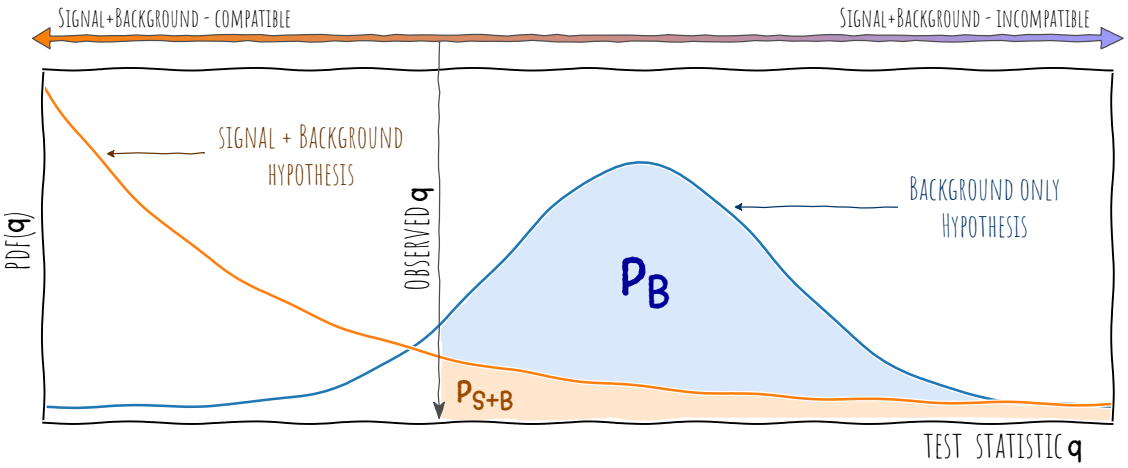


Figure 5.1: Sketch of the probability density distribution of the test statistic under the B (blue) and $S + B$ (orange) hypotheses (see footnote 5). The final CL_{S} is given by the ratio of the two shaded areas. The limit on μ is set such that this ratio is 0.05.

The 95% CL_{S} limit on the signal normalization μ is found by varying μ for each signal mass point until $CL_{\mathrm{S}}(\mu) = 0.05$ is found. This can be seen in figure 5.1 as the orange curve shifting left, and with a more pronounced decay, as μ gets larger. The limit is set when CL_{S}, given by the probabilities defined by the distributions and the observed $\tilde{q}_{\mathrm{obs}}$ (represented by the shaded areas), reach 5%. This signal strength value can then be converted to an upper limit to the production cross section $\sigma(X \to hh \to b\bar{b}b\bar{b})$ of the benchmark signal models, the quantity shown in limit plots such as the ones in figures 6.9 and 1.7.

The full distribution of the test statistic $\tilde{q}_{\mu}$ under the various $B + \mu S$ hypotheses (represented by orange and blue curves in 5.1 above) can be determined by running an ensemble of simulations known as *toy experiments*. In practice, a different

[5] The distribution shapes here are not exact representations of true ones the data would produce, but allow one to visualize the concepts clearly. The behaviour and concepts are still applicable to the analysis' test statistic. In fact, the full distributions are never simulated explicitly for the analysis, as the asymptotic approximation is used, see below.

approach (the Asymptotic Approximation Method [100]) is used for this analysis, which exploits the property of these distributions to asymptotically approximate a χ^2 distribution for large datasets. This saves computation time by avoiding large number of simulations.

The CL_S test is repeated for each of the resonance masses being tested, which allows us to plot the upper limits on production cross section as a function of the resonance mass predicted by a given model. A sketch of such plots and the components shown is depicted in figure 5.2. By convention, the solid line traces the upper limit on cross section that was found to be *incompatible* with the observed data (i.e. yielded $CL_S = 0.05$), for each of the mass points on the x-axis. Similarly, the dashed line shows an expected upper limit computed by using a background-only pseudo-dataset in place of the observed data (see section 5.1.4). The green (yellow) bands show a deviation of one (two) standard deviations from the expected limit. Points on the observed line are said to be excluded at the 95% Confidence Level (CL). Values above this line are then excluded as well.

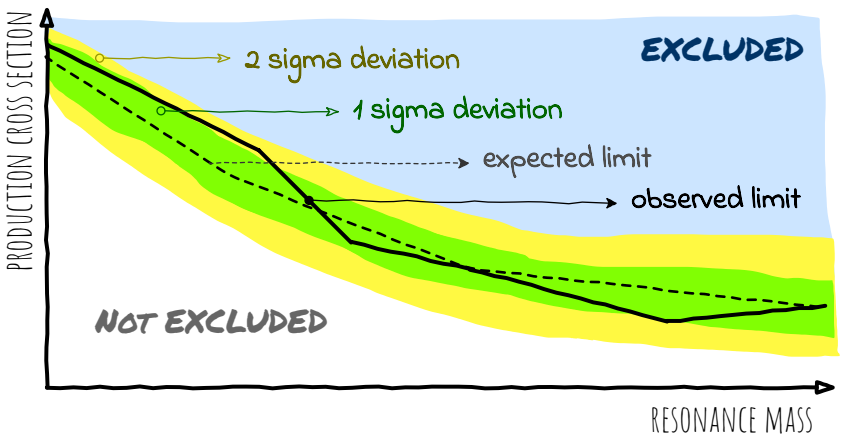


Figure 5.2: Sketch of the plots used to show the observed and expected CL_S limits on the resonance production cross section. The solid (dashed) line traces the observed (expected) limit for each of the mass points on the x-axis, and the coloured bands show the deviation from the expected limit. For each mass point, values of cross section above the observed line are excluded.

5.1.4 Expected Limits

Before unblinding, the statistical model is tested using an *Asimov dataset* [100] constructed to model the background estimate in the signal region. This dataset is actually a *pseudo-dataset* constructed by taking the nominal value of the parameters in the model, in this case it is the background only hypothesis (i.e. $\mu = 0$) and the nominal value for each of the NPs. This Asimov dataset then *exactly* resembles the nominal background prediction. The expected limit is computed by running the statistical analysis described above, but the profile likelihood fit is done on the Asimov data in place of the signal region data.

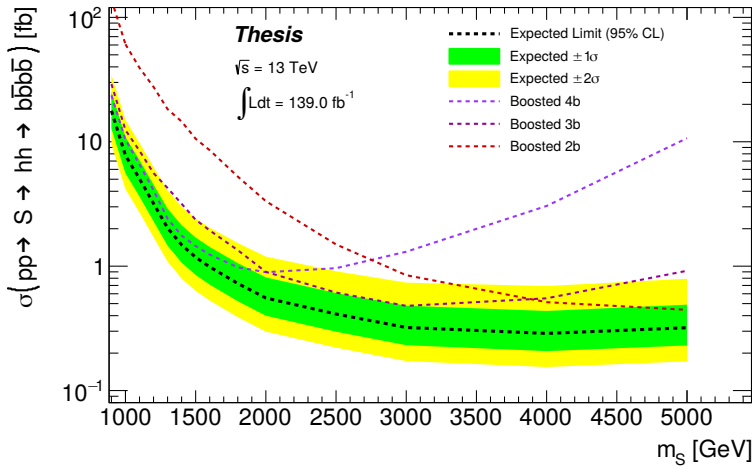
The expected limits are a good metric for the sensitivity of the analysis, independent of what is ultimately observed in the signal region. These are also helpful to understand how changes in the analysis would impact the final limit.

The expected limit on the production cross section of heavy scalar and RS graviton resonances computed with each of the b-tag regions, and all regions combined, are shown in figure 5.3. The black line shows the expected upper limit computed for the full model including the three regions, while the coloured lines show individual CLs fits of each b-tag region alone. The sensitivity of the $4b$ and $2b$ are best in the low and high resonance mass regions respectively, with the $3b$ region bridging the intermediate region between the two.

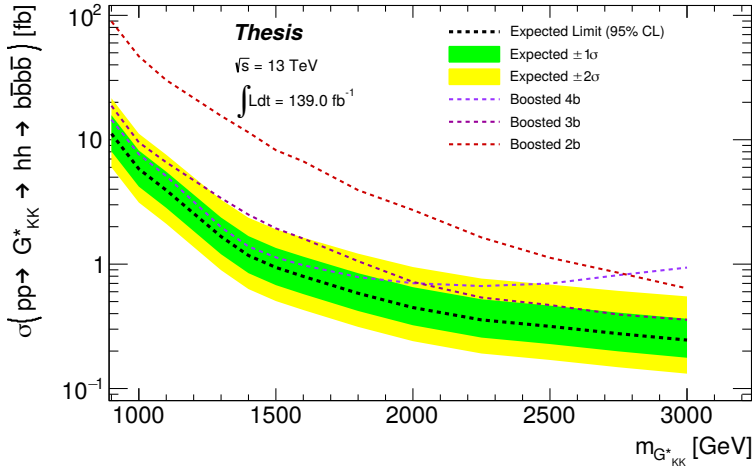
5.1.4.1 Impact of Systematic Uncertainty

Figure 4.16 in the previous chapter shows the expected 95% CL_{S} upper limits on the production cross section of resonances predicted by the heavy scalar model in two scenarios: taking only statistical uncertainty into account, and accounting for both statistical and systematic uncertainties. The impact of systematics is seen to be relatively small (within 1σ) when compared to the statistical uncertainty.

Further standard checks were done to test the ability of the fit to model the data, and to evaluate the impact of the various nuisance parameters (NPs) on this search. The tests in this subsection were done using the final unblinded signal region data in the likelihood fit.



(a)



(b)

Figure 5.3: Expected 95% CL upper limits on the production cross section times branching ratio of the (a) heavy scalar model and (b) RS graviton accounting for statistical uncertainties only, shown for the three b-tag regions ($4b$, $3b$, and $2b$).

Pulls of Nuisance Parameters

The *pulls* of NPs measure how well our model (with the included NPs and the assumptions on them) can describe the data. A fit on a dataset with the *exact nominal* values of all parameters would yield pulls of exactly 0 and uncertainty ± 1, since no new information would be present in such data, and the model describes it perfectly by construction. However, when fitting other data, the signal region signal region data in this case, only certain values of each NP are compatible with the dataset. This allows the fit to *pull* the value away from 0, or *constrain*

its uncertainty to be smaller than the original assumption. While pulls away from the assumed values are not necessarily problematic, if not understood, this could point to underlying issues in the analysis or fit model, making this a useful tool to evaluate the method.

The pulls of NPs with the highest impact are shown in figure 5.4. This plot shows the ranking of NPs of the 4b region model, with similar plots for the 3b and 2b region shown in appendix A. Most pulls are essentially zero, with the uncertainty on the pulls of all NPs being close to 1, which indicates that there are no major constraints the NPs and the model is adequate to fit the data. The most significant pull is seen in the 4b region extrapolation uncertainty, by close to 0.4. Considering the rest of NPs show little pulls and constrains, and that this is a NP calculated on a data-driven background in a region with very low data statistics, this is considered acceptable agreement. The fits for each b-tag region were restricted to ranges to account for sensitivity and empty bins in each m_{hh} spectrum. For ease of computation, these plots were computing accounting only for the 20 NP sets with the highest impact, taken from a ranking of the full set of NPs (∼100) in figure A.2.

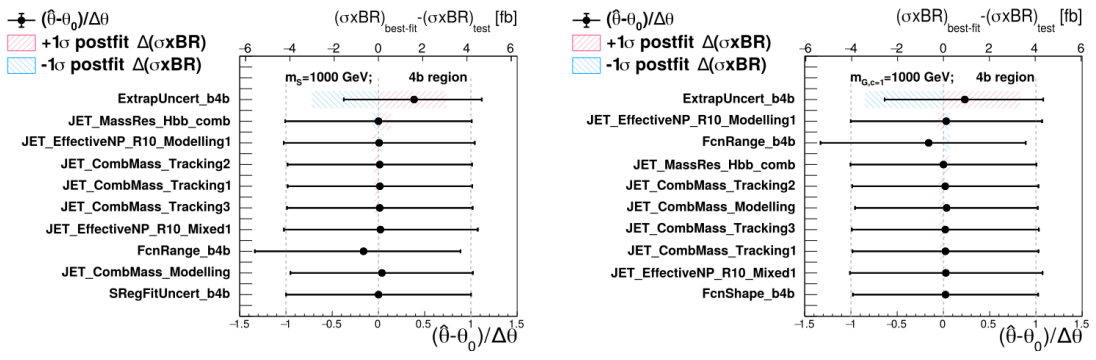


Figure 5.4: Pulls (circle markers, bottom axis) and impacts (bars, top axis) in femtobarns on the measured cross section times branching ratio of a 1 TeV (a) heavy scalar and (b) G^*_{kk} graviton. The plot shows the 10 nuisance parameters with most impact on the expected limit from the 4b region. NPs related to the background estimate and jet properties were ranked as the most impactful.

Impact of Nuisance Parameters

The *impact* of all NPs was also computed after the fit. The impact was calculated by

fixing the value of one NP at a time and running the fit with all other parameters free to float, and checking what impact it has on the cross section. Figure 5.4 also shows the pulls and impacts in femtobarns on the best-fit cross section times branching ratio of a 1 TeV scalar and graviton ranked by their post-fit impact, corresponding to the coloured bars and the top axis. The extrapolation uncertainty is found to be the dominant one in both signal models by a large margin. The ranking of subleading uncertainties was different for the graviton and scalar signals, which, in this regime limited by the statistical uncertainty, could be due to the otherwise small differences on how each uncertainty affects each signal's m_{hh} distributions. A limited set of systematics is included in this fit: the main background related systematics, the experimental systematics on jets and b-tagging, and luminosity. NPs associated with background estimation and jet systematics had the biggest impact on the fit.

It is worth noting that the flavour-tagging and jet-related nuisance parameters considered for the model in the results shown in this chapter are not the final recommendations of the relevant ATLAS performance groups. Due to the recent re-training of the b-tagging algorithms mentioned in section 6.2.1, ATLAS's flavour tagging group was able to provide only work-in-progress uncertainties by the time of this writing. However, studies done with the previous `MV2` algorithm show that flavour tagging uncertainties are not expected to have a strong impact on the limit (see appendix A for details). Jet uncertainties were seen to have a bigger impact, but for similar reasons they were not finalized by the time of this writing.

5.2 Results

The observed data in the signal region, as well as the fits to the unblinded signal regions and the 95% CL_{S} upper limits on the production cross section of a range of resonance masses predicted by the graviton G^*_{kk} and Heavy Scalar models, are presented below.

5.2.1 Unblinded Signal Region Distributions

Figure 5.5 shows the unblinded signal region distributions for the $4b$, $3b$, and $2b$ regions. The agreement between observed data and prediction is consistent with the validation region result in figure 4.14. The signal region histograms are passed through the statistical analysis described above, yielding the observed limits in section 5.2.2 below.

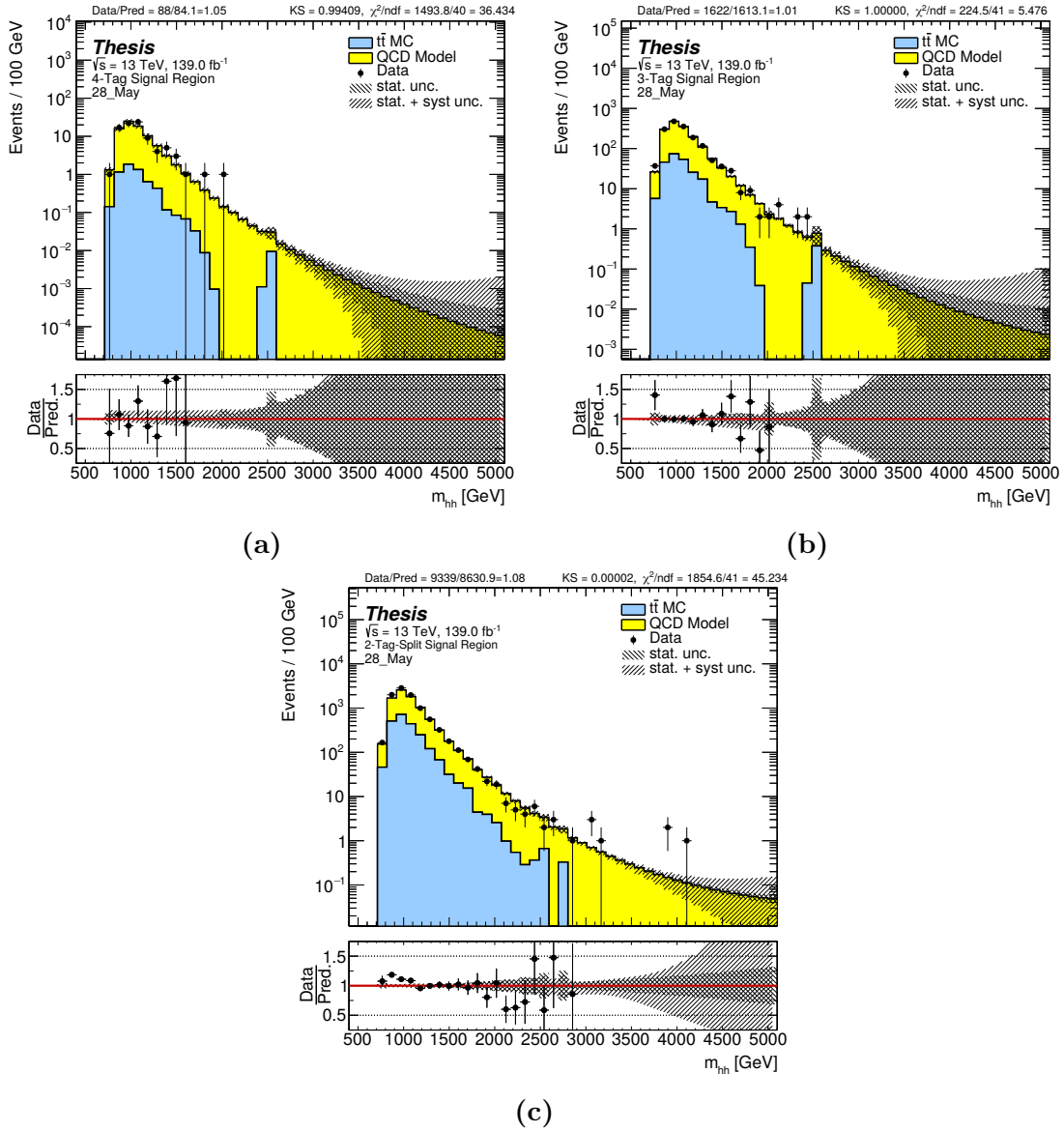


Figure 5.5: Di-Higgs mass in the unblinded signal region in the (a) $4b$, (b) $3b$, and (c) $2b$ regions.

