UNIVERSITY OF CALIFORNIA

SANTA CRUZ

AN INCLUSIVE SEARCH FOR THE DECAY OF BOOSTED HIGGS BOSONS IN THE $H \to b\bar{b}$ CHANNEL WITH THE ATLAS DETECTOR

A dissertation submitted in partial satisfaction of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PHYSICS

by

Jacob Martin Pasner

December 2019

The Dissertation of Jacob Martin Pasner
is approved:

Professor Jason Nielsen, Chair

Professor Abraham Seiden

Professor Michael Hance

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures **vii**

List of Tables **xiv**

Abstract **xvi**

Dedication **xviii**

Acknowledgments **xix**

1 Introduction **1**

I Theoretical Motivations and the Standard Model **3**

2 The Standard Model and Beyond **4**

2.1 The Standard Model 5

2.1.1 Bosons 6

2.1.2 Fermions 9

2.2 Electroweak Unification 10

2.3 Quantum Chromodynamics 13

2.4 The Higgs Mechanism 15

2.4.1 Electroweak Symmetry Breaking 16
2.4.2 Fermion Mass Terms 19
2.4.3 The Higgs Boson 21

3 Boosted Higgs at the LHC **22**
3.1 Higgs Production Mechanisms 23
3.2 Parton Distribution Function 26
3.3 Branching Ratios 27
3.4 Evidence for the SM Higgs 30
3.5 Boosted Higgs 30

II Experimental Apparatus and Associated Facilities **35**

4 The Large Hadron Collider **36**
4.1 Particle Injection Chain 37
4.2 LHC Layout and Design 39
4.3 Performance 42
4.4 Pile-up at the LHC 45

5 The ATLAS Detector **47**
5.1 ATLAS Coordinate System 50
5.2 Tracking with the Inner Detector 53
5.2.1 Pixel Detector 55
5.2.2 Semiconductor Tracker 56
5.2.3 Transition Radiation Tracker 57
5.3 Calorimetry 58
5.3.1 Electromagnetic Calorimeter 59

5.3.2 Hadronic Calorimeter 61
5.4 Muon Spectrometer 63
5.5 Trigger 65

III Search for Boosted $H \to b\bar{b}$ **67**

6 Data and Monte Carlo Simulation **68**
6.1 ATLAS Data Analyzed 69
6.2 Higgs Boson Signal Monte Carlo Samples 69
6.3 Background Monte Carlo 72

7 Physics Object Definitions **78**
7.1 Jets 79
7.1.1 Large-R Jets 81
7.1.2 Variable Radius Track Jets 82
7.2 Flavor Tagged Jets 84
7.3 Muons 92

8 Boosted $H \to b\bar{b}$ Event Selection **96**
8.1 Selected Triggers 96
8.2 Event Selection 98

9 Background Estimation **104**
9.1 Hadronic Vector Boson Background 105
9.2 Top Quark Pair Background 105
9.2.1 Constructing the $\mathrm{CR}_{t\bar{t}}$ Control Region 106
9.2.2 k-factor Estimation 109
9.3 Multijet QCD estimation 111

10 Systematic Uncertainties **115**

10.1 Uncertainties on Data-Driven QCD Modeling 116

10.2 Uncertainties on Resonant Backgrounds and Signal 116

10.3 Summary of Systematics 118

11 Statistical Fit **121**

11.1 Bayes' Theorem and Limit Setting 122

11.2 Implementation of Priors 126

11.3 Bayesian Analysis Toolkit 127

12 Results **131**

12.1 Measurement Procedure 132

12.2 Observation of Boosted $V \to b\bar{b}$ 132

12.3 Measurement of Boosted $H \to b\bar{b}$ 134

13 Conclusion **136**

Bibliography **138**

List of Figures

2.1 Summary of several Standard Model total and fiducial production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations [20]. All theoretical expectations were calculated at NLO or higher. The dark-color error bar represents the statistical uncertainty. The lighter-color error bar represents the full uncertainty, including systematics and luminosity uncertainties. The data/theory ratio, luminosity used and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers. They were not always evaluated using the same prescriptions for PDFs and scales. 7

2.2 Table of all observed fundamental particles of the Standard Model. [21] 8

2.3 A lower-dimensionality representation of the shape of the Higgs potential. The central peak represents a rotationally symmetric unstable state, while the trough represents the infinite number of minima that can be selected upon the spontaneous breaking of the symmetry. 17

3.1 Cross sections for the production of the SM Higgs boson signatures as a function of the center of mass energy ($\sqrt{s}$) at the LHC [23]. In order of decreasing cross section: the ggF process signature is H, the VBF process signature is qqH, the VH process signature is split into WH and ZH, and the $b\bar{b}$ / $t\bar{t}$ signatures are $t\bar{t}H$ / $b\bar{b}H$. 23

3.2 Feynman diagrams representing the dominant Higgs production modes at the LHC. 24

3.3 Feynman diagram for ggF Higgs + jet production. 25

3.4 MMHT2014 NNLO PDFs at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 with associated 68% confidence-level uncertainty bands [25]. The colored regions indicate the probability of finding the labeled parton with a momentum fraction given along the x axis. As expected the valence quarks contain the largest fraction of the momentum while the gluons are more likely to carry smaller fractions of the total momentum. Note that as Q^2 increases the contributions from sea quarks increases. 27

3.5 Feynman diagrams representing the leading Higgs decay channels. . . . 28

3.6 Branching ratios for the decay of the SM Higgs boson near $m_H = 125$GeV including theoretical uncertainty bands [23] 29

3.7 Best fit values of $\sigma_i \cdot B^f$ for each specific channel $i \to H \to f$, as obtained from the SM Higgs coupling parameterization with 23 parameters for the combination of the ATLAS and CMS measurements. The error bars indicate the 1σ intervals. The fit results are normalized to the SM predictions for the various parameters and the shaded bands indicate the theoretical uncertainties in these predictions. Only 20 parameters are shown because some are either not measured with a meaningful precision, in the case of the $H \to ZZ$ decay channel for the WH, ZH, and $t\bar{t}H$ production processes, or not measured at all and therefore fixed to their corresponding SM predictions, in the case of the $H \to b\bar{b}$ decay mode for the ggF and VBF production processes [26]. 31

3.8 Best fit results for the production signal strengths for the combination of ATLAS and CMS data. Also shown are the results from each experiment. The error bars indicate the 1σ (thick lines) and 2σ (thin lines) intervals. The measurements of the combined global signal strength μ are also shown [26]. 32

3.9 Best fit results for the decay signal strengths for the combination of ATLAS and CMS data. Also shown are the results from each experiment. The error bars indicate the 1σ (thick lines) and 2σ (thin lines) intervals [26]. 32

3.10 Feynman diagram for boosted Higgs decaying to $b\bar{b}$ 34

3.11 Cartoon showing collimated Higgs boson jet and ISR jet as a result of the large boost due to their mutual recoil [29]. 34

4.1 Sketch of the CERN accelerator complex [35]. 38

4.2 Labeled diagram of all the experiments at the LHC indicating the counter-circulating beams and the relative location of each major experiment along the circumference of the accelerator [38]. 40

4.3 Depiction of a LHC dipole magnet 2-in-1 design labeling the major components [39]. 41

4.4 Luminosity is monitored as both a running total known as the integrated luminosity as depicted in (a) and as an instantaneous quantity as shown in (b) [42]. 44

4.5 Pileup for data taking periods 2015 - 2018 [42]. 45

5.1 A cut-away side view of the ATLAS detector with the major components labeled [43]. Within each of these labeled components there may exist multiple different detector technologies. For scale two people in red are shown standing between the disk muon chambers on the left side of the figure. 48

5.2 This slice of the ATLAS detector depicts how different particles interact with each component of the detector they cross. A dashed line indicates no interaction while a solid line indicates interaction. Electrons (yellow) and charged hadrons (red) interact with the tracker and curve in the solenoid's magnetic field. Electron and photon showers (yellow/green) are absorbed by the electromagnetic calorimeter. All hadrons (red/yellow) are absorbed by the hadronic calorimeter. The muons (orange) curve in both the solenoid and toroid magnetic fields before exiting the detector. Finally, the neutrinos (white) pass through the entire detector without interacting [46]. 51

5.3 The standard Cartesian coordinate system is shown with its origin at the ATLAS interaction point, the positive x-axis towards the center of the LHC, the positive y-axis pointing upwards, and the positive z-axis pointing along the beamline [29]. 52

5.4 Modified from [47] this cartoon represents a selection of pseudorapidity (η) values overlaid with some Cartesian coordinates (dashed black lines). The red lines are drawn for $\eta = \pm 0.5,\ 1.0$ and 3.0. 53

5.5 Diagram of Inner Detector [50]. 54

5.6 Schematic of the Inner Detector including η lines. Each component shown is cylindrically symmetric leading to a multi-layered detector [51]. 55

5.7 A cutaway diagram of ATLAS sampling calorimeters [43]. 58

5.8 Sketch of LAr EMC barrel module where the lead and liquid argon layers are visible in an accordion like geometry. Looking from the foreground to the back there are 3 different layers of readout cells visible [43]. . . . 60

5.9 Schematic of a Tile Calorimeter module including a depiction of the connection between the scintillator tile to the photomultiplier via a wavelength-shifting fiber [43]. 62

5.10 A cut-away diagram of the ATLAS muon system and its many sub-detectors [43]. 64

7.1 Example clustering of a parton-level MC simulation event showing clustered jet shapes for $R = 1.0$ and (a) $P = 1$, (b) $P = 0$, and (c) $P = -1$ [71]. 82

7.2 Simulation of calorimeter clusters for decay of a boosted $Z' \to t\bar{t}$ clustered in a large-R jet [77]. 83

7.3 Labeling efficiency of subjet double b-tagging using truth information from Higgs decay as a function of Higgs jet p_{T}. The efficiency for $R = 0.2$ fixed radius track jests is included for comparison. Uncertainty bars include statistical uncertainties only [79]. 85

7.4 The average ΔR between either the two leading truth b-hadrons or the two leading subjets associated to a Higgs jet as a function of Higgs p_{T} [79]. 86

7.5 Distributions of ΔR between a truth matched b-hadron and the reconstructed leading subjet [79]. The uncertainties given reflect only statistical uncertanties. All algorithms are normalized to an area corresponding to the fraction of signal jets which contain a leading subjet. 86

7.6 Distributions of ΔR between a truth matched b-hadron and the reconstructed subleading subjet [79]. The uncertainties given reflect only statistical uncertainties. All algorithms are normalized to an area corresponding to the fraction of signal jets which contain a subleading subjet. 87

7.7 Cartoon of a b-jet decay containing a b-hadron decay vertex (blue ●) displaced from the primary pp vertex (red ●), and a c-hadron decay vertex (orange ●) further displaced and often close to the b-hadron flight axis [82]. The secondary (blue) and tertiary (orange) vertices have large impact parameters (green) with respect to the primary pp vertex. 88

7.8 Data-Monte Carlo comparisons of the transverse (d_0) and longitudinal (z_0) impact parameter significance values for IP3D selected charged tracks in the leading jet of a $Z \to \mu\mu$ + jets dominated sample [82]. 89

7.9 The distribution of the transverse impact parameter significance S_{d0} for the leading d_0 significance track and subleading d_0 significance track for b-jets (left) and light jets (right) [82]. 90

7.10 Normalized BDT response in simulated $t\bar{t}$ events of the SMT for reconstructed muons associated to b-jets (blue), c-jets, (green) and light-flavour jets (red) [82]. 91

7.11 Flowchart of inputs to the `MV2c10` b-tagging algorithm [29]. 91

7.12 Performance of the `MV2c10` BDT for the 2016 optimization in simulated $t\bar{t}$ events. The performance was evaluated on $t\bar{t}$ events simulated using POWHEG interfaced to PYTHIA6 [86]. 92

7.13 Total uncertainty in the efficiency scale factor for Medium muons as a function of p_{T} as measured in $Z \to \mu\mu$ (solid lines) and $J/\Psi \to \mu\mu$ (dashed lines) decays. The combined uncertainty is the sum in quadrature of the individual contributions [87]. 95

8.1 Diagram of the event selection process and the labeling scheme of the large-R jets in the signal candidate events [29]. 99

8.2 (a) Predicted flavor composition of the dijet background in the SR based on the truth-matched hadron content of the two leading-p_{T} track-jets associated to the signal candidate large-R jet, with the B/C labels indicating the presence of a b/c-quark and L indicating the presence of a light quark or a gluon. (b) The expected shape of the dijet background in the SR and CR normalized to the same event count between 70 GeV $< m_J <$ 230 GeV [88]. 100

9.1 Simulations of the non-resonant and resonant background contributions to the signal region for an integrated luminosity of 80.5 fb^{-1} 105

9.2 Comparison in p_{T} of the top quark for different generator setups used to assess the NLO+PS matching as well as the parton shower and hadronization uncertainty after optimization, compared to data at $\sqrt{s} = 13$ TeV [89]. 106

9.3 (a) Feynman diagram of semi-leptonic $t\bar{t}$ decay. (b) Cartoon depicting semi-leptonic $t\bar{t}$ decay to b-quarks in center of mass frame. 107

9.4 Comparison of the two variables, $\Delta\phi$(leading muon−signal) (a) and muon p_{T} (b), between PYTHIA8 multijet and POWHEG $t\bar{t}$ simulated events used to create the $t\bar{t}$ control region. The selection requires exactly one b-tagged track jet in the signal candidate large-R jet. The magenta line shows the expected $s/\sqrt{b}$ significance with 80.5 fb^{-1} [63]. 108

9.5 Diagram of the CR$_{t\bar{t}}$ selection criteria. 109

9.6 The pre-fit (left) and post-fit (right) data vs MC comparison for fitting the CR$_{t\bar{t}}$ region. 110

9.7 The pull distributions for the different nuisance parameters used in the CR$_{t\bar{t}}$ fit. Parameters with a error less than 1 indicate that the prior uncertainty has been constrained by the fit. 111

9.8 Empirical cumulative distribution function (CDF) for the test p-values for the two statistical tests applied to both function choices where the a priori threshold α is shown as a vertical black line. In both tests p-values below $\alpha = 0.1$ show preference for the model with more parameters, while p-values above $\alpha = 0.1$ indicate a preference for the model with fewer parameters [63]. 114

11.1 Cartoon representing the relationship between the signal normalization prior $\pi(\nu)$ (red), likelihood $\mathcal{L}(\nu|\text{data})$ (blue), and posterior $p(\nu|\text{data})$ (black) for the cases of (a) a small data sample and (b) a larger data sample. This simplified example contains no nuisance parameters. . . . 125

11.2 This figure illustrates a random walk in parameter space (θ_1, θ_2). The numbers indicate the number of iterations the chain remained at this point in parameter space, the blue arrows indicate accepted transitions, and the red arrows indicate rejected transitions. The marginalized posterior distributions obtained for the two parameters $p(\theta_1|\text{data})$ and $p(\theta_2|\text{data})$ are also shown, and the yellow bands correspond to the central 68% of the distributions [98]. 130

12.1 The top panel shows the post-fit comparison of the signal candidate large-R jet mass distribution for the combined SM Higgs boson, V + jets, $t\bar{t}$ and QCD model to the observed data [88]. The middle panel gives the ratio of the post-fit model and the data with the QCD and $t\bar{t}$ components subtracted, highlighting the large resonance from V + jets. The bottom panel gives the ratio of the post-fit model and the data with the QCD, V + jets, and $t\bar{t}$ components subtracted, highlighting a slight excess of events near $m_J = 125$ GeV. 133

12.2 Pulls of nuisance parameters in the final combined fit [63]. 134

12.3 Combined marginalized posterior distributions of μ_H and μ_V in the Signal Region [88]. It is seen that the best-fit values for the signal processes lie within the 68% Credibility Level (2σ) of the Standard Model prediction. 135

List of Tables

3.1 SM Higgs boson production cross sections in units of pb for $m_H = 125$ GeV in pp collisions for the current LHC center-of-mass energy, $\sqrt{s} = 13$ TeV. The predictions for the ggF channel include the latest N3LO results, which have reduced theoretical uncertainties by a factor around 2 compared to the NNLO results [23]. 25

3.2 The branching ratios and the relative uncertainty for a Standard Model Higgs boson with $m_H = 125$ GeV [23]. 29

4.1 Nominal design values of LHC operations parameters at ATLAS for 25 ns bunch crossing spacing [31, 40]. Design and ATLAS recorded values of the machine luminosity are also given for LHC Run II operations [41]. . 44

6.1 List of datasets used for Higgs boson Monte Carlo [63]. 71

6.2 List of datasets used for QCD Monte Carlo [63]. 73

6.3 List of datasets used for hadronically decaying W+jets and Z+jets Monte Carlo [63]. 75

6.4 List of datasets used for $t\bar{t}$ Monte Carlo [63]. 76

6.5 List of datasets used for single-top Monte Carlo [63]. 77

8.1 Summary of the large-R jet triggers used for the data taking periods of 2015, 2016, and 2017 and the offline p_{T} thresholds at which they become fully efficient. The recorded integrated luminosity for each trigger is also included [63]. 98

8.2 Cutflow showing the effect in simulation of selection criteria for each of the Higgs boson production mechanisms [63]. The numbers here can be compared to the number of events generated as shown in Table 6.1. . . . 101

8.3 Cutflow showing the effect of selection criteria on simulated background samples as well as data. The simulated QCD contribution is scaled by 0.74 which was derived from a fit to data in the CR_{QCD} [63]. The background simulation numbers here can be compared to the number of events generated as shown in Tables 6.2 to 6.4. 101

8.4 The efficiencies and yields in the 0-tag control region (CR_{QCD}) and signal region (SR) for the non-QCD background, the Higgs boson signal and data. The yields in the CR_{QCD} are scaled to the luminosity used for the background estimate of the non-resonant dijet process discussed in Section 9.3. The efficiencies are calculated relative to the leading large-R jet $p_T > 480$ GeV requirement [63]. 103

8.5 The fractional composition of the different resonant contributions in the 0-tag control region (CR_{QCD}) and the signal region (SR). The fraction is evaluated using the given contribution type as the total [63]. 103

9.1 The $t\bar{t}$ scale factors and their uncertainties from the three $t\bar{t}$ control regions. The value for $CR_{t\bar{t}}$ is used in the Signal Region. NOTE: $CR_{t\bar{t}2}$ fit failed due to low statistics but is almostly completely dominated by ttbar in this region. The value here is the simple ratio of number of events in ttbar vs. data, and the uncertainty is the $t\bar{t}$ statistical uncertainty. . . . 110

10.1 Summary of the impact ($\sqrt{\Delta\sigma_i^2}/\mu_s$) of the main systematic uncertainties on the total uncertainty, σ_{tot}, of the measured signal strength, μ_s, for the V + jets and Higgs boson + jets signals [88]. 120

12.1 Summary of pertinent boosted Higgs boson cross section and selection values in simulation. N_{SR} is the number of simulated Higgs bosons that pass into the SR. 135

Abstract

An Inclusive Search for the Decay of Boosted Higgs Bosons in the $H \to b\bar{b}$ Channel with the ATLAS Detector

by

Jacob Martin Pasner

A search for high-momentum Higgs bosons, produced with an associated jet and decaying to bottom quark pairs, is conducted using an integrated luminosity of 80.5 fb^{-1} of proton-proton collisions at $\sqrt{s} = 13$ TeV collected with the ATLAS detector at the Large Hadron Collider. The search for $H \to b\bar{b}$ is of particular importance, as current measurements of this decay channel contain large uncertainties and have not yet established evidence for the gluon-gluon fusion and vector boson fusion Higgs boson production modes. This is the first time this analysis has been performed with $\sqrt{s} = 13$ TeV at ATLAS and represents an important advancement in the use of boosted jet techniques including implementation of the Variable Radius jet algorithm. Furthermore, a novel $t\bar{t}$ control region strategy is implemented to correct the mismodeling of its normalization and is shown to constrain the associated large JES and JMR systematic uncertainties. For the Standard Model Higgs boson, the observed signal strength is $\mu_H = 5.8 \pm 3.1$ (stat.) ± 1.9 (syst.) ± 1.7 (theo.), which is 1.6 standard deviations higher than the background-only hypothesis, with an expected sensitivity of 0.28σ.

To my loving parents for giving me a beautiful life

To my friends and family for filling my life with joy

And to science for the miracle and mystery of life itself

Acknowledgments

There is no way I can fully express my gratitude for the people and places that have made this journey all that it has been. The path to this degree has been winding, uncertain and full of mistakes. The names below represent a small sample of the rich bounty this life has offered me for guidance and support. I am eternally grateful to them all.

Thank you Jason Nielsen for personally calling me and inviting me to work with you nearly seven years ago. I didn't know how lucky I was back then, but I can definitively say now I wouldn't have it any other way. Your patience, exuberance, curiosity and kindness have taught me the power of questions and the boundless possibilities of an inquisitive mind. Thank you for these gifts, and for supporting me in the pursuit of my dreams.

Thank you to my committee members, Michael Hance and Abraham Seiden, for bearing with the hectic timeline this dissertation was created on. You supported my less than traditional writing style and have helped me produce a document my family and I can be proud of for years to come.

Thank you to my collaborators in ATLAS and all my individual analysis groups. It has been an incredible honor to work with such a diverse and devoted group of brilliant scientists. Our incredible success as a field stands as a testament to the ingenuity and determination that can only come from cooperation. Particle physicists have shown me

what a world without boarders looks like, and I am grateful to count myself amongst their ranks.

Thank you to the teachers and professors who helped me realize my dream of becoming a scientist and more: Richard Toothman, John McDaniel (McD), Kelley Molitor, Mani Tripathi, Manuel Calderón de la Barca Sánchez, Randy Harris, and Joel Primack. You believed in me and I will never forget the power of that gift.

Thank you to Giordon Stark for seeing my potential and setting me on the path to policy. You have been a solid shoulder to lean on, a kind ear (eye?) to my worries and a powerful mind to sharpen my own against. Your friendship and brilliance make this world more friendly and interesting, everywhere you go. Also your baking is so good it should come with a warning label.

Thank you to the many post doctoral students at UCSC who have pulled me out of the academic mud time and time again: Matthew, Andrea, Simone, Hass, Chiara and Ryan. Navigating the ATLAS collaboration without your guidance and protection seems an impossible task, we are all in your debt.

Thank you to my physics cohort for all the late night homework sessions and your continuing friendship and support: Joey, Max, Ross, Dean, Cam, Mike, Dave, Noah, Peizei, Ahram, Skyler, Christoph, Eudald, Daniel, Philip. You're all brilliant and I can't wait to see where life takes you all.

Thank you to the UCSC graduate students who helped me along the way, sometimes

with words of encouragement, other times with a stiff drink but always with a smile: Sheena, Peyton, Eric, Jeff, Alexander, Monika, Carey, Doug, Brian, Muiris, Daniel, Jerah, Katie, Tayler, Daniel, Zippy, Carolyn, Cole, Natasha and many more. Graduate school is hard in a way very few people can understand, I'm so glad I didn't have to travel this path alone, thanks to you.

Thank you to my Santa Cruz housemates through the years: Elizabeth, Tessa, Stas, Dave, David, Rob, Roy, Zad, Brent, Philip, Graham, Kelly and Catherine. The house on Laurel Street will always be a special place in my memories and my heart. The music, food, dance and discussion we shared have taught me what true community looks like. My love and gratitude to all of you, I couldn't ask for better or more supportive friends.

Thank you to the Santa Cruz locals for all the shenanigans and memories: Cleveland, Gabe, Monica, Travis, Saul, Bob, Larry, Deidre, Brian, Philip, Sean, Lance, Chappy, Keelin, Angelica, Kage, Mikey (thanks for the laundry), Michael, Micha, Xander, Gabriel, Julian, Brooke, Maya, Rafferty, Allie, Sam, Kate, Carrie, Alex, Cesar, Nora, Chris and many more. Our time together gave me the moments of rest and relaxation I needed to keep me sane.

Thank you to my Geneva housemates during my time in Switzerland: Rob, Bijan, Leigh and Jennifer. Living abroad was scary and hard until you welcomed me into the Marigold Memory Palace and your hearts. Here's to the future and our continued friendship.

Thank you to everyone who made my time living abroad an exciting and eye opening experience: Fenton, Paolo, Ivan, Marika, Niko, Oliver, Ranveig, Tristan, Apostolos, Efi, Irene, Leo, Arizona, Soph, Elsa, Adrian, George, Kelvin, Dayane, Enrica, Sam, Galen, Kisa, Babette, Viktoria, Adam, Sam, Joel, Amal, Danii, Flavia, Julie, Cameron, Mor, Swathi, Gabriella (Biscoitinha), Duncan, Eddie, Murdo, Lewis, Stoyan, Sophie, Sun, Xanthe, the entire LTA, and many more. Together we danced, hiked, boxed, ran, snowboarded, flew, sailed, biked and rode our way through more adventures than I'd ever thought possible. My life is so much richer thanks to you.

Thank you to Josh and Emmanuel for putting up with having a crazy particle physicist as the intern for 10 weeks. You gave me the foot in the door I needed to set off on the next stage of my life in policy. Next round of coffee is on me.

Thank you to Heidi for being there for me in what has been the most trying period of my life thus far. You saw me through Giardia, injuries and crippling writing stress. Whatever life brings us, I am lucky to have had your strength during this time.

Thank you to Vin for seeing me through the lowest lows and the highest highs. You've done more for me than words can describe. I treasure your friendship and will always love you. Here's to our next adventure!

Thank you to my sister Yara for being my best friend. I can talk to you about anything, from group theory to relationships to politics and always appreciate your advice. I can't wait to see what you do in this world and am thankful to have you in my life, love you

moo moo.

Thank you to my parents Mike and Izzy. You gave me life, you fostered my imagination, and have bent over backwards to support my dreams. All of my accomplishments, all of my life, I owe in some part to you and your example. My love and gratitude towards you knows no bounds and will exist as long as I'm living.

Thank you to music for giving me a way to express myself when nothing made sense. In my times of greatest need I have found strength in your ethereal arms. A special thanks to the local Irish session, all my Lark friends and Nathaniel Berman and his concert choir and wind ensemble.

Thank you to science for giving me the tools to answer my never ending questions. You have given me the lens I needed to make sense of this world, and my what a beautiful world it is.

And finally, thank you to the Earth for giving us such a beautiful, bountiful and rich place to call home.

Chapter 1

Introduction

As children we are fascinated with the natural world that surrounds us. How was it made? What is it made of? Where did it come from? When was it made? These questions follow some of us into adulthood and lead to the study of the fundamental interactions of the Universe, the pursuit of the building blocks of reality. For the past century, experiments of increasing size and complexity have probed higher energies and smaller distances to answer these questions out of the pure desire to know the unknown. The results of these experiments have been interpreted and used to build the most predictive and successful model across all of science - the Standard Model (SM) of particle physics. At the current state-of-the-art facility, the Large Hadron Collider (LHC), an international team of scientists and engineers continue this legacy of discovery. There, protons are smashed together at energies rarely seen since the birth of the Universe some 13.8 billion years ago and the results are analyzed to answer the

outstanding questions of our time.

In 2012 the discovery of a Higgs-like boson [1, 2, 3, 4, 5] at CERN by the ATLAS and CMS [6, 7] collaborations answered one of these outstanding questions: *What is the origin of mass for particles in the SM*? This was a major triumph for both the theoretical and experimental particle physics communities and resulted in the Nobel Prize in Physics being awarded to Francois Englert and Peter W. Higgs for their contributions to the Brout-Englert-Higgs theory. However, the properties of this new particle are still under scrutiny, and new physics could be lurking in the large uncertainties of current observations of the SM Higgs boson. In particular, the decay of the Higgs boson to bottom quarks ($H \to b\bar{b}$) was recently confirmed in 2018, and the resulting measured coupling strength still has a large uncertainty [8, 9]. Furthermore, the $H \to b\bar{b}$ decay mode has yet to be observed for the vector boson fusion and gluon-gluon fusion production mechanisms. Finally, new physics could be accessible through observation of highly boosted Higgs bosons produced via gluon-gluon fusion, which is sensitive to possible anomalous couplings and new particles in the top-quark loop [10, 11, 12]. Thus, this analysis aims to directly measure the coupling of the Higgs boson to bottom quarks, focusing on the gluon-gluon fusion and vector boson fusion production modes.

Part I of this dissertation describes the Standard Model of particle physics including the Higgs mechanism and motivates the search for highly boosted $H \to b\bar{b}$ production and decay. Part II describes both the LHC and the ATLAS detector located at CERN. Lastly, Part III presents the analysis and the boosted Higgs boson measurement results.

Part I

Theoretical Motivations and the Standard Model

Chapter 2

The Standard Model and Beyond

The Standard Model (SM) of Particle Physics formalizes the force laws that describe the observed behavior of all known particles in the universe. Its formulation is a collection of Quantum Field Theories (QFT) that describe the following interactions of elementary matter in Nature: the electromagnetic force, the weak nuclear force and the strong nuclear force. Gravity is noticeably absent as currently there is no viable quantum theory for observed gravitational effects. The Glashow-Salam-Weinberg (GSW) [13, 14, 15] theory includes Quantum ElectroDynamics (QED) and describes the unification of the electromagnetic and weak forces, while Quantum ChromoDyanmics (QCD) [16] describes the strong force. These theories form the following symmetry group of the Standard Model:

$$\underbrace{\mathrm{SU_C}(3)}_{\mathrm{QCD}} \otimes \underbrace{\mathrm{SU_L}(2) \otimes \mathrm{U_Y}(1)}_{\mathrm{GSW}} . \tag{2.1}$$

The gauge principle states that the SM Lagrangian and its predictions must be invariant under local transformations using an operator from any of these constituent groups. Thus, any theory must only include transformations and terms that maintain the local invariance of the complete Lagrangian. In particular, this requirement is violated by any attempt to include an explicit mass term for the gauge bosons of the GSW electroweak model and for all fermions. Around 1960 a possible solution to this lack of mass was proposed in the form of the spontaneous breaking of the electroweak symmetry, now known as the Higgs mechanism. The following sections go into more detail about the Lagrangian formalism of the Standard Model, QCD, GSW electroweak unification and this recently verified Higgs mechanism.

2.1 The Standard Model

At the turn of the 20th century, humanity's understanding of the constituent matter of the universe was limited to what could be seen with microscopes and implied from the observations of light and electricity, giving evidence for both the photon and the electron. In the first half of the century the field of subatomic physics was revealed with Rutherford's 1911 gold foil scattering experiment [17] followed by the observation of the wave-particle duality of nature with Compton's scattering experiment in 1923

[18]. These were the first steps towards a particle physics representation of nature. In the second half of the century experiments delved deeper to discover that the nucleus contained structure, and thus the SM was developed to include the complex mechanics of quarks and gluons [19]. With the discovery of the Higgs in 2012, the Standard Model has become even more firmly established, as can be seen in the high level of agreement between theory and experiment in Figure 2.1.

The QCD and GSW theories predict two classes of particles - fermions and bosons - shown in Figure 2.2. These particles represent the quanta of the quantum fields of the Standard Model and the mediators of the fundamental forces of Nature.