Figure 5.6 shows the unblinded m_{h_1}– m_{h_2} plane for $4b$, $3b$, and $2b$ region data,

and it shows the expected smoothly falling background shape. A noticeable ridge is seen at the top quark mass in the leading Higgs boson candidate mass axis m_{h_1} in figures 5.6b and 5.6c.

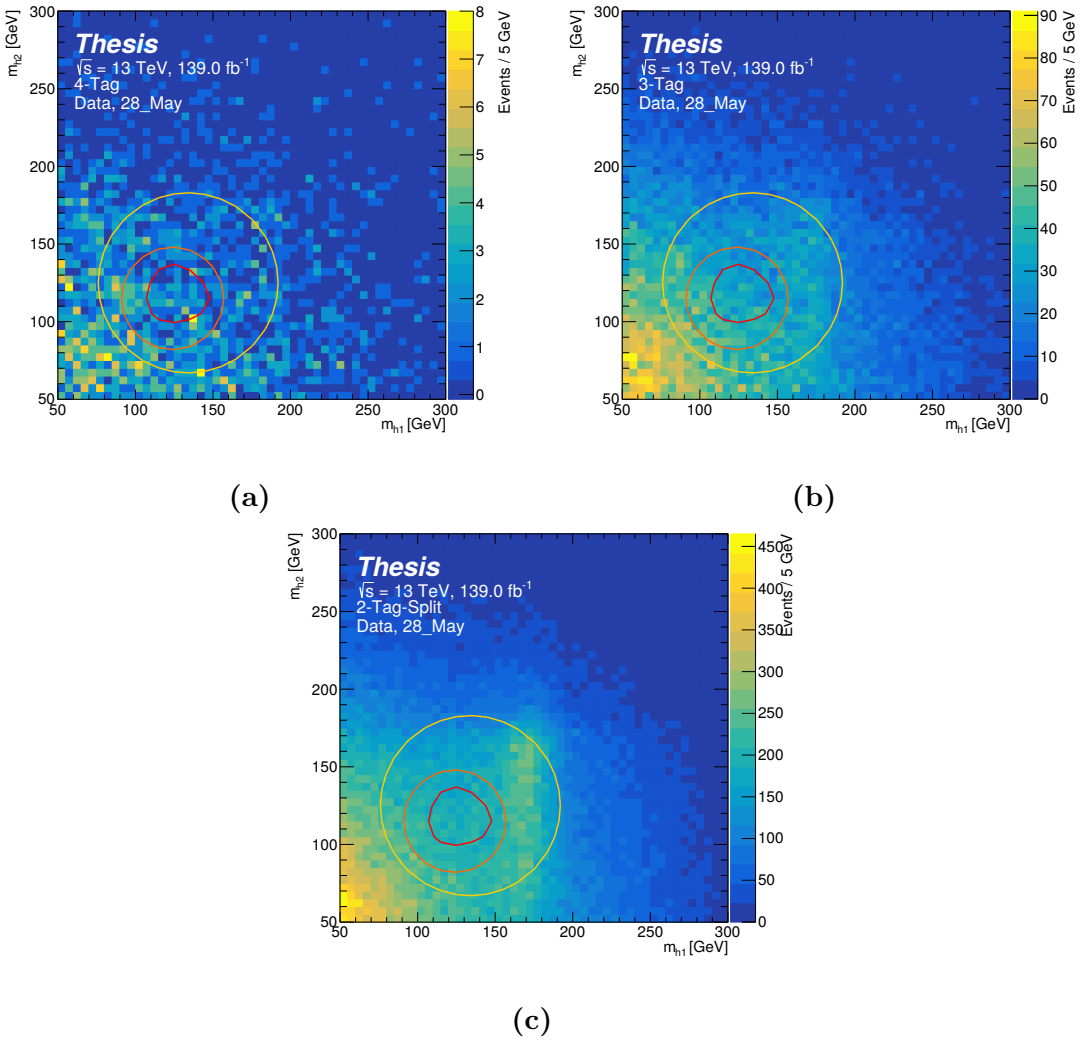
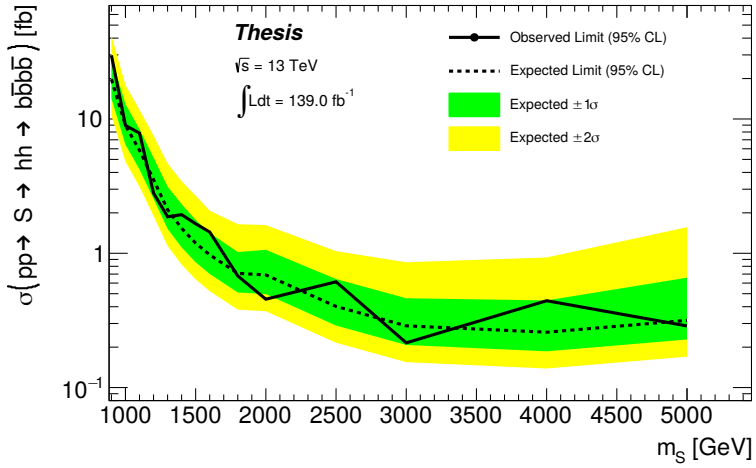


Figure 5.6: Scatter plots of the $m_{h_1} - m_{h_2}$ plane including the unblinded signal region for the (a) $4b$, (b) $3b$, and (c) $2b$ regions.

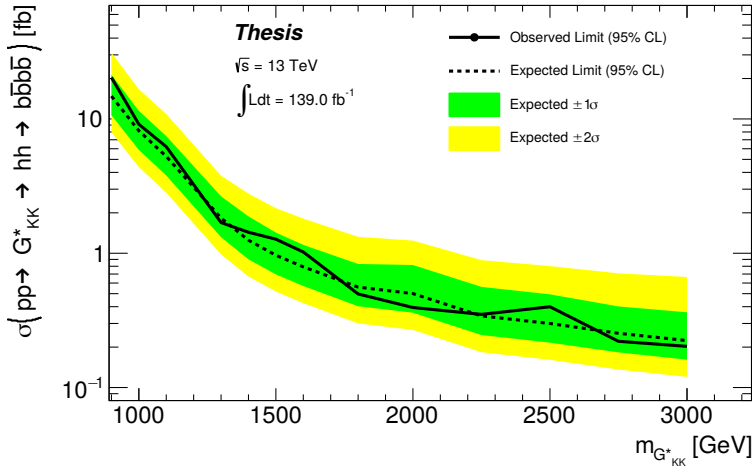
5.2.2 Observed Limits

The observed upper limit on the production cross section of the narrow scalar S and the RS graviton G^*_{kk} resonances are shown in figure 5.7. No significant deviation from the expected limits are observed.

Results of this boosted search are expected to be statistically combined with the resolved counterpart of this search, which is sensitive to low-mass resonances.



(a)



(b)

Figure 5.7: Observed 95% CL upper limits on the production cross section times branching ratio of the (a) heavy scalar model and (b) RS graviton accounting for both statistical and systematic uncertainties. No significant deviation from the expected limits are observed in either model.

6

Further Variable Radius Track Jet Studies

Contents

6.1 Collinear Track-Jets Veto 90
6.1.1 Veto Motivation and Definition 90
6.1.2 Veto Cause and Impact on the Analysis 90
6.1.3 Potential fix: Increase Minimum p_{T} to Trigger Veto . . 93
6.1.4 Performance of Variable Radius (VR) Track Jets 96
6.2 Impact of Variable Radius (VR) Track Jets on Sensitivity 100
6.2.1 Impact of VR Specific b-tagging Training 101

This section presents a series of studies done to test and validate the performance of variable radius (VR) track jets, and later implement them to be used in the $hh \to 4b$ analysis that were made as part of this thesis work. Previous iterations of this search have used fixed radius (FR) track jets with $R = 0.2$ and these jets are used for comparison in the following sections.

While the studies in this section were important in informing and guiding the boosted $hh \to 4b$ analysis, and most of what is discussed in this chapter was used in the final result, the level of detail was not deemed necessary to understand the analysis described in chapters 4 and 5. These are then presented as a separate chapter to complement the analysis chapters.

6.1 Collinear Track-Jets Veto

Following a recommendation from the flavour-tagging group, events with collinear, or very nearly collinear, variable-radius track jets were vetoed from this analysis. Some of these configurations lead to a problematic case when the soft jet's axis is also contained within the hard jet (see figure 4.8). While the set of tracks used to construct the jet is well-defined, in event of this problematic case, the set of tracks used as input to the b-tagging algorithm can encompass tracks from another jet. The b-tagging algorithm assigns tracks based purely on proximity to the jet axis and can incorrectly assign tracks when jets overlap. Having the correct tracks when forming a track jet is vital for b-tagging performance, since the algorithm heavily relies on finding a secondary vertex using the jet's tracks and other properties of specific tracks (see section 3.3). Events with such problematic collinear track jets are therefore vetoed from the analysis [1].

6.1.1 Veto Motivation and Definition

To remove these problematic cases from the analysis, any track jet with $p_{\mathrm{T}} > 5$ GeV was checked for overlap. Events were vetoed if one of the two leading track jets associated to a large-R jet met the following condition [2]:

$$\Delta R(\mathrm{jet}_i, \mathrm{jet}_j) < min(\mathrm{R}_i, \mathrm{R}_j) \tag{6.1}$$

where ΔR is the distance in the $\eta - \phi$ plane between jets, i and j run over all the jets fulfilling the p_{T} requirement, and $\mathrm{R}_{i,j}$ are the radius of each jet.

6.1.2 Veto Cause and Impact on the Analysis

To understand the characteristics of such collinear track jets, we study events with overlapping VR track jets [3]. The plot in figure 6.1 shows the frequency with which

[1] The final effect of the veto is detailed in table B.1 which details the yields of data and signal events after each step in the selection process.

[2] This checks whether a high p_{T} VR track jet contains the axis of a lower p_{T} VR jet.

[3] For this section *only*, events were only vetoed if the overlapping jets were associated to the leading Higgs boson candidate in order to easily study the characteristics of said jets.

various pairs of jets caused an event to be vetoed. The third-leading jet in p_{T} was the one overlapping with either the leading or subleading in most of the cases as shown by the prominent peaks in the h1_13 and h1_23 bins. This indicates that a large fraction of events were vetoed because the leading or subleading `VR` track jets fulfilled equation 6.1 when compared to the third leading track jet. Figure 6.2 shows the p_{T} distributions of the jets involved. The much lower transverse momentum of the third jet suggests that in these cases, the hard jet emitted some collinear radiation that was reconstructed as another, softer, jet.

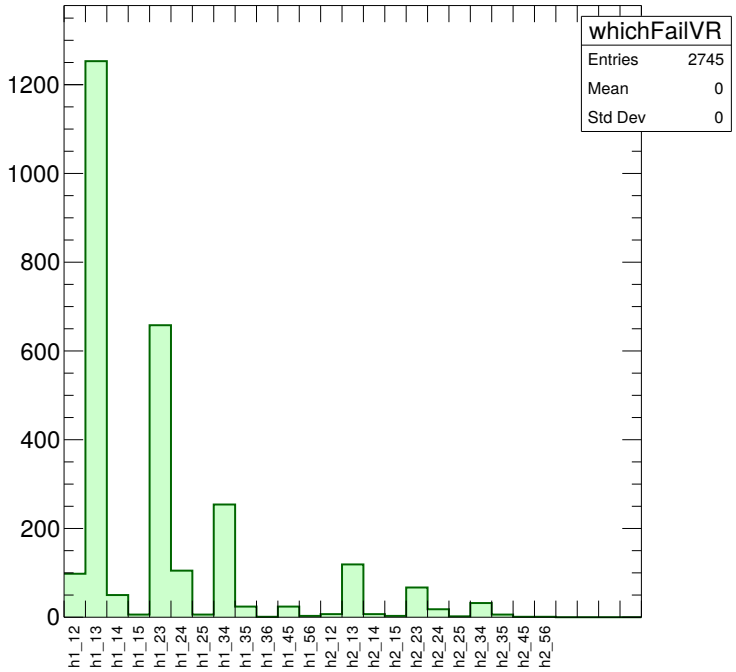


Figure 6.1: Histogram showing the frequency with which pair of track jets triggered a collinear veto. The prefix $h1$ or $h2$ in the x-axis labels indicates which Higgs boson candidate the overlapping track jets are associated with, while the second pair of numbers indicate which track jets overlapped (by their order in p_{T}).

In the highly boosted regime of this analysis, b-jets from the decay of a Higgs boson are expected to be collimated very close together, making this veto more likely. We studied the effect of removing such events on the signal selection efficiency. Figure 6.3 shows the impact of this veto on the acceptance of spin-2 signal events from MC, which ranges from around 10% to 20% in the range of resonance masses tested. The mass points which were most affected were the heaviest ones due to

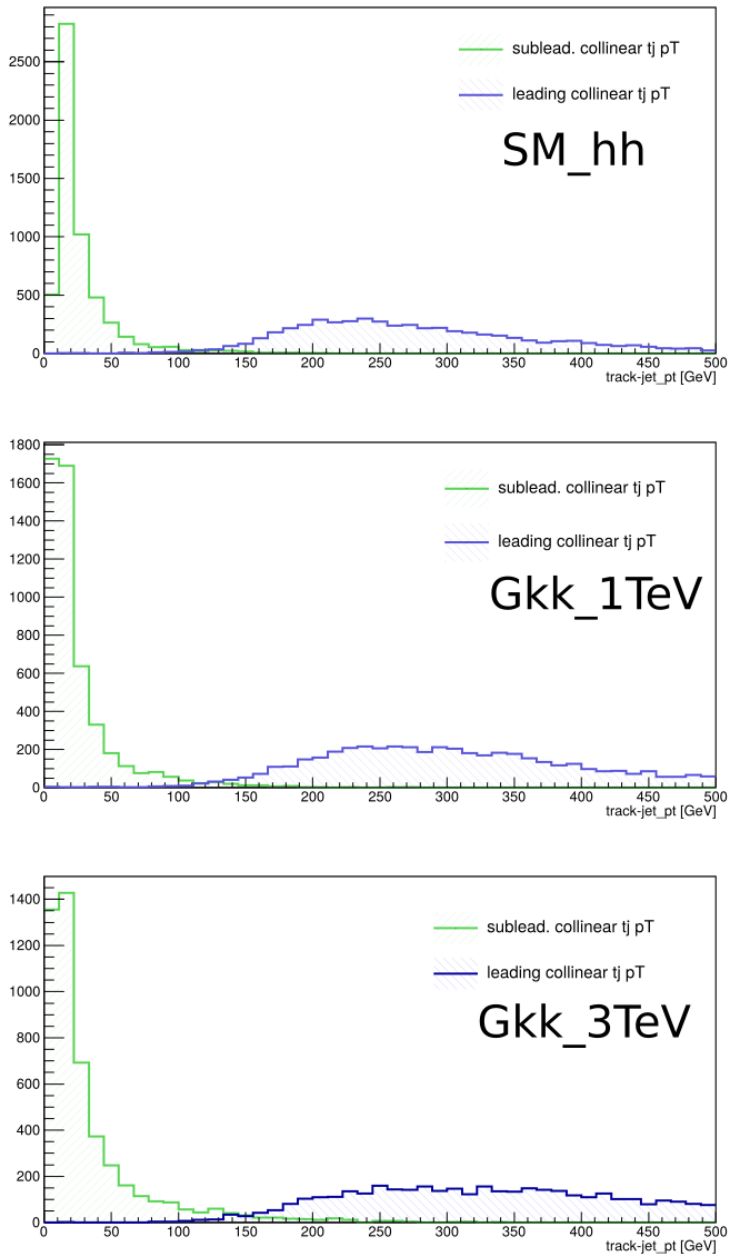


Figure 6.2: Distribution of the p_{T} of overlapping track jets that could cause an event to be vetoed. These plots are shown for a Standard Model non-resonant signal, and 1 TeV and 3 TeV spin-2 resonances, on each row. These plots only contain events with collinear VR track jets.

their tendency to favour more boosted jets. Although this effect is expected, the high mass resonances are also the most affected by reducing acceptance given that very few events are expected to reach such high m_{hh}.