2.1.1 Bosons

The spin-1 particles are known as the vector gauge bosons and are the force carriers of the SM. The most commonly known is the electromagnetic force's uncharged and massless photon (γ) which interacts with all particles charged under $U(1)_{\text{em}}$ and is often referred to as "light." The weak nuclear force is involved in nuclear interactions such as beta decays and is carried by 3 bosons, all of which have mass and couple to fermions. The $W^{\pm}$ bosons mediate the charged weak interaction and allow for flavor changing currents, while the Z boson mediates the neutral weak interaction. Finally there are 8 massless gluons which mediate the strong force and only interact with fermions that have a "color" charge such as the quarks contained inside the nucleons. The only spin-0 boson, the Higgs Boson (H) is the key to generating mass terms in the SM Lagrangian

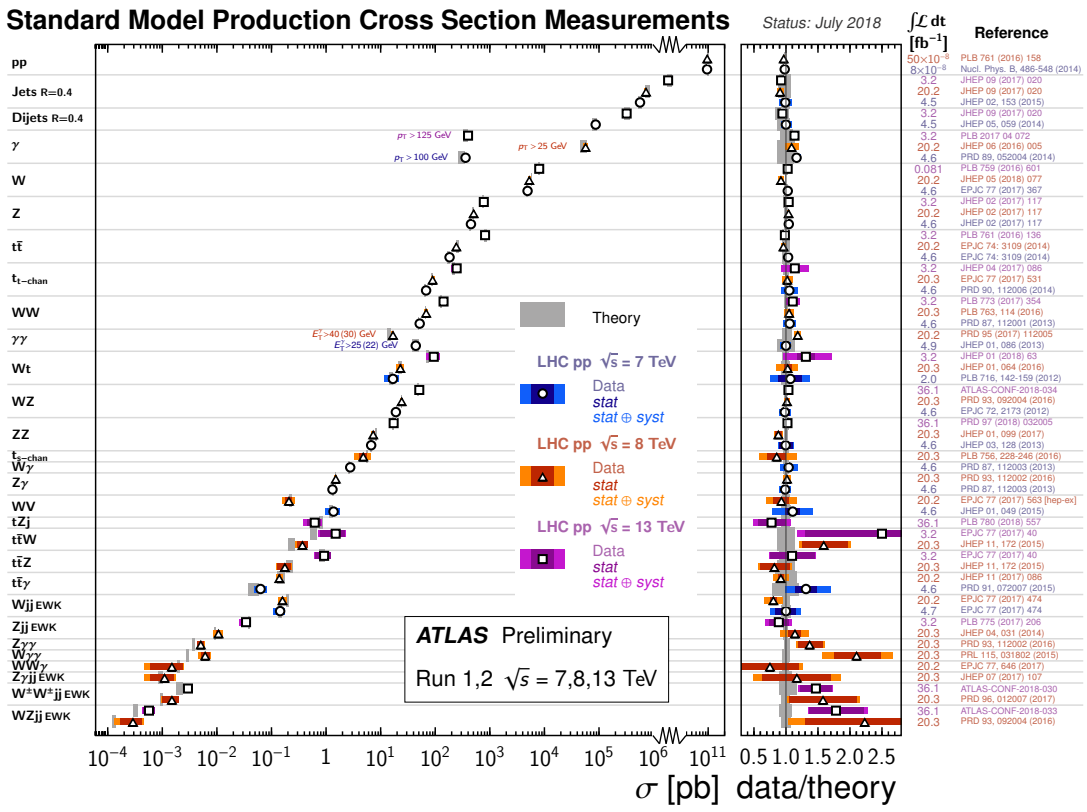


Figure 2.1: Summary of several Standard Model total and fiducial production cross section measurements, corrected for leptonic branching fractions, compared to the corresponding theoretical expectations [20]. All theoretical expectations were calculated at NLO or higher. The dark-color error bar represents the statistical uncertainty. The lighter-color error bar represents the full uncertainty, including systematics and luminosity uncertainties. The data/theory ratio, luminosity used and reference for each measurement are also shown. Uncertainties for the theoretical predictions are quoted from the original ATLAS papers. They were not always evaluated using the same prescriptions for PDFs and scales.

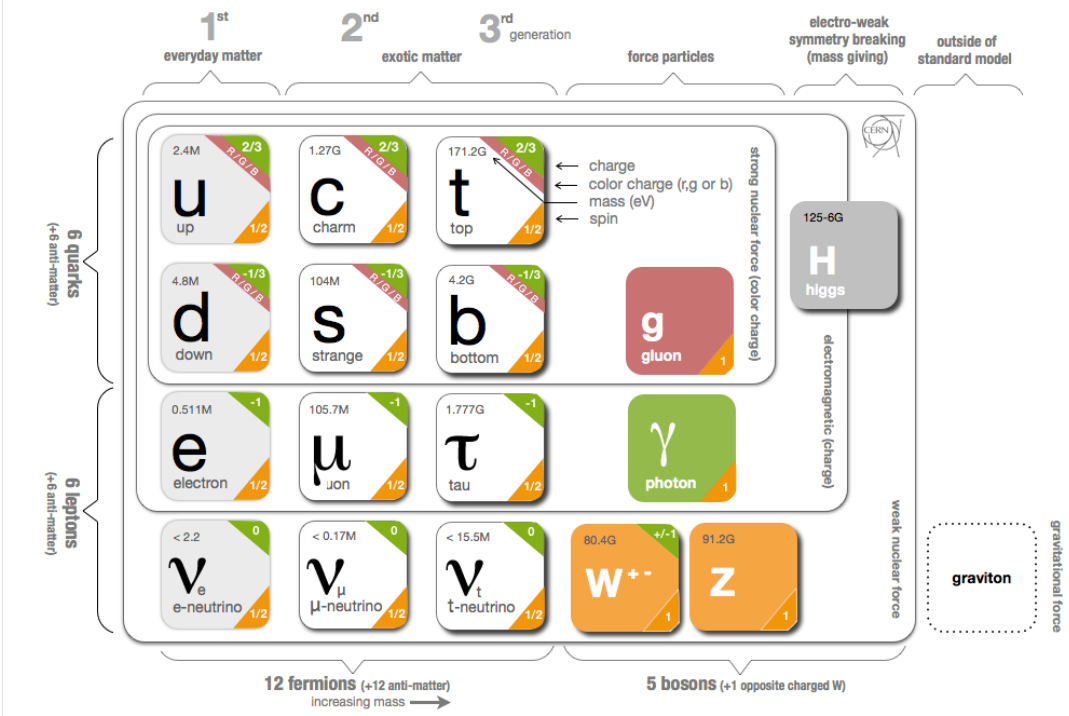


Figure 2.2: Table of all observed fundamental particles of the Standard Model. [21]

for the massive gauge bosons and for fermions. This is done through the so-called Higgs mechanism [22] and is discussed in more detail in Section 2.4.

2.1.2 Fermions

The spin-1/2 particles can be further broken up into two distinct families of particles, the leptons and the quarks, both of which contain three "generations" each with an "up"- and "down"-type particle where "up" and "down" differentiate the two components of the weak isospin doublet. The "down"-type leptons are the electrically charged electron (e), muon (μ) and tau (τ) while the "up"-type are their electrically neutral counterparts ν_e, ν_μ, ν_τ. The "up"-type quarks are the up (u), charm (c), and top (t), each with a $+2/3$ electric charge, while the "down"-type quarks are the down (d), strange (s), and bottom (b), all of which have a $-1/3$ electric charge. Each quark carries a "color" charge thus allowing them to couple to gluons and participate in strong force interactions. Due to the observed color confinement of the strong force these quarks are only observed in colorless bound states known as "mesons" (1 quark and 1 anti-quark) and "baryons" (3 quarks or anti-quarks). All of the above fermions have an anti-particle partner with the opposite weak isospin.

2.2 Electroweak Unification

In the SM the electromagnetic and weak forces are unified by GSW theory into the electroweak interaction which is represented by the $SU(2)_L \times U(1)_Y$ gauge group [**Goldstone:1962es**, 13, 15]. The L represents the experimental observation that the weak interaction, and thus the $SU(2)$ transformation, only acts on left-handed particle states. The Y indicates that this is the $U(1)$ symmetry for the weak hypercharge Y instead of the electromagnetic charge. The particle states for these interactions are solutions to the Dirac equation and are represented as isospin doublets ($\mathbf{\Psi}_L$) for the left-handed states, and as isospin singlets (Ψ_R) for the right-handed states. Thus a general transformation from the electroweak gauge group applied to the left-handed doublet is represented as

$$\mathbf{\Psi}_L \rightarrow \mathbf{\Psi}_L^{'} = \exp\left(\underbrace{ig^{'}\frac{Y}{2}\zeta(x)}_{U(1)_Y} + \underbrace{ig_W \boldsymbol{\alpha}(x)\cdot\mathbf{T}}_{SU(2)_L}\right)\mathbf{\Psi}_L. \tag{2.2}$$

For the right-handed singlet the $SU(2)_L$ transformation does not contribute, so the transformation is

$$\Psi_R \rightarrow \Psi_R^{'} = \exp\left(\underbrace{ig^{'}\frac{Y}{2}\zeta(x)}_{U(1)_Y}\right)\Psi_R. \tag{2.3}$$

Here the local gauge transformations have introduced space-time dependent terms $\boldsymbol{\alpha}(x)$ and $\zeta(x)$ into the electroweak Lagrangian. Due to the derivatives contained within

the kinetic term of this Lagrangian, this new configuration would introduce additional terms, thus violating the required local gauge invariance. These additional terms can be removed by replacing the standard derivative (∂_μ) with the covariant derivative (D_μ), as seen in Equation (2.4) for the left-handed states and Equation (2.5) for the right-handed states.

$$D_\mu = \partial_\mu - \underbrace{\frac{1}{2} i g^{'} B_\mu Y}_{U(1)_Y} - \underbrace{\frac{1}{2} i g_W \mathbf{W}_\mu \cdot \boldsymbol{\tau}}_{SU(2)_L} \tag{2.4}$$

$$D_\mu = \partial_\mu - \underbrace{\frac{1}{2} i g^{'} B_\mu Y}_{U(1)_Y} \tag{2.5}$$

This introduces two new gauge fields: the weak hypercharge field B_μ and the charged weak fields $\boldsymbol{W_\mu}$ as well as the associated coupling constants $g^{'}, g_W$, the hypercharge operator Y, and the $SU(2)$ generators $\boldsymbol{\tau}$. The transformation properties of these new fields are given by

$$\boldsymbol{W}_\mu(x) \rightarrow \boldsymbol{W}^{'}_\mu(x) = \boldsymbol{W}_\mu + \frac{1}{g_W} \partial_\mu \boldsymbol{\alpha}(x) + g_W \boldsymbol{W}_\mu(x) \times \boldsymbol{\alpha}(x) \tag{2.6}$$

$$B_\mu \rightarrow B^{'}_\mu = B_\mu + \frac{1}{g^{'}} \partial_\mu \zeta(x) \tag{2.7}$$

The form of these fields is chosen such that the final Lagrangian is invariant under

$SU(2)_L \times U(1)_Y$ transformations, and thus gauge invariance has been restored for the kinetic term of the electroweak Lagrangian. Inserting these new definitions into the Lagrangian for the field Ψ that satisfies the free-particle Dirac equation gives

$$\mathcal{L} = i\bar{\boldsymbol{\Psi}}_L \gamma^\mu \left(\partial_\mu - \frac{1}{2} i g^{'} B_\mu Y - \frac{1}{2} i g_W \mathbf{W}_\mu \cdot \boldsymbol{\tau} \right) \boldsymbol{\Psi}_L + i\bar{\Psi}_R \gamma^\mu \left(\partial_\mu - \frac{1}{2} i g^{'} B_\mu Y \right) \Psi_R \quad (2.8)$$

Next the gauge field self-interaction and mass terms are constructed

$$\mathcal{L} = -\frac{1}{4} \boldsymbol{F}_{\mu\nu} \boldsymbol{F}^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \frac{1}{2} M_W^2 \boldsymbol{W}_\mu \boldsymbol{W}^\mu + \frac{1}{2} M_B^2 B_\mu B^\mu \quad (2.9)$$

where the field tensors $\boldsymbol{F}^{\mu\nu}$ and $B^{\mu\nu}$ are defined to be

$$\boldsymbol{F}^{\mu\nu} = \partial^\mu \boldsymbol{W}^\nu - \partial^\nu \boldsymbol{W}^\mu + g_W \boldsymbol{W}^\mu \times \boldsymbol{W}^\nu \quad (2.10)$$

$$B^{\mu\nu} = \partial^\mu \boldsymbol{B}^\nu - \partial^\nu \boldsymbol{B}^\mu \quad (2.11)$$

The kinetic terms in Equation (2.9) are invariant under gauge transformations, but simply substituting Equation (2.6) or Equation (2.7) into the mass terms shows that these terms violate gauge invariance. This would imply that $M_W = 0$ and $M_B = 0$, in direct contradiction of the observed masses of the weak gauge bosons. This issue

arises again for fermion mass terms, as is evident below for the bottom quark field (b) expanded in its chiral basis.

$$m_b\bar{b}b = m_b \begin{pmatrix} b_R^\dagger & b_L^\dagger \end{pmatrix} \begin{pmatrix} b_L \\ \\ b_R \end{pmatrix} = m_b(b_R^\dagger b_L + b_L^\dagger b_R) \tag{2.12}$$

Remembering that the left- and right-handed components of the electroweak interaction transform differently shows that this mixture of right- and left- handed fields violates gauge invariance. This again forces the conclusion that $m_b = 0$, in contradiction to the observation that the bottom quark does indeed have mass. As mentioned in Section 2.1.1 the resolution to these mass mysteries lies in the Higgs mechanism discussed in Section 2.4.

2.3 Quantum Chromodynamics

Quantum Chromodynamics is the continuation of the mathematical framework established by the electroweak formalism in Section 2.2, this time for the strong force described by the $SU(3)_C$ gauge group, where C represents the "color" charge of QCD [16]. This color charge doesn't imply actual visible color, but is useful as an analogy to the visible spectrum where a combination of red, green, and blue generates white. For QCD the combination of red, green, and blue color charges can result in a colorless object. As mentioned in Section 2.1.2, the quarks have a color (anti-color) charge defining color

triplet field which transforms under the general $SU(3)$ transformation as

$$q = \begin{pmatrix} q_r \\ q_g \\ q_b \end{pmatrix} \rightarrow q^{'} = exp\left(ig_s \sum_{k=1}^{8} \eta_k(x) \frac{\lambda_k}{2} \right) q \tag{2.13}$$

Here the λ_k are the Gell-Mann generators for $SU(3)$, $\eta(x)_k$ is the space-time dependency for each generator, and g_s is the strong coupling constant. As with the electroweak Lagrangian, the introduction of these space-time dependent terms adds new terms into the kinematic portion of the Lagrangian and spoils the gauge invariance. Again, a covariant derivative is introduced

$$D_\mu = \partial_\mu - ig_s G_\mu^k \frac{\lambda_k}{2} \tag{2.14}$$

to restore gauge invariance. The G_μ^k are the new fields introduced for the 8 gluons. These new fields transform under $SU(3)$ as

$$G_\mu^k \rightarrow G_\mu^{'k} = G_\mu^k + \partial_\mu \eta_k(x) + g_s f_{klm} \eta_l(x) G_\mu^m \tag{2.15}$$

Given these definitions the QCD Lagrangian ($\mathcal{L}_{QCD}$) can be constructed as

$$\mathcal{L}_{QCD} = \bar{q}(i\gamma_\mu D^\mu - m_q)q - \frac{1}{4}G_k^{\mu\nu}G_{k\mu\nu} \tag{2.16}$$

where the gluon field tensor $G_k^{\mu\nu}$ is defined as

$$G_k^{\mu\nu} = \partial^\mu G_k^\nu - \partial^\nu G_k^\mu + g_s f_{klm} G_l^\mu G_m^\nu. \tag{2.17}$$

The strong force is peculiar in that experiments observe only colorless objects in the form of bound states of quarks known as hadrons. Qualitatively, the potential between quarks in a bound state (meson or baryon) gets stronger with separation, unlike the other forces. At the point where the system would separate into color-charged objects, it becomes energetically favorable to produce a quark/anti-quark pair in a process known as hadronization. In other words, attempting to separate a bound quark state into its colored constituents simply results in new colorless bound states. This requirement of colorless objects by the strong force is known as color confinement. For highly energetic strong interactions at hadron colliders the result is an expanding shower of hadronizing quarks and gluons and their decay products known as a jet.

2.4 The Higgs Mechanism

The Higgs mechanism is the system by which the gauge bosons and fermions gain mass through the spontaneous breaking of the electroweak symmetry of the Higgs potential

[1, 3, 22]. This section will also discuss briefly the couplings of the Higgs boson to massive particles, as well as its self couplings.

2.4.1 Electroweak Symmetry Breaking

The Higgs field is expressed as a complex doublet, $\mathbf{\Phi}$, and thus has four components defined as

$$\mathbf{\Phi}(x) = \begin{pmatrix} \phi^+ \\ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ \\ \phi_3(x) + i\phi_4(x) \end{pmatrix}. \tag{2.18}$$

The four components of this field each represent a degree of freedom which become the longitudinal polarizations of the $W^{\pm}, Z$ gauge bosons and the Higgs boson. The resulting Lagrangian for the Higgs includes a kinetic term (K) as well as the Higgs potential (V), all of which are invariant under the electroweak gauge symmetry $SU(2)_L \times U(1)_Y$. The definition is

$$\mathcal{L}_{\text{Higgs}} = \underbrace{(D_\mu \mathbf{\Phi})^\dagger D^\mu \mathbf{\Phi}}_{\text{K}} - (\underbrace{\mu^2 \mathbf{\Phi}^\dagger \mathbf{\Phi} + \lambda(\mathbf{\Phi}^\dagger \mathbf{\Phi})^2}_{\text{V}}). \tag{2.19}$$

Here the $\mu^2 < 0$ and $\lambda > 0$ are constrained such that the potential forms a ring stable minima. The shape of this potential is shown in Figure 2.3 and is often described as the "Mexican-hat" or "wine-bottle" potential.

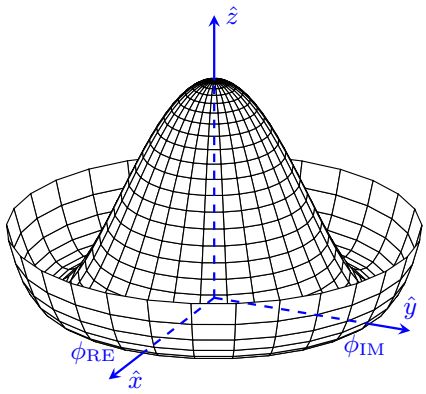


Figure 2.3: A lower-dimensionality representation of the shape of the Higgs potential. The central peak represents a rotationally symmetric unstable state, while the trough represents the infinite number of minima that can be selected upon the spontaneous breaking of the symmetry.

The normed value of this minima can be calculated by taking the derivative of V with respect to $\mathbf{\Phi}$ and setting it equal to 0. This value, also known as the vacuum expectation value (vev) has been found to be $v \equiv \sqrt{-\mu^2/\lambda} = 246$ GeV. The arbitrary vev of the ground state Higgs field is acquired when the symmetry of the Higgs potential is spontaneously broken. For ease of calculation the coordinate system is oriented such that

$$\langle \mathbf{\Phi}(x) \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.20}$$

Next small perturbations around the minimum of the Higgs potential are parametrized as:

$$\langle \mathbf{\Phi}(x) \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \exp\left(i \frac{\tau^i}{2} \theta^i(x) \right) \tag{2.21}$$

Here the real scalar field $h(x)$ corresponds to radial perturbations of the minima, and the three $\theta^i(x)$ are the Nambu-Goldstone fields with values determined by the choice of gauge. Choosing the unitary gauge of $\theta^i(x) = 0$ and expanding the kinetic term of Equation (2.19) around the vev gives

$$\mathcal{L}_{\text{Higgs,K}} = \frac{g^2 v^2}{8} \left((W_\mu^-)^\dagger W^{-\mu} + (W_\mu^+)^\dagger W^{+\mu} \right) + \frac{1}{2} \begin{pmatrix} W_\mu^{3\dagger} & B_\mu^\dagger \end{pmatrix} \boldsymbol{M}^2 \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix} + \dots \tag{2.22}$$

Here the first term is the physical mass term for the $W^\pm$ bosons where these charge eigenstates have been constructed out of the $W^{1,2}$ fields as such $W^\pm = \frac{1}{\sqrt{2}}(W^1 \mp iW^2)$. The second term represents the mixture of the W^3 and B fields through the mass matrix $\boldsymbol{M}$. Diagonalizing this matrix (M_D) and identifying the mass eigenstates gives the physical fields of the photon (γ) and the Z boson

$$\boldsymbol{M}_D^2 = \begin{pmatrix} 0 & 0 \\ 0 & \frac{v^2}{4}(g_W^2 + g'^2) \end{pmatrix} \tag{2.23}$$

The upper left diagonal element corresponds to the massless photon while the lower right

diagonal element gives the mass of the massive Z boson. This results in the following masses for the four electroweak bosons

$$m_W = \frac{1}{2} g_W v \quad , \quad m_Z = \frac{1}{2} v \sqrt{g_W^2 + g'^2} \quad , \quad m_\gamma = 0 \tag{2.24}$$

The masses of the $W^\pm$ and Z gauge bosons can be related through the Weinberg mixing angle defined as

$$\theta_W = \cos^{-1}\left(\frac{g_W}{\sqrt{g_W^2 + g'^2}}\right) \rightarrow m_Z = \frac{m_W}{\cos\theta_W} \tag{2.25}$$

Using this definition one can write out the exact mixture of B and W^3 that make up the photon and Z boson as

$$\gamma = \cos(\theta_W)B + \sin(\theta_W)W^3 \tag{2.26}$$

$$Z = -\sin(\theta_W)B + \cos(\theta_W)W^3 \tag{2.27}$$

2.4.2 Fermion Mass Terms

Section 2.2 shows how a simple fermion mass terms violate gauge invariance due to the mixing of the left and right chiral states. The Higgs mechanism, however, allows for

a gauge-invariant method of generating mass terms through the Yukawa coupling of the Higgs field to the fermion fields. An example is the Yukawa coupling term for a quark doublet ($\mathbf{\Psi}_L$) and singlet (Ψ_R) coupling to the Higgs field ($\mathbf{\Phi}$) after spontaneous symmetry breaking, with the form shown in Equation (2.21), when the unitary gauge $\Phi^i(x) = 0$ is chosen.

$$\mathcal{L}_{\text{Yukawa}} = -g_b \left[\bar{\mathbf{\Psi}}_{\boldsymbol{L}} \mathbf{\Phi} \Psi_R + \bar{\Psi}_R \mathbf{\Phi}^\dagger \mathbf{\Psi}_{\boldsymbol{L}} \right] \tag{2.28}$$

$$= -\frac{g_b}{\sqrt{2}} \left[\begin{pmatrix} \bar{t} & \bar{b} \end{pmatrix}_L \begin{pmatrix} 0 \\ v+h \end{pmatrix} b_R + \bar{b}_R \begin{pmatrix} 0 & (v+h) \end{pmatrix} \begin{pmatrix} t \\ b \end{pmatrix}_L \right] \tag{2.29}$$

$$= -\underbrace{\frac{g_b}{\sqrt{2}} v}_{m_b} \left(\bar{b}_L b_R + \bar{b}_R b_L \right) - \underbrace{\frac{g_b}{\sqrt{2}}}_{g_{b,h}} h \left(\bar{b}_L b_R + \bar{b}_R b_L \right) \tag{2.30}$$

In this way mass terms are generated for the fermion field, and the gauge invariance of the Lagrangian is maintained via the proper combination of covariant derivatives and fields. This operation also produces the second term which represents the coupling of the bottom quark to the Higgs itself and thus gives the form of its coupling constant $g_{b,h}$. Using this newly found mass of the bottom quark m_b, the coupling can be written as

$$g_{b,h} = \frac{g_b}{\sqrt{2}} = \frac{m_b}{v}. \tag{2.31}$$

Thus the coupling of the Higgs boson to a fermion is proportional to the mass of the fermion itself.

2.4.3 The Higgs Boson

It has been shown that the Higgs mechanism properly mixes the gauge fields to provide the correct gauge-invariant mass terms, and it also properly combines the left and right chiral states of fermions to produce their mass terms. During spontaneous symmetry breaking of the electroweak potential three of the four degrees of freedom in the Higgs doublet (Equation (2.18)) become Goldstone bosons. Since this theory is gauged these Goldstone bosons are "eaten" by the 3 gauge bosons $W^{\pm}$ and Z to form their longitudinal components thus give them their mass. However, the final broken degree of freedom is absorbed by the new massive scalar particle, the Higgs boson [2].

Focusing on the Higgs potential term (V) of Equation (2.19) and substituting in the definition for $\mathbf{\Phi}$ given in Equation (2.21) gives

$$\mathcal{L}_{\text{Higgs,V}} = \frac{1}{2}\mu^2 v^2 - \mu^2 h^2 + \lambda v h^3 + \frac{1}{4}\lambda h^4 \tag{2.32}$$

The first term is constant and thus can be ignored. The second term is the mass term for the Higgs boson, $m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda}v$. Because $h = h(x)$ was used for small radial perturbations of the Higgs field the Higgs boson can be identified as a radial excitation of the Higgs field. Finally, the third and fourth terms represent the Higgs boson self-couplings. With these couplings and mass terms in hand the next step is to verify this theory experimentally as discussed next in Chapter 3.

Chapter 3

Boosted Higgs at the LHC

The Higgs mechanism, as described in Section 2.4, solves the problem of generating gauge boson and fermion mass terms while also maintaining gauge invariance. To understand the search for the SM Higgs boson requires the discussion of how to produce and detect it. In order to gather sufficient data to validate the theory a collider is required that is capable of putting enough energy into a collision to rapidly produce Higgs bosons for stud. To this end the Large Hadron Collider (LHC) discussed in Chapter 4 was laboriously designed, funded, and constructed by the largest international collaboration of collider physicists and engineers on the planet. This chapter will include a discussion of the relevant Higgs boson production mechanisms available at the LHC and the various decay modes of the Higgs boson that are used to measure its properties including the final state that is the focus of this dissertation.

3.1 Higgs Production Mechanisms

At the LHC the dominant production mechanisms for the Higgs boson in order of decreasing cross section are: gluon-fluon fusion (ggF), vector boson fusion (VBF), vector boson associated production or "Higgsstrahlung" (VH), and associated production with $t\bar{t}$ ($t\bar{t}H$) and $b\bar{b}$ ($b\bar{b}H$). The cross sections for the signatures of these processes with associated theoretical uncertainties for each are shown as a function of the center-of-mass energy $\sqrt{s}$ in Figure 3.1, and the leading order (LO) Feynman diagrams can be seen in Figure 3.2.

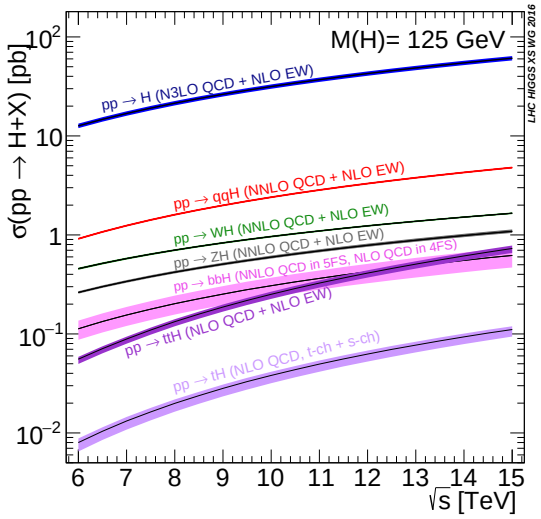
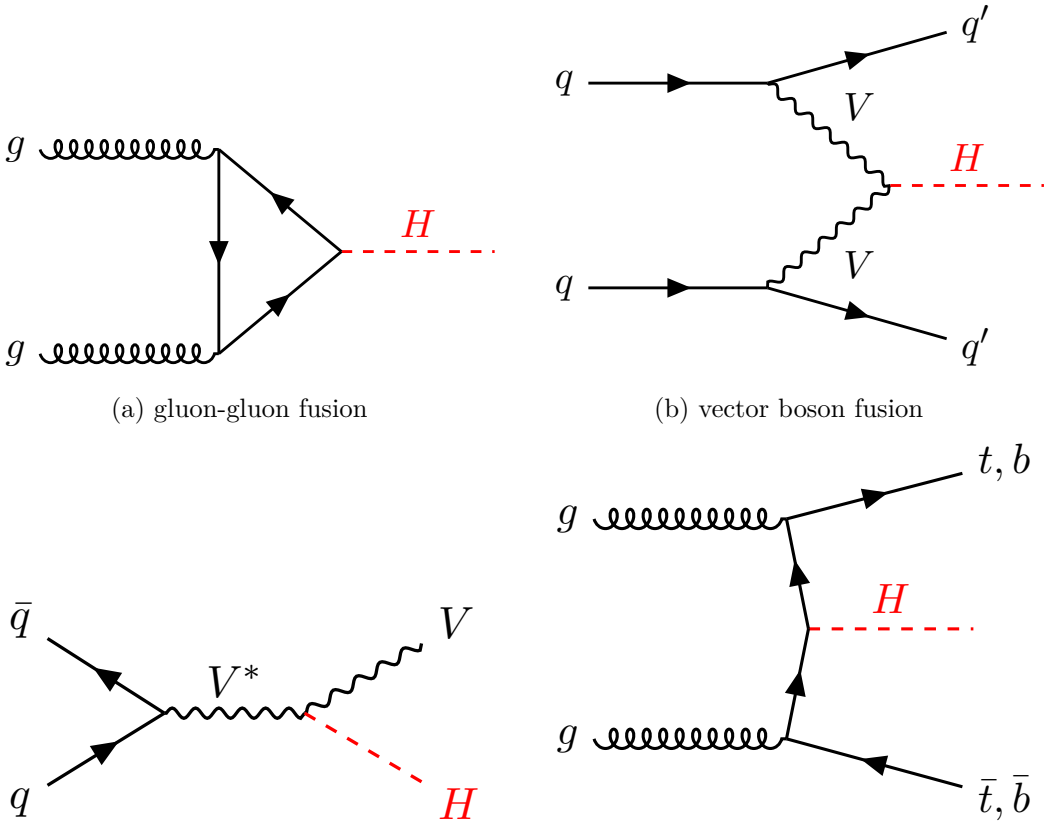


Figure 3.1: Cross sections for the production of the SM Higgs boson signatures as a function of the center of mass energy ($\sqrt{s}$) at the LHC [23]. In order of decreasing cross section: the ggF process signature is H, the VBF process signature is qqH, the VH process signature is split into WH and ZH, and the $b\bar{b}$ / $t\bar{t}$ signatures are $t\bar{t}H$ / $b\bar{b}H$.

The LO Feynman diagram contains the least number of vertices, and thus coupling constants, making it the largest contribution to the cross section calculation. Adding

an additional vertex, represents a higher order correction to the LO calculation known as next-to-leading order (NLO) where each additional vertex adds a "next" (NNLO, N3LO, etc.). For reference, the current best estimates of the production cross sections for the leading production mechanisms are detailed in Table 3.1.



(a) gluon-gluon fusion

(b) vector boson fusion

(c) associated production

(d) $t\bar{t}$ ($t\bar{t}H$) and $b\bar{b}$ ($b\bar{b}H$)

Figure 3.2: Feynman diagrams representing the dominant Higgs production modes at the LHC.

The dominant Higgs production mechanism at hadron colliders is gluon-gluon fusion. This may seem strange as gluons are massless and thus do not couple directly to the Higgs boson. Instead the gluons indirectly couple to the Higgs boson via a quark loop.

Table 3.1: SM Higgs boson production cross sections in units of pb for $m_H = 125$ GeV in pp collisions for the current LHC center-of-mass energy, $\sqrt{s} = 13$ TeV. The predictions for the ggF channel include the latest N3LO results, which have reduced theoretical uncertainties by a factor around 2 compared to the NNLO results [23].

$\sqrt{s}$ (TeV)	ggF	VBF	WH	ZH	$t\bar{t}H$	Total (pb)
13	$48.6^{+5\%}_{-5\%}$	$3.78^{+2\%}_{-2\%}$	$1.37^{+2\%}_{-2\%}$	$0.88^{+5\%}_{-5\%}$	$0.50^{+9\%}_{-13\%}$	55.1

As discussed in Section 2.4.2, the coupling of a fermion is proportional to m_f, so the dominant contribution to this quark loop comes from the top quark. It is important to note that the ggF cross section in Table 3.1 is inclusive in number of final state jets and thus will include diagrams like the one shown in Figure 3.3. There has been considerable effort to calculate exclusive H + jet(s) production process at NLO and NNLO [24] for use in analysis where there is an explicit requirement for an associated jet such as the one presented in this dissertation in Part III.

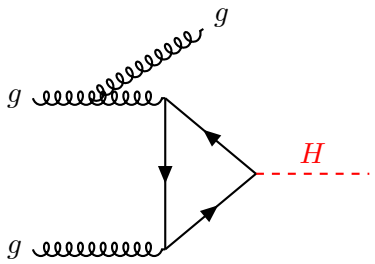


Figure 3.3: Feynman diagram for ggF Higgs + jet production.

The second-largest cross section for Higgs production at the LHC comes from the VBF mechanism. In VBF the initial state quarks scatter via the exchange of a $W^{\pm}$ or Z boson which subsequently radiates the Higgs boson. Unlike ggF this production mechanism scatters the initial state quarks which allows them to be observed as part of the interaction. The presence of these extra quarks makes these interactions easier to

select during analysis.

The third-largest cross section for Higgs production is in association with a vector boson. The cross section for this is even smaller than the above two, but remains important due to the easily selected signature of the decaying vector boson. The largest background at the LHC is multijet events coming from interactions that produce strong force objects. Thus the leptons from the boson's decay act as a discriminator against this multijet background, greatly reducing its effect on sensitivity. Note that the W/Z can also decay hadronically giving a final state that looks like H + jet.

The lowest cross section of the four methods discussed is the production of the Higgs boson in association with either $b\bar{b}$ or $t\bar{t}$. This channel is important because it allows direct measurement of the ttH coupling, unlike the ggF method where the quark in the loop is never directly observed.

3.2 Parton Distribution Function

The LHC collides protons, however in the Feynman diagrams in Figure 3.2 it is quarks and gluons (a.k.a partons) that produce these fundamental interactions. This is an indicator that when the production cross section is calculated for a process at the LHC, one must not only consider the hard-scatter probability of the specific diagram, but also consider the composition of the proton itself. Furthermore, the calculation must consider the fraction of the total momentum of the proton held by each of its constituent

partons. This concept is described by Parton Distribution Functions (PDFs) which give the probability that the indicated parton carries momentum fraction x of the proton when probed at energy scale Q. An example PDF for $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 is shown in Figure 3.4.

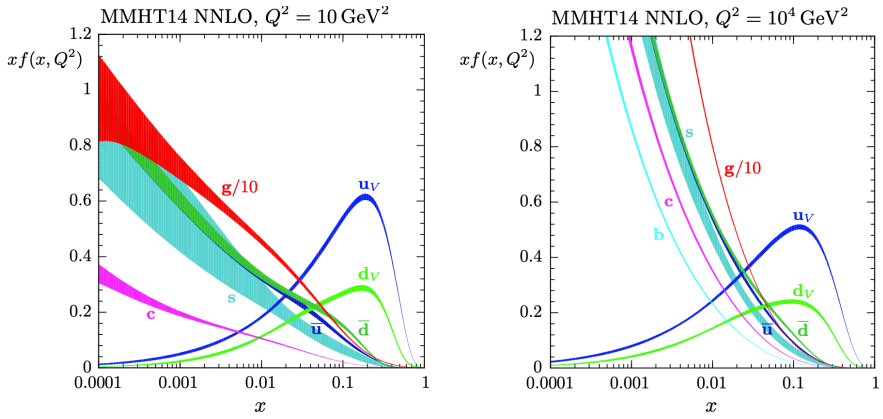
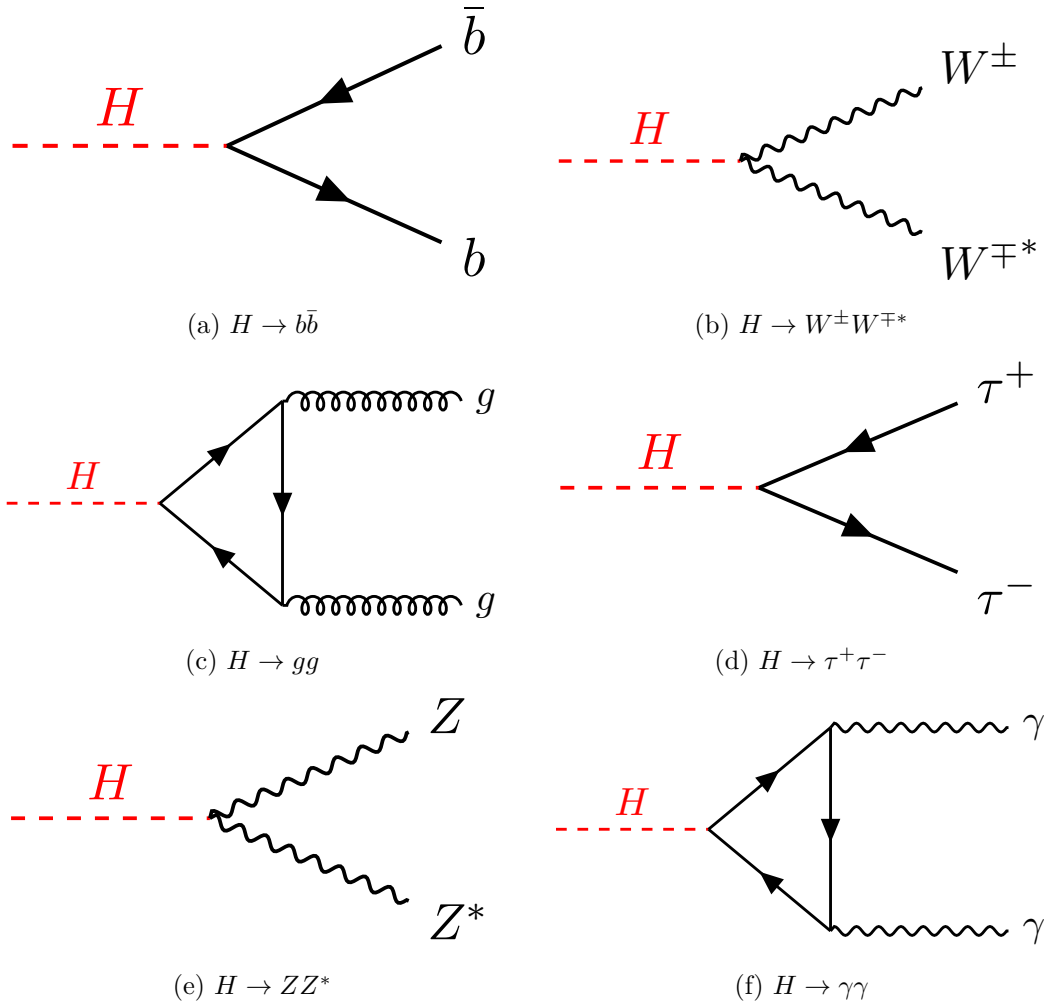


Figure 3.4: MMHT2014 NNLO PDFs at $Q^2 = 10$ GeV^2 and $Q^2 = 10^4$ GeV^2 with associated 68% confidence-level uncertainty bands [25]. The colored regions indicate the probability of finding the labeled parton with a momentum fraction given along the x axis. As expected the valence quarks contain the largest fraction of the momentum while the gluons are more likely to carry smaller fractions of the total momentum. Note that as Q^2 increases the contributions from sea quarks increases.

3.3 Branching Ratios

The coupling of the SM Higgs with the gauge bosons and fermions has been shown to give these particles their mass; however, it also means that the Higgs boson can decay into all of these particles. In order of most to least likely final states of a Higgs decay is the decay to a pair of b-quarks ($b\bar{b}$), a pair of weak vector bosons where one is off-shell (VV^*), two gluons (gg), a duo of tau leptons ($\tau^+\tau^-$), or a pair of photons ($\gamma\gamma$). Similar

to the ggF production mechanism discussed in Section 3.1 the decays to massless gauge bosons (photons and gluons) are facilitated through loops of massive particles. The exact Feynman diagrams depicting the above processes are shown in Figure 3.5, while information about their branching ratios is detailed in Table 3.2.



(a) $H \to b\bar{b}$ (b) $H \to W^{\pm}W^{\mp *}$

(c) $H \to gg$ (d) $H \to \tau^{+}\tau^{-}$

(e) $H \to ZZ^{*}$ (f) $H \to \gamma\gamma$

Figure 3.5: Feynman diagrams representing the leading Higgs decay channels.

In Table 3.2 the order is determined by two distinct effects: the proportionality of the Higgs couplings to the mass of the decay product, and whether or not the rest mass of

Table 3.2: The branching ratios and the relative uncertainty for a Standard Model Higgs boson with $m_H = 125$ GeV [23].

Decay Channel	Branching Ratio	Relative Uncertainty
$H \to b\bar{b}$	5.84×10^{-1}	$^{+3.2\%}_{-3.3\%}$
$H \to W^+W^-$	2.14×10^{-1}	$^{+4.3\%}_{-4.2\%}$
$H \to \tau^+\tau^-$	6.27×10^{-2}	$^{+5.7\%}_{-5.7\%}$
$H \to ZZ$	2.62×10^{-2}	$^{+4.3\%}_{-4.1\%}$
$H \to \gamma\gamma$	2.27×10^{-3}	$^{+5.0\%}_{-4.9\%}$
$H \to Z\gamma$	1.53×10^{-3}	$^{+9.0\%}_{-8.9\%}$
$H \to \mu^+\mu^-$	2.18×10^{-4}	$^{+6.0\%}_{-5.9\%}$

the Higgs boson is sufficient to produce the two final state objects. Figure 3.6 shows that the cross section for $H \to WW$ grows as the mass of the Higgs boson gets closer to $2m_W$.

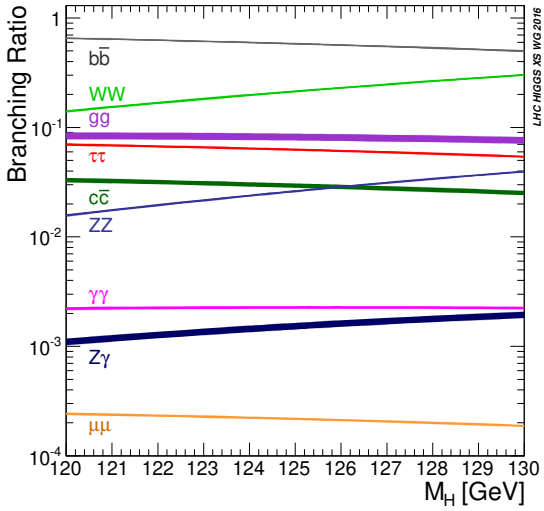


Figure 3.6: Branching ratios for the decay of the SM Higgs boson near $m_H = 125$GeV including theoretical uncertainty bands [23]

3.4 Evidence for the SM Higgs

Using the above information about predicted final states, the CMS and ATLAS collaborations analyzed 5 fb^{-1} of LHC Run 1 data, where $\sqrt{s} = 7$ and 8 TeV, to make measurements of the SM Higgs production cross sections and branching ratios [26]. The combined results of these studies can be seen in Figures 3.7 to 3.9. Given the uncertainties on the measurements, these results show good agreement between the predictions of the Standard Model and experiment, with all best fit values falling within 2σ of the SM theoretical prediction. Note that there is currently no measurement of the $H \to b\bar{b}$ decay mode for the ggF and VBF production processes.

3.5 Boosted Higgs

The strong agreement between the theoretical predictions for the SM Higgs boson and the experiment shown in Section 3.4 represents the fulfillment of a generation of incredible technological and theoretical achievement. The next step is to complete the set of Higgs coupling measurements, like adding ggF $H \to b\bar{b}$, and to push the search for deviations from the Standard Model that might hint at the new physics of mysteries like the matter / anti-matter asymmetry of the universe, dark matter, the particle nature of gravity, and dark energy.

Minimal Supersymmetric Standard Model (MSSM) and Standard Model Effective Field Theory (SMEFT) arguments for extensions of the SM suggest that precise measure-

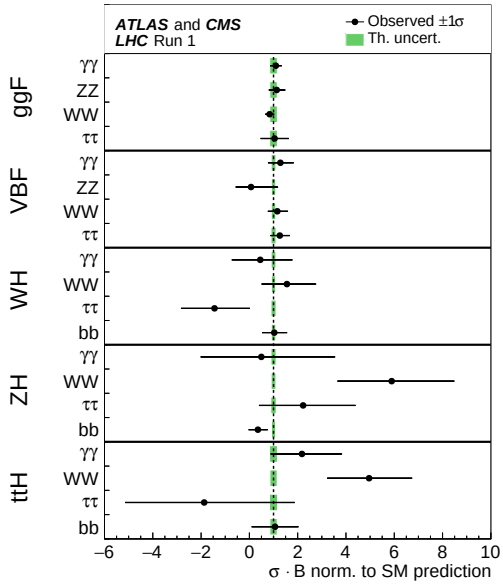


Figure 3.7: Best fit values of $\sigma_i \cdot B^f$ for each specific channel $i \to H \to f$, as obtained from the SM Higgs coupling parameterization with 23 parameters for the combination of the ATLAS and CMS measurements. The error bars indicate the 1σ intervals. The fit results are normalized to the SM predictions for the various parameters and the shaded bands indicate the theoretical uncertainties in these predictions. Only 20 parameters are shown because some are either not measured with a meaningful precision, in the case of the $H \to ZZ$ decay channel for the WH, ZH, and $t\bar{t}H$ production processes, or not measured at all and therefore fixed to their corresponding SM predictions, in the case of the $H \to b\bar{b}$ decay mode for the ggF and VBF production processes [26].

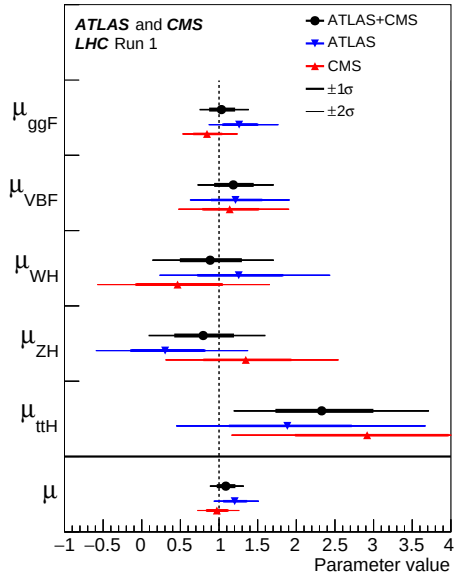


Figure 3.8: Best fit results for the production signal strengths for the combination of ATLAS and CMS data. Also shown are the results from each experiment. The error bars indicate the 1σ (thick lines) and 2σ (thin lines) intervals. The measurements of the combined global signal strength μ are also shown [26].

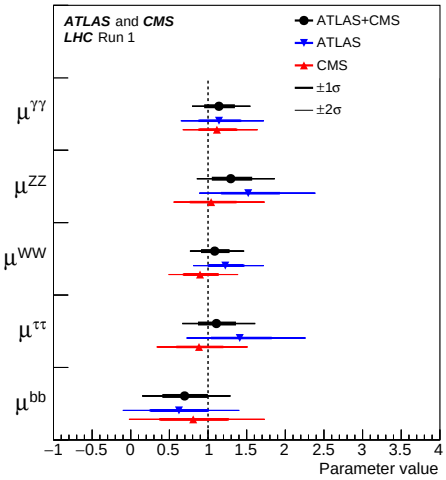


Figure 3.9: Best fit results for the decay signal strengths for the combination of ATLAS and CMS data. Also shown are the results from each experiment. The error bars indicate the 1σ (thick lines) and 2σ (thin lines) intervals [26].

ments of the shape of the momentum distribution for highly boosted (high momentum) Higgs bosons offer the opportunity to see the effects of new physics. For example, in MSSM the cross section for ggF production of Higgs bosons with $p_{\mathrm{T},H} \geq 500$ GeV could be enhanced or suppressed as a result of new heavy particles participating in the top quark loop [10]. The effect of these new particles on the observed Higgs boson p_{T} spectrum would be limited to the high momentum region where there is enough energy in the quark loop to allow these heavy particles to contribute significantly. Furthermore, SMEFT models predict modifications to the high p_{T} spectrum for boosted Higgs bosons as a result of new physics effects that would modify the top and bottom Yukawa couplings. These new physics effects are parameterized by Wilson coefficients which can vary the predicted cross section, and thus the predicted p_{T} spectrum, up and down while maintaining agreement with current experimental results [12]. Both of these examples would modify the effective ggH interaction and thus give an anomalous result compared to the SM. The effect on the production cross section for boosted Higgs through ggF can exceed 50% for $p_{\mathrm{T},H} \geq 500$ GeV [10, 11, 12, 27].

Searches for boosted Higgs bosons are also easier because the LHC produces a large number of soft (low momentum) QCD interactions as seen in Figure 2.1. Because of this a boosted signal is easier to differentiate from the common QCD interactions which fall off exponentially as a function of momentum. However, to achieve this boost, the Higgs must recoil off of a high energy jet or photon [28] produced through initial state radiation (ISR). This dissertation only considers strongly produced ISR due to the low

cross section for the $H + \gamma$ final state. Thus, the analysis presented in Part III searches for a final state signature of H + jet in the $H \to b\bar{b}$ decay mode. An example leading order Feynman diagram for this process is shown in Figure 3.10.

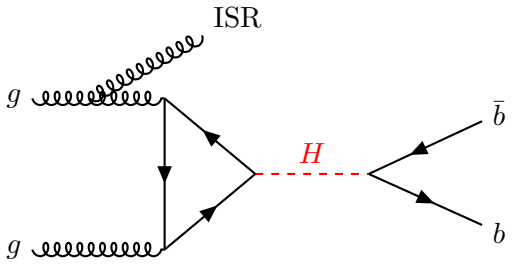


Figure 3.10: Feynman diagram for boosted Higgs decaying to $b\bar{b}$

As a result of this boost, the decay products of the Higgs boson and the hadronization products of the ISR become highly collimated as shown in Figure 3.11. The two pronged structure of the jet that results from the $H \to b\bar{b}$ decay provides a unique signature that can be used to differentiate the Higgs signal from other QCD processes.

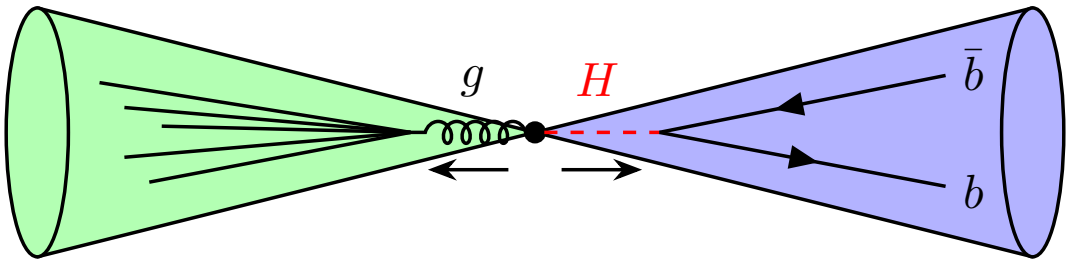


Figure 3.11: Cartoon showing collimated Higgs boson jet and ISR jet as a result of the large boost due to their mutual recoil [29].

In pursuit of the rich physics potential discussed above, an analysis of a boosted Higgs decaying to a $b\bar{b}$ pair with an associated ISR jet was undertaken and is discussed further in Part III. However, before giving more details about the ATLAS data analysis, it will be useful to describe the experimental apparatus and reconstruction methods.

Part II

Experimental Apparatus and Associated Facilities

Chapter 4

The Large Hadron Collider

Lying an average of 100 meters under the Swiss / French border, the 26.7 kilometer circumference Large Hadron Collider (LHC) is the world's largest piece of scientific apparatus built to date [30, 31]. This apparatus is used to produce proton and heavy ion collisions for observation by the four major experiments at the LHC: ATLAS, CMS, LHCb, and ALICE. The system was designed for a maximum center-of-mass energy of $\sqrt{s} = 14$ TeV and a peak instantaneous luminosity of $L = 10^{34}\text{cm}^{-2}\text{s}^{-1}$.

The first LHC workshop was held in 1984 in Lausanne at the European Organization for Nuclear Research (CERN) [32]. The nearly 30-year-old case for a machine that would push towards the discovery of the elusive Higgs Boson was presented using the existing CERN accelerator facilities and the Large Electron Positron (LEP) Collider tunnel. The proposal became reality on September 10, 2008 when the first proton beams were circulated, only to have calamity strike 9 days later in the form of a catastrophic

electrical fault [33]. The repairs and improvements lasted until November 2009 when the LHC restarted. Since then, this modern marvel has worked wonderfully and, as hoped, lead to the discovery of the Higgs Boson by the ATLAS and CMS collaborations, announced on July 4th, 2012 [6, 7, 34].

The following chapter provides a brief introduction to the world's most powerful accelerator starting with the little red bottle of hydrogen, and ending with the interaction point where protons collide at the highest energies ever produced by humans.

4.1 Particle Injection Chain

We begin with the most common element in the Universe, hydrogen, as the source of protons. A bottle of hydrogen gas provides 100 microsecond pulses of raw H_2 which is then injected into a Duoplasmatron [36]. There, a strong electric field and free electrons from a cathode ionize the molecule into bare H^+ - a proton! These protons are then accelerated by a 90 kV electric field, leaving the Duoplasmatron at 1.4% the speed of light ($\sim$4000 km/s) or, in particle physics units, $\sim$ 90 keV [37]. The bare protons are then fed into the accelerating Radio Frequency (RF) cavities of Linear Accelerator 2 (LINAC2). Inside, conductors charged by a powerful oscillating electromagnetic field accelerate the protons to an energy of 50 MeV. Along the way, small quadrupole magnets shape the proton bunch to ensure they remain in a tight beam. This pattern of acceleration with RF cavities and shaping/tuning with magnets is then repeated with CERN's first

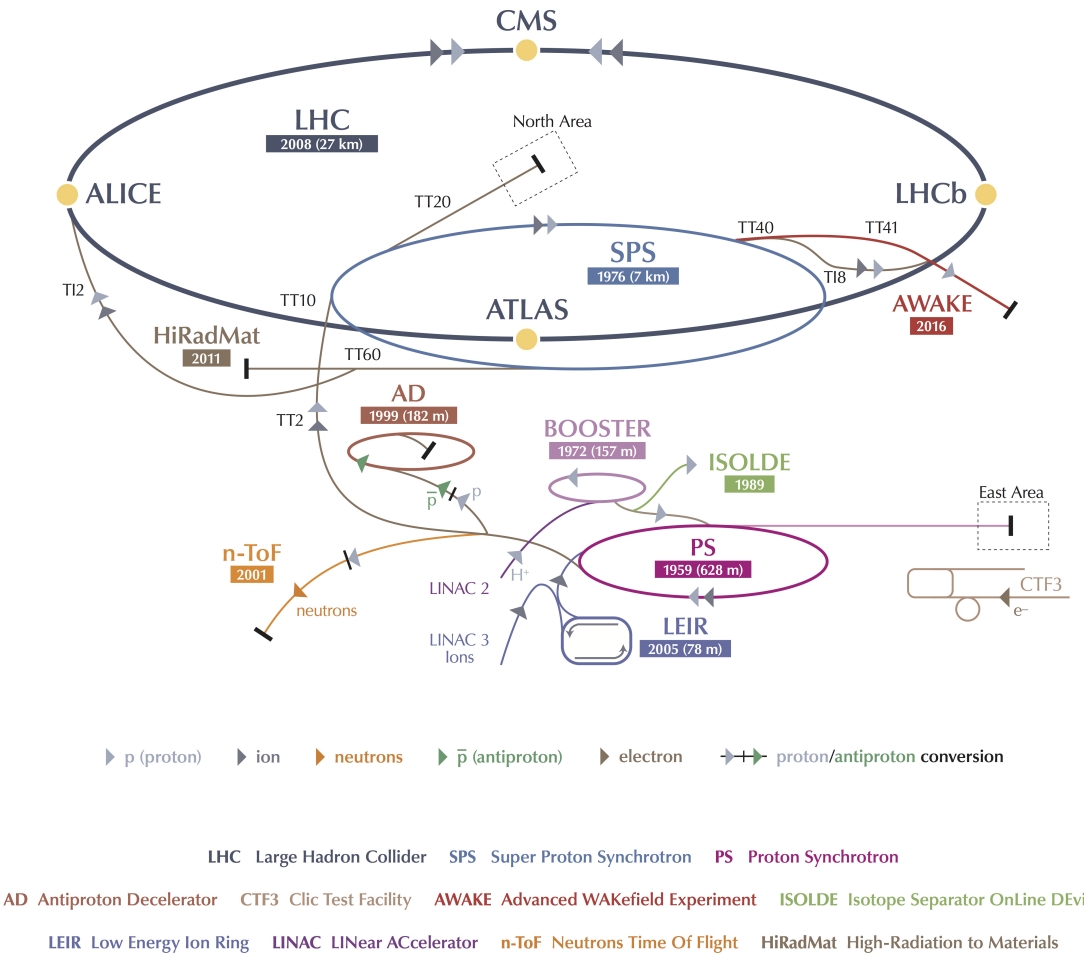


Figure 4.1: Sketch of the CERN accelerator complex [35].

synchrotron, the Proton Synchrotron (PS), rendering a 1.4 GeV proton beam. The final step before the LHC comes with the Super Proton Synchrotron where the same technologies are implemented to produce 450 GeV protons, ready for injection into the LHC. A diagrammatic representation of this chain can be seen in Figure 4.1.

In order to produce proton-proton collisions, the LHC uses two beams circulating in opposite directions. The beams are not continuous, but instead consist of bunches of around 120 billion protons with a spacing of 25 ns. The LHC circumference allows for 3564 bunches, however only 2808 are filled per beam due to safety requirements and injection limitations. Each beam takes 4 minutes and 20 seconds to fill and then 20 additional minutes for the protons to reach their maximum energy of 7 TeV, or 99.99999991% the speed of light! Under normal operating conditions these beams can be used for 8 to 10 hours.

4.2 LHC Layout and Design

While often depicted as a perfect circle the LHC is in reality an octagon with rounded edges, called arcs, as can be seen in Figure 4.2. The counter-circulating beams of protons are depicted in red and blue. These beams are focused and collided at the 4 dedicated interaction points at rates of up to 40 MHz. Two of these points are occupied by the ATLAS and CMS experiments, both of which are high luminosity, multi-purpose experiments.

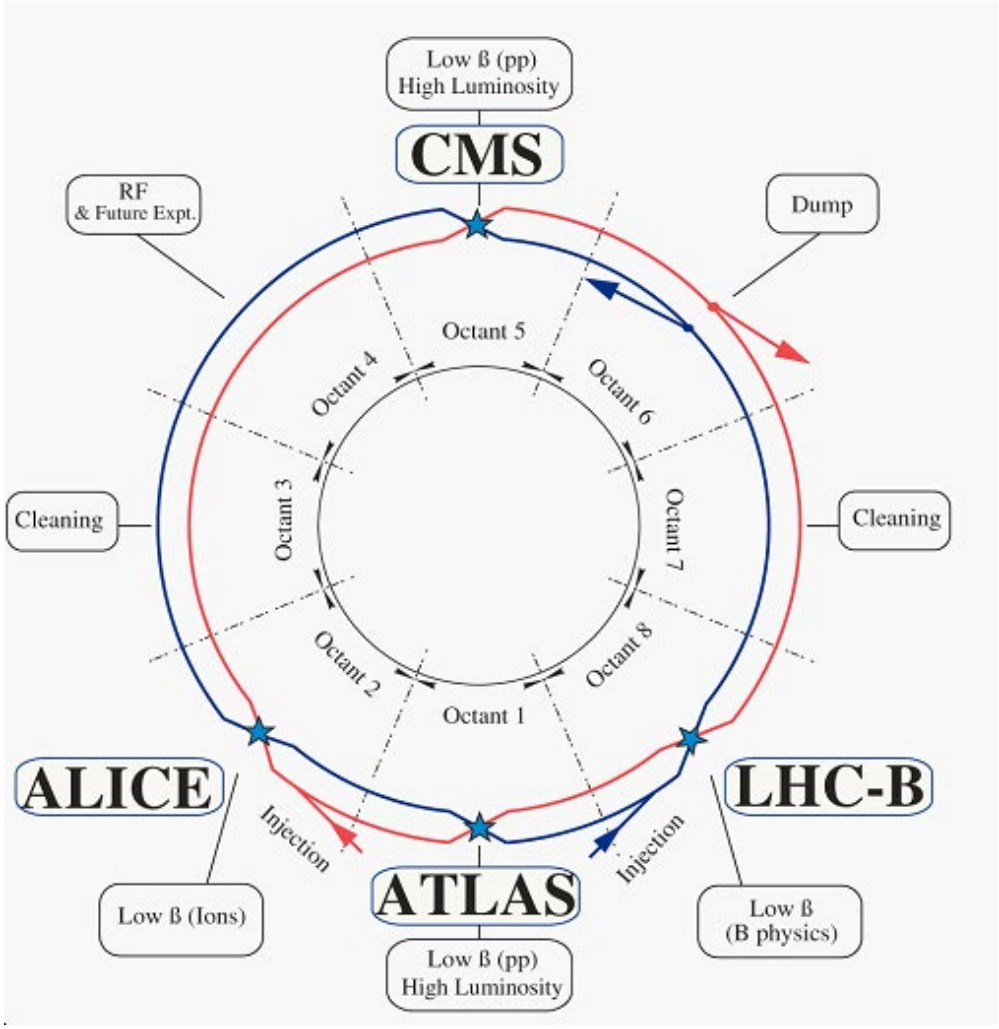


Figure 4.2: Labeled diagram of all the experiments at the LHC indicating the counter-circulating beams and the relative location of each major experiment along the circumference of the accelerator [38].

The exact design of the LHC tunnel is due to the experimental constraints of the original machine for which it was built, the Large Electron Positron (LEP) Collider. For the $\sim 2,000$ times lighter electron the maximum energy was limited by synchrotron radiation, proportional to $\frac{1}{m^4}$, requiring long straight sections of accelerating RF cavities to recuperate the lost energy. Because this effect is $\mathcal{O}(10^{13})$ times smaller for the proton, the LHC is instead limited by the ability to design and construct magnets strong enough to bend the beam given the already determined curvature of the 8 arcs.