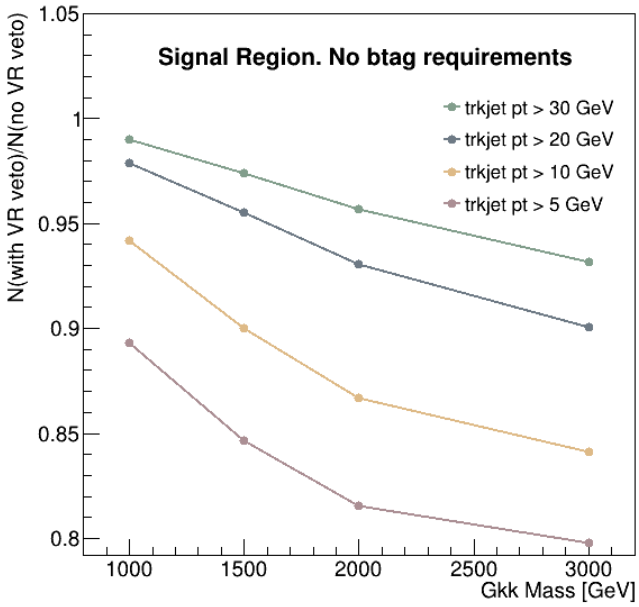


Figure 6.3: Impact of the collinear VR veto on various spin-2 signal mass points. The y axis shows the fraction of events in the signal region (before any b-tagging selection) that pass the veto. Such impact is shown for various minimum track jet p_{T} values to trigger a veto.

6.1.3 Potential fix: Increase Minimum p_{T} to Trigger Veto

A method to recover some of the lost collinear VR track-jet events was tested. This method involves raising the minimum p_{T} threshold of track jets considered for the veto (nominally this is $p_{\mathrm{T}} > 5$ GeV), and is motivated by the large difference in p_{T} observed between the collinear jets causing the veto. Figure 6.3 shows the effect of raising such threshold to 10, 20, or 30 GeV, reducing the loss in signal acceptance from 10% - 20% down to around 1% - 5% throughout the mass range of the analysis.

Validation of b-tagging on Recovered Events

Although raising the p_{T} threshold to trigger a veto recovers a significant fraction of vetoed events, it is not immediately clear that the b-tagging performance was not affected by the presence of the overlap. Small test samples of simulated G^*_{kk} signal data were created to test the b-tagging performance on recovered events. These test samples were identical to the nominal ones with the exception of the minimum p_{T} required for a track jet to be reconstructed, which was set to 10, 20 or 50 GeV in this

case [4]. The new samples allow us to evaluate the behaviour of b-tagging algorithms on recovered jets, and compare it with the behaviour on non-collinear jets or on jets that are collinear where the soft jet was ignored (the case described in section 6.1.3).

Figures 6.4 and 6.5 show these b-tagging score comparisons. Figure 6.4a shows the normalized MV2c10 score [5] distributions of all jets in a 3 TeV spin-2 signal taken from the nominal sample (minimum track-jet p_{T} of 5 GeV) and the varied signal samples. The larger peak in high b-tagging score suggests that most of the recovered jets are b-tagged. The plot in figure 6.4b shows the MV2c10 score distribution of the leading Higgs boson candidate's leading track jet from the same samples. By comparing the same track jet's b-tagging score for the different samples it is possible to more closely compare the algorithm's behaviour, with good agreement found across the variations, especially in the high score range where jets are considered to be b-tagged. The lower panels on these two plots show the ratio of each variation to the nominal case [6].

To evaluate any differences in b-tagging scores between jets recovered by ignoring soft track jets (method suggested in 6.1.3) as opposed to ones recovered by rising the minimal p_{T} for track jet reconstruction (new samples), we compare recovered jets from a test signal sample [7] to the same jets in the vetoed event from the nominal sample. Figure 6.5 shows the MV2c10 score distribution of recovered jets, shown with the equivalent jet's score taken from vetoed events in the nominal case (6.5a), and with the score of jets that were never collinear (6.5b). In the latter case, the never-collinear jets are the leading track jet from the leading Higgs boson candidate on events vetoed by collinear jets being present in the subleading Higgs boson candidate [8].

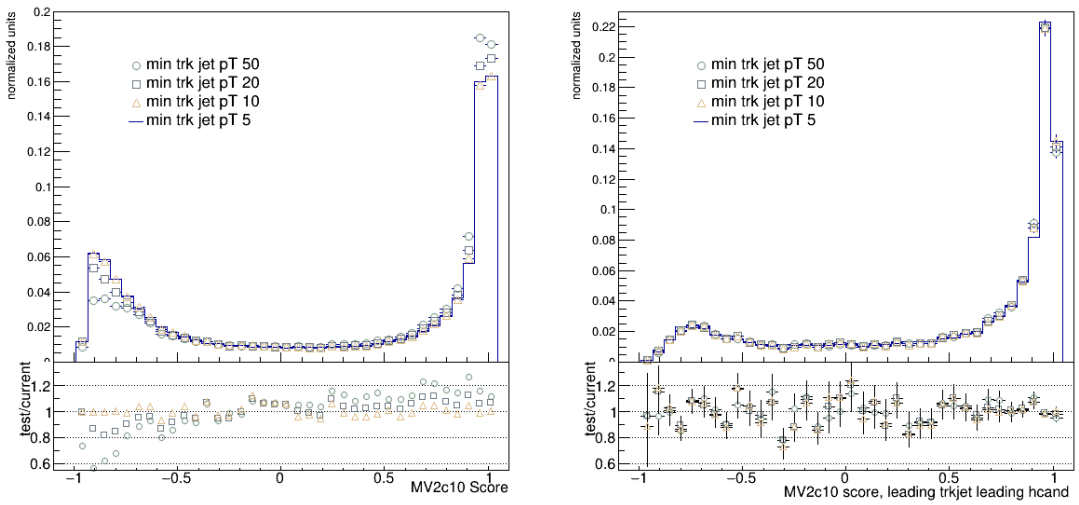
[4] Note that in the tests in section 6.1.3 `VR` track jets below the threshold were ignored when checking for collinear jets, while in the new samples they would be clustered *into* the higher p_{T} jet.

[5] MV2c10 score here, or DL1 score elsewhere, refers to the output of the b-tagging algorithm when applied to a jet.

[6] The nominal case being the sample with the veto threshold of 5 GeV, plotted as a line.

[7] The test sample used was the one with minimum track jet p_{T} of 50 GeV as it provides the largest sample of recovered jets.

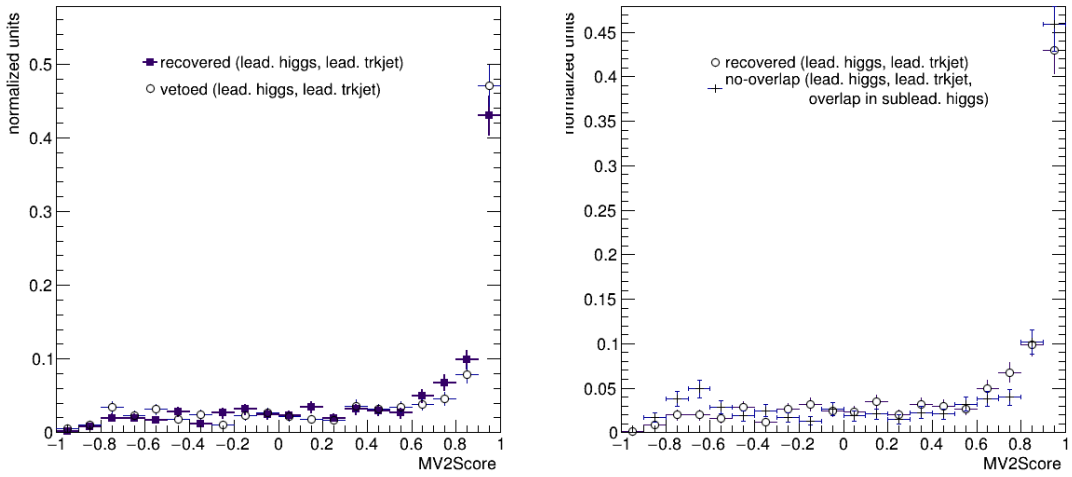
[8] Here, events are required to have been vetoed by overlap in the large-R jet that is *not* being inspected in order to to select events with similar topologies to the ones with recovered jets.



(a) **(b)**

Figure 6.4: Plots of the MV2c10 score for various sets of jets in a 3 TeV spin-2 signal sample. The (a) MV2c10 score for all jets and (b) the leading track jet associated to the leading Higgs boson candidate, from various samples with increasing minimum track-jet p_{T} requirement are plotted. The lower panels show the ratio of each variation to the nominal case, where the increased acceptance of b-tagged jets (a), and the small impact of varying the p_{T} threshold for vetoing (b) can be seen.

Even though the low number of events in the test derivations lead to large statistical uncertainties in some of the plots above, b-tagging behaviour is found to be sufficiently close on jets recovered both by ignoring soft collinear track jets or by rising the minimum p_{T} threshold of track jets. This indicates that the jets from recovered events could safely be added back to the dataset.



(a) **(b)**

Figure 6.5: MV2c10 score of the jets recovered by raising the minimum track-jet p_{T} requirement from 5 GeV to 50 GeV, shown with (a) the equivalent score taken from the sample of vetoed events, and (b) with the score of jets that were never collinear.

6.1.4 Performance of Variable Radius (VR) Track Jets

From the results in reference [80], which are shown in figure 3.5, it is expected that by using VR track jets the efficiency of finding both b-jets in a Higgs boson candidate would improve, which in turn would raise the acceptance of events to the 4b region. Unexpectedly, the opposite effect was observed.

Significant drops in signal acceptance were observed in the 4b and 3b regions (see figure 6.6a,6.6b [9]) after the first implementation of the analysis using variable-radius track jets compared to fixed radius (FR) track jets, even before applying the veto detailed in the previous section 6.1.2. Acceptance is increased for VR in the 2b region (figure 6.6c), which suggests that events migrated from the higher b-tag regions to the lower ones.

In collaboration with the flavour tagging group, the event migration was traced to slightly suboptimal b-tagging performance on VR track jets, as a consequence of the training being optimized using a different type of jet, which can be seen in the signal acceptance plot in figure 6.7. The impact on acceptance of this per-jet

[9] These plots show all available b-tagging operating points to more thoroughly verify the behaviour of the algorithm.

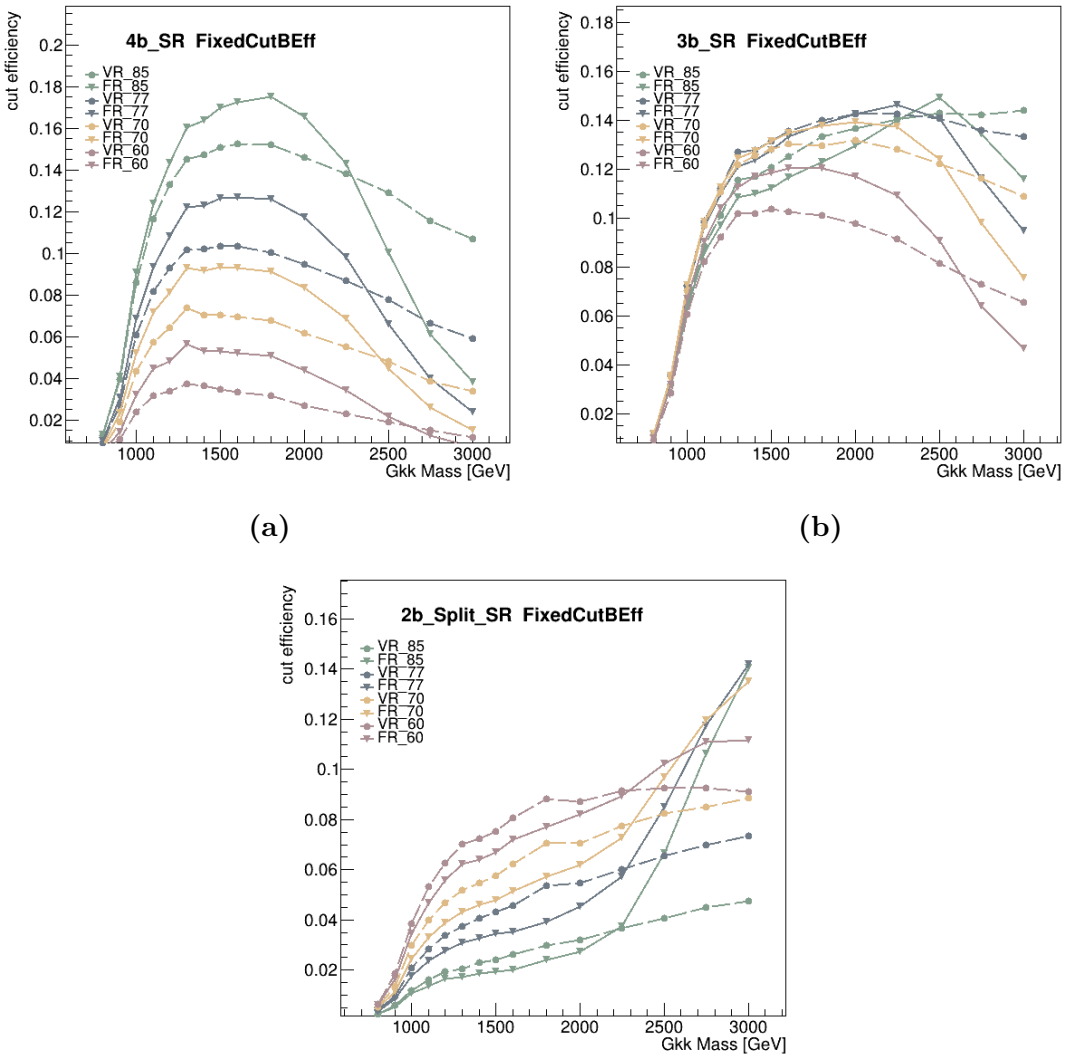


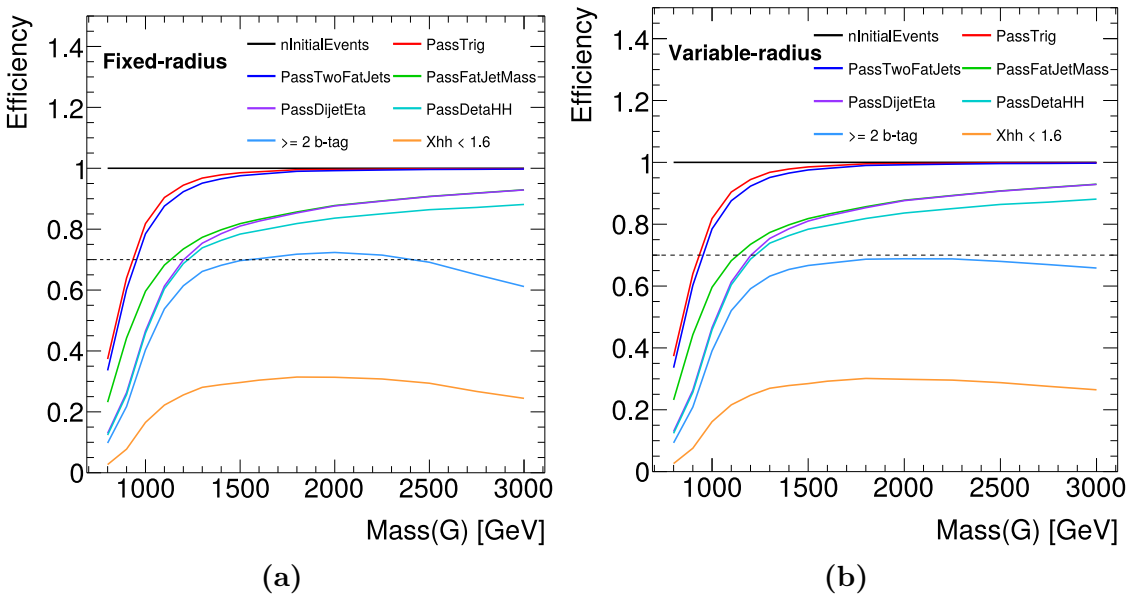
Figure 6.6: Selection efficiency per signal region, for a range of spin-2 signal mass points, shown for `FR` track jets (solid lines, triangular markers) and `VR` track jets (dashed lines and circle markers). All available b-tagging working points are shown (corresponding to the number after the underscore). All analysis selections have been applied.

underperformance increases geometrically with the number of b-tags required by the analysis, making the $hh \rightarrow 4b$ analysis specially vulnerable.