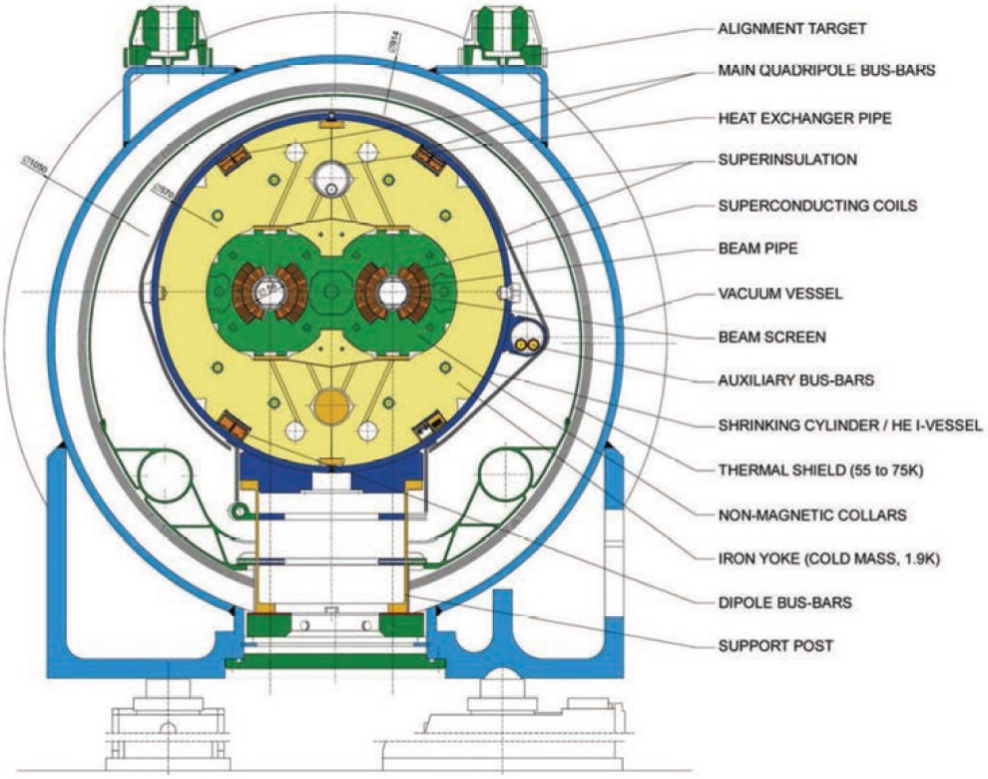


Figure 4.3: Depiction of a LHC dipole magnet 2-in-1 design labeling the major components [39].

The oppositely circulating beams must each have their own ring and magnetic field

which led to the creation of a twin-bore (i.e. "two-in-one") magnet design, a cross section of which can be seen in Figure 4.3. These magnets are constructed using NbTi superconductors which are cooled to 2 K using superfluid helium. These magnets are design to provide the needed 8.33 T magnetic field required to bend the proton trajectories at the designed beam energy of 7 TeV. In total 1231 of these 15 m bending dipole magnets are used, in association with 392 5-7 m quadrupole magnets which are responsible for keeping the proton bunches in a tight beam by squeezing them both horizontally and vertically.

4.3 Performance

Since the begining of its stable running in 2010 the LHC has performed well, exceeding expectations. While the LHC operation itself is incredibly complex, the performance of the machine, for the purposes of analysis, can be reduced to three numbers; the familiar center of mass energy of the beams, the integrated luminosity and the instantaneous luminosity.

For particle physics the integrated luminosity is proportional to the total number of collisions recorded during a specified time period, while the instantaneous luminosity is a function of the number of protons per bunch, the bunch crossing rate and the cross section of a proton-proton interaction and represents the particle flux. Knowing this one can see that the integrated luminosity, L_{int} is simply the integral of the instantaneous

luminosity $L_{\text{inst.}}$ for a chosen data period as seen in Equation (4.1).

$$L_{\text{int}} = \int L_{\text{inst.}} dt \tag{4.1}$$

For a standard Gaussian beam, $L_{\text{inst.}}$ can be written as

$$L_{\text{inst.}} = \frac{N_b^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F \tag{4.2}$$

where N_b is the number of particles per bunch, n_b the number of bunches per beam, f_{rev} the revolution frequency, γ_r the relativistic gamma factor, ϵ_n the normalized transverse beam emittance, β^* the beta function at the collision point, and F the geometric luminosity reduction factor due to the crossing angle at the interaction point given by

$$F = \left(1 + \left(\frac{\theta_c \sigma_z}{2\sigma^*}\right)^2\right)^{-1/2} \tag{4.3}$$

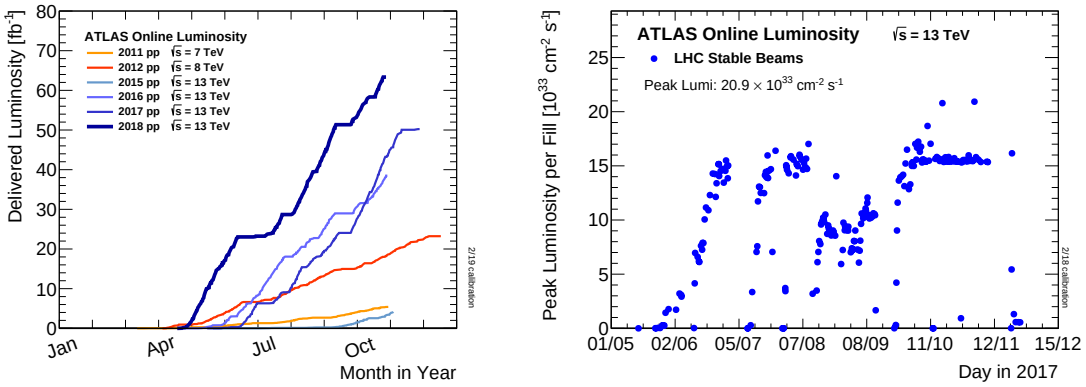
where θ_c is the full crossing angle at the interaction point, σ_z is the RMS bunch length, and σ^* is the transverse RMS beam size at the interaction point [31]. Nominal values for the above quantities are shown in Table 4.1 which also demonstrates the incredible success of the LHC operators and accelerator teams during the LHC Run II data taking period.

The ATLAS experiment integrated luminosity for each year can be seen in Figure 4.4a

Table 4.1: Nominal design values of LHC operations parameters at ATLAS for 25 ns bunch crossing spacing [31, 40]. Design and ATLAS recorded values of the machine luminosity are also given for LHC Run II operations [41].

Parameter	Symbol	LHC Run II Value
LHC circumference		26,659 m
LHC design beam energy		7 TeV
LHC beam energy in Run II		6.5 TeV
Number of protons per bunch	N_b	1.15×10^{11}
Number of proton bunches per beam	n_b	2,808
Revolution frequency	f_{rev}	11.245 kHz
Lorentz factor (design)	γ_r	7462.69
Lorentz factor at $\sqrt{s} = 13$ TeV		6929.64
Normalized transverse beam emittance	ϵ_n	3.75 μm
Collision point beta function	β^*	0.55 m
Full crossing angle	θ_c	285 μrad
RMS bunch length	σ_z	7.55×10^{-2} m
Transverse RMS beam size	σ^*	16.6 μm
Peak design machine luminosity at 13 TeV	$L_{\text{inst.}}$	9 $\text{nb}^{-1}\text{s}^{-1}$
Peak ATLAS recorded machine luminosity		21 $\text{nb}^{-1}\text{s}^{-1}$

along with an example of the instantaneous luminosity for 2017 in Figure 4.4b which is used in the presented analysis.



(a) Integrated Luminosity 2011 - 2018 (b) 2017 Peak Instantaneous Luminosity

Figure 4.4: Luminosity is monitored as both a running total known as the integrated luminosity as depicted in (a) and as an instantaneous quantity as shown in (b) [42].

4.4 Pile-up at the LHC

To maximize the probability of having a hard scatter proton-proton interaction the bunches are filled with a large number of protons and squeezed into a small area before being collided. This results in multiple inelastic proton-proton interactions per bunch crossing, known as "pile-up," which make it more difficult to identify the vertex of the primary hard scatter interaction of interest. The average number of proton collisions per bunch crossing is known as μ. The time-averaged μ for a given time period is reported as $\langle\mu\rangle$. For the data taking periods used for this dissertation, 2015-2017, the $\langle\mu\rangle = 31.9$ [42].

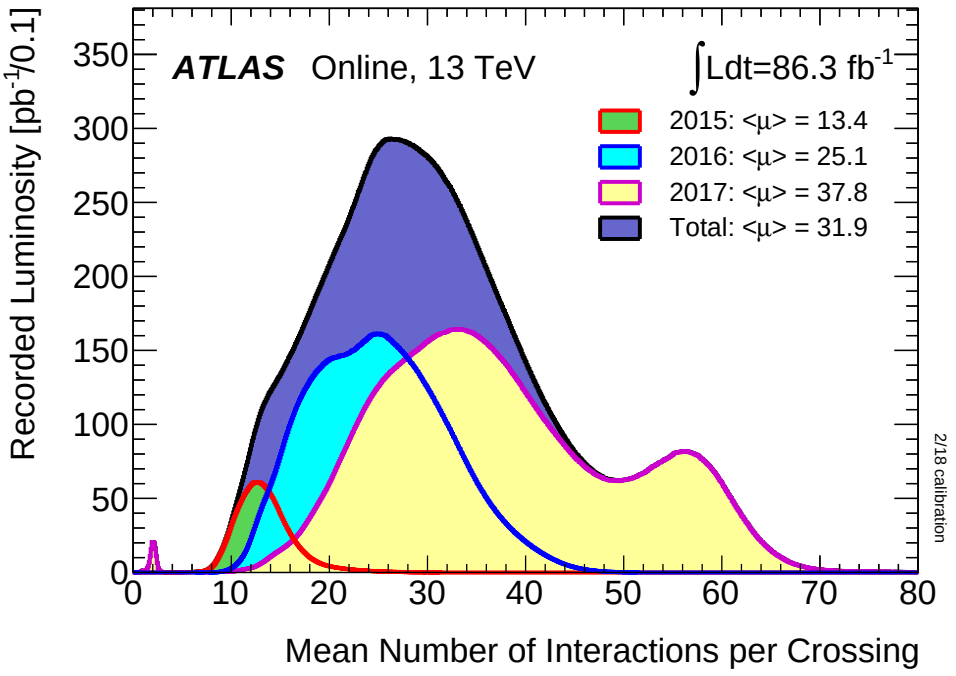


Figure 4.5: Pileup for data taking periods 2015 - 2018 [42].

The pile-up profile for past years can be seen in Figure 4.5. The width of this distribution is due to a combination of Poisson statistics, the decrease in number of protons per bunch over the lifetime of a single run, and optimization tweaks to the beam's profile during the LHC's operation. Understanding and eliminating the unwanted contributions from these pile-up events is crucial to reconstructing physics variables that describe the primary interaction being observed.

Chapter 5

The ATLAS Detector

The immense energies available at the LHC and the veritable zoo of particles require a general-purpose experiment in order to fully exploit the full range of physics opportunities. Two international collaborations rose to this challenge, the CMS (Compact Muon Solenoid) and ATLAS (A Toroidal LHC ApparatuS) experiments. While both have similar physics goals and each of them strengths and weaknesses, this dissertation will focus on the ATLAS experiment and the intricacies of its three sub-detectors and two massive magnet systems depicted in Figure 5.1.

Originally proposed in 1994, the ATLAS detector saw its first events in 2008 [44, 45]. On July 4th, 2012 in a joint announcement the ATLAS and CMS experiments presented the discovery of the long predicted Higgs Boson [6, 7, 34]. The ATLAS collaboration now boasts over 3000 physicists from 175 institutions spread across 38 countries and continues to probe the limits of the Standard Model in pursuit of answers to some of

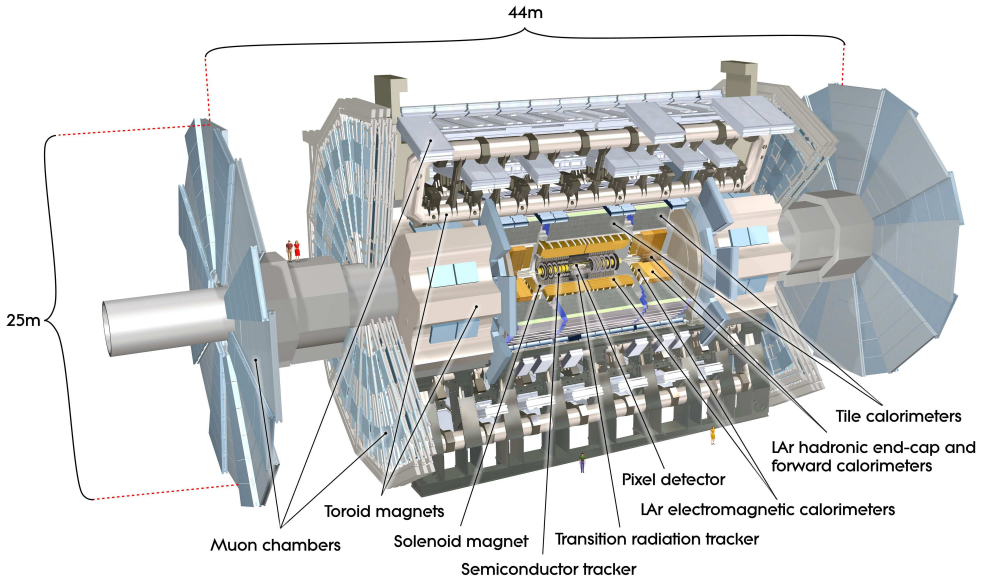


Figure 5.1: A cut-away side view of the ATLAS detector with the major components labeled [43]. Within each of these labeled components there may exist multiple different detector technologies. For scale two people in red are shown standing between the disk muon chambers on the left side of the figure.

humanity's deepest questions.

Located approximately 100 meters underground in a vast excavated chamber, the ATLAS detector rests its 7000 metric tons on a bed of steel-reinforced concrete. Out of it flow the signals from 100 million electronic channels through a zip-tied mass of 3000+ kilometers of cabling. At its very center is one of the four interaction points of the LHC, Point 1, where the two counter-circulating proton beams are shaped by a series of magnets. The energetic particles resulting from this collision then fly out in all directions into the bulk of the ATLAS detector.

The first sub-system they meet is the Inner Detector (ID) and its many layers of strip and pixel silicon detectors, along with a transition radiation gaseous wire detector, all bathed in the 2 T magnetic field from the surrounding superconducting solenoidal magnet. This system exploits the ionization signal of charged particles to track their curved trajectory through the magnetic field. This curvature gives charge information, a momentum measurement, and precisely-located 3D vertices crucial to the identification of the secondary vertices of a b-hadron decay.

Outside of the solenoid the particles encounter the electromagnetic and then the hadronic sampling calorimeters. Here, layers of scintillator and high-radiation-length materials are interleaved to measure the energy of electrons, photons, and hadrons. As the goal is to completely absorb the energy of all outgoing particles, the calorimeter has a nearly 4π solid angle coverage.

The outermost subsystem, the muon spectrometer, surrounds the calorimeter and is equipped with its own toroidal magnet system. Here the charged muons bend in the magnetic field while leaving a trail of ionization in the Muon Spectrometer (MS) before exiting the detector completely. Neutrinos are the only other Standard Model particle that leave the detector; however, they do so without detection. A depiction of the various particle interactions with the different detector sub-systems can be seen in Figure 5.2.

5.1 ATLAS Coordinate System

Choosing the nominal interaction point as the origin, ATLAS defines a right-handed coordinate system where the positive x-axis points towards the center of the LHC ring, the positive y-axis points upwards, and the positive z-axis is defined by the counter-clockwise circulating beam direction as viewed from above, as shown in Figure 5.3 [43].

Using these coordinates the physical momentum of the objects measured is defined as $\vec{p} = (p_{\mathrm{T}}, p_z)$ with p_{T} being the momentum of the object in the transverse plane and p_z the momentum along the beam axis. Considering the cylindrical symmetry of ATLAS it is desirable to define the polar angle θ from the beam axis with the r-ϕ plane being perpendicular to that axis. The radial distance is defined as $r = \sqrt{x^2 + y^2}$ and the azimuthal angle is defined with $\phi = 0$ corresponding to the x-axis.

The initial state of a collision is inevitably boosted in the z-axis, but it is difficult to accurately measure this boost. Instead it is preferable to use geometric quantities which

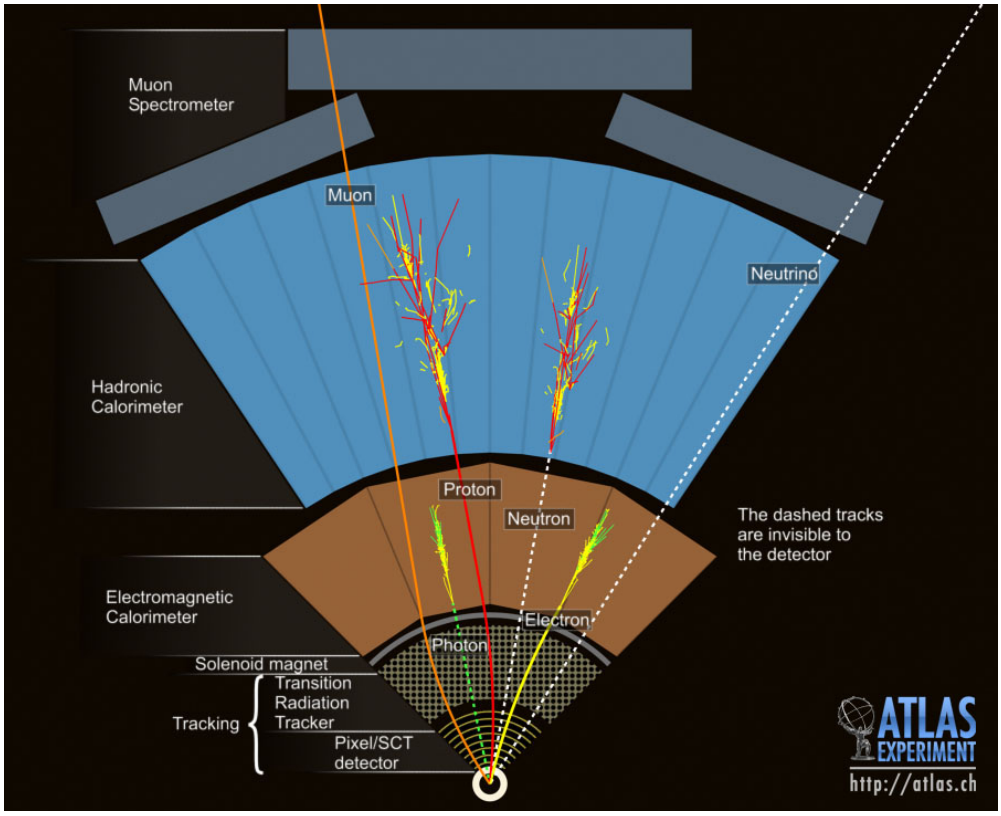


Figure 5.2: This slice of the ATLAS detector depicts how different particles interact with each component of the detector they cross. A dashed line indicates no interaction while a solid line indicates interaction. Electrons (yellow) and charged hadrons (red) interact with the tracker and curve in the solenoid's magnetic field. Electron and photon showers (yellow/green) are absorbed by the electromagnetic calorimeter. All hadrons (red/yellow) are absorbed by the hadronic calorimeter. The muons (orange) curve in both the solenoid and toroid magnetic fields before exiting the detector. Finally, the neutrinos (white) pass through the entire detector without interacting [46].

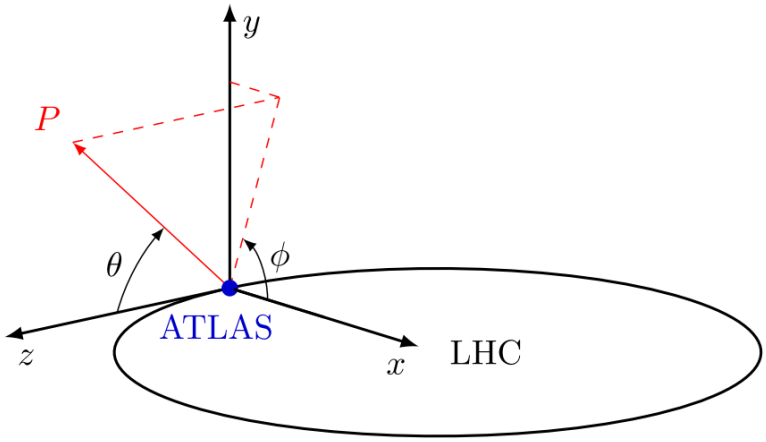


Figure 5.3: The standard Cartesian coordinate system is shown with its origin at the ATLAS interaction point, the positive x-axis towards the center of the LHC, the positive y-axis pointing upwards, and the positive z-axis pointing along the beamline [29].

are invariant under these boosts. In the high energy regime, the pseudorapidity

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \tag{5.1}$$

is a good approximation of the rapidity (y) of a particle — a measurement of the velocity of a particle parallel to the beam axis (z-axis)

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right) \tag{5.2}$$

where E is the energy of the particle and p_z is defined above. Differences in rapidity (Δy) are Lorentz invariant under boosts along the z-axis. However, pseudorapidity is preferred as it is an entirely geometric quantity independent of particle energy. Fig-

ure 5.4 gives a sense of the distribution of η in the y-z plane. Since ϕ is defined in the x-y plane, differences in the azimuthal angle ($\Delta\phi$) are also invariant under Lorentz boosts along the z-axis. This allows for the definition of the functionally Lorentz invariant angular separation

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{5.3}$$

between objects in the detector.

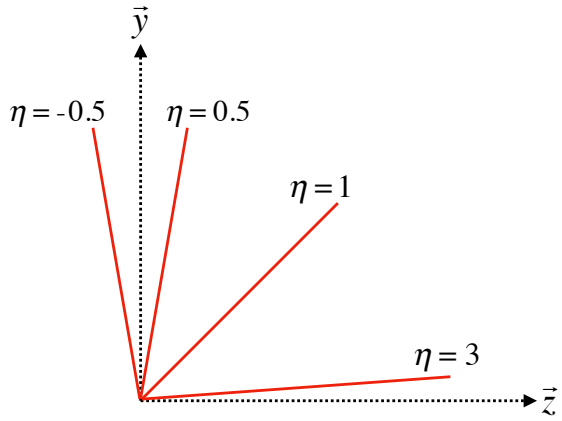


Figure 5.4: Modified from [47] this cartoon represents a selection of pseudorapidity (η) values overlaid with some Cartesian coordinates (dashed black lines). The red lines are drawn for $\eta = \pm 0.5,\ 1.0$ and 3.0.

5.2 Tracking with the Inner Detector

The Inner Detector (ID), shown in Figure 5.5, is responsible for providing precise momentum resolution and identification of both primary and secondary vertex measure-

ments of charged particle tracks [48, 49]. The ID's innermost subsystem, the Insertable B-Layer (IBL), is only 3.3 cm from the interaction point [50] and thus must be able to perform its crucial duties while enduring the high fluence of radiation from the colliding beams.

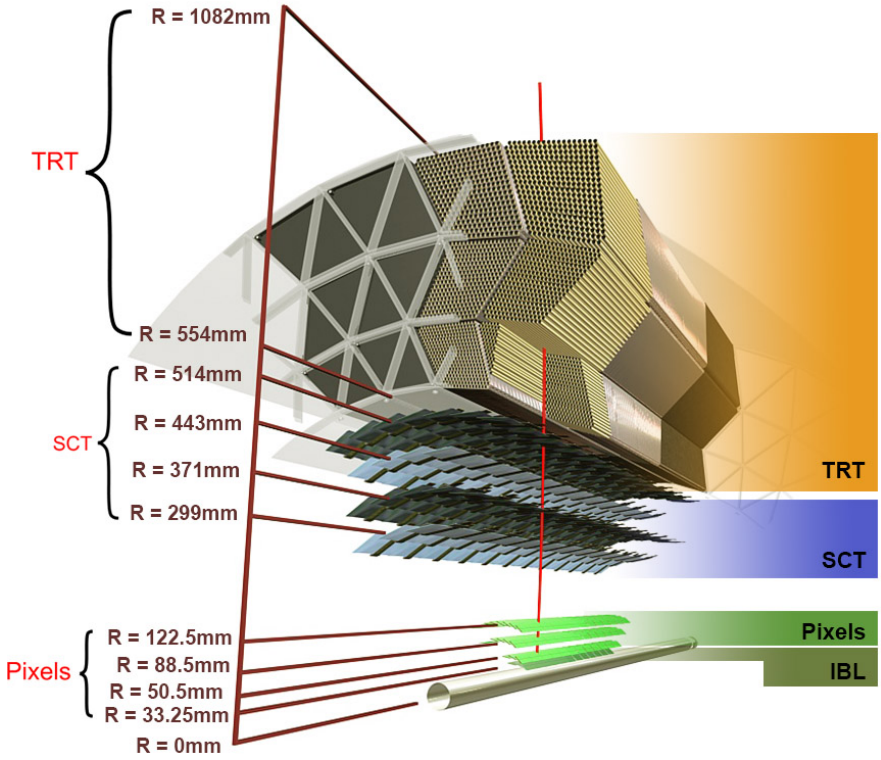


Figure 5.5: Diagram of Inner Detector [50].

The ID is designed to be very compact to reduce the probability of a particle decaying inside and to give precision measurements of the particles' curvature in the 2 T solenoidal magnetic field. This leads to excellent momentum resolution above the nominal p_{T} threshold of 0.5 GeV and within the pseudorapidity range of $|\eta| < 2.5$ as shown in Figure 5.6.

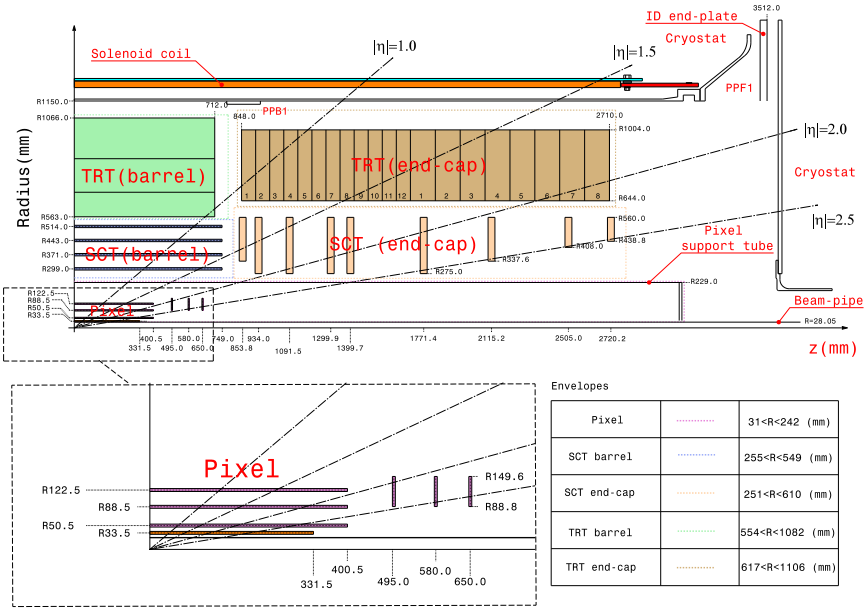


Figure 5.6: Schematic of the Inner Detector including η lines. Each component shown is cylindrically symmetric leading to a multi-layered detector [51].

The ID is composed of three different detector technologies for particle trajectory reconstruction: Pixel Detector (including the IBL), Semiconductor Tracker (SCT) and the Transition Radiation Tracker (TRT). These will be discussed in the following sections.

5.2.1 Pixel Detector

The ATLAS Pixel Detector [43], the innermost subdetector of the ID, is designed to give the best resolution possible as close as possible to the interaction point. This is accomplished using the 4 barrel layers and 3 disks per end-cap as indicated in Figure 5.6. The innermost barrel layer, the IBL, has pixel dimensions of 50 µm($\hat{\phi}$) × 250 µm($\hat{z}$) × 200 µm($\hat{r}$). For the other layers the dimensions are 50 µm($\hat{\phi}$) × 400 µm($\hat{z}$) for about

90% of the pixels and 50 µm($\hat{\phi}$) × 600 µm($\hat{z}$) for the others, all with a thickness of 250 µm($\hat{r}$). This gives a total active area of 1.88 m^2 collected through 92.4 million readout channels, more than half of the total number of channels for ATLAS. This detailed charged particle information very close to the interaction point is crucial not only for pattern recognition and track reconstruction, but also for the reconstruction of the primary and secondary vertices intrinsic to the decay of b-hadrons, a critical element of the analysis presented in this thesis.

5.2.2 Semiconductor Tracker

Surrounding the Pixel Detector, the Semiconductor Tracker (SCT) [43] is composed of 4088 double-sided silicon microstrip modules. Each side of a module is constructed out of two silicon strip sensors that are daisy-chained together. The result is 768 composite strips each 12.6 cm long with an inter-strip pitch of 80 µm. In the barrel the strips are aligned with the $\hat{z}$ direction, while in the end-caps they are aligned with the $\hat{r}$ direction. In both cases the separation of the strips is constant in $\hat{\phi}$. The two sides are rotated with respect to each other by 40 mrad to allow for position measurement along the length of the strip. These modules are then used to tile the 4 barrel layers and 9 disks per end-cap (18 disks in total) as seen in Figure 5.6. This design is chosen to ensure that each charged track interacts with 8 strip layers (equivalent to four space points). This information is used to further measure the momentum and impact parameter and to identify the vertices of charged particles.

5.2.3 Transition Radiation Tracker

The Transition Radiation Tracker (TRT) [43], the outermost subdetector of the ID, provides tracking information by detecting the ionization of gas caused by charged particles in $\eta < 2.0$ as they pass through the drift tube channels also known as straws. The 4 mm diameter straws are filled with a 70% Xe, 27% CO_2, and 3% O_2 gas mixture [1] and a 31 µm diameter gold-plated tungsten wire anode at the center for the collection of the ionization signal. In the barrel, 73 azimuthally symmetric layers of 144 cm straws are oriented parallel to the beam pipe with an electrical division in the center of each allowing the two sides to be read out separately. For each end-cap the straws are radially oriented in 160 symmetric planes each containing 768 37 cm-long drift tubes shown in Figure 5.6. There are typically 36 TRT ionization hits per charged particle track.

The TRT also provides electron identification functionality through the detection of transition radiation. In both the barrel and the end-caps, polypropylene fibers (barrel) or foils (end-caps) function as the transition radiation material which causes the relativistic charged particles to radiate and thus ionize the gas in the straw. The amount of transition radiation produced is proportional to the Lorentz factor meaning that for particles with equal energy, lighter particles (e.g. electrons) will produce more radiation. Thus, by defining a high and low charge threshold, tracks belonging to electrons can be identified by requiring a larger fraction of the TRT hits along the track to pass

[1] During Run 2 straws belonging to modules with large gas leaks were filled with a mixture of 70% Ar, 27% CO_2, and 3% O_2. Ar is less efficient at absorbing transition radiation but maintains similar tracking capabilities as Xe [52].

the high threshold.

5.3 Calorimetry

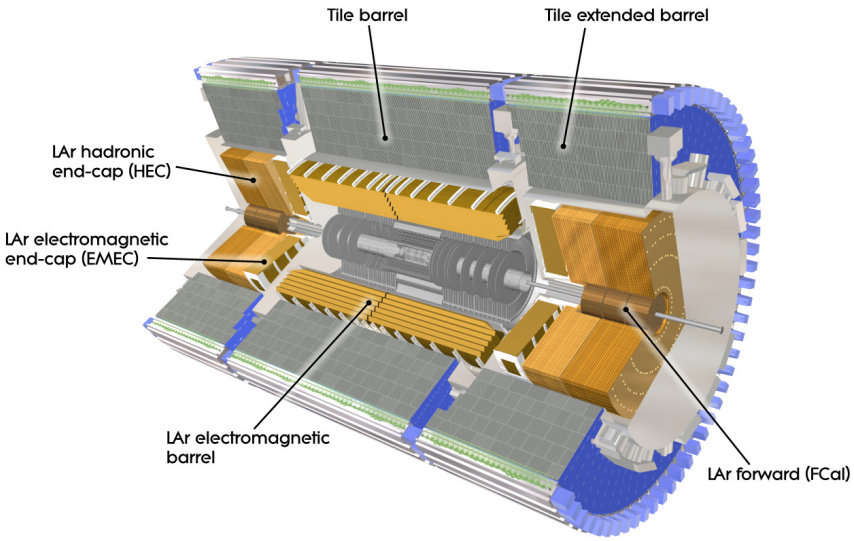


Figure 5.7: A cutaway diagram of ATLAS sampling calorimeters [43].

Once the proton-proton collision remnants have passed through the ID and its surrounding solenoid they enter into the ATLAS calorimeters depicted in Figure 5.7. Sampling calorimeter technologies were chosen for their compact geometry and lower cost point. These are constructed by alternating layers of absorber, a dense material which reduces the incident particle's energy, and active material which produces a detectable signal when a particle passes through. This means that the detected signal is only a fraction of the total energy of the particle and thus must be calibrated in a study of the calorimeter [53]. The first system, the Electromagnetic Calorimeter (EMC), is designed

to measure the energy of electrons and photons which primarily lose their energy via bremsstrahlung and pair production electromagnetic interactions. Outside of the EMC is the hadronic calorimeter which is designed to measure the energy of jets of hadrons through their electromagnetic and strong interactions. These detectors cover the entire $|\eta| < 4.9$ range [2] and provide complete containment of both electromagnetic and hadronic showers with higher granularity in the EMC for $|\eta| < 2.5$, the region matched to the ID, for precision measurements of electrons and photons.

5.3.1 Electromagnetic Calorimeter

The innermost calorimeter, the Liquid Argon (LAr) Electromagnetic Calorimeter (EMC) [43], uses lead as the absorber and liquid argon as the active material in an "accordion geometry" as seen in Figure 5.8. This geometry was chosen for uniform coverage in $\hat{\phi}$ due to its lack of un-instrumented cracks in the radial direction. The barrel region covers $|\eta| < 1.475$ and an end cap on each side covers $1.375 < |\eta| < 3.2$. The barrel and end-cap calorimeters are each housed in their own cryostat. The barrel is composed of two half barrels with a 4 mm gap at $z = 0$ and both end caps are divided into an inner wheel covering $2.5 < |\eta| < 3.2$ and an outer wheel covering $1.375 < |\eta| < 2.5$.

In the $|\eta| < 2.5$ region the EMC has 3 radial layers for precision physics measurements. Layer 1 consists of strip cells which are finely segmented with $\Delta\eta = 0.0031$ and $\Delta\phi = 0.0245$ allowing for precision position resolution which gives discrimination

[2] EMC performance is degraded in the region $1.37 \leq |\eta| \leq 1.52$ due to the transition between the barrel and the end-cap calorimeters.

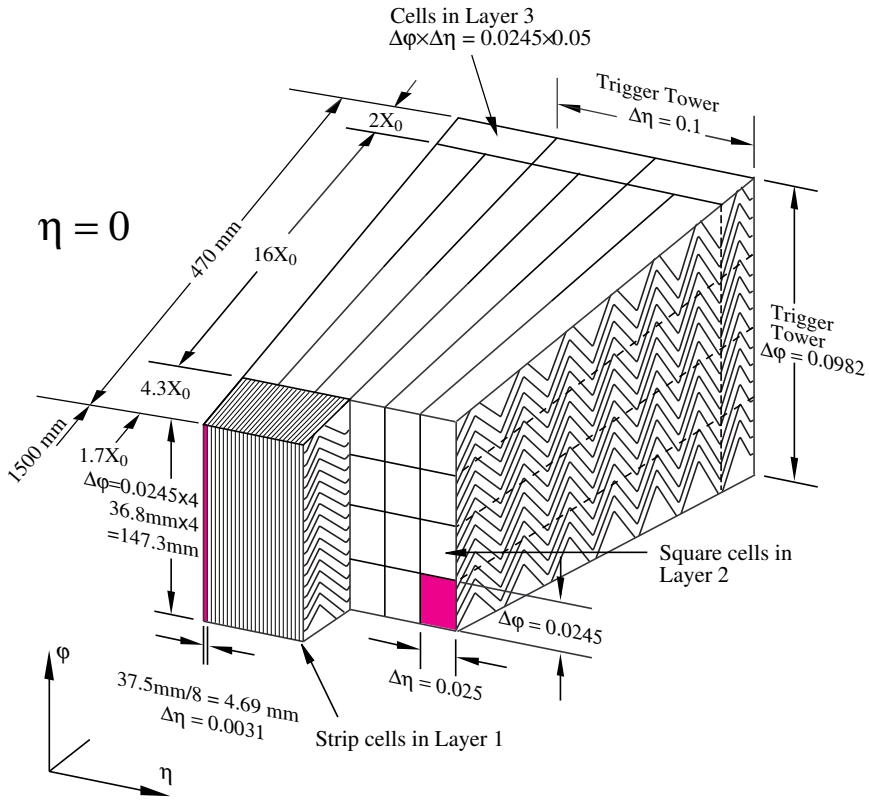


Figure 5.8: Sketch of LAr EMC barrel module where the lead and liquid argon layers are visible in an accordion like geometry. Looking from the foreground to the back there are 3 different layers of readout cells visible [43].

power between a single γ deposit and the π^0 characteristic $\gamma\gamma$ deposit. Layer 2, which collects the largest fraction of energy from electromagnetic shower, is segmented with $\Delta\eta = 0.025$ and $\Delta\phi = 0.0245$. Layer 3 collects the tail of the electromagnetic shower using a coarser segmentation of $\Delta\eta = 0.05$ and $\Delta\phi = 0.0245$. Additionally, in the region $|\eta| < 1.8$ a thin pre-sampler, which contains no lead absorber, was placed in front of Layer 1 to allow for energy corrections due to losses upstream of the EMC. Combined the EMC is > 22 radiation lengths (X_0) in the barrel and > 24 X_0 in the end-caps, where a radiation length is the average distance an electron travels in a given material before losing $1/e$ of its original energy E_0 via bremsstrahlung radiation.

5.3.2 Hadronic Calorimeter

Directly outside the EMC envelope is the hadronic calorimeter system [43] which consists of three sampling calorimeter technologies: the Tile Calorimeter, the LAr Hadronic End-cap Calorimeter (HEC) and the LAr Forward Calorimeter (FCal). Combined, these three subsystems give measurements of hadronic jet energies in the $0 < |\eta| < 4.9$ range. The Tile Calorimeter uses steel as the absorber layer and scintillating tiles as the active material and covers the region $|\eta| < 1.7$ with a barrel section flanked by two barrel extensions each divided azimuthally into 64 modules. These scintillator tiles are read out on two sides by wavelength-shifting fibers connected to photomultiplier tubes as seen in Figure 5.9. At $\eta = 0$ the total Tile Calorimeter thickness is 9.7 nuclear interaction lengths (λ), where λ is the average distance a hadron travels before

interacting inelastically with a nucleus.

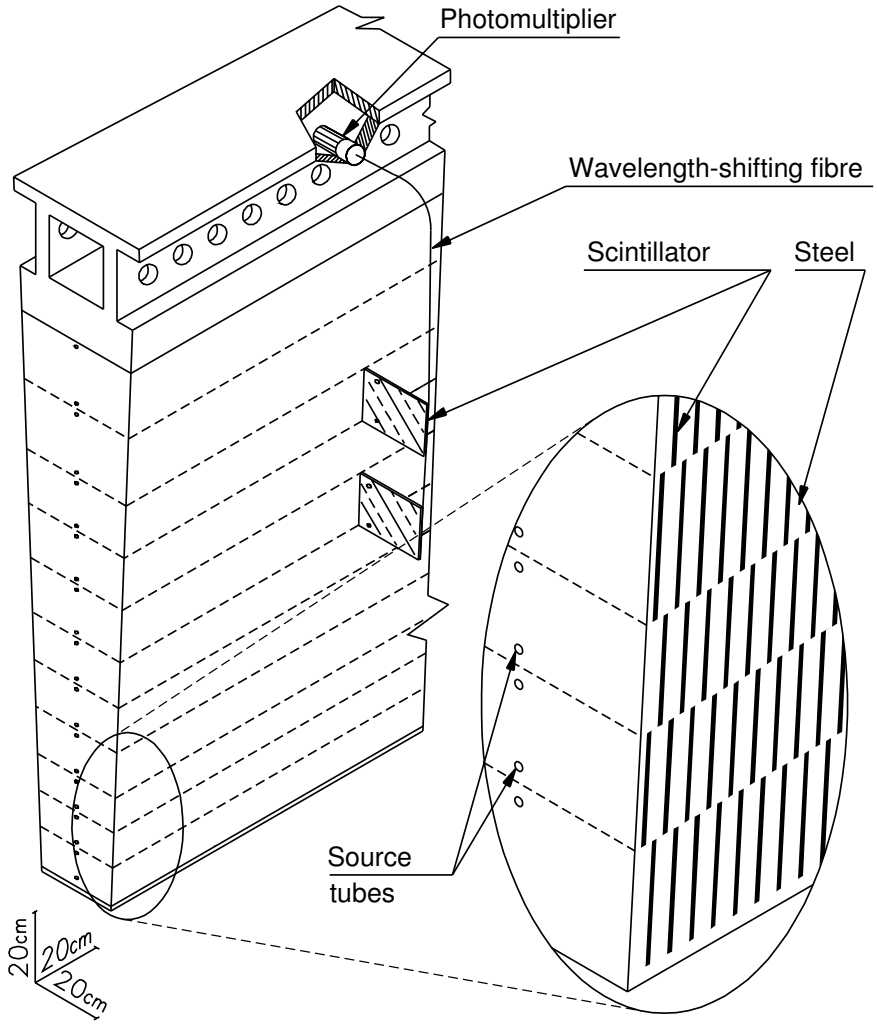


Figure 5.9: Schematic of a Tile Calorimeter module including a depiction of the connection between the scintillator tile to the photomultiplier via a wavelength-shifting fiber [43].

The high η region receives a larger radiation dose that would damage scintillator tiles. Thus, technologies that are more radiation hard were chosen for the HEC [54]. The HEC is composed of two independent wheels per end-cap located just past the EMC

end-cap but sharing the same cryostat. This system uses copper as an absorber and liquid argon for the active material and covers the $1.5 < |\eta| < 3.2$ range using 32 wedge-shaped modules per wheel. Finally, the FCal shares the same cryostat as the EMC and HEC end-caps and acts to extend the coverage of the combined calorimeter system to include the $3.1 < |\eta| < 4.9$ range. Each end-cap contains 3 modules, the first an electromagnetic module (copper/liquid-argon) which is followed by two hadronic modules (tungsten/liquid-argon).

5.4 Muon Spectrometer

The ATLAS Muon Spectrometer (MS) [43], see Figure 5.10, accomplishes tracking of muons in the $|\eta| < 2.7$ region for momentum reconstruction while also triggering on charged particles in the $|\eta| < 2.4$ region. The magnetic field necessary for momentum reconstruction is provided by 3 air-core toroid systems, one barrel toroid covering $|\eta| < 1.4$ and two end-cap toroid systems which are inserted into the inner radius of the barrel toroid to cover the $1.6 < |\eta| < 2.7$. The so-called transition region $1.4 < |\eta| < 1.6$ between these two magnet systems is covered by a combination of the barrel and end-cap toroid magnets. Similar to the ID the resolution in the MS is inversely proportional to the particle's incident momentum. Any muon with p_{T} lower than ~ 3 GeV will never make it to the MS due to energy loss in the calorimeters [43].

Precision tracking measurements for momentum reconstruction is accomplished using

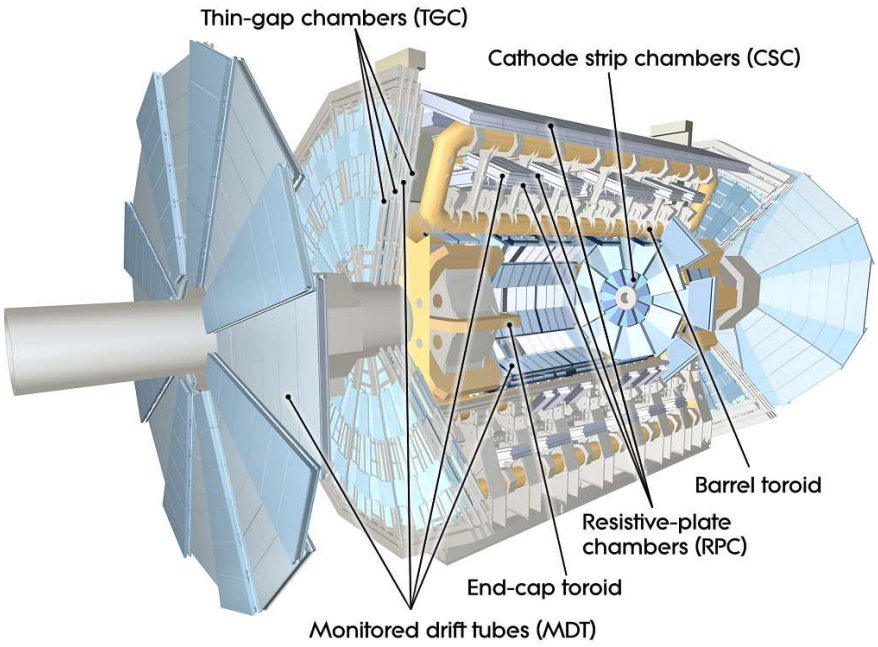


Figure 5.10: A cut-away diagram of the ATLAS muon system and its many sub-detectors [43].

the Monitored Drift Tube chambers (MDTs) for $|\eta| < 2.0$. The MDT system consists of 1163 drift tube chambers arranged in three to eight layers for varying η. The Cathode-Strip Chambers (CSCs) span the range $2.0 < |\eta| < 2.7$ and are designed to withstand the higher rate at high η and retain good time resolution using multiwire proportional chambers with orthogonal segmented cathode planes.

The MS also gives nanosecond tracking information for triggering on muon tracks. This is accomplished using Resistive Plate Chambers (RPC) in the barrel region $|\eta| < 1.05$ and Thin Gap Chambers (TGC) in the end-cap $1.05 < |\eta| < 2.4$ region. Both chamber systems deliver a triggerable signal with a spread of $15-25$ ns, thus providing the ability to tag individual beam-crossings.

5.5 Trigger

The 25 ns spacing of proton bunches at the LHC results in a crossing rate of proton bunches of 40 MHz with an average of 31.9 *pp* interaction events per crossing as seen in Figure 4.5. Given that each event takes up 1.6 MB of storage this would result in a data rate of 64 TB/s. Because of the limits of technology and budget, this rate of data is far beyond what could be read out, much less stored for analysis. The solution is the ATLAS trigger system which decides in real time whether or not to record the current event, or dump all sensor output and prepare for the next bunch crossing.

ATLAS uses a two-level trigger system to reduce the event rate while making sure to

not skip any event of interest. The Level-One (L1) trigger is hardware based and uses coarse granularity inputs from the muon spectrometer and calorimetery systems to make a decision within 2.5 µs of the bunch crossing for the event in question. This results in an event rate reduction from 40 MHz to 100 kHz. The second level of the system, the High Level Trigger (HLT), is software based and uses the full detector output to make its decision within a 200 ms window. This final triggering step reduces the event rate down to a more manageable 1 kHz which is then written to disk at a rate of over a gigabyte per second.

Part III

Search for Boosted $H \to b\bar{b}$

Chapter 6

Data and Monte Carlo Simulation

This analysis focuses on the data collected by the ATLAS detector from pp collisions produced by the LHC at the center-of-mass energy of 13 TeV. In particular the analysis uses datasets collected in 2015, 2016, and 2017 which correspond to an integrated luminosity of 80.5 fb^{-1} after beam, detector and data-quality requirements are taken into account.

In order to compare our findings with theory, the predictions of the SM are used to produce Monte Carlo (MC) simulated events to model the signal and background processes. These MC samples go through a full simulation of the ATLAS detector and are reconstructed using the same algorithms as used on data, so the MC and data have the same format at analysis level. This allows the MC and data to be analyzed using the same framework thus giving a direct comparison between theory and measurement as the final product.

The following sections describe the systems for selecting the data used in the analysis and the software packages, developed in collaboration with theorists, used to simulate the signal and background processes relevant to the analysis.

6.1 ATLAS Data Analyzed

As mentioned earlier, the data are checked to make sure they are of high quality, meaning that the beam, detector and data collection systems were all fully operational for the given bunch crossing event. These data quality requirements are enforced by choosing only events from each respective year's Good Runs List (GRL), an XML file produced by the ATLAS data quality monitoring team that lists all luminosity blocks that have met the data quality criteria. This analysis uses three such GRLs - one for each year of data taking (2015, 2016, 2017) - corresponding to annual integrated luminosities of 3.2 fb^{-1}, 33 fb^{-1}, and 44.3 fb^{-1}.

6.2 Higgs Boson Signal Monte Carlo Samples

In order to simulate Higgs boson events, the three leading production mechanisms at the LHC were considered, shown in Figure 3.2: gluon-gluon fusion, vector boson fusion and Higgsstrahlung. These three production modes represent approximately 88%, 7% and 4% of the total Higgs signal, respectively, before analysis cuts are applied. In all samples the simulated Higgs boson is forced to decay to the $b\bar{b}$ final state.

The ggF H + jet events were generated using the HJ+MiNLO [55] prescription, assuming a finite top quark mass, with the POWHEG-BOX 2 generator [56] and the NNPDF30 NNLO parton distribution functions [57]. After generation the events were showered using PYTHIA 8.212 [58] with the AZNLO tune and the CTEQ6L1 parton distribution functions [59]. Any b-hadrons produced during this process were decayed using EVTGEN [60]. During generation a parton level filter of $k_{\mathrm{T}} > 200$ GeV was used to select high momentum events. This filtering is reflected in the cross section in Table 6.1.

The VBF H + jet events were also generated using the POWHEG-BOX generator [61] with the NNPDF30 NLO parton distribution functions [57]. Again the showering was performed with PYTHIA 8.212 [58] using the AZNLO tune and the CTEQ6L1 parton distribution functions [59]. The decay of b-hadrons was again performed using EVTGEN [60]. During generation the produced Higgs boson was required to have a $p_{\mathrm{T}} > 250$ GeV. This filtering is reflected in the filtering efficiency in Table 6.1.

Higgsstrahlung events were generated using the PYTHIA 8.212 generator [58], the AZNLO tune and the CTEQ6L1 parton distribution functions [59]. Again the decay of b-hadrons is handled by EVTGEN [60]. Unfortunately the PYTHIA ZH process does not include the $gg \to ZH$ contribution. To account for this the cross section is corrected to the LHC Higgs cross section working group's (LHCHXSWG) recommended $\sigma^{\mathrm{ggZH}}_{\mathrm{NLO+NLL}}$ with NLO + Next to Leading Log (NLL) accuracy [62]. During generation a filter of $p_{\mathrm{T}} > 350$ GeV was used to select for high momentum events. This filtering is reflected in the filtering efficiency in Table 6.1.

Dataset	Campaign	xsec (pb)	Filter Efficiency	Events Generated	Effective Luminosity (fb^{-1})
309450.PowhegPy8EG_NNLOPS_nnlo_30_ggH125_bb_kt200	MC16a	0.4763	1.0	400000	839.1
345931.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_bb	MC16a	3.7296	0.026877	499500	4794
309451.Pythia8EvtGen_A14NNPDF23LO_WH125_bb_fj350	MC16a	0.63487	0.01306	100000	12061
309452.Pythia8EvtGen_A14NNPDF23LO_ZH125_bb_fj350	MC16a	0.34656	0.013036	100000	22135
309450.PowhegPy8EG_NNLOPS_nnlo_30_ggH125_bb_kt200	MC16d	0.4763	1.0	499000	1047
345931.PowhegPy8EG_NNPDF30_AZNLOCTEQ6L1_VBFH125_bb	MC16d	3.7296	0.026877	650000	6238
309451.Pythia8EvtGen_A14NNPDF23LO_WH125_bb_fj350	MC16d	0.63487	0.01306	130000	15679
309452.Pythia8EvtGen_A14NNPDF23LO_ZH125_bb_fj350	MC16d	0.34656	0.013036	130000	28775

Table 6.1: List of datasets used for Higgs boson Monte Carlo [63].

6.3 Background Monte Carlo

Modeling the expected contributions of backgrounds to the analysis is done with a mix of data driven methods and Monte Carlo simulated samples as discussed in Chapter 9. The MC samples were used for the development of the modeling of the non-resonant QCD multijet backgrounds and for the estimation of the major resonant backgrounds from V+jet, $t\bar{t}$, and single-top production. For the multijet background the final estimation is data driven, but the MC was used to develop the background model.

The QCD dijet events were simulated by the Pythia 8.186 generator [64] with the A14 tune and the NNPDF23 LO PDF [65] using EvtGen to decay the resulting b-hadrons [60]. To maintain a constant statistical precision over a large momentum range, the weighted events were generated with a flat p_{T} spectrum. The QCD samples used to develop the background model are summarized in Table 6.2.

Hadronically decaying W and Z events were produced with a maximum of four additional partons at leading order (LO). This was accomplished with the Sherpa 2.1.1 generator [66] and the CT10 parton distribution functions [67]. For leptonically decaying W and Z events samples were produced with a maximum of two additional partons at LO and a maximum of two at next-to-leading order (NLO). Next the LO hadronic W and Z cross sections were corrected by applying multiplicative "k-factors" derived from the corresponding NLO leptonic W and Z samples. These corrections were 1.28 for the W+jets and 1.37 for the Z+jets [28]. An alternate sample of hadronically decaying W

Dataset	Campaign	xsec (pb)	Filter Efficiency	Events Generated	Effective Luminosity (fb^{-1})
361020.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0W	MC16a	78420e+06	0.9755	16000000	2.0915e-07
361021.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1W	MC16a	78420e+06	0.00067143	15998000	7.5762e-05
361022.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2W	MC16a	24332e+05	0.00033423	15989500	0.0053309
361023.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3W	MC16a	26454e+03	0.00032016	15879500	0.44989
361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W	MC16a	25463e+01	0.00053138	15925500	29.858
361025.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5W	MC16a	4553.5	0.00092409	15993500	1103.2
361026.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6W	MC16a	257.53	0.00094242	17834000	25916
361027.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7W	MC16a	16.215	0.0003928	15983000	35888
361028.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8W	MC16a	0.62503	0.010176	15999000	25154e+02
361029.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9W	MC16a	0.019639	0.012076	13995500	59013e+03
361030.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10W	MC16a	0.0011962	0.0059087	13985000	19786e+05
361031.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11W	MC16a	4.2259e-05	0.0026761	15948000	1.4102e+11
361032.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12W	MC16a	1.0367e-06	0.00042592	15995600	3.6226e+13
361020.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ0W	MC16d	78420e+06	0.9755	15987000	2.0898e-07
361021.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ1W	MC16d	78420e+06	0.00067143	15997000	7.5728e-05
361022.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ2W	MC16d	24332e+05	0.00033423	15981000	0.005331
361023.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ3W	MC16d	26454e+03	0.00032016	15878500	0.45043
361024.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ4W	MC16d	25463e+01	0.00053138	15974500	29.968
361025.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ5W	MC16d	4553.5	0.00092409	15991500	1103.6
361026.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ6W	MC16d	257.53	0.00094242	17880400	25979
361027.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ7W	MC16d	16.215	0.0003928	15116500	33886e+01
361028.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ8W	MC16d	0.62503	0.010176	15987000	25136e+02
361029.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ9W	MC16d	0.019639	0.012076	14511500	61189+03
361030.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ10W	MC16d	0.0011962	0.0059087	15988000	22620e+05
361031.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ11W	MC16d	4.2259e-05	0.0026761	15993000	1.4142e+11
361032.Pythia8EvtGen_A14NNPDF23LO_jetjet_JZ12W	MC16d	1.0367e-06	0.00042592	15640000	3.5421e+13

Table 6.2: List of datasets used for QCD Monte Carlo [63].

and Z events was produced for cross checks using the Herwig++2.7.1 generator [68] with the UEEE4 tune [69] and the CTEQ6L1 [59]. Unlike the SHERPA samples these HERWIG samples only contained one additional parton in the matrix element. The hadronic W+jets and Z+jets samples used in this analysis [1] are summarized in Table 6.3.

Our $t\bar{t}$ samples were generated using POWHEG-BOX 2 [56] and the NNPDF30 NLO parton distribution functions [70]. After generation the events were showered using PYTHIA 8.230 [58] using the A14 tune and the NNPDF23 LO parton distribution functions [65] with all b-hadron decays performed by EVTGEN [60]. The samples were then broken up according to the decay mode of the two top quarks into three categories; all-hadronic, semi-leptonic, all-leptonic. Additional $t\bar{t}$ events were generated with the SHERPA 2.2.1 generator [66] using the NNPDF30 NNLO parton distribution functions [70]. This second sample was used as a cross check of the main samples generated with POWHEG-BOX 2 + PYTHIA 8. The $t\bar{t}$ samples used in this analysis are summarized in Table 6.4.

Single-top samples, containing a single top/anti-top quark and a W^-/W^+, were generated with POWHEG-BOX 2 [56] with the NNPDF30 parton distribution functions [67]. This process was showered using PYTHIA 8.230 [58] configured with the A14 tune and the NNPDF23 parton distribution functions [65] with all resulting b-hadrons decayed via EVTGEN [60]. The single-top samples used in this analysis are summarized in Table 6.5.

[1] Note that the leptonically decaying W and Z samples were only used to derive correction factors and thus are not presented.

Dataset	Campaign	xsec (pb)	Filter Efficiency	Events Generated	Effective Luminosity (fb^{-1})
304307.Sherpa_CT10_Wqq_Pt280_500	MC16a	29.487	1.0	150000	4.8429
304308.Sherpa_CT10_Wqq_Pt500_1000	MC16a	2.162	1.0	30000	12.646
304309.Sherpa_CT10_Wqq_Pt1000	MC16a	0.046218	1.0	15000	273.61
304307.Sherpa_CT10_Wqq_Pt280_500	MC16d	29.481	1.0	140000	4.5533
304308.Sherpa_CT10_Wqq_Pt500_1000	MC16d	2.1642	1.0	30000	13.431
304309.Sherpa_CT10_Wqq_Pt1000	MC16d	0.046152	1.0	15000	287.55
304707.Sherpa_CT10_Zqq_Pt280_500	MC16a	12.623	1.0	75000	5.8079
304708.Sherpa_CT10_Zqq_Pt500_1000	MC16a	0.90771	1.0	15000	16.090
304709.Sherpa_CT10_Zqq_Pt1000	MC16a	0.018344	1.0	10000	251.86
304707.Sherpa_CT10_Zqq_Pt280_500	MC16d	12.644	1.0	75000	5.6273
304708.Sherpa_CT10_Zqq_Pt500_1000	MC16d	0.90886	1.0	15000	16.278
304709.Sherpa_CT10_Zqq_Pt1000	MC16d	0.018264	1.0	10000	524.45
304673.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_280_500	MC16a	13.491	1.0	150000	11.119
304674.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_500_700	MC16a	0.91592	1.0	39000	42.580
304675.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_700_1000	MC16a	0.17413	1.0	30000	172.29
304676.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_1000_1400	MC16a	0.021936	1.0	30000	1367.6
304677.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_1400	MC16a	0.0022502	1.0	20000	8888.1
304673.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_280_500	MC16d	13.491	1.0	190000	14.083
304674.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_500_700	MC16d	0.91592	1.0	50000	54.590
304675.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_700_1000	MC16d	0.17413	1.0	40000	229.71
304676.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_1000_1400	MC16d	0.021936	1.0	40000	1823.5
304677.Herwigpp_UEEE5CTEQ6L1_Wjhadronic_1400	MC16d	0.0022502	1.0	30000	13332
304678.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_280_500	MC16a	5.4672	1.0	75000	13.718
304679.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_500_700	MC16a	0.37032	1.0	20000	54.007
304680.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_700_1000	MC16a	0.070961	1.0	15000	211.38
304681.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_1000_1400	MC16a	0.0087931	1.0	15000	1705.9
304682.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_1400	MC16a	0.00089549	1.0	10000	11167
304678.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_280_500	MC16d	5.4672	1.0	100000	18.291
304679.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_500_700	MC16d	0.37032	1.0	30000	81.011
304680.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_700_1000	MC16d	0.070961	1.0	20000	281.84
304681.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_1000_1400	MC16d	0.0087931	1.0	20000	2274.5
304682.Herwigpp_UEEE5CTEQ6L1_Zjhadronic_1400	MC16d	0.00089549	1.0	20000	22334

Table 6.3: List of datasets used for hadronically decaying W+jets and Z+jets Monte Carlo [63].

Dataset	Campaign	xsec (pb)	Filter Efficiency	Events Generated	Effective Luminosity (fb^{-1})
410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad	MC16a	729.77	0.54384	60413000	148.78
410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad	MC16a	729.78	0.45627	19994000	59.101
410472.PhPy8EG_A14_ttbar_hdamp258p75_dil	MC16a	729.77	0.10546	39974000	511.18
410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad	MC16d	729.77	0.54384	74486000	184.71
410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad	MC16d	729.78	0.45627	24714000	73.047
410472.PhPy8EG_A14_ttbar_hdamp258p75_dil	MC16d	729.77	0.10546	44876000	573.88
410249.Sherpa_221_NNPDF30NNLO_ttbar_AllHadronic_MEPS_NLO	MC16a	330.54	1.0	9993000	4.0608
410250.Sherpa_221_NNPDF30NNLO_ttbar_SingleLeptonP_MEPS_NLO	MC16a	158.76	1.0	14987000	8.9811
410251.Sherpa_221_NNPDF30NNLO_ttbar_SingleLeptonM_MEPS_NLO	MC16a	158.97	1.0	14991000	5.0114
410252.Sherpa_221_NNPDF30NNLO_ttbar_dilepton_MEPS_NLO	MC16a	76.277	1.0	9995000	1.2000
410249.Sherpa_221_NNPDF30NNLO_ttbar_AllHadronic_MEPS_NLO	MC16d	330.44	1.0	9988000	2.6319
410250.Sherpa_221_NNPDF30NNLO_ttbar_SingleLeptonP_MEPS_NLO	MC16d	158.74	1.0	14979000	15.336
410251.Sherpa_221_NNPDF30NNLO_ttbar_SingleLeptonM_MEPS_NLO	MC16d	159.01	1.0	14986000	12.569
410252.Sherpa_221_NNPDF30NNLO_ttbar_dilepton_MEPS_NLO	MC16d	76.09	1.0	9983000	0.48186

Table 6.4: List of datasets used for $t\bar{t}$ Monte Carlo [63].

Dataset	Campaign	xsec (pb)	Filter Efficiency	Events Generated	Effective Luminosity (fb^{-1})
410646.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_top	MC16a	37.936	1.0	4996000	130.45
410647.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_antitop	MC16a	37.905	1.0	4999000	130.65
410646.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_top	MC16d	37.936	1.0	6243000	162.99
410647.PowhegPythia8EvtGen_A14_Wt_DR_inclusive_antitop	MC16d	37.905	1.0	6240000	163.10

Table 6.5: List of datasets used for single-top Monte Carlo [63].