Input from the boosted $hh \rightarrow 4b$ and other analyses prompted the re-training of b-tagging algorithms to be optimal on `VR` track jets. While the results in the main analysis chapters, 4 and 5, contain the improved training, all the studies in this chapter 6 do not include this updated training [10] since they were done

[10]Except section 6.2.1 dedicated to the difference between old and new training.

prior to the labour and resource intensive retraining campaign. The final result is expected to use the updated training as well.



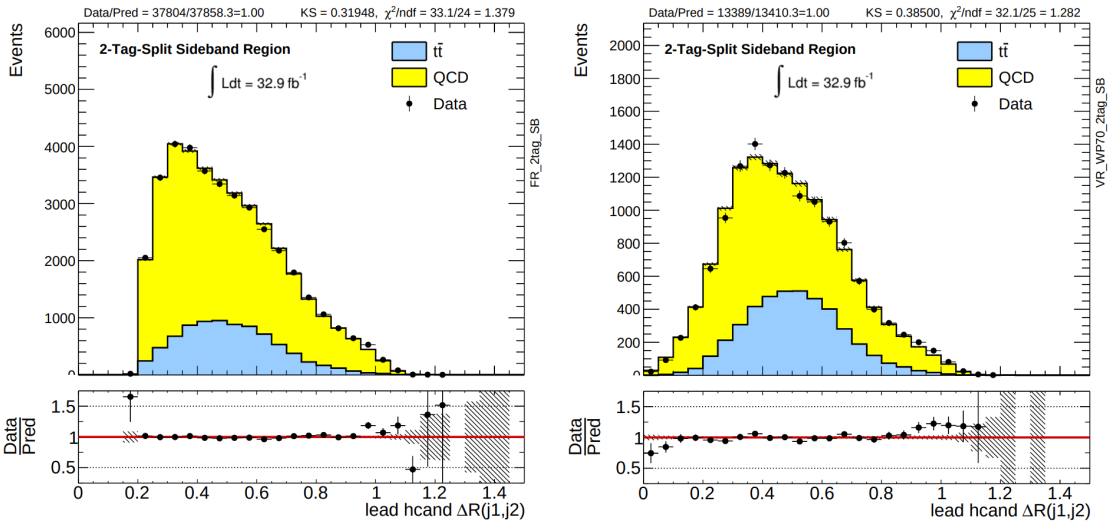
(a) (b)

Figure 6.7: Efficiency of each sequential selection in the analysis for a range of mass points of G^*_{kk} resonance signal samples, made using (a) fixed-radius and (b) variable-radius track jets.

Comparison of Variable Radius (VR) Analysis to Fixed Radius FR

To test the performance of VR in this di-Higgs search, the analysis was re-run using FR track jets. One interesting difference can be observed in the distribution of the separation ΔR between the two track jets associated to a Higgs boson candidate in figure 6.8, which shows this quantity calculated in the CR of both versions. While FR track jets force events with $\Delta R < 0.2$ to be dropped since these would be overlapping, VR jets' radii can shrink allowing the recovery of events inaccessible to fixed-radius algorithms. Note that while these plots both use the same amount of luminosity and are the same b-tag region ($2b$), the different track-jet clustering allows for events to migrate from one b-tag region to another which yields different numbers of events for each case.

The analysis selection efficiency of the sequential filters, computed on graviton signal samples, is shown in figure 6.7. The trigger requirement (leading large-R jet $p_{\mathrm{T}} > 450$ GeV) limits acceptance below 1 TeV in G_{KK} mass. Up to the b-tag



(a) **(b)**

Figure 6.8: Angular separation ΔR between the two track jets in the leading Higgs boson candidate in a signal-free side-band region, showing the difference between Higgs boson candidates with `FR` (a) and `VR` (b) track jets. Data in this region are plotted over the stacked estimates of $t\bar{t}$ and QCD multijet backgrounds. Figure 6.8a used fixed radius and 6.8b variable radius track jets.

requirement (> 2 b-tag in light blue), both plots have identical efficiency. After this selection, `VR` track jets help keep flat efficiency throughout the mass range, while with `FR`, the efficiency drops at higher mass. However, in the range between 1.2 TeV and 2.5 TeV, `FR` can be seen to be slightly more efficient than `VR`, as mentioned above at the introduction to this section.

`VR` Veto Impact on the Background

The studies in this section were primarily focused on the effect the collinear `VR` track-jet veto would have on the signal acceptance due to the very limited number of events expected from the heavy resonances in this search. However, the effect on the background also plays a key role. The veto was observed to have similar impact on the background to that seen on the signal (table B.1).

6.2 Impact of Variable Radius (VR) Track Jets on Sensitivity

While it is not easy to see the impact of VR track jets in the signal acceptance in figure 6.7, the effect is visible in the expected sensitivity of the search. Figure 6.9 shows the expected 95% confidence level (CL) upper limit on the production cross-section times branching ratio (see section 5.1) of $G^*_{kk} \to hh \to b\bar{b}b\bar{b}$ from the analyses using FR and VR using the 2016 ATLAS dataset, accounting for statistical uncertainty only. Despite the inefficiency mentioned above in section 6.1.4, VR nearly matches the performance of FR performance in the range where it is less efficient (as seen in figure 6.6), and performs better for higher mass points. The good performance of VR in the low mass range, where FR proved more efficient, was found to be due to VR more effectively rejecting background in this region, while the improved sensitivity in high masses was found to come from the increased acceptance of VR for highly boosted Higgs boson candidates. These results were obtained *prior* to the re-training of the b-tagging algorithms.

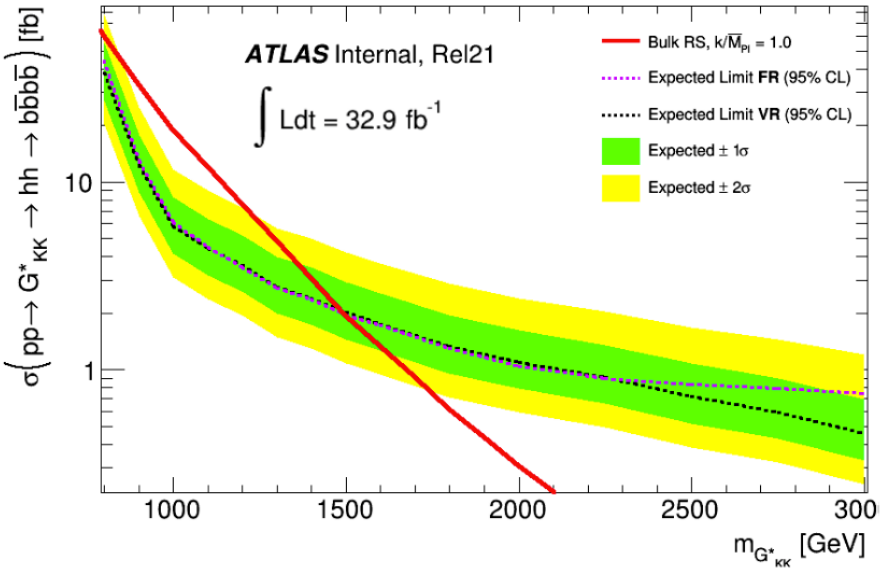


Figure 6.9: Expected 95% CL upper limits on the production cross section times branching ratio of graviton making use of fixed and variable radius track jets. The uncertainties in the limits include statistical uncertainties only and do not account for systematic uncertainties. The red line represents the expected cross section from theoretical predictions of the graviton model. This limit does *not* make use of the updated b-tagging training for VR track jets (see section 6.2.1).

6.2.1 Impact of VR Specific b-tagging Training

The improvement on the search sensitivity from dedicated training of the b-tagging algorithms on VR track jets, is shown in figure 6.10. The expected limit on the production cross section times branching ratio of the heavy scalar model is shown for two versions of the analysis. Both versions use VR track jets, but the updated version makes use of the re-trained b-tagging algorithms. Better sensitivity is seen for masses below 2.5 TeV, the region where VR was seen to underperform as compared to FR. [11].

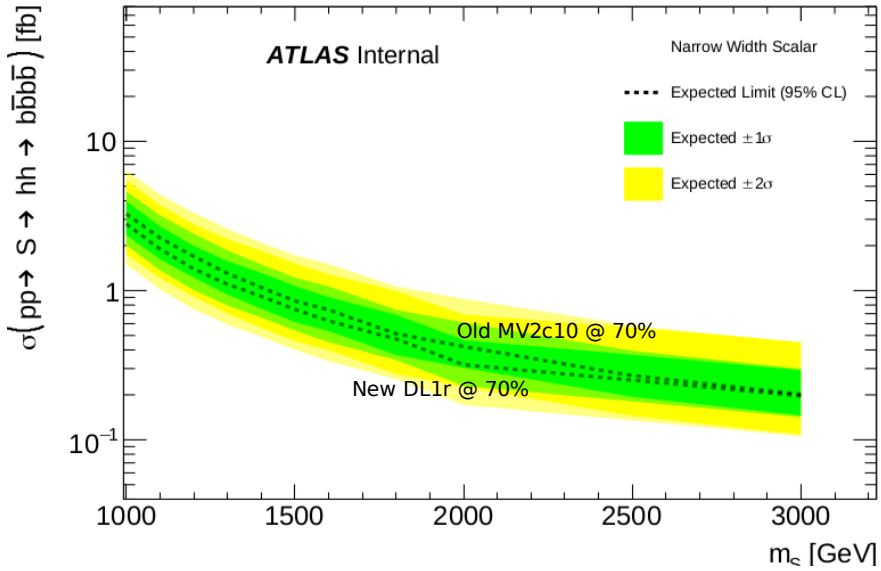


Figure 6.10: Expected 95% CL upper limits on the production cross section times branching ratio of the heavy scalar model with the old (MV2c10) and new (DL1r) b-taggers. The uncertainties in the limits include statistical uncertainties only and do not account for systematic uncertainties.

[11] The MV2 algorithm was outdated in favour of the DL1r tagger which has better performance. MV2 did not get the VR-specific re-training, so this comparison is the closest evidence of the impact of the re-training.

7
Future Prospects

Contents

7.1 Di-Higgs to $b\bar{b}b\bar{b}$ HL-LHC Phenomenology Study . . . 104
7.1.1 Analysis Overview . . . 105
7.1.2 Neural-Network Analysis . . . 105
7.1.3 Feature Importance . . . 107
7.1.4 Network correlation plots . . . 112
7.1.5 Analysis Results . . . 115
7.1.6 Phenomenology Study Final Remarks . . . 116

As more data are collected by the LHC experiments, the di-Higgs channel's sensitivity will expand from heavy resonances of BSM models, to a viable search for SM non-resonant Higgs boson pair production. Even further into the future, with new colliders, di-Higgs searches will in all likelihood, move from being a search for exotic particles to a precision measurement of the SM.

A recent ATLAS result combines di-Higgs and single-Higgs searches, with either the partial or total run 2 dataset depending on the analysis, narrowing the constraints on κ_λ to $-2.3 < \kappa_\lambda < 10.3$ [1], with $hh \rightarrow 4b$ being one of the most constraining channels close to $\kappa_\lambda = 1$ [24]. Figure 7.1 shows the expected

[1] As defined in equation 1.4, κ_λ is the ratio of a value of the triple-Higgs coupling to the SM expected value

constraints on the κ_λ spectrum to be set by each of single- and double-Higgs-boson channels in this combination, as measured by the profile-likelihood test statistic defined in [100], here labeled Λ. In general, the double-Higgs channels have better sensitivity, with the $hh \to b\bar{b}\gamma\gamma$ setting the tightest constraint [2], while $hh \to 4b$ sets the third best constraints. Single-Higgs-boson channels, such as $h \to \gamma\gamma$ or $h \to ZZ^*$, contribute significantly on negative κ_λ values. The most recent measurement from CMS comes from combining several di-Higgs searches and constrains the self-coupling to $-11.8 < \kappa_\lambda < 18.8$ [40].

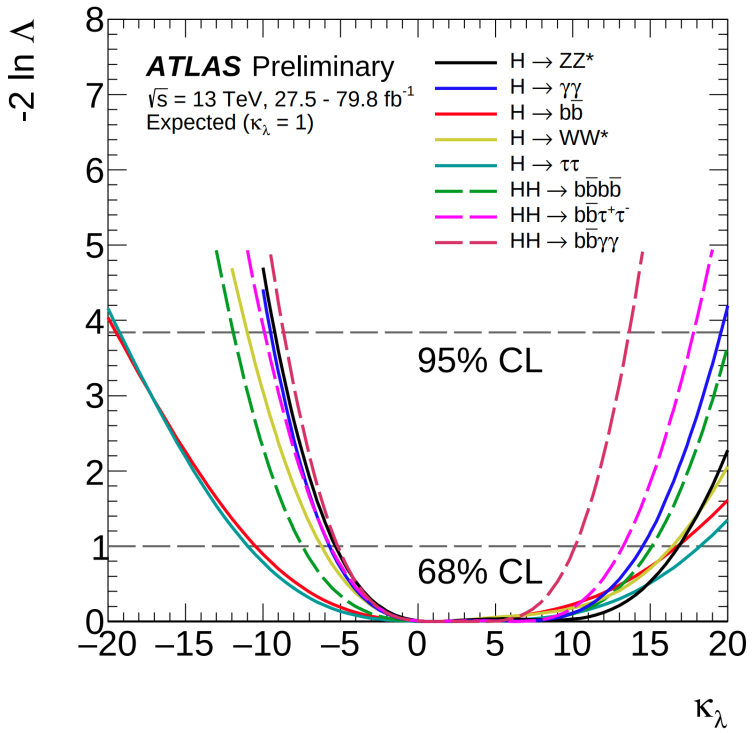


Figure 7.1: Expected value of $-2\ln\Lambda$ as a function of κ_λ (all other couplings set to their SM value) obtained in the $\kappa_\lambda = 1$ hypothesis for the single-and double-Higgs-boson decay modes[24].

Looking further into the future, ATLAS predicts sensitivity of 3σ when searching for the process, and around 50 % precision on constraints of κ_λ with the full dataset of the High-Luminosity LHC (HL-LHC) [102]. Current searches have focused on the search for Higgs pair production, and use the resulting cross section measurement

[2]Note the luminosity is not equal for all channels as full run 2 results for these channels are still being produced. See reference [24] for details.

to constrain κ_λ [31–37]. ATLAS also only considers the resolved channel in the combination mentioned above, ignoring any boosted or semi-boosted/intermediate [3] topology. However, these approaches are not necessarily optimal to constrain the Higgs boson self-coupling. How to optimize the search to best constrain κ_λ, and the impact of new approaches this analysis could implement in the future, were studied in a phenomenology paper that was part of the work for this dissertation [1].

7.1 Di-Higgs to $b\bar{b}b\bar{b}$ HL-LHC Phenomenology Study

Parallel to the ATLAS $hh \to 4b$ study presented in this document, I also worked on a phenomenology study of the process in conditions similar to the HL-LHC [1]. This work aimed to test various analysis strategies to search for $hh \to 4b$ in the HL-LHC era with specific focus on exploring the optimal path to constrain κ_λ. The paper presents an analysis strategy based on separating not only boosted and resolved event topologies (as presented in the ATLAS analysis in chapters 4-5), but also adding a third *intermediate* or *semi-boosted* category, where only one of the Higgs boson candidates is reconstructed as a large-R jet and the other by two smaller jets.

The study introduces a machine-learning based approach to separate signal and the different background events, shows potential sensitivity gains from the introduction of the intermediate channel, and highlights important experimental and theoretical factors for constraining κ_λ (notably low p_{T} jet trigger thresholds and constraints on the SM top Yukawa coupling y_t). It finds modest improvement in sensitivity from a simple neural-network analysis which suggests more gains could be obtained with more optimized approaches, and it finds that current knowledge of y_t can have up to a 20% impact on κ_λ constraints at the HL-LHC.

A brief summary of reference [1], with focus on my contributions to the study, is presented in this section.

[3] Semi-boosted or intermediate here refers to events where one Higgs boson can be reconstructed as a large jet, and the other as two smaller jets.

7.1.1 Analysis Overview

Various Monte Carlo (MC) simulated data sets were generated to model the signal and backgrounds: signal samples of $gg \to h \to hh$ with varied Higgs boson self-coupling κ_λ and top-quark Yukawa coupling [4] κ_t, irreducible 4 b-jet background ($4b$), reducible 2-jet plus 2-b-jet background ($2b2j$), backgrounds with top-quarks ($t\bar{t}$, $t\bar{t}b\bar{b}$), background single-Higgs plus heavy-flavour quarks ($t\bar{t}h$, $b\bar{b}h$), and vector-boson background processes (ZZ, Zh, Wh).