Chapter 7

Physics Object Definitions

In order to analyze the data one must first reconstruct the detector-level output into representations of real physical particles (physics objects). This process begins once an event is accepted by the ATLAS trigger system and recorded to disk, as discussed in Section 5.5. This raw detector output, such as energy deposits in the calorimeters or hits in the inner detector, is then fed into a system of algorithms which build objects of interest such as jets and muons. These complex objects can then be used as representations of the true SM particles in the analysis. Since these representations are only as good as the measurement and reconstruction algorithms, they are then calibrated and given an uncertainty as discussed in Chapter 10. This chapter gives an overview of the methodology used to reconstruct the objects required for the boosted $H \to b\bar{b}$ analysis.

7.1 Jets

The particles that carry QCD color charge do not exist by themselves in isolation, but instead combine with other colored particles to form colorless composite hadrons making it impossible to observe individual quarks or gluons in ATLAS. This process of hadronization, discussed in Section 2.3, creates a shower of charged and neutral particles that leave tracks in the inner detector and energy deposits in the calorimeters. Each pp collision event at the LHC tends to produce a lot of these hadronic showers which are then clustered into objects known as "jets." These clustered jet objects thus represent the reconstructed signature of a gluon or quark after it hadronizes.

Jets can be formed from tracks using the hits of charged particles in the ID, from energy deposits left by both neutral and charged particles in the calorimeters, or from the truth particles produced via MC. Since there is no unique procedure for clustering these signatures, several different approaches have been developed. The most common clustering options are the k_t, Cambridge-Aachen, and anti-k_t algorithms [71]. These algorithms work by iteratively applying the following rules to the chosen collection of objects:

1. Define the distance d_{iB} between the object i and the beamline B in terms of the transverse momentum p_{T} and P which determines the effect of the energy in the clustering algorithm:

$$d_{iB} = p_{\mathrm{T}i}^{2P}$$

2. Compute the pairwise distance d_{ij} between objects i and j:

$$d_{ij} = \min\left(p_{\mathrm{T}i}^{2P}, p_{\mathrm{T}j}^{2P}\right) \frac{\Delta R_{ij}^2}{R^2}$$

where $\Delta R_{ij}^2 = (\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2$ is the geometric term dependent on rapidity η and the azimuthal angle ϕ. R is a parameter which controls the size of the jet.

3. Choose the smallest distance $d_{\min}$ out of the list of distances:

$$d_{\min} \in \{\{d_{ij}\}, \{d_{iB}\}\}$$

4. If $d_{\min}$ is the distance between an object, i, and the beamline:

$$d_{\min} \in \{d_{iB}\}$$

label this object a jet and remove it from the list

5. If $d_{\min}$ is the distance between some objects i and j:

$$d_{\min} \in \{d_{ij}\}$$

merge objects i and j into a new object k, add k to the object collection and remove objects i and j.

6. Repeat until all objects have been clustered into jets and the object collection is

empty.

In the above iterative process the choice of parameter P corresponds to three choices of jet clustering algorithm:

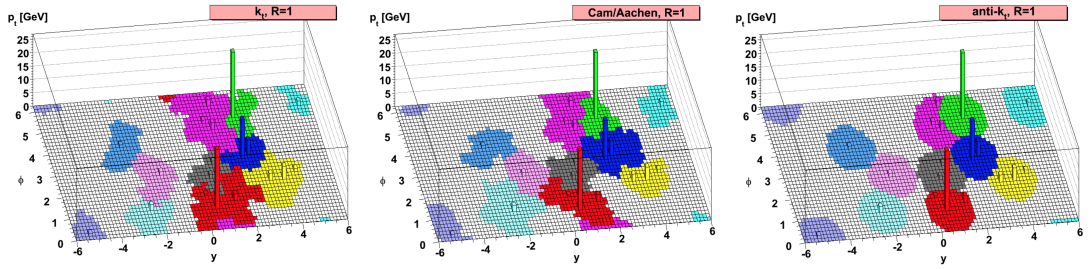
$P = 1$ This defines the k_t algorithm [72]. Here the algorithm prioritizes the clustering of soft objects first and then gradually moves on to harder objects. This has the effect of creating irregular shaped jets that evolve from the "outside-in" as seen in Figure 7.1a.

$P = 0$ This defines the Cambridge-Aachen algorithm [73]. In this case the effect of the energy of the jet is ignored, and instead the jets are clustered based on their geometric distance only. Again this results in irregular shaped jets as seen in Figure 7.1b.

$P = -1$ This defines the anti-k_t algorithm which was used in this analysis [71]. This algorithm prioritizes the clustering of hard objects first and then clusters the surrounding softer objects. This results in mostly cone-like jets which center on the hardest object(s) as seen in Figure 7.1c.

7.1.1 Large-R Jets

This dissertation focuses on measuring the boosted signature of the Higgs boson as it decays to $b\bar{b}$. As discussed in Section 3.5, the resulting decay products are highly



(a) k_t (b) Cambridge-Aachen (c) anti-k_t

Figure 7.1: Example clustering of a parton-level MC simulation event showing clustered jet shapes for $R = 1.0$ and (a) $P = 1$, (b) $P = 0$, and (c) $P = -1$ [71].

collimated and thus can be reconstructed using a single large radius ("large-R") jet. These large-R jets are reconstructed from topological calorimeter clusters using the anti-k_t algorithm with a radius parameter of $R = 1.0$ resulting in jets similar to the one shown in Figure 7.2 [74, 75]. After clustering the large-R jets are "trimmed" to improve mass resolution and reduce dependence on pile-up. Trimming is done by first using the k_t algorithm to recluster the constituents into subjets with $R = 0.2$. Then any subjet with p_{T} less than 5% of the parent large-R jet's energy is removed [76].

7.1.2 Variable Radius Track Jets

After capturing the decay of the Higgs in a large-R jet the next step is to identify the two sub-jets that represent the b and $\bar{b}$ children of the Higgs. In this dissertation the Variable Radius (VR) jet was chosen [78, 79], defined by a radius parameter R_{eff} which decreases as a function of the jet p_{T}:

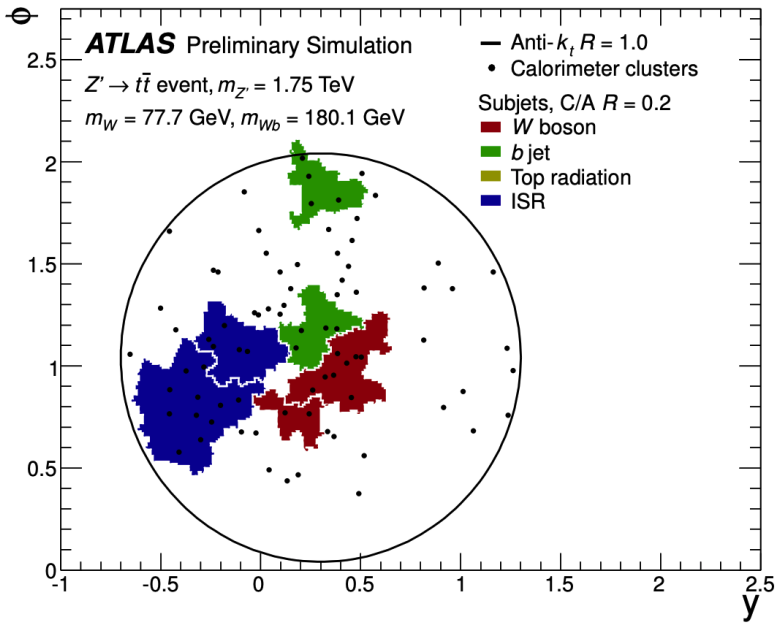


Figure 7.2: Simulation of calorimeter clusters for decay of a boosted $Z' \to t\bar{t}$ clustered in a large-R jet [77].

$$R_{\text{eff}}(p_{\text{T}}) = \frac{\rho}{p_{\text{T}}}.$$

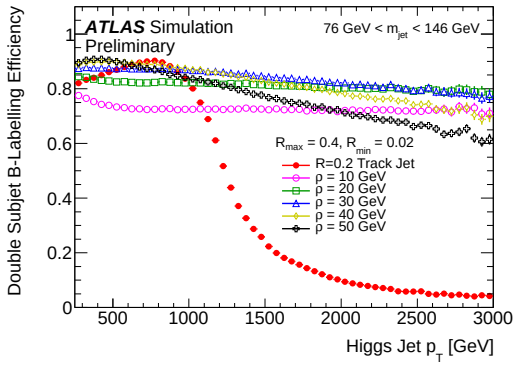
The constant ρ determines how quickly the effective size of the VR jet decreases with respect to the transverse momentum contained inside the jet. In this definition the choice of ρ should be proportional to the mass of the resonance you are attempting to reconstruct and should correctly reproduce the size of jets as long as $\rho \lesssim 2p_{\text{T}}$. In addition to ρ the VR jet algorithm requires two parameters, R_{min} and R_{max}, to impose lower and upper cut-offs on the jet size, respectively. These parameters are scanned to optimize the reconstruction of track jets from $H \to b\bar{b}$. Using these bounds the R_{eff} becomes

$$R_{\text{eff}}\left(p_{\text{T}}\right) = \max\left[\min\left(\frac{\rho}{p_{\text{T}}}, R_{\text{max}}\right), R_{\text{min}}\right] .$$

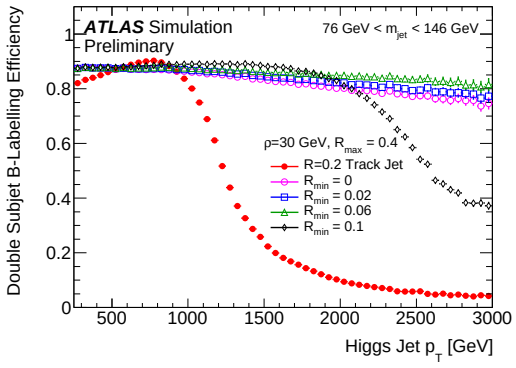
In reconstructing the Higgs boson with $m_H = 125$ GeV, the variable radius track jet parameters were chosen to be $\rho = 30$ GeV, $R_{\text{min}} = 0.02$, and $R_{\text{max}} = 0.4$, in order to maximize the truth subjet double b-tagging efficiency [79], as seen in Figure 7.3. In Figure 7.4 the VR Track Jets properly describe the truth ΔR distribution for b-hadrons while the $R = 0.2$ fixed track jets deviate for higher Higgs jet p_{T}. This deviation is caused by the subjets being so collimated that they begin to overlap. Furthermore, in Figure 7.5 and Figure 7.6 the VR track jets do an equally good or better job of identifying the subjets associated with b-hadrons when compared to the $R = 0.2$ fixed radius track jets. This ability to give a flat efficiency across the entire Higgs p_{T} spectrum while accurately describing the topology makes these VR track jets the perfect choice for this analysis.

7.2 Flavor Tagged Jets

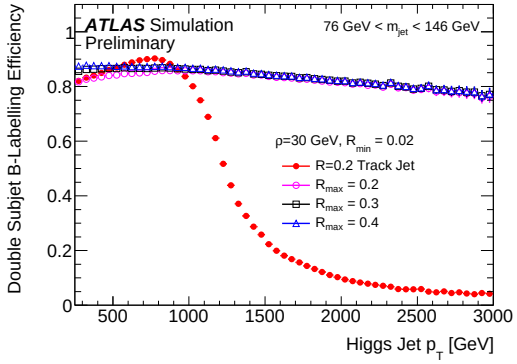
In general reconstruction algorithms for jets are agnostic to the "flavor" label - light, charm (c), or bottom (b) - of the hadrons produced in the shower. However, flavor tagging is a powerful tool for discriminating the $b\bar{b}$ decay products of the Higgs from the large, predominantly light-flavor, multijet background [80]. These b-quark initiated



(a) Study of ρ VR track jet parameter with $R_{\text{min}} = 0.02$ and $R_{\text{max}} = 0.4$



(b) Study of R_{min} VR track jet parameter with $\rho = 30$ GeV and $R_{\text{max}} = 0.4$



(c) Study of R_{max} VR track jet parameter with $\rho = 30$ GeV and $R_{\text{min}} = 0.0.02$

Figure 7.3: Labeling efficiency of subjet double b-tagging using truth information from Higgs decay as a function of Higgs jet p_{T}. The efficiency for $R = 0.2$ fixed radius track jests is included for comparison. Uncertainty bars include statistical uncertainties only [79].

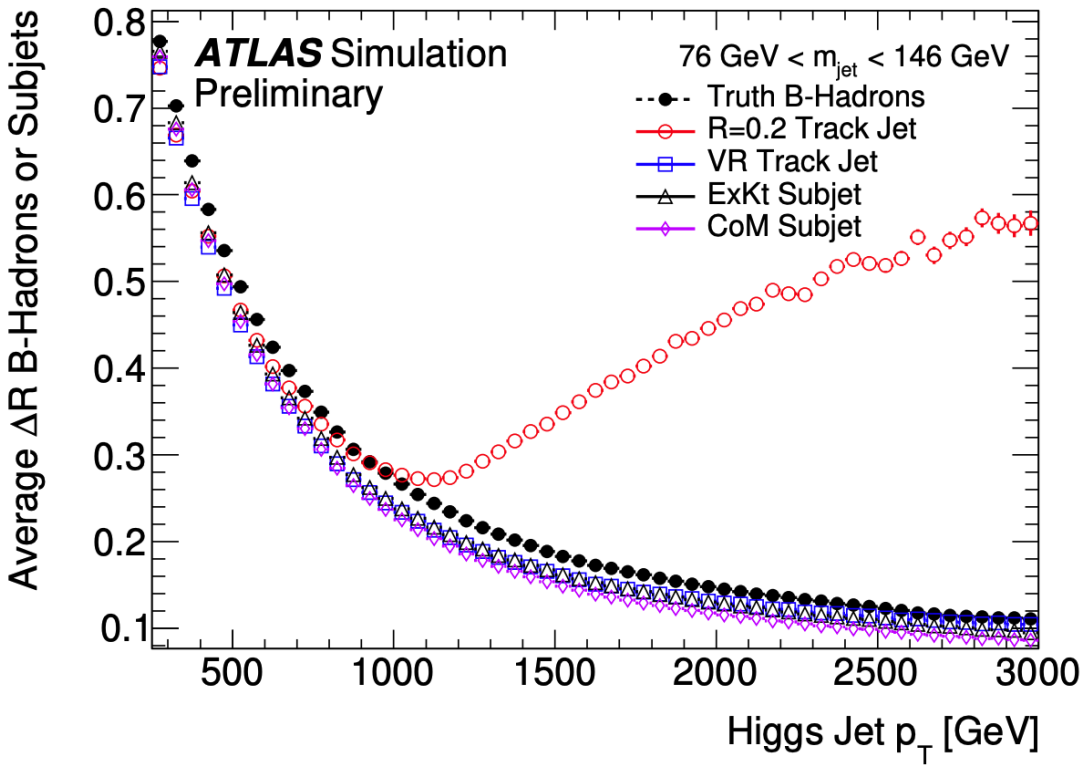
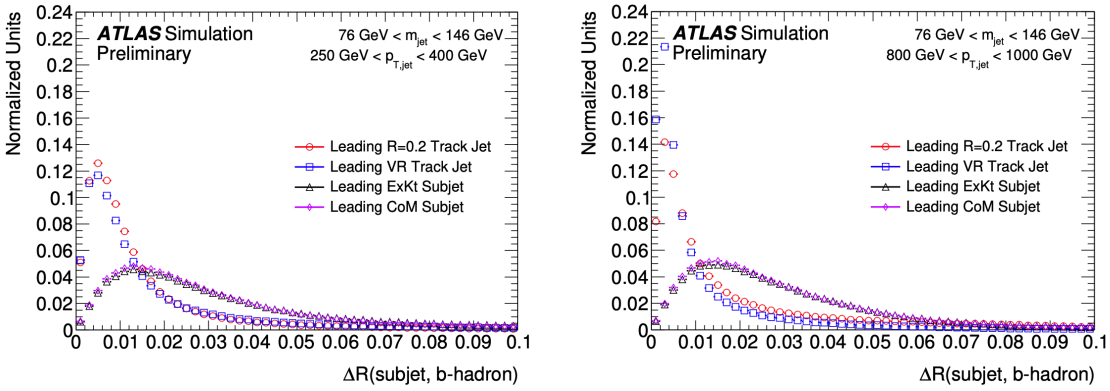


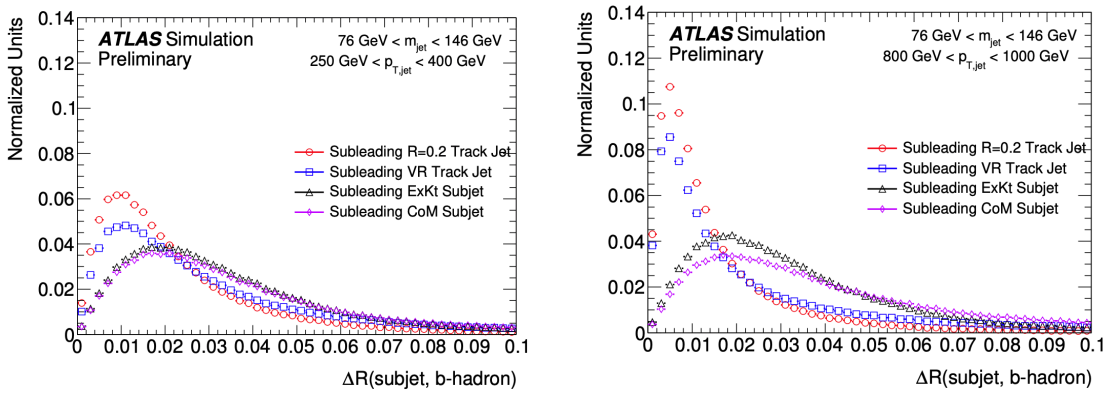
Figure 7.4: The average ΔR between either the two leading truth b-hadrons or the two leading subjets associated to a Higgs jet as a function of Higgs p_{T} [79].



(a) 250 GeV < Higgs p_{T} < 400 GeV

(b) 800 GeV < Higgs p_{T} < 1000 GeV

Figure 7.5: Distributions of ΔR between a truth matched b-hadron and the reconstructed leading subjet [79]. The uncertainties given reflect only statistical uncertanties. All algorithms are normalized to an area corresponding to the fraction of signal jets which contain a leading subjet.



(a) 250 GeV < Higgs p_{T} < 400 GeV (b) 800 GeV < Higgs p_{T} < 1000 GeV

Figure 7.6: Distributions of ΔR between a truth matched b-hadron and the reconstructed subleading subjet [79]. The uncertainties given reflect only statistical uncertainties. All algorithms are normalized to an area corresponding to the fraction of signal jets which contain a subleading subjet.

jets are identified using the `MV2c10` [81] Boosted Decision Tree (BDT) [1], trained with a machine learning algorithm that uses the weighted score from a series of decision trees to give a discriminant for how similar any given jet is to a b-jet. The BDT uses inputs from the kinematics of the jet (p_{T} and $|\eta|$) and the outputs of tracking algorithms, discussed below, to look for signatures consistent with a b-hadron decay as shown in Figure 7.7. Tracking information is crucial to flavor tagging, and thus the flavor tagging can only be applied within the tracking volume ($|\eta| < 2.5$).

The relatively long lifetime of b-hadrons ($\approx$ 1.5 ps) gives them a characteristic length scale of $c\tau \sim .45$ mm. This means the b-hadron travels the non-negligible distance of $\sim$ 5 mm from the primary interaction vertex before decaying, assuming $\gamma = 10$. This macroscopic flight distance is large enough that this decay can be identified as a

[1] The name `MV2c10` means that this multivariate algorithm had a training sample with roughly $\sim 10\%$ c-jets and $\sim 90\%$ light jets in order to define a good balance of c-jet and light jet rejection.

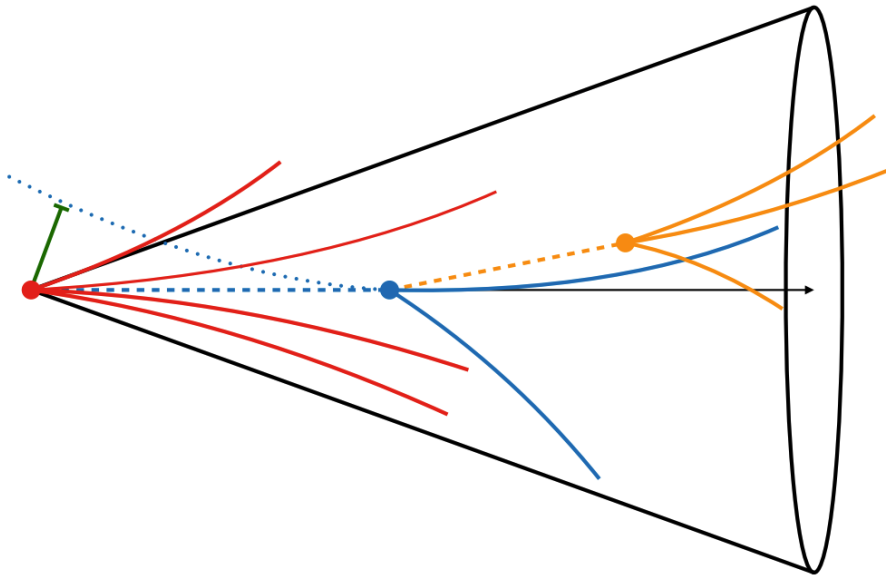
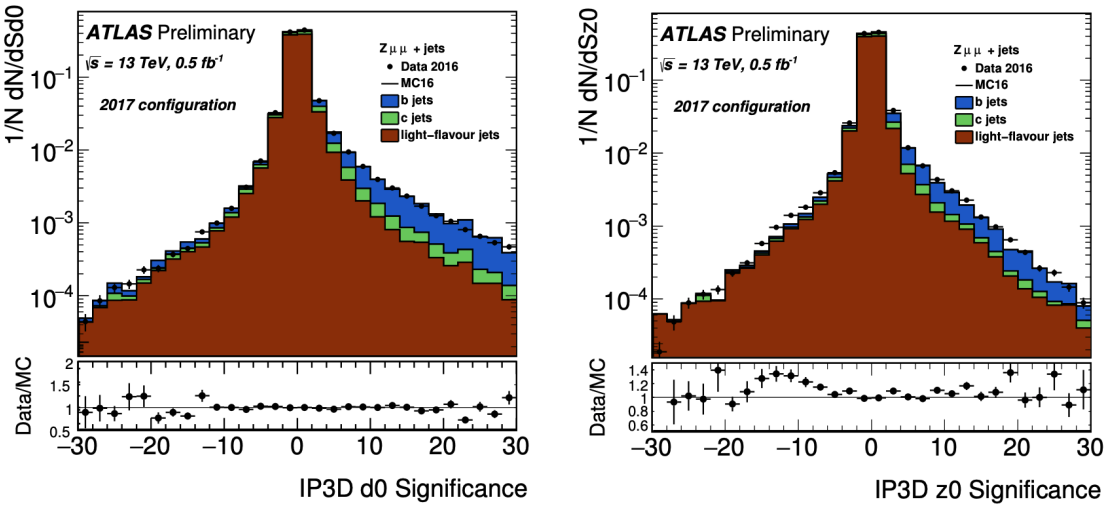


Figure 7.7: Cartoon of a b-jet decay containing a b-hadron decay vertex (blue ●) displaced from the primary pp vertex (red ●), and a c-hadron decay vertex (orange ●) further displaced and often close to the b-hadron flight axis [82]. The secondary (blue) and tertiary (orange) vertices have large impact parameters (green) with respect to the primary pp vertex.

secondary vertex (SV) displaced from the original primary vertex (PV). Furthermore, roughly 90% of b-jets will contain a c-jet which will create a tertiary vertex when it decays (TV) [82]. The secondary vertex finding algorithm (SV1), and Kalman filter algorithm (JetFitter) look for events matching this characteristic b-hadron decay chain.

Using tracking information the two-dimensional and three-dimensional impact parameter algorithms, IP2D and IP3D, determine the transverse and longitudinal impact-parameters - d_0 and z_0 - respectively. Looking at Figure 7.8 the impact parameters of the b-flavor jets tend to be positive, while those of c-jets and light jets tend to be distributed more symmetrically around 0.

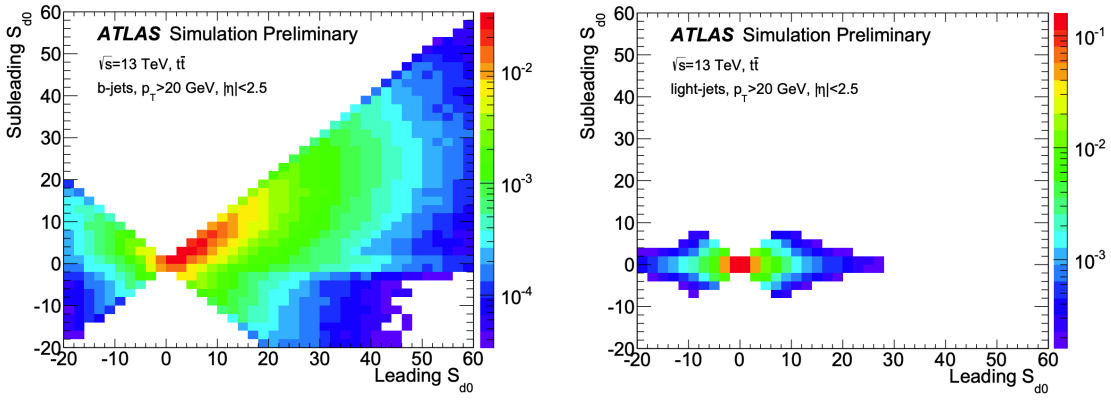


(a) Transverse Impact Parameter d_0 (b) Longitudinal Impact Parameter z_0

Figure 7.8: Data-Monte Carlo comparisons of the transverse (d_0) and longitudinal (z_0) impact parameter significance values for IP3D selected charged tracks in the leading jet of a $Z \to \mu\mu +$ jets dominated sample [82].

In 2017 two new tools were incorporated into the tagger [83]; a Recurrent Neural Network (RNN) impact parameter tagger (RNNIP) and a Soft Muon Tagger (SMT). The

RNNIP [84] exploits the fact that b-jets tend to have many tracks with highly significant impact parameters, while light jets do not, as seen in Figure 7.9. The SMT [85] searches for muons coming from the semi-leptonic decays of b-hadrons and c-hadrons to discriminate against light jets which have much softer muons or none at all. The separation power of the SMT is shown in Figure 7.10.



(a) b-jets (b) light jets

Figure 7.9: The distribution of the transverse impact parameter significance S_{d0} for the leading d_0 significance track and subleading d_0 significance track for b-jets (left) and light jets (right) [82].

All of the above algorithms' outputs, along with jet kinematic information, are used as input to the `MV2c10` algorithm as seen in Figure 7.11. The output is a discriminant score which indicates how b-jet-like or how un-b-jet-like the jet in question is, compared to the training sample used, as shown in Figure 7.12. The performance is calibrated in data using jets containing a muon, indicating the semileptonic decay of the b-hadron, and a correction is derived for the simulated events [86].

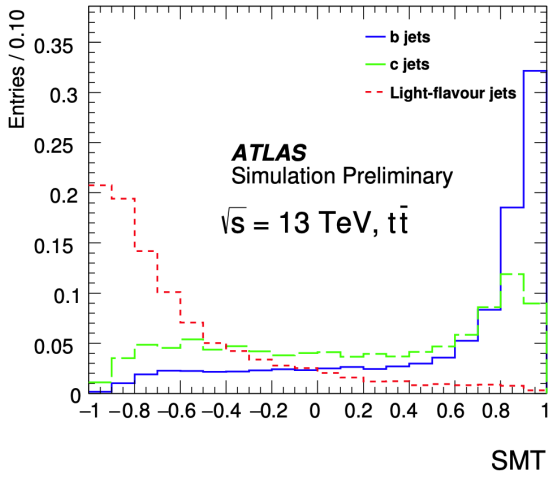


Figure 7.10: Normalized BDT response in simulated $t\bar{t}$ events of the SMT for reconstructed muons associated to b-jets (blue), c-jets, (green) and light-flavour jets (red) [82].

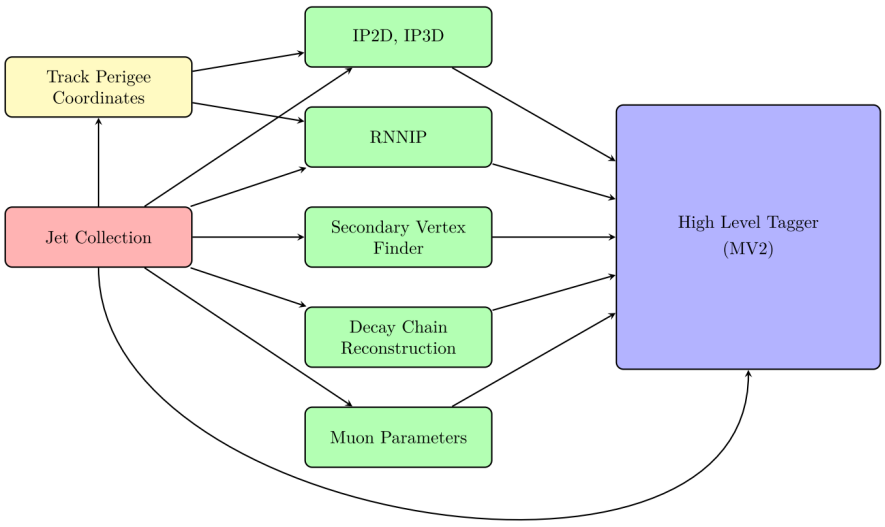
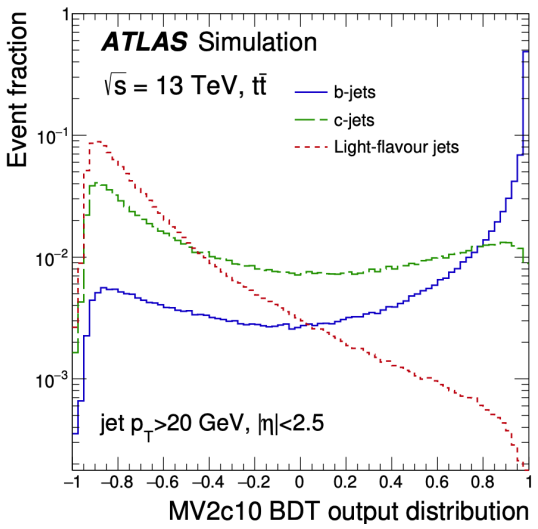
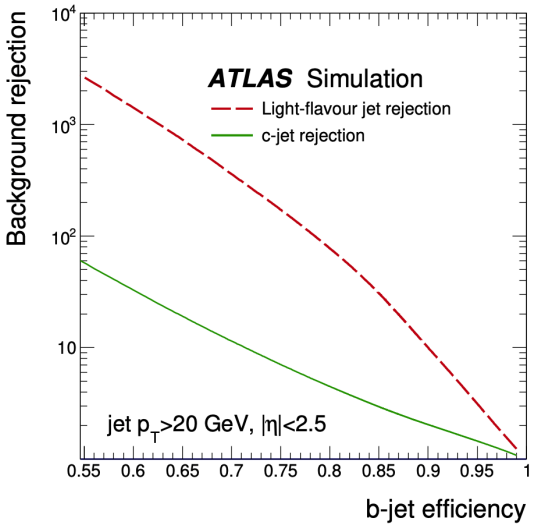


Figure 7.11: Flowchart of inputs to the `MV2c10` b-tagging algorithm [29].



(a) `MV2c10` discriminant for b-jets compared to c-jets and light jets in simulated $t\bar{t}$ events.



(b) light jet and c-jet rejection factors as a function of b-jet tagging efficiency of the `MV2c10` BDT.