The datasets were then separated into three regimes (resolved, intermediate, and boosted) depending on the number of large-R jets present in each event (0, 1, or $\geq$ 2). The data are then selected using the *baseline analysis* selection, summarized in table 7.1. The selection was loosely based on the ATLAS $hh \to 4b$ selection in [31].

A *deep neural network* (DNN) trained to discriminate signal events from the $t\bar{t}$ and multijet backgrounds is used to include a final selection requirement additional to the baseline selection. The two approaches are then compared. Developing, training and implementing the *neural-network analysis* was part of my DPhil work, and is presented in the following sections.

7.1.2 Neural-Network Analysis

A single network architecture, depicted in figure 7.2, based on the Keras library [103] was used for the resolved, intermediate, and boosted regimes. These are feed-forward networks [104–106] with two internal ('hidden') layers each with 200 nodes, densely connected to each other and to the 20 input and 3 output nodes. This network classifies events as *signal*-like, *ttbar*-like, or *multijet*-like.

The 20 variables used as input to the networks are summarised in Table 7.2.

A special *high-statistics* signal dataset [5] was used to train the networks, together with half the multijet and $t\bar{t}$ MC events. A fraction of these training datasets (20%) was reserved for validation, and are not used for actual training. The

[4] Similarly to κ_λ, the top-quark Yukawa coupling is studied as the ratio to the Standard Model expectation of the coupling $\kappa_t = y_t / y_t^{\mathrm{SM}}$.

[5] The signal samples used in the analysis have the Higgs bosons decaying inclusively, while the high-statistics sample has each Higgs boson decaying exclusively to $b\bar{b}$.

Observable	Preselection		
Large jet j_L	$R = 1.0, p_{\mathrm{T}} > 250$ GeV, $\lvert\eta\rvert < 2.0$		
Small jet j_S	$R = 0.4, p_{\mathrm{T}} > 40$ GeV, $\lvert\eta\rvert < 2.5$		
Track jet j_T	$R = 0.2, p_{\mathrm{T}} > 20$ GeV, $\lvert\eta\rvert < 2.5$		
$j_T \in j_L$	$\Delta R(j_T, j_L) < 1.0$		
	Resolved	Intermediate	Boosted
$N(j_L)$	$= 0$	$= 1$	$= 2$
$N(j_S)$	≥ 4	≥ 2	≥ 0
h_1^{cand}	$j_S^{(i)}$ pair	j_L	$j_L^{(1)}$
h_2^{cand}	$j_S^{(i)}$ pair	$j_S^{(i)}$ pair, $\Delta R(j_S^{(i)}, j_L) > 1.2$	$j_L^{(2)}$
ΔR_{jj}	m_{hh}-dependent cut	—	—
	Signal region		
$j_T \in h_1^{\mathrm{cand}}$	—	≥ 2	≥ 2
$j_T \in h_2^{\mathrm{cand}}$	—	—	≥ 2
b-tagging	Two b-tags for each h_i^{cand}		
$\lvert\Delta\eta(h_1, h_2)\rvert$	< 1.5		
$E_{\mathrm{T}}^{\mathrm{miss}}$	< 150 GeV		
$p_{\mathrm{T}}^{\ell}, \lvert\eta_{\ell}\rvert$	> 10 GeV, < 2.5		
N_{ℓ}	$= 0$		
$p_{\mathrm{signal}}^{\mathrm{DNN}}$	> 0.75 (*neural-network analysis* only)		
	Resolved	Intermediate	Boosted
$m(h_1)$ [GeV]	[90, 140]	[90, 140]	[90, 140]
$m(h_2)$ [GeV]	[90, 140]	[90, 140]	[90, 140]

Table 7.1: Overview of event selection for the *baseline analysis* in the resolved, intermediate and boosted categories. The requirements above the upper double rule are the same as the preselection used for the *neural-network analysis* training. The requirements below the upper double rule are the signal region requirements.

Reconstructed objects	Variables used for training
Higgs boson candidates $h_{1,2}^{\mathrm{cand}}$	$(p_{\mathrm{T}}, \eta, \phi, m)$
Subjets $\in h_{1,2}^{\mathrm{cand}}$	$\Delta R(j_1, j_2)$
Missing transverse momentum	$E_{\mathrm{T}}^{\mathrm{miss}}, \phi(\mathbf{p}_{\mathrm{T}}^{\mathrm{miss}})$
Leptons	N_e, N_μ
b-tagging	Boolean for $j_i \in h_{1,2}^{\mathrm{cand}}$
Di-Higgs system	$p_{\mathrm{T}}^{hh}, m_{hh}$

Table 7.2: Input variables used to train the neural network.

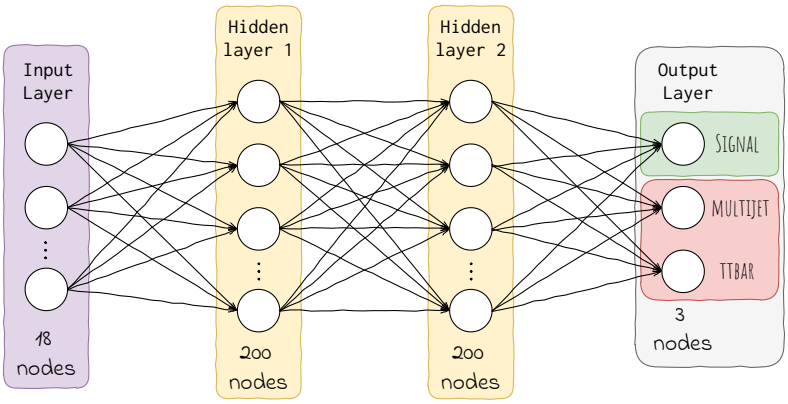


Figure 7.2: Architecture of the neural network.

networks use dropout[6] [107] of 30% of the nodes in each internal layer to avoid over-training. The learning rate, training batch size and dropout rate were optimised using a random search method.

Figure 7.3 shows the signal $p_{\text{signal}}^{\text{DNN}}$ score distributions, for the networks trained with signals with $\kappa_\lambda = 1$, and $\kappa_\lambda = 5$, in the three regimes. Signal–background discrimination is somewhat improved across the categories, suggesting that the networks capture kinematic information beyond that which is used in the *baseline analysis*. However, this depends on the value of κ_λ. For example, the upper-left plot shows that the DNN trained on $\kappa_\lambda = 1$ adds substantial discrimination power for a $\kappa_\lambda = 1$ signal, but not for a $\kappa_\lambda = 5$ signal. The *neural-network analysis* adds the same requirement of $p_{\text{signal}}^{\text{DNN}} > 0.75$ across all regimes for simplicity.

7.1.3 Feature Importance

To quantify *how much impact* on the signal score each variable (or *model feature*) fed to the network had, the SHapley Additive exPlanations (SHAP) [108] framework was used. Interpretation of neural networks is a complex and rapidly developing field in computational sciences, and the SHAP framework is one of the few tools available to evaluate *what* a network has learned. This framework combines several feature importance tests available for machine learning models in the literature into a single value, which are detailed in Ref. [108].

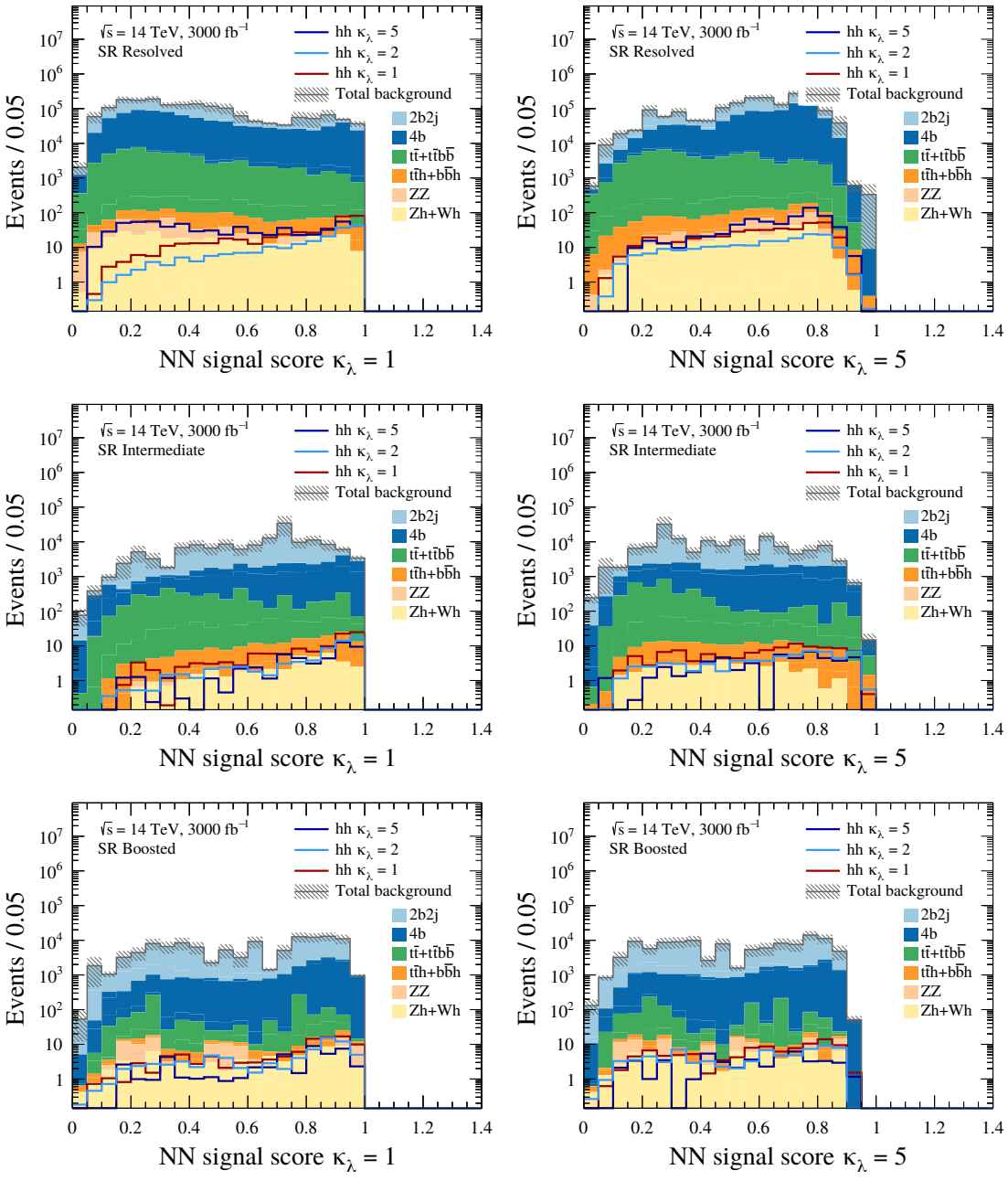
[6]This means 30% of internal nodes are randomly masked during each training iteration.

The SHAP framework requires that the trained model is applied to a specific set of events, and it only evaluates the importance of each feature for this given set of events. A new subset of events was constructed with different fractions of signal and background to reflect the importance of distinguishing them from each other. This set contains half signal events and half background events. The large proportion of signal is chosen to prioritise its discrimination above background, rather than discrimination between different sources of background. The background sample used for this evaluation is composed of 80% $2b2j$, 15% $4b$ and 5% $t\bar{t}$ events, to approximate the background composition of the resolved, intermediate and boosted analyses.

Figure 7.4 shows the 20 input variables of the neural networks, ranked by their impact on the final signal score for the $\kappa_\lambda = 5$ training. This impact is measured by their magnitude of their SHAP values averaged over the whole data set given to the framework, plotted on the x axis. Each point plotted per a row corresponds to one event fed to the framework, and its location along the x axis represents what impact that variable had on the signal score of the event. Points with a negative SHAP value lower the signal score while positive ones increase it. The relative magnitude represents how much the value of that specific variable changes the signal score compared to all other variables in all events. The absolute scale of the SHAP value is arbitrary in these plots. The colour scale indicates the value of the feature on the specific event e.g. a blue dot on the b-tag(h_1^{cand}, j_1) indicates the leading subjet associated to the leading Higgs boson candidate was not b-tagged. Meanwhile, a pink dot in the m_{hh} row indicates that event had high di-Higgs invariant mass for which the SHAP value is plotted.

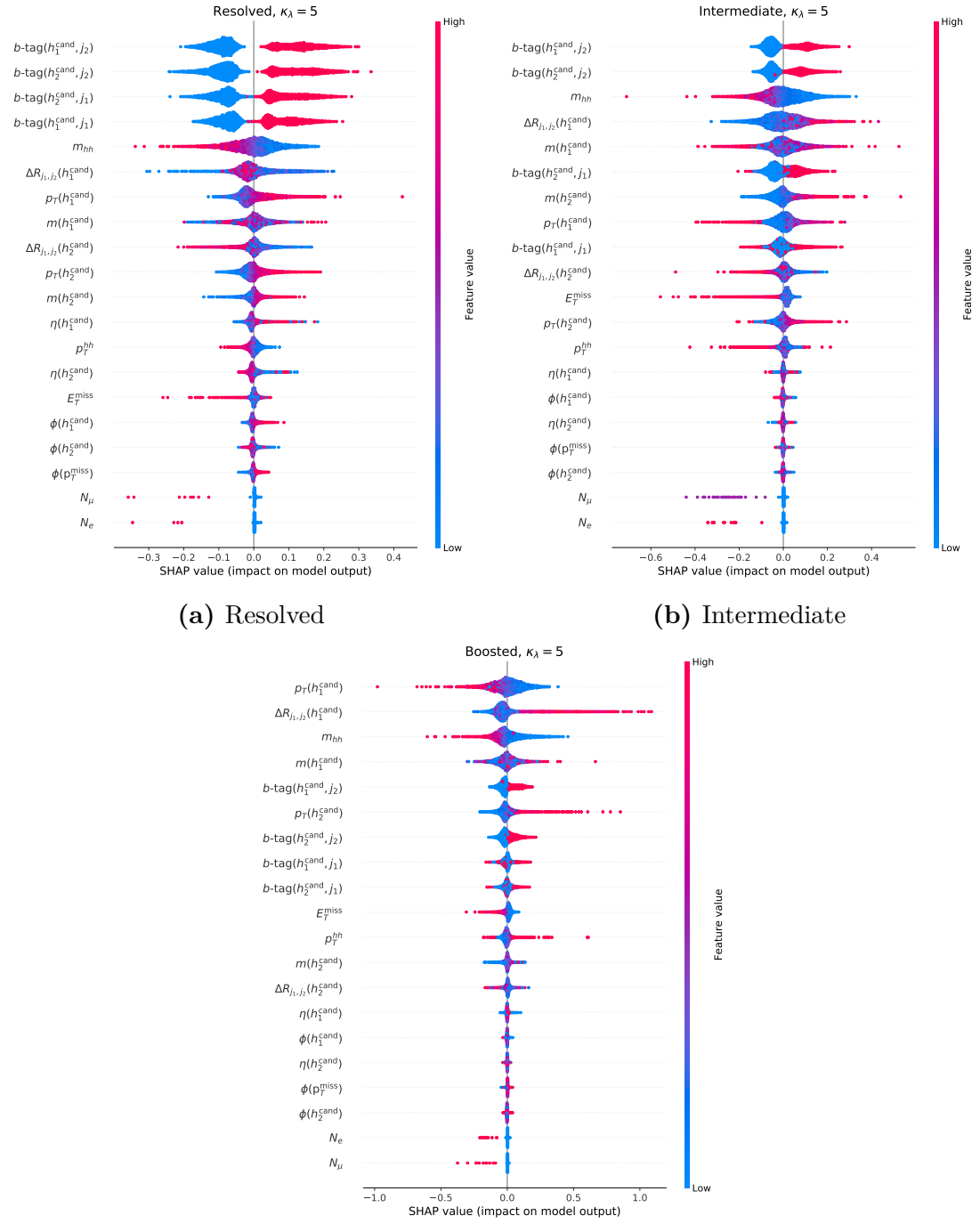
The b-tagging state of (sub-)jets are among the most important features in the resolved and intermediate categories. The boosted analysis may be less sensitive to due to lower b-tagging efficiencies at high p_{T}, which could potentially be improved by future work, for example the novel b-tagging techniques in reference [109]. Angular and mass variables are stronger discriminants against multijet processes in this regime. The opening angles between these (sub-)jets and the invariant mass of

the di-Higgs system carry a large amount of information in all three categories. Variables sensitive to semi-leptonic b-hadron decays, such as the number of leptons or missing transverse momentum, were effective at ruling out background (large negative SHAP value for high feature value) and less so at identifying signal (small positive SHAP values for any feature value).