Figure 7.12: Performance of the `MV2c10` BDT for the 2016 optimization in simulated $t\bar{t}$ events. The performance was evaluated on $t\bar{t}$ events simulated using POWHEG interfaced to PYTHIA6 [86].

7.3 Muons

As muons traverse the entire detector they leave a track of charge deposits in the Inner Detector (ID), Calorimeter and Muon Spectrometer (MS) which is then reconstructed to represent the path of the muon [87]. This results in four different muon types, dependent upon which subdetectors were used in the reconstruction:

1. Combined (CB) muons: First the charged track is reconstructed independently in the ID and MS. Then a global refit combines the hits from both subdetectors. This global fit may add or subtract MS hits from the track to achieve the best fit quality.

2. Segment-tagged (ST) muons: A track is developed in the ID and then extrapolated to the MS. If this extrapolation finds at least one local track segment in the MDT or CSC it is labeled a muon. This is generally used for low p_{T} muons which may only traverse one layer of the MS.

3. Calorimeter-tagged (CT) muons: A track formed in the ID is labeled a muon if it can be associated with a calorimeter deposit consistent with a minimum-ionizing particle. This is the least pure muon type, but it allows for reconstruction of muons that pass through the partially instrumented region of the MS.

4. Extrapolated (ME) muons: These muons are reconstructed using only MS track information and the loose requirement that the hits are compatible with a trajectory originating from the interaction point. This type is useful for extending the muon acceptance into the region not covered by the ID.

After the muon type is determined the quality of the muon is categorized by requiring a specific number of hits in each subcomponent. These quality requirements are provided to address the specific needs of different physics analysis. The four muon quality levels are defined:

`loose` The lowest quality is designed to maximize the reconstruction efficiency for muons by allowing all muon types to be used. This is primarily useful for analyses of multi-leptonic final states such as $H \to 4\ell$.

`medium` The medium quality is designed to minimize the systematic uncertainties associ-

ated with muon reconstruction and calibration. Only CB and ME tracks are used, with at least 3 CB track hits and at least 3 ME layers. This is the default quality selection in ATLAS and the one used for muons in this analysis.

`tight` This selection maximizes the purity of muons but reduces the reconstruction efficiency. Only CB muons with at least 2 layers of the MS that also satisfy the `medium` selection requirements are allowed.

`high-`p_{T} Designed to optimize the momentum resolution for tracks with $p_{\mathrm{T}} > 100$ GeV. This selection only includes CB muons with hits in at least two layers of the MS that also satisfy the `medium` selection requirements. This quality level is mostly used for high-mass W' and Z' analyses.

The final step for muon reconstruction is to check that the muon is well isolated, in order to suppress muons resulting from meson and heavy-flavor decays. This is done using both the track-based ($p_{\mathrm{T}}^{\mathrm{varcone30}}$) and the calorimeter-based ($E_{\mathrm{T}}^{\mathrm{topocone20}}$) isolation-based variables which represent the scalar sum of p_{T} inside a $\Delta R < 0.3$ cone and the scalar sum of E_{T} inside a $\Delta R < 0.2$ cone. Taking the ratio of these variables to the total p_{T} of the candidate muon gives a sense of how much radiation is surrounding the core of the muon in question. Many different isolation working points are established by cutting on these ratio distributions. For this analysis the `loose` working point was chosen which gives a 99% muon reconstruction efficiency constant in η and p_{T} as measured in both a simulated $Z \to \mu\mu$ sample and data [87].

After reconstruction, these muons are calibrated to data using the well understood decay $J/\Psi \to \mu^+\mu^-$ to cover the low p_{T} spectrum and $Z \to \mu^+\mu^-$ for the high p_{T} spectrum, as shown in Figure 7.13.

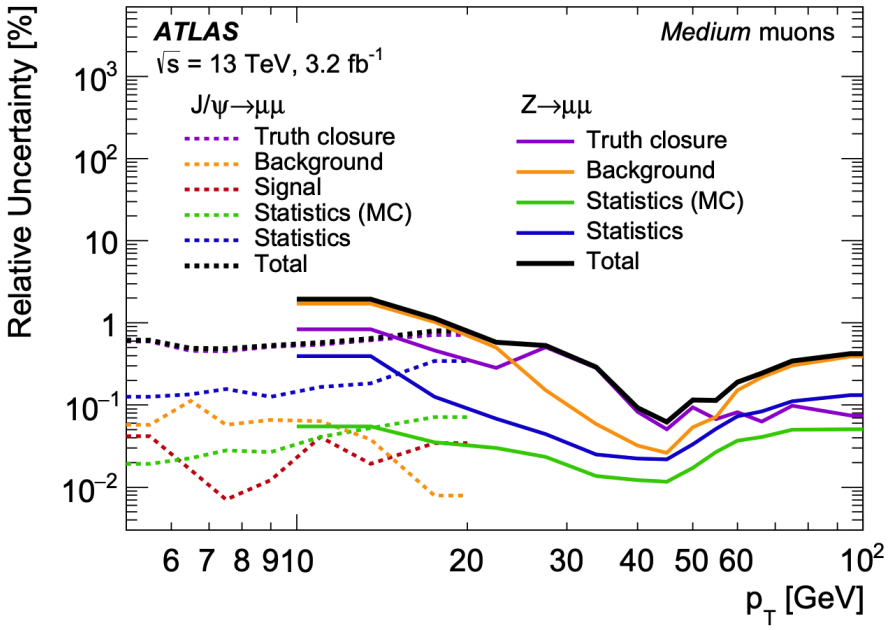


Figure 7.13: Total uncertainty in the efficiency scale factor for Medium muons as a function of p_{T} as measured in $Z \to \mu\mu$ (solid lines) and $J/\Psi \to \mu\mu$ (dashed lines) decays. The combined uncertainty is the sum in quadrature of the individual contributions [87].

Chapter 8

Boosted $H \to b\bar{b}$ Event Selection

Having discussed the methods used to reconstruct the physics objects needed to accurately capture the boosted $H \to b\bar{b}$ final state, the next consideration is the actual requirements made on these reconstructed objects. These selections and quality criteria are applied to define the set of baseline and candidate physics objects that will be used as part of the analysis. This includes the definition of multiple orthogonal control regions (CR) used during the statistical fitting procedure for the characterization of backgrounds, as well as the signal region (SR) where the actual Higgs search is realized.

8.1 Selected Triggers

This analysis uses exclusively large-R jet triggers in order to maximally capture the boosted decay products of the Higgs boson. However, pile-up increased over the course

of LHC Run 2 as shown in Figure 4.5. This increase in pile-up corresponds to an increase in the underlying event energy that can be clustered into jets and thus an increase in the rate of the large-R jet trigger. As a result the High Level Trigger (HLT) p_{T} requirement was increased over the years in order to maintain a low trigger rate for data recording. This results in a different triggering algorithm for the 2015, 2016, and 2017 data taking years. While all the triggers require a large-R jet to be reconstructed in the HLT, the 2015 and 2016 triggers used an ungroomed (a10) large-R jet algorithm while in 2017 a trimmed (a10t) version was used. In 2015 the large-R trigger looked for an ungroomed large-R jet with $p_{\mathrm{T}} > 360$ GeV. In 2016 the p_{T} threshold was raised to 420 GeV. In 2017 the requirement changed to look for a trimmed large-R jet with $p_{\mathrm{T}} > 480$ GeV. This information is summarized in Table 8.1, which also details the integrated luminosity for the various triggers as well as the offline p_{T} threshold. In all cases the offline threshold differs from the HLT threshold. This is due to the fact that the trigger must make a quick calculation of the jet energy in order to keep data rates low. However, this online calculation sacrifices accuracy, so some jets that should have passed are instead ignored. This results in a turn-on curve in the p_{T} distribution near the online threshold where the trigger is not yet fully efficient. Thus, a higher offline threshold is chosen corresponding to when the trigger is fully efficient. All triggers used in this analysis are fully efficient for an offline threshold of $p_{\mathrm{T}} > 480$ GeV. In order make sure the triggers are fully efficient in the offline selection, all events are required to contain a trimmed large-R jet with $p_{\mathrm{T}} > 480$ GeV.

Table 8.1: Summary of the large-R jet triggers used for the data taking periods of 2015, 2016, and 2017 and the offline p_{T} thresholds at which they become fully efficient. The recorded integrated luminosity for each trigger is also included [63].

Year	Trigger Name	Offline p_{T} Threshold (GeV)	Luminosity $\left(\mathrm{fb}^{-1}\right)$
2015	HLT_j360_a10_lcw_sub_L1J100	410	3.2
2016	HLT_j420_a10_lcw_L1J100	450	33.0
2017	HLT_j460_a10t_lcw_jes_L1J100	480	44.3

8.2 Event Selection

Starting from the events passing the large-R jet trigger discussed in Section 8.1, further selections are made to maximize signal sensitivity, reject background, and construct regions where the backgrounds can be characterized.

First, all events are preselected to make sure they contain at least two large-R jets with $p_{\mathrm{T}} > 250$ GeV [1] that fall entirely within the ID instrumented region $|\eta| < 2.0$. This ensures that charged tracks can be reconstructed for flavor tagging purposes. Events are further preselected by requiring the leading p_{T} large-R jet to have $p_{\mathrm{T}} > 480$ GeV and the sub-leading p_{T} large-R jet to have $p_{\mathrm{T}} > 250$ GeV. The leading jet p_{T} cut ensures that all events pass the offline p_{T} threshold for all three triggers. The sub-leading jet p_{T} cut is an explicit requirement for the presence of the ISR jet.

Next, the signal candidate large-R jet, assumed to contain the decay products of the Higgs boson, is selected from the list of large-R jets passing the following requirements. The signal candidate must have $p_{\mathrm{T}} > 480$ GeV and be sufficiently boosted such that

[1] The 250 GeV cut off is due to the p_{T} requirements in the EXOT8 derivations used in this analysis.

$p_{\mathrm{T},J} > 2m_J$. It must contain at least 2 VR track-jets with $p_{\mathrm{T}} > 10$ GeV, as required for b-tagging. Furthermore, the distance between the two leading VR track-jets (ΔR_{VR}) must be greater than the radius of the smaller of the two VR track-jets ($\min R_{VR}$). This requirement helps avoid b-tagging anomalies and prevent signal contamination from gluon splitting. The highest-p_{T} jet that passes all of the above requirements is labeled the signal candidate, and the next highest-p_{T} large-R jet in the list is taken to be the ISR jet. Any events containing a muon with $p_{\mathrm{T}} > 40$ GeV opposite the signal candidate large-R jet, $\Delta\phi > 2\pi/3$, are removed to ensure no overlap with the $t\bar{t}$ control region ($\mathrm{CR}_{t\bar{t}}$) discussed in Section 9.2 and to reduce $t\bar{t}$ contribution in the SR. The event selection process is illustrated in Figure 8.1.

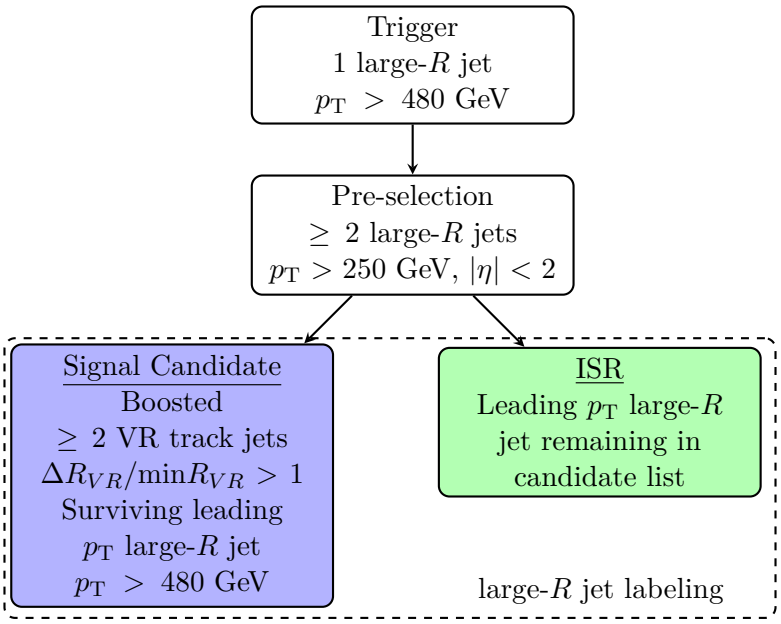
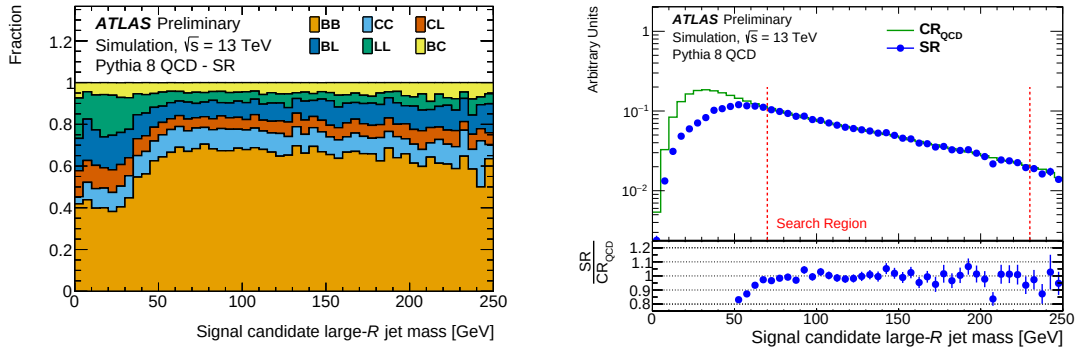


Figure 8.1: Diagram of the event selection process and the labeling scheme of the large-R jets in the signal candidate events [29].

The signal candidate events are classified based on how many of the two leading VR

trackjets pass b-tagging criteria. Two b-tagging criteria were used: the "loose" working point with 85% efficiency and the "tight" working point with 77% efficiency [2]. Events with exactly 0 "loose" b-tagged track-jets form the control region used to estimate the non-resonant QCD background ($\mathrm{CR_{QCD}}$) discussed in chapter Section 9.3. Events with exactly 2 "tight" b-tagged track-jets form the signal region (SR) where the final fit is performed. The 77% working point used to define the SR was chosen to optimize signal significance. For reference, the simulated flavor composition of the SR is shown in Figure 8.2a. The cutflow for the $\mathrm{CR_{QCD}}$ and SR are shown for the Higgs boson signal in Table 8.2 and for data and backgrounds in Table 8.3.



(a) (b)

Figure 8.2: (a) Predicted flavor composition of the dijet background in the SR based on the truth-matched hadron content of the two leading-p_{T} track-jets associated to the signal candidate large-R jet, with the B/C labels indicating the presence of a b/c-quark and L indicating the presence of a light quark or a gluon. (b) The expected shape of the dijet background in the SR and CR normalized to the same event count between 70 GeV $< m_J <$ 230 GeV [88].

The QCD estimate is only valid for regions where the SR and $\mathrm{CR_{QCD}}$ have a similar dijet mass shape. This results in a lower bound cut on the signal candidate mass of

[2] These working points are determined by applying the b-tagging algorithm to a Monte Carlo sample of $t\bar{t}$ events and then checking the output against truth information to determine efficiencies.

	Cut	ggF	VBF	WH	ZH	Total
Preselection	Trigger	2749	1063	665	408	4884
	Jet Cleaning	2748	1062	665	408	4882
	Lead large-R jet $p_{\mathrm{T}} > 480$ GeV	593	258	219	132	1201
	Sublead large-R jet $p_{\mathrm{T}} > 250$ GeV	546	229	185	115	1076
	At least one signal candidate	530	222	181	112	1045
	No opposite muon ($\mathrm{CR}_{t\bar{t}}$ veto)	530	222	179	111	1042
	signal candidate $p_{\mathrm{T}} > 480$ GeV	470	189	166	102	927
	signal candidate mass$>$ 40 GeV	401	153	154	94	802
$\mathrm{CR}_{\mathrm{QCD}}$	0 b-tagged VR subjets (85% WP)	140	41	71	40	292
SR	2 b-tagged VR subjets (77% WP)	115	53	26	21	216

Table 8.2: Cutflow showing the effect in simulation of selection criteria for each of the Higgs boson production mechanisms [63]. The numbers here can be compared to the number of events generated as shown in Table 6.1.

	Cut	QCD	W+jets	Z+jets	$t\bar{t}$	Total	Data
Preselection	Trigger	385385408	1419912	629320	1624282	389193728	159567488
	Jet Cleaning	385195264	1418308	628717	1623173	389000192	159354608
	Lead large-R jet $p_{\mathrm{T}} > 480$ GeV	77069624	427349	182546	427728	78151368	77183760
	Sublead large-R jet $p_{\mathrm{T}} > 250$ GeV	69628144	403870	172806	354653	70596512	71117008
	At least one signal candidate	61771844	385898	165464	340425	62699272	64152952
	No opposite muon ($\mathrm{CR}_{t\bar{t}}$ veto)	61770952	385864	165439	321558	62677836	64110376
	signal candidate $p_{\mathrm{T}} > 480$ GeV	52034768	348965	148796	288012	52851096	53996920
	signal candidate mass$>$ 40 GeV	38050452	295753	124293	258416	38756440	39555696
$\mathrm{CR}_{\mathrm{QCD}}$	0 b-tagged VR subjets (85% WP)	29435344	219353	84389	110905	29863870	29883336
SR	2 b-tagged VR subjets (77% WP)	400020	1506	6173	10553	419087	484551

Table 8.3: Cutflow showing the effect of selection criteria on simulated background samples as well as data. The simulated QCD contribution is scaled by 0.74 which was derived from a fit to data in the $\mathrm{CR}_{\mathrm{QCD}}$ [63]. The background simulation numbers here can be compared to the number of events generated as shown in Tables 6.2 to 6.4.

$m_J > 70$ GeV due to the higher turn-on curve of the SR shown in Figure 8.2b. This low mass turn on is due to the b-tagging algorithm not being able to resolve the b-jets as they begin to overlap. Furthermore, for $m_J > 230$ GeV the boost from a p_{T} of 480 GeV is no longer sufficient to merge two hadrons into a single large-R jet. Thus the SR signal candidate masses considered in the analysis range from 70 GeV to 230 GeV.

Given the above criteria, the efficiencies and yields in the SR and $\mathrm{CR_{QCD}}$ for the resonant backgrounds and the Higgs boson signal are shown in Table 8.4. The composition of the vector boson, $t\bar{t}$ and $H \to b\bar{b}$ resonant components of the SR and $\mathrm{CR_{QCD}}$ are given in Table 8.5. For the vector boson background, the $W+$jets contribution dominates in the $\mathrm{CR_{QCD}}$ due to its larger cross section. However, the $Z+$jets contribution dominates in the SR, as the Z boson can decay to two b-quarks. The $t\bar{t}$ contribution is roughly the same in the SR and $\mathrm{CR_{QCD}}$ with $\sim 60\%$ of events decaying entirely to hadrons (all hadronic), $\sim 40\%$ of events with one W boson decaying leptonically (semi-leptonic), and a small percentage of events with both W bosons decaying leptonically (dileptonic). In the SR the dominant $H \to b\bar{b}$ production mechanism is ggF, contributing 53% of the signal, followed by VBF production with 25% and Higgstrahlung at 22%.

Table 8.4: The efficiencies and yields in the 0-tag control region (CR_{QCD}) and signal region (SR) for the non-QCD background, the Higgs boson signal and data. The yields in the CR_{QCD} are scaled to the luminosity used for the background estimate of the non-resonant dijet process discussed in Section 9.3. The efficiencies are calculated relative to the leading large-R jet $p_T > 480$ GeV requirement [63].

Process	CR_{QCD} Eff. (%)	CR_{QCD} Yield in 1.4 fb^{-1}	SR Eff. (%)	SR Yield in 80.5 fb^{-1}
$W \to q\bar{q} + \text{jets}$	51.3	3810	0.4	1500
$Z \to q\bar{q} + \text{jets}$	46.2	1470	3.4	6200
$t\bar{t}$	25.9	1929	2.5	10550
$H \to b\bar{b}$	24.3	5	17.9	216
ggF	23.6	2	19.4	115
VBF	15.8	1	20.7	53
WH	32.4	1	12.0	26
ZH	30.5	1	15.8	21
Data	38.7	519710	0.6	484600

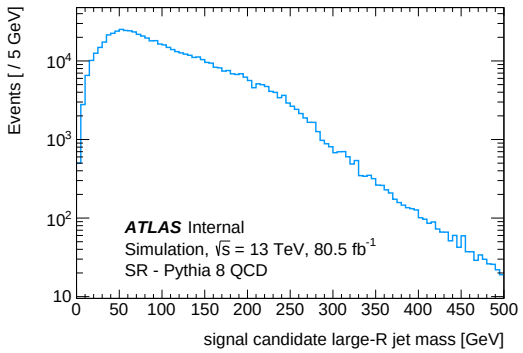
Table 8.5: The fractional composition of the different resonant contributions in the 0-tag control region (CR_{QCD}) and the signal region (SR). The fraction is evaluated using the given contribution type as the total [63].

Process	CR_{QCD} Fraction	SR Fraction
$V + \text{jets}$		
$Z + \text{jets}$	0.28	0.80
$W + \text{jets}$	0.72	0.20
$t\bar{t}$		
hadronic	0.58	0.63
semi-leptonic	0.38	0.34
dileptonic	0.04	0.03
$H \to b\bar{b}$		
ggF	0.50	0.53
VBF	0.17	0.25
WH	0.21	0.12
ZH	0.12	0.10

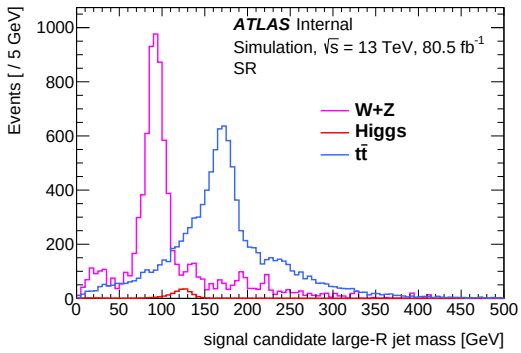
Chapter 9

Background Estimation

The estimated contribution of the resonant and non-resonant backgrounds is critical to the final statistical interpretation discussed in Chapter 11 . The resonant $V+$jets and $t\bar{t}$ processes are estimated using the MC samples discussed in Section 6.3, whereas a data-driven technique is used to estimate the non-resonant QCD background. The final fit is done in the invariant mass spectrum of the signal candidate in the SR over the range 70 GeV $< m_J <$ 230 GeV. As shown in Figure 9.1, the Higgs boson signal is small, and it is on the tails of the comparably larger resonant backgrounds. This means that any statistical fluctuation in the MC templates could hide the signal. To avoid this, smooth parametric shapes are fitted to the resonant MC histograms and then used as the inputs to the fit. This chapter covers the creation of the $V+$jets template; the $t\bar{t}$ template, including the derivation of a correction to its normalization using a fit in the $\mathrm{CR}_{t\bar{t}}$; and the data-driven modeling of the QCD background in the $\mathrm{CR}_{\mathrm{QCD}}$.



(a) Simulation of the contribution of non-resonant multijet QCD background in the signal region.



(b) Simulation of resonant V+jets and $t\bar{t}$ backgrounds in the signal region. The Higgs simulation is included for comparison.

Figure 9.1: Simulations of the non-resonant and resonant background contributions to the signal region for an integrated luminosity of 80.5 fb^{-1}

9.1 Hadronic Vector Boson Background

The V + jets template is constructed by fitting the generated MC histogram with the sum of three Gaussians plus a constant term. The systematic variations, discussed in Chapter 10, of this template are re-fit using the same functional choices. This results in smooth histograms to be used as input to the fit discussed in Chapter 11.

9.2 Top Quark Pair Background

The substantial contribution to the SR by boosted $t\bar{t}$ production shown in Table 8.4 makes its modeling a top priority [1]. Unfortunately, current MC generators are not able to predict the $t\bar{t}$ cross section well in this boosted regime as seen in Figure 9.2. This long-standing issue is likely due to missing higher-order diagrams rather than due to

[1] Pun!

suboptimal generator setup [89]. To compensate for this mismodeling the $t\bar{t}$ yield in the SR is corrected with a normalization scale factor. The scale factor, or k-factor, is derived by fitting the $t\bar{t}$ normalization in the $t\bar{t}$ enriched control region ($\mathrm{CR}_{t\bar{t}}$). In the final fit the SR $t\bar{t}$ MC sample is fit by a double-sided Crystal Ball function [90] to smooth out statistical fluctuations, and then its normalization is constrained with the derived flat k-factor using its uncertainty.

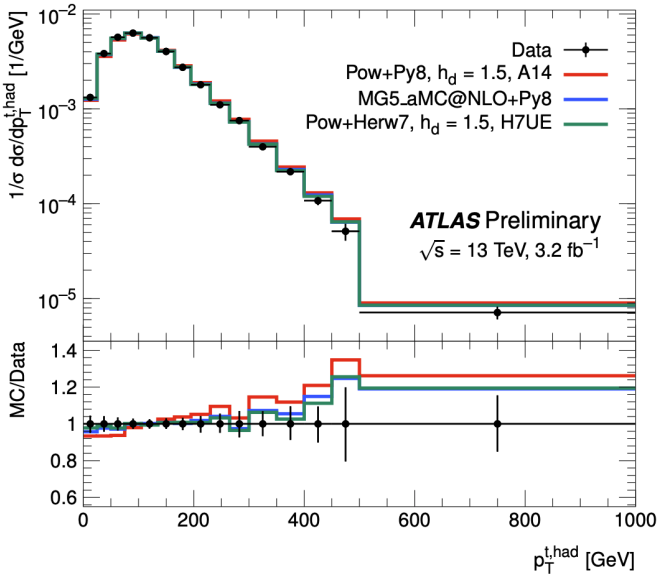
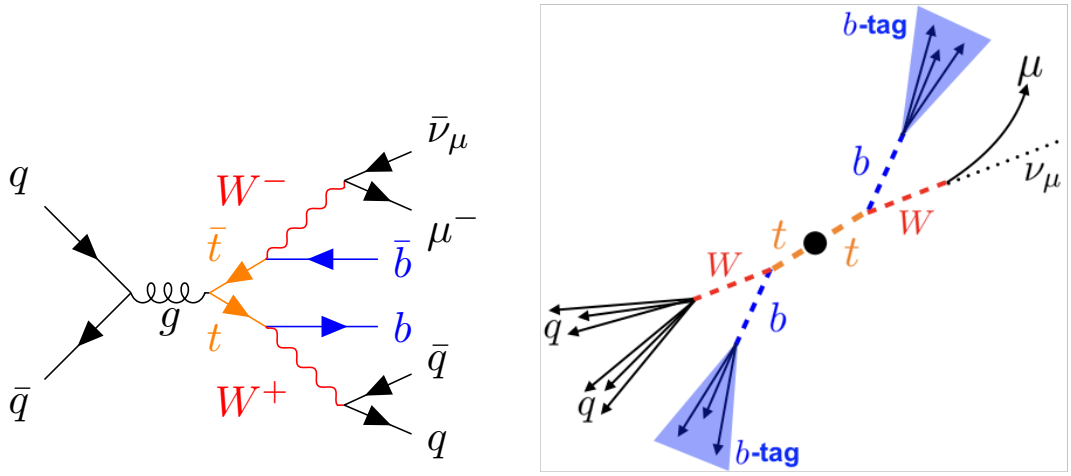


Figure 9.2: Comparison in p_{T} of the top quark for different generator setups used to assess the NLO+PS matching as well as the parton shower and hadronization uncertainty after optimization, compared to data at $\sqrt{s} = 13$ TeV [89].

9.2.1 Constructing the $\mathrm{CR}_{t\bar{t}}$ Control Region

The $t\bar{t}$-enriched control region uses the same p_{T} selections as the signal candidate large-R jet in addition to the following criteria to select $t\bar{t}$ events like the one shown in Figure 9.3. Three regions are defined by requiring zero ($\mathrm{CR}_{t\bar{t}0}$), one ($\mathrm{CR}_{t\bar{t}}$) or two ($\mathrm{CR}_{t\bar{t}2}$) tight

quality b-tags in the two leading VR track jets of the signal candidate. The configuration with exactly one b-tag, $\mathrm{CR}_{t\bar{t}}$, was chosen to exploit the single b-quark that results from the dominant decay mode of the top quark $t \to bW$. This one b-tag region has contributions from $t\bar{t}$, single-top, Z and H when one b-tag is lost, and multijet when the b-tag is faked. The other two regions are used to validate the extrapolation of the k-factor into the $\mathrm{CR}_{\mathrm{QCD}}$ and SR.



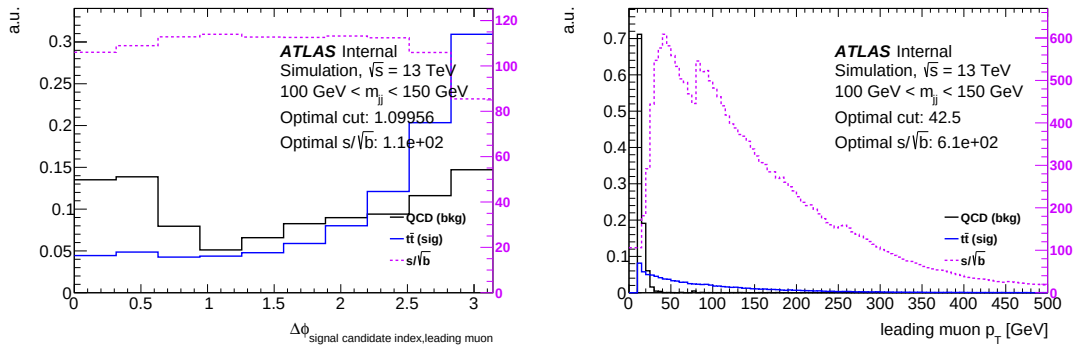
(a) (b)

Figure 9.3: (a) Feynman diagram of semi-leptonic $t\bar{t}$ decay. (b) Cartoon depicting semi-leptonic $t\bar{t}$ decay to b-quarks in center of mass frame.

To reduce multijet contamination in the sample the second top quark in the opposite hemisphere of the signal candidate is required to decay leptonically in the muon channel. Figure 9.4a compares the $\Delta\phi$ between the leading muon in the event and the signal candidate jet for the multijet and $t\bar{t}$ MC samples. This distribution shows the expected back-to-back topology of the $t\bar{t}$ system depicted in Figure 9.3b versus the multijet sample where the muons come from the signal candidate due to hadron decays in flight. Thus, a

cut of $\Delta\phi(\text{muon} - \text{signal}) > \frac{2\pi}{3}$ was chosen[2] to reduce the QCD contribution by several orders of magnitude while only reducing the $t\bar{t}$ contribution by roughly a factor of five.

The QCD contribution is further suppressed by requiring the muon $p_{\mathrm{T}} > 40$ GeV. Figure 9.4b compares the p_{T} spectra for $t\bar{t}$ and QCD before the $\Delta\phi$ cut selection. The larger mass of the W boson leads to higher p_{T} muons than the softer x muons from heavy flavor jets.



(a) $\Delta\phi$(leading muon $-$ signal) study (b) Leading muon p_{T} study

Figure 9.4: Comparison of the two variables, $\Delta\phi$(leading muon $-$ signal) (a) and muon p_{T} (b), between PYTHIA8 multijet and POWHEG $t\bar{t}$ simulated events used to create the $t\bar{t}$ control region. The selection requires exactly one b-tagged track jet in the signal candidate large-R jet. The magenta line shows the expected $s/\sqrt{b}$ significance with 80.5 fb^{-1} [63].