(a) DNN trained on $\kappa_\lambda = 1$ **(b)** DNN trained on $\kappa_\lambda = 5$

Figure 7.3: Neural network score distributions $p_{\text{signal}}^{\text{DNN}}$ of benchmark signals (solid lines) and background processes (filled stacked) displayed in the legend. All event selection criteria of the *neural-network analysis* except the $p_{\text{signal}}^{\text{DNN}} > 0.75$ requirement are imposed. The neural networks are trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signals. These are displayed for (upper) resolved, (middle) intermediate and (lower) boosted regimes. The plots are normalised to $\mathcal{L} = 3000$ fb^{-1}.



(a) Resolved

(b) Intermediate

(c) Boosted

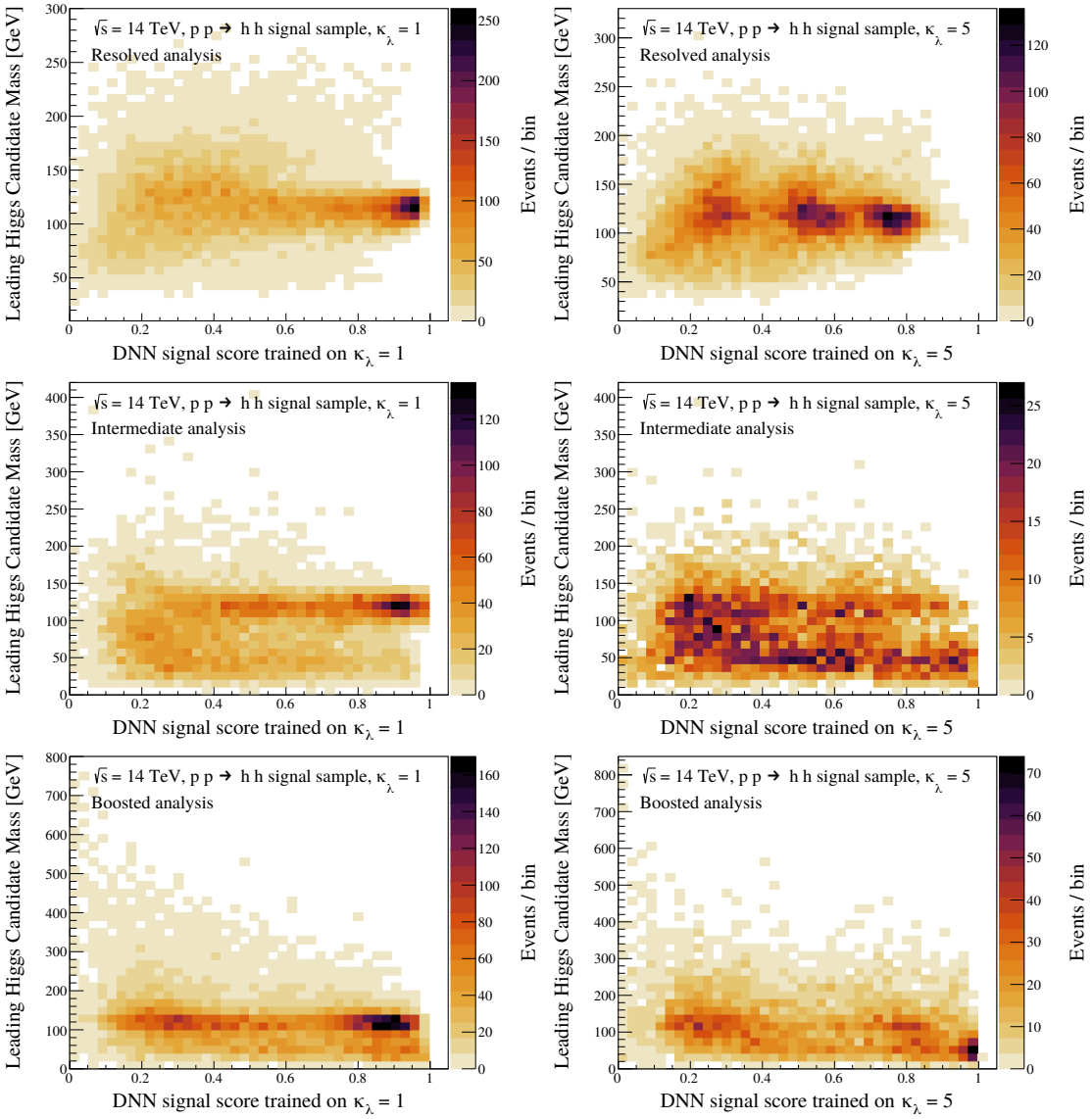
Figure 7.4: Feature importance for the networks trained on $\kappa_\lambda = 5$ signals for the (a) resolved, (b) intermediate and (c) boosted categories. The model features are ranked by their average absolute SHAP value, which is plotted on the x axis. The colour scale indicates the value of the feature on the specific event for which the SHAP value is plotted.

7.1.4 Network correlation plots

To validate the architecture and test that the training of the neural-network analysis has the expected behaviour, the networks' performance was tested on a statistically independent set of signal and background events. Figure 7.5 shows the leading Higgs boson candidate mass as a function of the signal score trained on $\kappa_\lambda = 1$ and $\kappa_\lambda = 5$ signals when tested on the corresponding $pp \to hh$ signal that exclusively decays to b-quarks. For the resolved category, a prominent peak appears at high signal score trained on $\kappa_\lambda = 1$ and around $m(h_1^{\text{cand}}) \sim 125$ GeV. There is a tail of events at lower signal score where it is difficult to kinematically distinguish the signal from background.

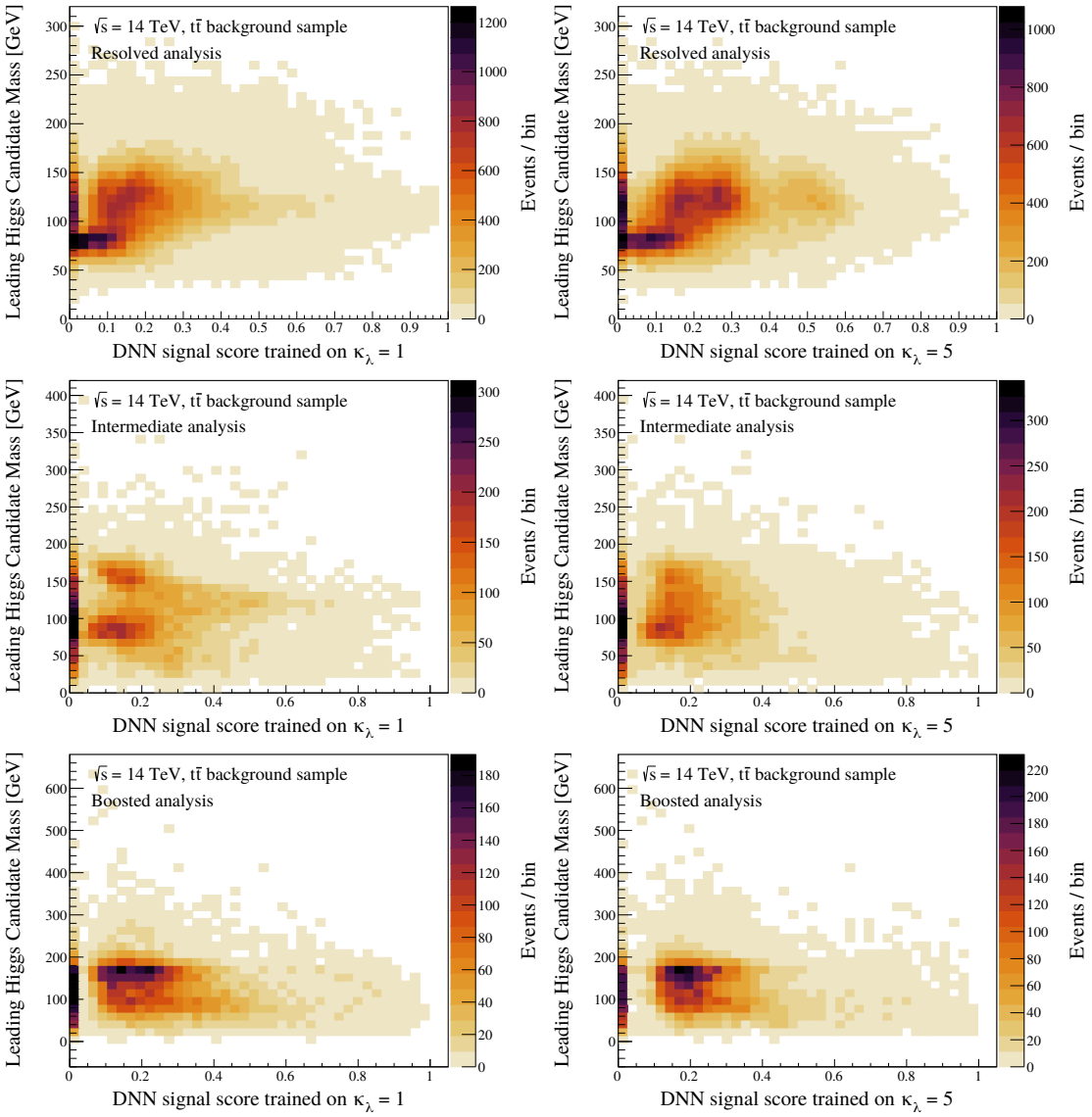
An interesting feature is observed in Figure 7.5 depicting the $\kappa_\lambda = 5$ DNN score computed in the $\kappa_\lambda = 5$ signal sample versus the leading Higgs boson candidate mass. Three distinct peaks can be seen. We find each peak corresponds to events with different numbers of b-tags. The most prominent peak with the highest signal score corresponds to events with three or four b-tags. Meanwhile, the two peaks with signal scores around 0.3 and 0.5 correspond to events with one and two b-tags respectively.

Figures 7.6 show similar correlation plots, now testing on a $t\bar{t}$ sample. These feature low values of signal scores, indicating that the networks are effective at classifying this background. More structure is seen in these two-dimensional plots due to the inclusive decay of the b-quarks in this sample. Several of these plots feature two distinct peaks, corresponding to semi-leptonic and hadronic b-decays. Hadronic decays correspond to the peak with higher signal score, since these are more similar to the $hh \to 4b$ signal. Furthermore, the plots for the intermediate and boosted categories feature two distinct peaks in the leading Higgs mass around ~ 80 GeV and ~ 170 GeV. This suggests that the large jet is capturing the boosted decay products of the W boson and top quarks.



(a) DNN trained on $\kappa_\lambda = 1$ **(b)** DNN trained on $\kappa_\lambda = 5$

Figure 7.5: The leading Higgs boson candidate mass vs neural network scores trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal sample for the (upper) resolved, (middle) intermediate, and (lower) boosted analyses. The test samples used to make these distributions are an independent set of (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal events.



(a) DNN trained on $\kappa_\lambda = 1$ **(b)** DNN trained on $\kappa_\lambda = 5$

Figure 7.6: The leading Higgs boson candidate mass vs neural network scores trained on (a) $\kappa_\lambda = 1$ and (b) $\kappa_\lambda = 5$ signal sample for the (upper) resolved, (middle) intermediate, and (lower) boosted analyses. The test samples used to make these distributions are an independent set of $t\bar{t}$ events.

7.1.5 Analysis Results

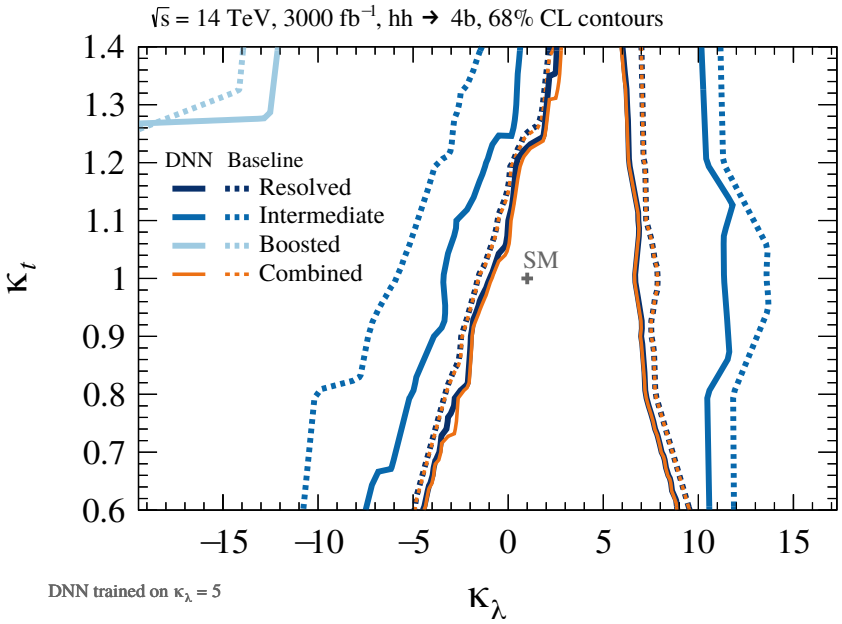


Figure 7.7: Summary of 68% CL ($\chi^2 = 1$) contours in the κ_t vs κ_λ plane. These are displayed for the resolved (dark blue), intermediate (medium blue), boosted (light blue) categories and their combination (orange) for the *baseline analysis* (dashed) and *neural-network analysis* (DNN) trained on $\kappa_\lambda = 5$ (solid). A luminosity of 3000 fb^{-1} is assumed along with systematic uncertainties of 0.3%, 1% and 5% for the resolved, intermediate and boosted categories, respectively. The cross indicates the SM prediction.

Figure 7.7 overlays two-dimensional 68% CL contours in κ_t and κ_λ (defined by ($\chi^2 = 1$)) to compare the baseline analysis (dashed lines) with neural-network analysis trained on $\kappa_\lambda = 5$ (solid lines) using a χ^2 test with:

$$\chi^2 = \sum_i \left[\frac{(S - S_{\mathrm{SM}})^2}{S + B + (\zeta_b B)^2 + (\zeta_s S)^2} \right]_i . \tag{7.1}$$

where B is the total background rate, S_{SM} is the SM signal rate, $\zeta_s(\zeta_b)$ corresponds to the uncertainty on the signal(background), and S is the signal rate we wish to distinguish from the SM counterpart. This figure of merit is designed to capture how effective the analysis is at detecting deviations of λ_{hhh} from its Standard Model prediction. These are evaluated in each m_{hh} bin. Statistical and systematic uncertainties on the signal are neglected, as background uncertainties dominate for $S \ll B$ regimes. For simplicity, the statistical combination is performed by

summing the χ^2 values assuming constant and uncorrelated uncertainties across bins $\{i\}$. First, the yields in the m_{hh} bins are combined separately for the resolved, intermediate and boosted categories to explore their complementary sensitivity. The three categories are then combined to obtain the final combined constraint.

This figure also separates the contours by category, but as the boosted has little constraining power (in part due to the conservative 5% systematic), only comparisons between resolved and intermediate are of interest. Sensitivity is driven by the resolved category across $\kappa_t \neq 1$ with the intermediate category contribution being non-negligible for negative κ_λ. These contours capture how uncertainties on the measurement of κ_t affect the constraints on κ_λ, and it shows the improvement obtained by the neural-network analysis over the baseline.

7.1.6 Phenomenology Study Final Remarks

This study of the phenomenology of the $hh \rightarrow 4b$ process [1] expands current techniques in several directions to improve this search at future colliders. It highlights the importance of the low energy channels (resolved and intermediate) for constraining λ_{hhh}, and it shows that analyses optimized for the discovery of SM hh production are suboptimal for constraining λ_{hhh} in this channel.

The simple neural network analysis implemented here (which improves the signal significance from 0.69 (0.39) to 0.75 (0.53) without (with) the nominal systematics assumed) serves as proof-of-concept of further optimizations that could be obtained in the channel. The study also finds that the current uncertainty on the top-quark Yukawa coupling constraints can modify λ_{hhh} by $\sim 20\%$, and sharpens the importance of maintaining low p_{T} jet triggers (crucial for this search), specially around the boundary of expected sensitivity for positive κ_λ $\sim$5 values [1].

Although other di-Higgs decay channels may provide better sensitivity to λ_{hhh}, such as $bb\gamma\gamma$, the study finds that contribution from the $bbbb$ channel is significant and would improve sensitivity in combination with other channels.

8
Conclusions

This thesis provides a detailed account of the search for boosted $hh \rightarrow 4b$ in ATLAS with the full run 2 datasets. The 139 fb^{-1} dataset of $\sqrt{s} = 13$ TeV proton-proton collision events was analyzed. The expanded dataset, improved b-tagging techniques, and the use of variable-radius track jet algorithms to reconstruct subjets within the larger Higgs boson candidate provide better sensitivity than previous studies of the process.

The observed data were used to search for two BSM benchmark resonances: a heavy scalar S, and Kaluza-Klein gravitons G^*_{kk} predicted in the bulk Randall-Sundrum (RS) model. The analysis finds no significant deviation from the Standard Model expectation. Upper limits were set in the production cross section of these spin-0 and spin-2 resonances (figure 5.7). This boosted $hh \rightarrow 4b$ search is expected to be statistically combined with its resolved counterpart targeting low-mass resonances in an upcoming ATLAS paper.

The best constrains on the Higgs boson self-coupling by the ATLAS experiment set κ_λ between $-2.3 < \kappa_\lambda < 10.3$ by combining several single and di-Higgs analyses [24]. These constrains are expected to improve the sensitivity to test $\kappa_\lambda \sim 1$ with future runs of the LHC and HL-LHC [22] and this process could be even further constrained in future "Higgs Factory" lepton colliders.

At present di-Higgs production is a powerful tool to search for any exotic particle that can couple to the Higgs boson, such as the resonant production studies presented throughout this work. As the LHC continues on its schedule and gathers more data, searches for $hh \to 4b$, and di-Higgs in general, will enter a new phase of sensitivity to Standard Model Higgs-boson-pair production and λ_{hhh}, where this precise model will be put to the test once again.

Appendices

A

Full Pull and Impact of NPs

Figure A.1 shows the pulls and impacts of NPs on the G^*_{kk} and scalar samples for the $3b$ and $2b$ regions (only the $4b$ region plots were featured in the main text) in mass points relevant to each region's sensitivity range. Similar to figure 5.4, these plots and the fits were restricted to the m_{hh} range relevant to their sensitivity range, and show few significant pulls away from zero or constrains in the width of the pulls (one NP is pulled to around -0.4 in A.1a and two NPs were constrained by 30% in A.1c) which indicates the model is adequate to fit the data.

Figure A.2 shows the pulls and impacts of all the nuisance parameters included for the results shown in chapter 5, and was used to choose the top 20 most impactful NPs only. Note that these consider all b-tag regions, even the ones with suboptimal sensitivity for this mass region, which drives pulls away from zero and constrains below one. When fitting the individual regions in m_{hh} where each is impactful showed no major pulls or constrains.

Figure A.3 shows the pulls and impacts of NPs on the 800 GeV G^*_{kk} graviton sample on an old version of the model. As referenced in chapter 5, this statistical analysis includes the flavour tagging related nuisance parameters as it used the `MV2` tagger and the associated uncertainties available. Flavour tagging systematic uncertainties have small impact on the model, occupying the 13^{th} spot in the ranking

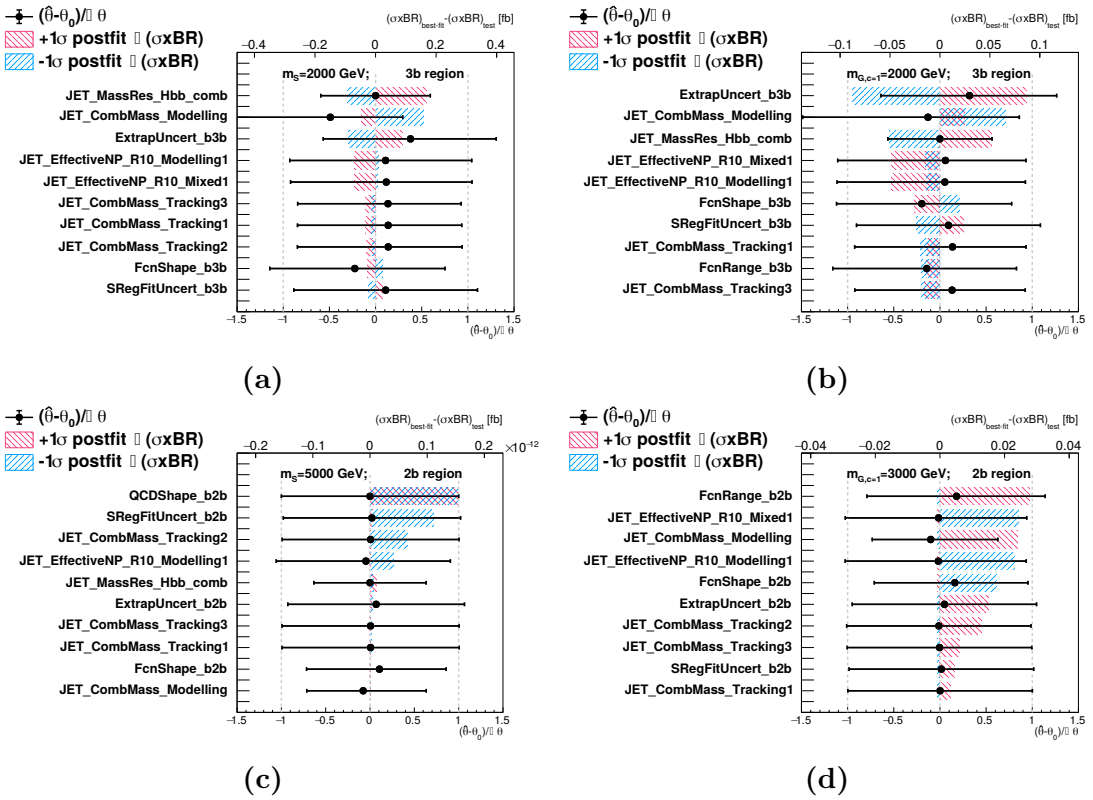


Figure A.1: Pulls (circle markers, bottom axis) and impacts (bars, top axis) in femtobarns on the measured cross section times branching ratio of a heavy scalar (left column) and spin-2 (right column) resonance. The plot shows the 10 nuisance parameters with most impact on the expected limit from the $3b$ (top row) and $2b$ (bottom row) regions. NPs related to the background estimate and jet properties ranked as the most impactful.

by impact. The work-in-progress flavour tagging uncertainties provided for the `DL1r` tagger are ranked even lower in the full impact plot in figure A.2.

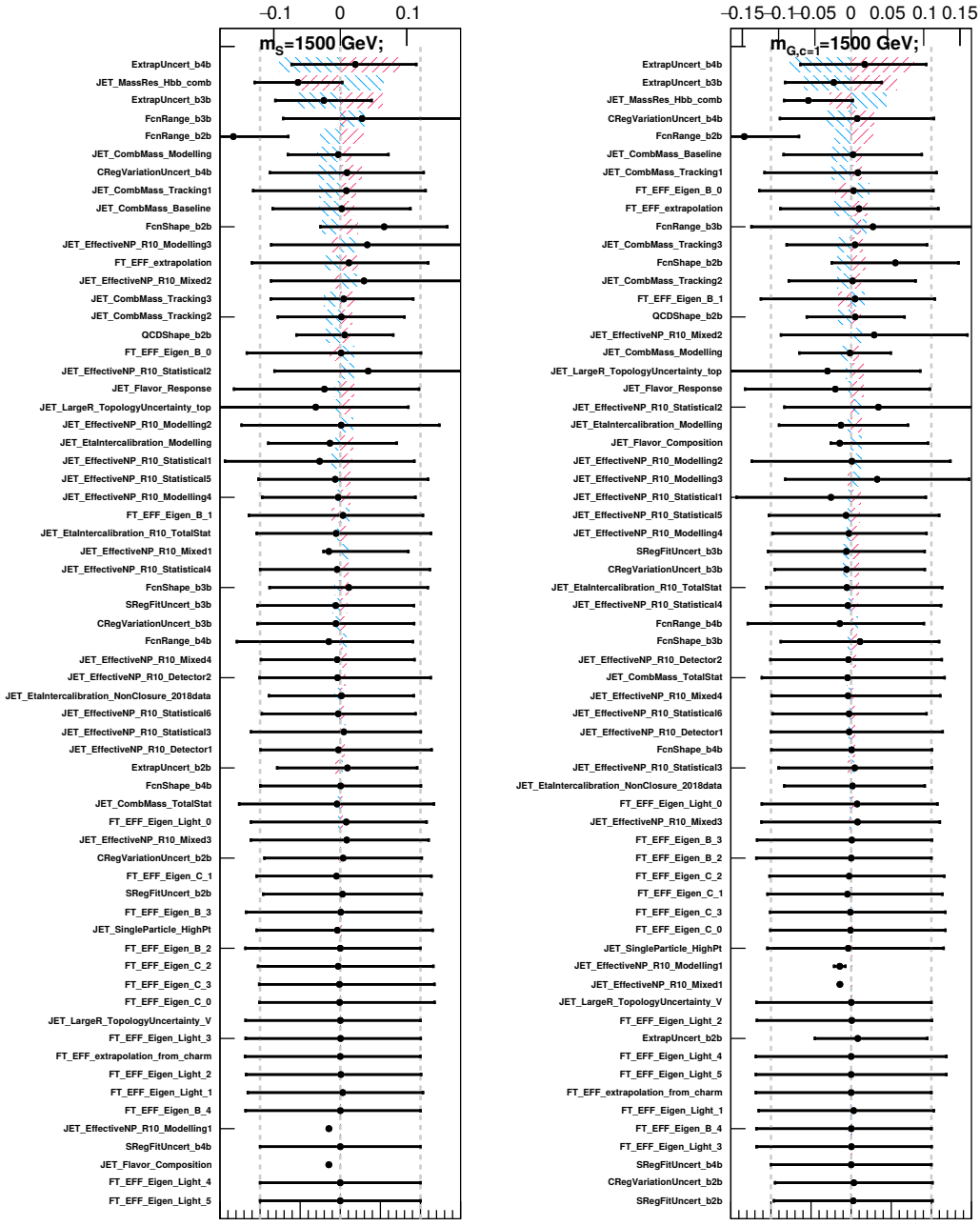


Figure A.2: Pulls (circle markers, bottom axis) and impacts (bars, top axis) in fb on the measured cross section times branching ratio of a 1.5 TeV (a) scalar and (b) graviton. The axes show the same quantities as all plots in this appendix and have been omitted to enlarge the fonts on the plots. Fits for these plots were made with all b-tag regions considered.

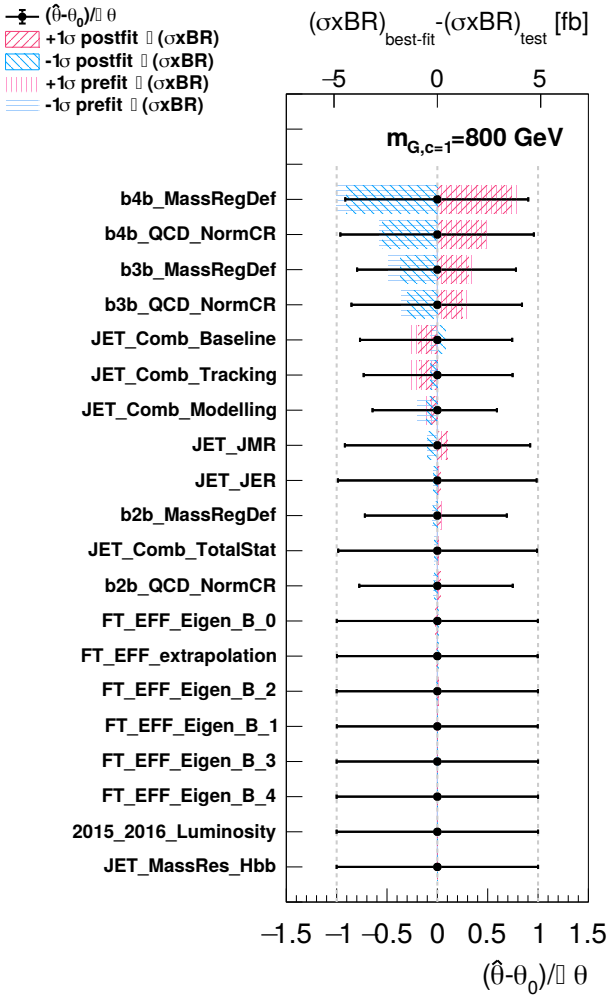


Figure A.3: Pulls (circle markers, bottom axis) and impacts (bars, top axis) in fb on the measured cross section times branching ratio of a 800 GeV G^*_{kk} graviton. NPs related to the background estimate and jet properties ranked as the most impactful. This model was a work-in-progress version used to inform the analysis and not used for any results.

B
Event Selection Yields

Table B.1 shows the event yield at each step in the selection process described in section 4.3 for the 2017 dataset, as well as the 2 TeV signal scalar and spin-2 resonances.

	2017 Data		G^*_{kk} m = 2 TeV		scalar m = 2 TeV	
	Yield	% lost	Yield	% lost	Yield	% lost
Initial Events	224364211.0		209943.0		17956.0	
Pass Trigger	20872544.0	90.70	208325.0	0.77	17147.0	4.51
≥ 2 large-R jets	18991091.0	9.01	186622.0	10.42	14763.0	13.90
Large-R jet mass	18921042.0	0.37	183970.0	1.42	14558.0	1.39
Large-R jet p_{T}	13495276.0	28.68	183559.0	0.22	14389.0	1.16
$\Delta\eta_{hh} < 1.3$	8665412.0	35.79	160018.0	12.82	9649.0	32.94
Resolved Veto	8665223.0	< 0.01	159603.0	0.26	9633.0	0.17
VR Veto	7381149.0	14.82	131998.0	17.30	8003.0	16.92
≥ 2 b-tags	219863.0	97.02	110972.0	15.93	6584.0	17.73
Signal Region	BLIND	-	51118.0	53.94	3171.0	51.84

Table B.1: Event yield and percentage of events filtered by each step in the event selection process for data collected in 2017 and the two benchmark signal models.

Bibliography

[1] Santiago Paredes Saenz et al. *Higgs self-coupling measurements using deep learning and jet substructure in the $b\bar{b}b\bar{b}$ final state*. 2020. arXiv: 2004.04240 [hep-ph].

[2] Santiago Rafael Paredes Saenz. *Evaluation and Monitoring of the MET Trigger Performance in 2017*. Tech. rep. ATL-COM-DAQ-2018-014. Geneva: CERN, Mar. 2018. URL: https://cds.cern.ch/record/2307316.

[3] ATLAS Collaboration. *Performance of the missing transverse momentum triggers for the ATLAS detector during Run-2 data taking*. 2020. arXiv: 2005.09554 [hep-ex].

[4] R. P. Feynman. "Mathematical Formulation of the Quantum Theory of Electromagnetic Interaction". In: *Phys. Rev.* 80 (3 Nov. 1950), pp. 440–457. DOI: 10.1103/PhysRev.80.440. URL: https://link.aps.org/doi/10.1103/PhysRev.80.440.

[5] Sheldon L. Glashow. "Partial-symmetries of weak interactions". In: *Nuclear Physics* 22.4 (1961), pp. 579–588. ISSN: 0029-5582. DOI: https://doi.org/10.1016/0029-5582(61)90469-2. URL: http://www.sciencedirect.com/science/article/pii/0029558261904692.

[6] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (21 Nov. 1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264. URL: https://link.aps.org/doi/10.1103/PhysRevLett.19.1264.

[7] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc. C* 680519 (1968), pp. 367–377. DOI: 10.1142/9789812795915_0034.

[8] M. Y. Han and Y. Nambu. "Three-Triplet Model with Double SU(3) Symmetry". In: *Phys. Rev.* 139 (4B Aug. 1965), B1006–B1010. DOI: 10.1103/PhysRev.139.B1006. URL: https://link.aps.org/doi/10.1103/PhysRev.139.B1006.

[9] G. Zweig. "An SU(3) model for strong interaction symmetry and its breaking. Version 2". In: *DEVELOPMENTS IN THE QUARK THEORY OF HADRONS. VOL. 1. 1964 - 1978*. Ed. by D.B. Lichtenberg and Simon Peter Rosen. Feb. 1964, pp. 22–101.

[10] P.W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Physics Letters* 12.2 (1964), pp. 132–133. ISSN: 0031-9163. DOI: https://doi.org/10.1016/0031-9163(64)91136-9. URL: http://www.sciencedirect.com/science/article/pii/0031916364911369.

[11] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[12] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[13] ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[14] The CMS Collaboration. *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*. 2012. arXiv: 1207.7235 [hep-ex].

[15] Arthur Kosowsky, Michael S. Turner, and Richard Watkins. "Gravitational waves from first-order cosmological phase transitions". In: *Phys. Rev. Lett.* 69 (14 Oct. 1992), pp. 2026–2029. DOI: 10.1103/PhysRevLett.69.2026. URL: https://link.aps.org/doi/10.1103/PhysRevLett.69.2026.

[16] Peisi Huang, Andrew J. Long, and Lian-Tao Wang. "Probing the electroweak phase transition with Higgs factories and gravitational waves". In: *Phys. Rev. D* 94 (7 Oct. 2016), p. 075008. DOI: 10.1103/PhysRevD.94.075008. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.075008.

[17] Bastian Bergerhoff and Christof Wetterich. "Electroweak Phase Transition in the Early Universe?" In: *Current Topics in Astrofundamental Physics: Primordial Cosmology*. Ed. by N. Sánchez and A. Zichichi. Dordrecht: Springer Netherlands, 1998, pp. 211–240. ISBN: 978-94-011-5046-0. DOI: 10.1007/978-94-011-5046-0_6. URL: https://doi.org/10.1007/978-94-011-5046-0_6.

[18] Arttu Rajantie. *Higgs cosmology*. 2018. URL: http://doi.org/10.1098/rsta.2017.0128.

[19] James M. Cline. *Baryogenesis*. 2006. arXiv: hep-ph/0609145 [hep-ph].

[20] Fa Peng Huang et al. "Testing the electroweak phase transition and electroweak baryogenesis at the LHC and a circular electron-positron collider". In: *Phys. Rev. D* 93 (10 May 2016), p. 103515. DOI: 10.1103/PhysRevD.93.103515. URL: https://link.aps.org/doi/10.1103/PhysRevD.93.103515.

[21] Marcela Carena, Zhen Liu, and Marc Riembau. "Probing the electroweak phase transition via enhanced di-Higgs boson production". In: *Phys. Rev. D* 97 (9 May 2018), p. 095032. DOI: 10.1103/PhysRevD.97.095032. URL: https://link.aps.org/doi/10.1103/PhysRevD.97.095032.

[22] ATLAS and CMS Collaborations. *Report on the Physics at the HL-LHC and Perspectives for the HE-LHC*. 2019. arXiv: 1902.10229 [hep-ex].

[23] Massimiliano Grazzini et al. *Higgs boson pair production at NNLO with top quark mass effects*. 2018. arXiv: 1803.02463 [hep-ph].

[24] *Constraints on the Higgs boson self-coupling from the combination of single-Higgs and double-Higgs production analyses performed with the ATLAS experiment*. Tech. rep. ATLAS-CONF-2019-049. Geneva: CERN, Oct. 2019. URL: https://cds.cern.ch/record/2693958.

[25] Felix Kling, Jose Miguel No, and Shufang Su. *Anatomy of Exotic Higgs Decays in 2HDM*. 2016. arXiv: 1604.01406 [hep-ph].

[26] Benoit Hespel, David Lopez-Val, and Eleni Vryonidou. *Higgs pair production via gluon fusion in the Two-Higgs-Doublet Model*. 2014. arXiv: 1407.0281 [hep-ph].

[27] Lisa Randall and Raman Sundrum. *A Large Mass Hierarchy from a Small Extra Dimension*. 1999. arXiv: hep-ph/9905221 [hep-ph].

[28] G. C. Branco et al. *Theory and phenomenology of two-Higgs-doublet models*. 2011. arXiv: 1106.0034 [hep-ph].

[29] A.D. Sakharov. "Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe". In: *Sov. Phys. Usp.* 34.5 (1991), pp. 392–393. DOI: 10.1070/PU1991v034n05ABEH002497.

[30] C. Patrignani et al. "Review of Particle Physics". In: *Chin. Phys.* C40.10 (2016), p. 100001. DOI: 10.1088/1674-1137/40/10/100001.

[31] ATLAS Collaboration. "Search for pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector". In: *JHEP* 01 (2019), p. 030. DOI: 10.1007/JHEP01(2019)030. arXiv: 1804.06174 [hep-ex].

[32] ATLAS Collaboration. *Search for resonant and non-resonant Higgs boson pair production in the $b\bar{b}\tau^+\tau^-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2018. arXiv: 1808.00336 [hep-ex].

[33] ATLAS Collaboration. *Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment*. 2018. arXiv: 1807.04873 [hep-ex].

[34] ATLAS Collaboration. *Search for Higgs boson pair production in the $b\bar{b}WW^*$ decay mode at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2018. arXiv: 1811.04671 [hep-ex].

[35] ATLAS Collaboration. *Search for non-resonant Higgs boson pair production in the $bb\ell\nu\ell\nu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV*. 2019. arXiv: 1908.06765 [hep-ex].

[36] ATLAS Collaboration. *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*. 2018. arXiv: 1807.08567 [hep-ex].

[37] ATLAS Collaboration. *Search for Higgs boson pair production in the $WW^{(*)}WW^{(*)}$ decay channel using ATLAS data recorded at $\sqrt{s} = 13$ TeV*. 2018. arXiv: 1811.11028 [hep-ex].

[38] CMS Collaboration. *Search for a massive resonance decaying to a pair of Higgs bosons in the four b quark final state in proton-proton collisions at $\sqrt{s} = 13$ TeV*. 2017. arXiv: 1710.04960 [hep-ex].

[39] A L Read. "Presentation of search results: theCLstechnique". In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313. URL: https://doi.org/10.1088%2F0954-3899%2F28%2F10%2F313.

[40] CMS Collaboration. *Combination of searches for Higgs boson pair production in proton-proton collisions at $\sqrt{s} = 13$ TeV*. 2018. arXiv: 1811.09689 [hep-ex].

[41] ATLAS Collaboration. *Search for Higgs boson pair production in the $b\bar{b}b\bar{b}$ final state from pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*. 2015. arXiv: 1506.00285 [hep-ex].

[42] Lyndon R Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008). This report is an abridged version of the LHC Design Report (CERN-2004-003), S08001. 164 p. DOI: 10.1088/1748-0221/3/08/S08001. URL: https://cds.cern.ch/record/1129806.

[43] ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[44] ALICE Collaboration. "The ALICE experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08002–S08002. DOI: 10.1088/1748-0221/3/08/s08002. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08002.

[45] CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08004.

[46] LHCb Collaboration. "The LHCb Detector at the LHC". In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08005–S08005. DOI: 10.1088/1748-0221/3/08/s08005. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08005.

[47] ATLAS Collaboration. *Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC*. 2013. arXiv: 1302.4393 [hep-ex].

[48] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*. 2016. arXiv: 1608.03953 [hep-ex].

[49] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. Tech. rep. ATLAS-CONF-2019-021. Geneva: CERN, June 2019. URL: https://cds.cern.ch/record/2677054.

[50] ATLAS Collaboration. *Pileup Interactions and Data Taking Efficiency*. Accessed: 2020-02-21. URL: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[51] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331063.

[52] ATLAS Collaboration. *ATLAS Insertable B-Layer Technical Design Report*. Tech. rep. CERN-LHCC-2010-013. ATLAS-TDR-19. Sept. 2010. URL: https://cds.cern.ch/record/1291633.

[53] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Pixel Detector*. Tech. rep. CERN-LHCC-2017-021. ATLAS-TDR-030. Geneva: CERN, Sept. 2017. URL: https://cds.cern.ch/record/2285585.

[54] ATLAS Collaboration. *Technical Design Report for the ATLAS Inner Tracker Strip Detector*. Tech. rep. CERN-LHCC-2017-005. ATLAS-TDR-025. Geneva: CERN, Apr. 2017. URL: https://cds.cern.ch/record/2257755.

[55] Vasiliki A. Mitsou. *ATLAS silicon microstrip tracker: operation and performance*. 2011. arXiv: 1110.1983 [physics.ins-det].

[56] V Lacuesta. "Track and vertex reconstruction in the ATLAS experiment". In: *Journal of Instrumentation* 8.02 (Feb. 2013), pp. C02035–C02035. DOI: 10.1088/1748-0221/8/02/c02035. URL: https://doi.org/10.1088%2F1748-0221%2F8%2F02%2Fc02035.

[57] Giles Barr et al. *Particle physics in the LHC era*. Oxford master series in particle physics, astrophysics and cosmology. Oxford: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780198748557.001.0001. URL: https://cds.cern.ch/record/2034442.

[58] A Vogel. *ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates*. Tech. rep. ATL-INDET-PROC-2013-005. Geneva: CERN, Apr. 2013. URL: https://cds.cern.ch/record/1537991.

[59] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061.

[60] *ATLAS tile calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331062.

[61] ATLAS Collaboration. "Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1". In: *The European Physical Journal C* 77.7 (July 2017), p. 490. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5004-5. URL: https://doi.org/10.1140/epjc/s10052-017-5004-5.

[62] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/331068.

[63] M. zur Nedden. "The LHC Run 2 ATLAS trigger system: design, performance and plans". In: *Journal of Instrumentation* 12.03 (Mar. 2017), pp. C03024–C03024. DOI: 10.1088/1748-0221/12/03/c03024. URL: https://doi.org/10.1088%2F1748-0221%2F12%2F03%2Fc03024.

[64] Theodor Christian Herwig. *ATLAS jet trigger performance in 2016 data*. Tech. rep. ATL-DAQ-PROC-2016-020. Geneva: CERN, Nov. 2016. DOI: 10.22323/1.282.0854. URL: https://cds.cern.ch/record/2229226.

[65] R Achenbach et al. "The ATLAS Level-1 Calorimeter Trigger". In: *Journal of Instrumentation* 3.03 (Mar. 2008), P03001–P03001. DOI: 10.1088/1748-0221/3/03/p03001. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F03%2Fp03001.

[66] ATLAS Collaboration. *Performance of the ATLAS Trigger System in 2015*. 2016. arXiv: 1611.09661 [hep-ex].

[67] ATLAS Collaboration. *The performance of the jet trigger for the ATLAS detector during 2011 data taking*. 2016. arXiv: 1606.07759 [hep-ex].

[68] ATLAS Collaboration. *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*. 2017. arXiv: 1704.07983 [hep-ex].

[69] ATLAS Collaboration. *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC*. 2016. arXiv: 1611.10235 [physics.ins-det].

[70] Gavin P. Salam. "Towards jetography". In: *The European Physical Journal C* 67.3 (June 2010), pp. 637–686. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-010-1314-6. URL: https://doi.org/10.1140/epjc/s10052-010-1314-6.

[71] ATLAS Collaboration. *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. 2017. arXiv: 1703.09665 [hep-ex].

[72] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jet Trimming". In: *JHEP* 02 (2010), p. 084. DOI: 10.1007/JHEP02(2010)084. arXiv: 0912.1342 [hep-ph].

[73] ATLAS Collaboration. "Performance of jet substructure techniques for large-R jets in proton-proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector". In: *JHEP* 09 (2013), p. 076. DOI: 10.1007/JHEP09(2013)076. arXiv: 1306.4945 [hep-ex].

[74] *Jet mass reconstruction with the ATLAS Detector in early Run 2 data*. Tech. rep. ATLAS-CONF-2016-035. Geneva: CERN, July 2016. URL: https://cds.cern.ch/record/2200211.

[75] ATLAS Collaboration. *In situ calibration of large-R jet energy and mass in 13 TeV proton-proton collisions with the ATLAS detector*. 2018. arXiv: 1807.09477 [hep-ex].

[76] Matteo Cacciari and Gavin P. Salam. "Pileup subtraction using jet areas". In: *Phys. Lett.* B659 (2008), pp. 119–126. DOI: 10.1016/j.physletb.2007.09.077. arXiv: 0707.1378 [hep-ph].

[77] *Flavor Tagging with Track Jets in Boosted Topologies with the ATLAS Detector*. Tech. rep. ATL-PHYS-PUB-2014-013. Geneva: CERN, Aug. 2014. URL: http://cds.cern.ch/record/1750681.

[78] *Expected performance of the ATLAS b-tagging algorithms in Run-2*. Tech. rep. ATL-PHYS-PUB-2015-022. Geneva: CERN, July 2015. URL: https://cds.cern.ch/record/2037697.

[79] David Krohn, Jesse Thaler, and Lian-Tao Wang. "Jets with Variable R". In: *JHEP* 06 (2009), p. 059. DOI: 10.1088/1126-6708/2009/06/059. arXiv: 0903.0392 [hep-ph].

[80] *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*. Tech. rep. ATL-PHYS-PUB-2017-010. Geneva: CERN, June 2017. URL: https://cds.cern.ch/record/2268678.

[81] *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*. Tech. rep. ATL-PHYS-PUB-2016-012. Geneva: CERN, June 2016. URL: http://cds.cern.ch/record/2160731.

[82] ATLAS Collaboration. *ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV*. 2019. arXiv: 1907.05120 [hep-ex].

[83] ATLAS Collaboration. *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$=13 TeV*. 2016. arXiv: 1603.05598 [hep-ex].

[84] J. Allison et al. "Recent developments in Geant4". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 835 (2016), pp. 186–225. ISSN: 0168-9002. DOI: https://doi.org/10.1016/j.nima.2016.06.125. URL: http://www.sciencedirect.com/science/article/pii/S0168900216306957.

[85] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (Nov. 2007), pp. 070–070. ISSN: 1029-8479. DOI: 10.1088/1126-6708/2007/11/070. URL: http://dx.doi.org/10.1088/1126-6708/2007/11/070.

[86] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (June 2008), pp. 852–867. ISSN: 0010-4655. DOI: 10.1016/j.cpc.2008.01.036. URL: http://dx.doi.org/10.1016/j.cpc.2008.01.036.

[87] D. J. Lange. "The EvtGen particle decay simulation package". In: *Nucl. Instrum. Meth.* A462 (2001), pp. 152–155. DOI: 10.1016/S0168-9002(01)00089-4.

[88] Richard D. Ball et al. "Parton distributions with LHC data". In: *Nuclear Physics B* 867.2 (Feb. 2013), pp. 244–289. ISSN: 0550-3213. DOI: 10.1016/j.nuclphysb.2012.10.003. URL: http://dx.doi.org/10.1016/j.nuclphysb.2012.10.003.

[89] ATLAS Collaboration. *Summary of ATLAS Pythia 8 tunes*. Tech. rep. ATL-PHYS-PUB-2012-003. Geneva: CERN, Aug. 2012. URL: http://cds.cern.ch/record/1474107.

[90] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *Journal of High Energy Physics* 2014.7 (July 2014). ISSN: 1029-8479. DOI: 10.1007/jhep07(2014)079. URL: http://dx.doi.org/10.1007/JHEP07(2014)079.

[91] Manuel Bähr et al. "Herwig++ physics and manual". In: *The European Physical Journal C* 58.4 (Nov. 2008). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-008-0798-9. URL: http://dx.doi.org/10.1140/epjc/s10052-008-0798-9.

[92] Roman Kogler et al. "Jet substructure at the Large Hadron Collider". In: *Reviews of Modern Physics* 91.4 (Dec. 2019). ISSN: 1539-0756. DOI: 10.1103/revmodphys.91.045003. URL: http://dx.doi.org/10.1103/RevModPhys.91.045003.

[93] ATLAS Collaboration. *Search for new phenomena in dijet events using 37 fb^{-1} of pp collision data collected at $\sqrt{s}$ =13 TeV with the ATLAS detector*. 2017. arXiv: 1703.09127 [hep-ex].

[94] ROBERT M. HARRIS and KONSTANTINOS KOUSOURIS. "SEARCHES FOR DIJET RESONANCES AT HADRON COLLIDERS". In: *International Journal of Modern Physics A* 26.30n31 (Dec. 2011), pp. 5005–5055. ISSN: 1793-656X. DOI: 10.1142/s0217751x11054905. URL: http://dx.doi.org/10.1142/S0217751X11054905.

[95] G. Avoni et al. "The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS". In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.

[96] ATLAS Collaboration. *Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV*. 2018. arXiv: 1805.01845 [hep-ex].

[97] Andrew Gordon Wilson. "Covariance kernels for fast automatic pattern discovery and extrapolation with Gaussian processes". PhD thesis. University of Cambridge, 2014.

[98] Andrew Gordon Wilson and Ryan Prescott Adams. *Gaussian Process Kernels for Pattern Discovery and Extrapolation*. 2013. arXiv: 1302.4245 [stat.ML].

[99] Wouter Verkerke and David Kirkby. *The RooFit toolkit for data modeling.* 2003. arXiv: physics/0306116 [physics.data-an].

[100] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: http://dx.doi.org/10.1140/epjc/s10052-011-1554-0.

[101] J. Neyman and E. S. Pearson. "On the Problem of the Most Efficient Tests of Statistical Hypotheses". In: *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933), pp. 289–337. ISSN: 02643952. URL: http://www.jstor.org/stable/91247.

[102] *Measurement prospects of the pair production and self-coupling of the Higgs boson with the ATLAS experiment at the HL-LHC.* Tech. rep. ATL-PHYS-PUB-2018-053. Geneva: CERN, Dec. 2018. URL: https://cds.cern.ch/record/2652727.

[103] François Chollet et al. *Keras.* https://keras.io. 2015.

[104] Richard Lippmann. "An introduction to computing with neural nets". In: *IEEE ASSP Magazine* 4 (1987), pp. 4–22.

[105] K. Hornik, M. Stinchcombe, and H. White. "Multilayer Feedforward Networks Are Universal Approximators". In: *Neural Netw.* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: 10.1016/0893-6080(89)90020-8. URL: http://dx.doi.org/10.1016/0893-6080(89)90020-8.

[106] Kurt Hornik. "Approximation capabilities of multilayer feedforward networks". In: *Neural Networks* 4 (1991), pp. 251–257.

[107] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[108] Scott M Lundberg and Su-In Lee. "A Unified Approach to Interpreting Model Predictions". In: *Advances in Neural Information Processing Systems 30.* Ed. by I. Guyon et al. Curran Associates, Inc., 2017, pp. 4765–4774. URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

[109] B. Todd Huffman, Charles Jackson, and Jeff Tseng. "Tagging b quarks at extreme energies without tracks". In: *J. Phys.* G43.8 (2016), p. 085001. DOI: 10.1088/0954-3899/43/8/085001. arXiv: 1604.05036 [hep-ex].