Finally, at least one tight b-tagged track-jet is required to be within $\Delta R < 1.5$ of the chosen muon. This requirement exploits the collimation of the decay products of the boosted top quark $t \to b\mu\nu_{\mu}$. This reduces the contamination from $V(ll)$+jets and VV events. The above CR$_{t\bar{t}}$ selections are summarized visually in Figure 9.5.

[2]Note that while the dotted magenta line in Figure 9.4a technically indicates an optimal $\Delta\phi$(muon $-$ signal) cut of 1.09956, practically a cut of $\frac{2\pi}{3}$ was chosen to reject as much multijet background as possible.

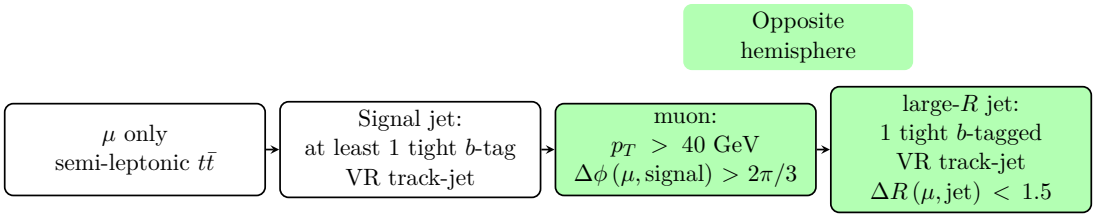


Figure 9.5: Diagram of the $CR_{t\bar{t}}$ selection criteria.

9.2.2 k-factor Estimation

Finally the $t\bar{t}$ MC template is fit to the data in the $CR_{t\bar{t}}$ over the mass range 100 GeV to 200 GeV. Single top (Wt) and $W \to l\nu$ templates were included in the fit with their normalizations kept constant. The full statistical and systematic uncertainty is determined by running the Bayesian Analysis Toolkit (BAT) [91], discussed in Chapter 11, with the large-R jet energy scale (JES), jet mass resolution (JMR), luminosity, $t\bar{t}$ modeling and flavor tagging systematic uncertainties discussed in Chapter 10.

The pre- and post-fit distributions for the $CR_{t\bar{t}}$ are shown in Figure 9.6 with the pull distributions shown in Figure 9.7. The JES and JMR systematics are constrained by information about the peak in $t\bar{t}$ which was not available in the dataset used to derive the recommendations[3]. A systematic uncertainty dominated k-factor of 0.84 ± 0.11 is found for $CR_{t\bar{t}}$, showing that the MC overestimates the $t\bar{t}$ yield as expected (see Figure 9.2). This value is used to constrain the $t\bar{t}$ contribution in the final fit to the SR. The results for all three $t\bar{t}$ control regions are given in Table 9.1. All three regions are consistent with each other and the results presented in reference [92].

[3]Note that the $CR_{t\bar{t}}$ and final fit are not simultaneous so these JES and JMR constraints do not affect the final fit.

Table 9.1: The $t\bar{t}$ scale factors and their uncertainties from the three $t\bar{t}$ control regions. The value for $\mathrm{CR}_{t\bar{t}}$ is used in the Signal Region. NOTE: $\mathrm{CR}_{t\bar{t}2}$ fit failed due to low statistics but is almostly completely dominated by ttbar in this region. The value here is the simple ratio of number of events in ttbar vs. data, and the uncertainty is the $t\bar{t}$ statistical uncertainty.

Region	scale factor	uncertainty
$\mathrm{CR}_{t\bar{t}0}$	0.87	0.12
$\mathrm{CR}_{t\bar{t}}$	0.84	0.11
$\mathrm{CR}_{t\bar{t}2}$	0.96	0.21

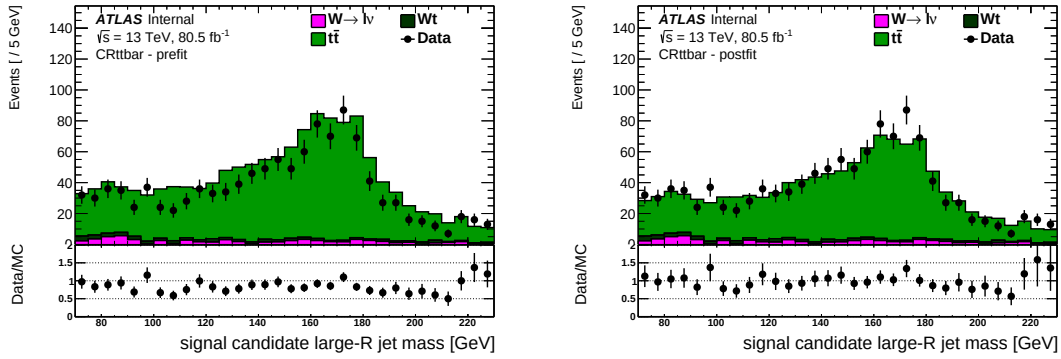


Figure 9.6: The pre-fit (left) and post-fit (right) data vs MC comparison for fitting the $\mathrm{CR}_{t\bar{t}}$ region.

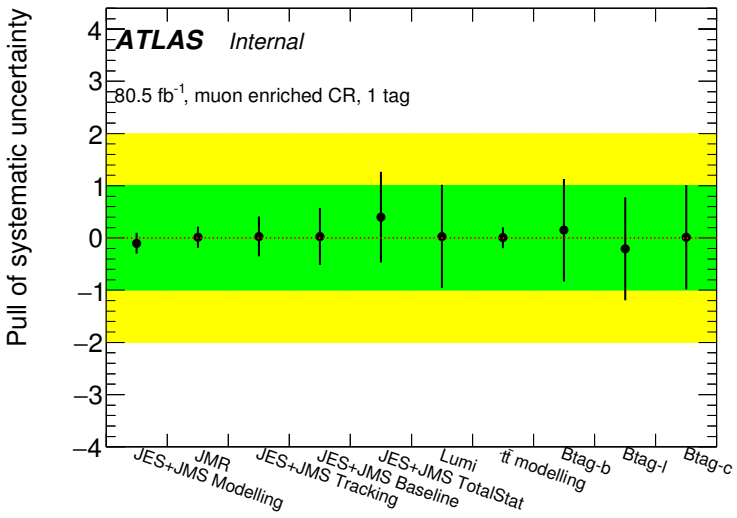


Figure 9.7: The pull distributions for the different nuisance parameters used in the $\mathrm{CR}_{t\bar{t}}$ fit. Parameters with a error less than 1 indicate that the prior uncertainty has been constrained by the fit.

9.3 Multijet QCD estimation

The dominant background contribution to the SR comes from the non-resonant multijet QCD process. Unfortunately the estimation of this process through MC is not reliable due to the statistical precision of available samples and the underlying inaccuracies of the event generation. Thus, a data-driven estimate was employed by fitting the large-R jet mass distribution, m_J, in the SR with a parametric function after validation of the procedure in the $\mathrm{CR}_{\mathrm{QCD}}$. This approach is further motivated by the good agreement in shape of the $\mathrm{CR}_{\mathrm{QCD}}$ and SR over the mass range of the fit, 70 GeV to 230 GeV, as seen in Figure 8.2b. A brief summary of this process is presented below, with in-depth details available in Reference [29].

It was found that the number of events in a $\sim$ 1 fb^{-1} dataset from the CR$_{\text{QCD}}$ is comparable to the number of events in the full 80.5 fb^{-1} SR dataset. Thus, the CR$_{\text{QCD}}$ was broken into 60 slices constructed using adjacent data runs resulting in an average of 1.2 fb^{-1} of data per slice.

Two families of parametric functions were used for this modeling. The polynomial exponential function was chosen to be the nominal model

$$f_n\left(x\middle|\vec{\theta}\right) = \theta_0 \exp\left(\sum_{i=1}^{n} \theta_i\, x^i\right), \quad x = \frac{m_J - 150 \text{ GeV}}{80 \text{ GeV}}, \tag{9.1}$$

and the alternative model, the formal Laurent series,

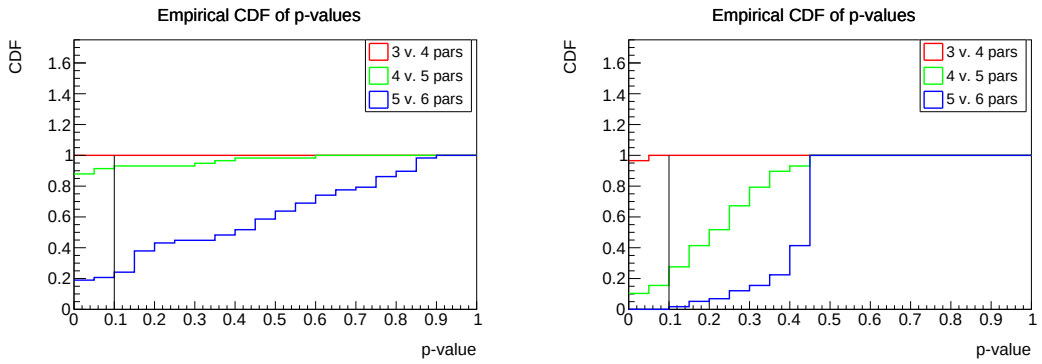
$$f_n\left(x\middle|\vec{\theta}\right) = a \sum_{i=0}^{n} \frac{\theta_i}{x^{i+1}}, \quad a = 10^5,\, x = \frac{m_J + 90 \text{ GeV}}{160 \text{ GeV}}. \tag{9.2}$$

The θ coefficients are determined by the fit, a is empirically chosen to keep the scale of parameters at $\mathcal{O}(1)$, and the independent variable x parameterizes the fit range $m_J \in [70, 230]$ GeV to $x \in [-1, 1]$ for Equation (9.1) and $x \in [1, 2]$ for Equation (9.2). This reparameterization was empirically seen to provide improved numerical stability in the fit.

Both functions are tested on a random CR$_{\text{QCD}}$ slice to determine the minimum number of model parameters needed to describe the shape of the distribution. The Z + jets, W + jets, and k-factor corrected $t\bar{t}$ contributions are all scaled by their cross sections

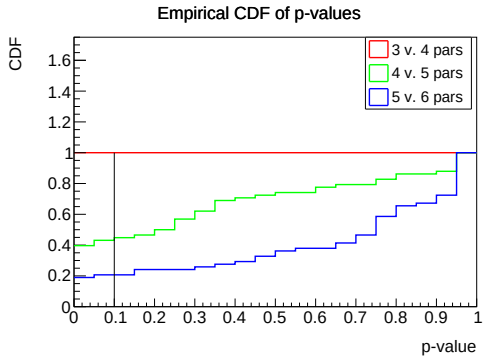
times the luminosity and then subtracted from the slice to remove bias from the fit. The results of a likelihood ratio test with Wilks' theorem [93] and the F-test [94] were both used to determine the minimum number of model parameters for both function choices. The two statistical tests were performed for all $\mathrm{CR}_{\mathrm{QCD}}$ data slices and the relative frequencies of the observed p-values were iteratively summed as shown in Figure 9.8 for both function choices. The Wilks test preferred a five-parameter model for both while the F-test preferred a four-parameter model for both. To be conservative in the final estimate, the five-parameter model was chosen for both the polynomial exponential function and the formal Laurent series.

In order to determine the robustness of these fit functions, they were validated using all of the $\mathrm{CR}_{\mathrm{QCD}}$ slices. For these validation studies the fit included properly scaled Z + jets, W + jets, $t\bar{t}$ and single top contributions in addition to the $\mathrm{CR}_{\mathrm{QCD}}$ slice. Within the statistical precision given by the different data slices, the χ^2/ndf from the individual fits is found to follow the expected distribution of a good fit.

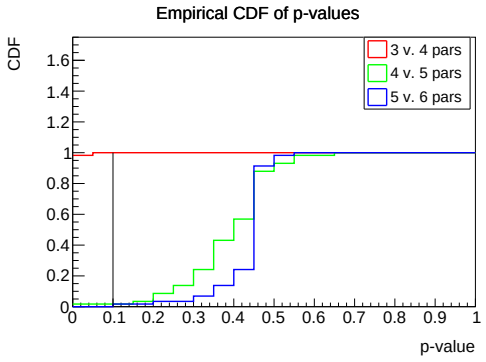


(a) Polynomial exponential Wilks test

(b) Polynomial exponential F test



(c) Formal Laurent Wilks test



(d) Formal Laurent F test

Figure 9.8: Empirical cumulative distribution function (CDF) for the test p-values for the two statistical tests applied to both function choices where the a priori threshold α is shown as a vertical black line. In both tests p-values below $\alpha = 0.1$ show preference for the model with more parameters, while p-values above $\alpha = 0.1$ indicate a preference for the model with fewer parameters [63].

Chapter 10

Systematic Uncertainties

This chapter presents the sources of systematic uncertainty in the analysis that could result in biases to the final fit. These sources of uncertainty can come from experimental calibrations or procedures and from the MC modeling of signal and background. They can contribute to uncertainties in the overall yield ("normalization") and to the differential shape of the signal candidate mass observable used in the statistical procedure for the final fit. The sources of uncertainty on the derivation of the QCD multijet background are separated from the uncertainties on the resonant backgrounds as the former is estimated using data while the later are estimated using MC simulation. Each of the relevant sources of uncertainty are briefly discussed and a summary is presented.

10.1 Uncertainties on Data-Driven QCD Modeling

The QCD background estimate discussed in Section 9.3 is based on a direct fit to data. The choice of function used to fit the QCD background is empirically based so it is important to consider the uncertainty introduced by this choice. The uncertainty is modeled by considering the differences between the nominal polynomial exponential function in Equation (9.1) and the alternate formal Laurent series Equation (9.2) in a series of pseudoexperiments. Samples are produced via Poisson sampling of the QCD component of the fit to the SR which contains all components of the full model (QCD, Z + jets, W + jets, $t\bar{t}$, and the Higgs boson signal) but no nuisance parameters. These samples are then fit with the nominal and alternative function choices. The uncertainty bands from these two sets of fits provide a measure of the statistical uncertainty on the multijet parameterization derived from the spread of fit parameters, and a systematic uncertainty on the choice of fitting function derived from the difference between the two fitted shapes.

10.2 Uncertainties on Resonant Backgrounds and Signal

All MC templates in this analysis contain uncertainties derived from the large-R jet energy and mass calibrations [74] and the calibration of the `MV2c10` b-tagging algorithm [86] which affects b, c, and l jet flavors differently. The jet energy and jet mass calibration uncertainties affect both normalization and shape of the templates. This means their

impact on the analysis must be determined by varying the jet energy and jet mass up and down within their uncertainties and propagating the varied templates through the entire analysis procedure. The effect of the jet energy resolution uncertainty was also tested but found to be negligible. The impacts of uncertainties on the calibration of the `MV2c10` algorithm have been found to be independent of the large-R jet mass for all MC templates and thus only affect the signal normalization.

To account for the modeling uncertainties associated with the choice of MC generator, additional shape uncertainties are applied to the V + jets and $t\bar{t}$ templates[1]. The systematic uncertainty is determined by taking the difference between the large-R jet mass shapes from two different MC generators. For V + jets the nominal shape was generated using Sherpa 2.1.1 and then compared to the alternate shape generated with Herwig++ 2.7. For $t\bar{t}$ the nominal shape was generated with Powheg-Box 2 and compared with the alternate shape generated with Sherpa 2.1.1.

The $t\bar{t}$ normalization in the SR is constrained by the k-factor derived from a fit to data in the $\mathrm{CR}_{t\bar{t}}$ as discussed in Section 9.2. The uncertainty on the measurement was found to be 13% and is treated as a systematic uncertainty on the $t\bar{t}$ normalization.

In order to have an accurate simulation the normalization of all MC templates must be scaled by the integrated luminosity of the dataset in question. The systematic uncertainty on the integrated luminosity measurement is derived following the methodology presented in Ref. [95]. For this analysis it was found to be 2.1%.

[1]Note that no dedicated uncertainty for Higgs shape was generated because the theory uncertainty is so large

Theoretical systematic uncertainties on the normalization of the V + jets and Higgs components are added to the fit in the SR. Theory uncertainties on the V + jets processes result from the impact of higher order electroweak and QCD corrections to their differential cross sections [96]. For Higgs boson production the dominant theoretical uncertainty is from ggF production which is taken to be 30%. This is consistent with the cross section uncertainty calculated with the MiNLO procedure and includes the effects of the non-zero top-mass for Higgs bosons with $p_{\mathrm{T}} > 400$ GeV [97]. If the finite top mass is implemented correctly the resulting uncertainty does not depend on p_T^H. This same uncertainty is then applied for the other production mechanisms for Higgs' with $p_{\mathrm{T}} > 400$ GeV. The result is a total theory uncertainty on the Higgs boson production cross section of 30%.

10.3 Summary of Systematics

The signal strength (μ_s) is defined as the scale factor which multiplies the cross section times branching fraction that is predicted by the signal hypothesis being considered. Thus, the final fit, discussed in Chapter 11, of the theory-based MC to the SR data provides a direct measurement of this signal strength and therefore a direct measurement of the model's prediction given the data. However, to understand the significance and error on this result the statistical and systematic uncertainties must be taken into account during the fitting procedure.

To get a sense of the individual impact of each systematic on the total uncertainty for the measurement of μ_s, the following ad hoc comparison is made. First, the total uncertainty on μ_s (σ_{tot}) is determined by running the fit with all sources of error (systematic and statistical) allowed to float within their defined constraints. By definition:

$$\sigma^2_{\text{tot}} = \underbrace{\sum_{i}^{n} \sigma_i^2}_{\substack{\text{Contributions from} \\ n\text{systematics}}} + \underbrace{\sigma^2_{\text{stat.}}}_{\substack{\text{Contribution from} \\ \text{statistics}}} \tag{10.1}$$

Next, the fit is re-run with all systematics fixed to their pre-fit values except for the systematic being investigated for its impact. This gives σ_i, the uncertainty on μ_s when only considering the effect of the i^{th} nuisance parameter. Note that the statistical uncertainty (σ_{stat}) is still included as it is inherent to the fitting procedure.

Finally, the difference in quadrature between σ_{tot} and σ_i is found and divided by the measured μ_s value found in the final fit which includes all sources of uncertainty.

$$\frac{\sqrt{\Delta\sigma_i^2}}{\mu_s} = \frac{\sqrt{\sigma^2_{\text{tot}} - \sigma_i^2}}{\mu_s} \tag{10.2}$$

This ad hoc comparison represents the impact of the i^{th} nuisance parameter on the total uncertainty of the measurement of μ_s. A summary of the impact of each systematic for the two considered signal models (Higgs + jets and V + jets) is presented in Table 10.1.

Table 10.1: Summary of the impact ($\sqrt{\Delta\sigma_i^2}/\mu_s$) of the main systematic uncertainties on the total uncertainty, σ_{tot}, of the measured signal strength, μ_s, for the V + jets and Higgs boson + jets signals [88].

Source	Type	V + jets	Higgs + jets
Jet energy and mass scale	Norm. & Shape	15%	14%
Jet mass resolution	Norm. & Shape	20%	17%
V + jets modeling	Shape	9%	4%
$t\bar{t}$ modeling	Shape	$< 1\%$	1%
b-tagging (b)	Normalization	11%	12%
b-tagging (c)	Normalization	3%	1%
b-tagging (l)	Normalization	4%	1%
$t\bar{t}$ k-factor	Normalization	2%	3%
Luminosity	Normalization	2%	2%
Alternate QCD function	Norm. & Shape	4%	4%
W/Z and QCD (Theory)	Normalization	14%	
Higgs (Theory)	Normalization		30%

Chapter 11

Statistical Fit

The final significance of V + jets and Higgs + jets production are quantified through a Bayesian implementation of a likelihood function. This procedure uses the information known about the systematic uncertanties and normalization of the signal before the fit, called priors, along with the template histograms representing the impact of each prior on the distribution of interest m_J. These priors are then iteratively updated by the data, contained in a likelihood function, using a Markov Chain Monte Carlo procedure. This results in a marginalized probability density function (p.d.f) which represents the probability of there being a number of signal events present in the data. This p.d.f is then used to determine the signal strength μ_s and its uncertainty as well as a 95% credibility level (CL). The following chapter will discuss the theory and implementation of these ideas in the context of the systematics presented in Chapter 10 using the Bayesian Analysis Toolkit (BAT) [91].

11.1 Bayes' Theorem and Limit Setting

In the Bayesian approach, the data are considered to be fixed, and a probability is assigned to the parameters representing the predictions of the theory. Thus, given the data, a probability is assigned to the number of signal events which could be present in the data. In other words, Bayes' theorem enables the calculation of the conditional probability of event A, theory predictions, occurring given that event B, the data, has occurred. This is given by

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{11.1}$$

where $P(B|A)$ is the conditional probability of event B occurring, given that event A has occurred, $P(A)$ is the probability of event A occurring, and $P(B)$ is the probability of event B occurring. For events where the outcome is continuous rather than discrete, the probabilities (P) are replaced by probability density functions (p or p.d.f).

In this analysis the parameters of the theory are the normalization of the signal template ν, which is equivalent to the number of signal events, and the nuisance parameters $\boldsymbol{\theta}$, which are the systematics discussed in Chapter 10. Utilizing the continuous form of Bayes' theorem given in Equation (11.1), the p.d.f can be calculated for the theory parameters $(\nu, \boldsymbol{\theta})$, given the data. This particular p.d.f is referred to as the posterior, as it reflects the knowledge about the theory parameters after the data has been analyzed;

its formula is given in Equation (11.2).

$$p(\nu, \boldsymbol{\theta}|\text{data}) = \frac{\mathcal{L}(\nu, \boldsymbol{\theta}|\text{data})\pi(\nu, \boldsymbol{\theta})}{p(\text{data})} \tag{11.2}$$

Here the likelihood of the theory parameters, given the data, is denoted $\mathcal{L}(\nu, \boldsymbol{\theta}|\text{data})$ and is equivalent to $p(\text{data}|\nu, \boldsymbol{\theta})$. The exact form of the likelihood is a choice of the analyzer and is discussed further in Section 11.3. The information known about the theory parameters before the data is analyzed is contained in the prior probability density $\pi(\nu, \boldsymbol{\theta})$. This can be re-written as $\pi(\nu, \boldsymbol{\theta}) = \pi(\nu) \prod_i \pi(\theta_i)$ since the parameter of interest ν and each of the nuisance parameters θ_i are independent from each other. The form of the priors for the signal and each systematic is chosen by the analyzer and is discussed in Section 11.2. Note that the denominator of Equation (11.2), $p(\text{data})$, is simply an overall normalization factor and therefore can be dropped during the limit setting procedure.

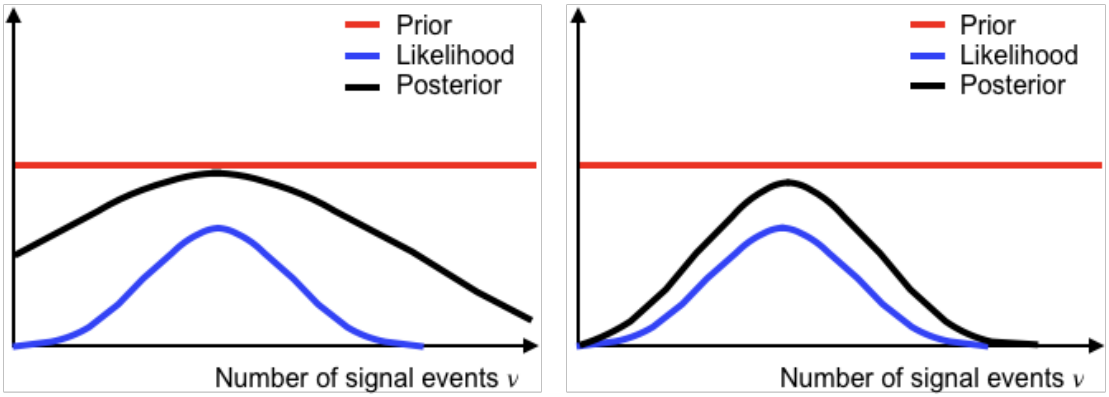
The ultimate goal is to obtain the p.d.f for a single parameter, the number of signal events ν, given the data. This is known as the marginalized posterior $p(\nu|\text{data})$. This is done by integrating out the dependence of Equation (11.2) on $\boldsymbol{\theta}$, as shown in Equation (11.3).

$$p(\nu|\text{data}) = \int p(\nu, \boldsymbol{\theta}|\text{data}) \text{d}\boldsymbol{\theta} \tag{11.3}$$

This multi-dimensional integral over the nuisance parameters $\boldsymbol{\theta}$ is known as marginalization. Note that the space of nuisance parameters may be very large, making this integral extremely computationally expensive to calculate. To expedite this process a Markov Chain Monte Carlo procedure is implemented as discussed in Section 11.3. By substituting the definition of the posterior (Equation (11.2)) into the marginalized posterior (Equation (11.3)), utilizing the expanded prior probability density $\pi(\nu, \boldsymbol{\theta}) = \pi(\nu) \prod_i \pi(\theta_i)$ and dropping the $p(\text{data})$ normalization factor gives the following definition of Bayes' equation:

$$p(\nu|\text{data}) \propto \int \mathcal{L}(\nu, \boldsymbol{\theta}|\text{data})\pi(\nu) \prod_i \pi(\theta_i) \text{d}\boldsymbol{\theta} \tag{11.4}$$

This form of Bayes' equation can be interpreted as follows: the prior knowledge of the analyzer encoded into $\pi(\nu)$ and $\pi(\theta_i)$ is updated by the outcome of the experiment contained in the likelihood $\mathcal{L}(\nu, \boldsymbol{\theta}|\text{data})$ by marginalizing the posterior with respect to $\boldsymbol{\theta}$ in order to obtain the p.d.f of the parameter of interest given the data, $p(\nu|\text{data})$. This is shown diagrammatically for the simplified case (neglecting nuisance parameters) for a small data sample in Figure 11.1a and a large data sample in Figure 11.1b, both with a flat signal prior. The flat prior in these figures represents an uninformative prior that reflects ignorance about the value of the parameter, so that it has minimal influence on the obtained posterior. This choice is consistent with the prior used in a similar dijet analysis [98] and simplifies the comparison of results.



(a) (b)

Figure 11.1: Cartoon representing the relationship between the signal normalization prior $\pi(\nu)$ (red), likelihood $\mathcal{L}(\nu|\text{data})$ (blue), and posterior $p(\nu|\text{data})$ (black) for the cases of (a) a small data sample and (b) a larger data sample. This simplified example contains no nuisance parameters.

By integrating the marginalized posterior $p(\nu|\text{data})$ across a range of ν values, the probability that the true value of ν lies in the chosen region can be determined. A common choice in particle physics is to define the upper limit ν_{upper} as the value below which 95% of the marginalized posterior lies, referred to as the 95% credibility level (CL) as shown in Equation (11.5) below.

$$0.95 = \int_{-\infty}^{\nu_{\text{upper}}} p(\nu|\text{data})\text{d}\nu \tag{11.5}$$

Here p(ν|data) is given by Equation (11.4). This means there is a 95% probability that the number of signal events in the data is less than or equal to ν_{upper}. This definition is used in the final fit to define the 95% CL.

11.2 Implementation of Priors

As discussed in Section 11.1 the priors representing the nuisance parameters and the signal normalization are chosen to represent the analyzer's knowledge before the data is considered. Generally speaking, the shape of the prior distribution for systematic uncertainties is chosen to give decreased probability as the fit tries to pull the value away from their nominal value, while for the signal parameter the choice is made to have as little influence on the marginalized posterior as possible.

For the $t\bar{t}$ signal strength a Gaussian prior is used with the mean and width determined by the fit to data in the $\mathrm{CR}_{t\bar{t}}$. The Gaussian prior is then normalized to unity and defined over the range $[-\nu_{5\cdot t\bar{t}}, \nu_{5\cdot t\bar{t}}]$, where $\nu_{5\cdot t\bar{t}}$ is five times the expected value for $t\bar{t}$, and is set to zero elsewhere.

The remaining nuisance parameters are included using a Gaussian prior for each source. The Gaussian for the QCD fit function choice uncertainty is defined in the range $[0\sigma, 1\sigma]$, where 0σ corresponds to the nominal fit function and 1σ corresponds to the alternate fit function. All other sources of systematic uncertainty are defined using a Gaussian over the large range $[-3\sigma, 3\sigma]$ to allow ample room for fluctuations.

The two signal models, V + jets and Higgs + jets, are each included in the combined fit by utilizing a uniform prior to represent the parameter corresponding to the number of events. This is done to remove analyzer bias, thus allowing the final result to more accurately reflect the data. Furthermore, both priors are normalized to unity over their

respective ranges for reasons discussed in Section 11.3. The V + jets uniform prior is defined to be $1/(2\cdot\nu_{5\cdot V})$ over the range $[-\nu_{5\cdot V}, \nu_{5\cdot V}]$ and zero everywhere else. Here $\nu_{5\cdot V}$ is defined to be five times the expected value for V + jets. The Higgs + jets uniform prior is defined to be $1/\nu_{H\text{max}}$ over the range $[0, \nu_{H\text{max}}]$ and zero elsewhere. This $\nu_{H\text{max}}$ is the number of Higgs + jets events corresponding to the point where the likelihood is a factor of 10^5 times smaller than its maximum value. In both cases the ranges were chosen to be very large so as to not influence the result.

11.3 Bayesian Analysis Toolkit

The Bayesian Analysis Toolkit (BAT) [91, 98] is used to obtain the marginalized posterior distribution in Equation (11.4) discussed in Section 11.1. It takes as input the data, the parameters ν and $\boldsymbol{\theta}$ along with their corresponding prior distributions discussed in Section 11.2, and the chosen likelihood function $\mathcal{L}(\nu, \boldsymbol{\theta}|\text{data})$. Here the likelihood function is given by the product of the Poisson probability in each bin:

$$\mathcal{L}(\nu, \boldsymbol{\theta}|\text{data}) = \prod_{i=1}^{N} \frac{(s_i(\nu, \boldsymbol{\theta}) + b_i(\nu, \boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(s_i(\nu,\boldsymbol{\theta})+b_i(\nu,\boldsymbol{\theta}))} \tag{11.6}$$

In the above the product runs over all N bins in the histogram being fit, s_i and b_i are the expected number of signal and background events expected in bin i dependent upon ν and $\boldsymbol{\theta}$, and n_i is the number of data events in bin i. Note that in Section 11.2 the signal parameter priors are normalized to unity such that ν corresponds to the number of signal

events. Now Equation (11.6) can be used to calculate the probability of a given set of parameter values ν and $\boldsymbol{\theta}$ given the data. Plugging this definition into Equation (11.4) gives the final form of Bayes' equation used to calculate the marginalized posterior $p(\nu|\text{data})$ given below.

$$p(\nu|\text{data}) \propto \pi(\nu) \int \prod_{i=1}^{N} \frac{(s_i(\nu,\boldsymbol{\theta}) + b_i(\nu,\boldsymbol{\theta}))^{n_i}}{n_i!} e^{-(s_i(\nu,\boldsymbol{\theta})+b_i(\nu,\boldsymbol{\theta}))} \prod_j \pi(\theta_j) \mathrm{d}\boldsymbol{\theta} \qquad (11.7)$$

The final step is to calculate the integral in Equation (11.7) across the multi-dimensional space of $(\nu, \boldsymbol{\theta})$. However, calculating the integral for such a large space is not computationally feasible, so BAT instead employs a Markov Chain Monte Carlo (MCMC) [99, 98] in order to sample the space efficiently.

The basic idea of a MCMC is to perform a random walk in the parameter space $(\nu, \boldsymbol{\theta})$ making sure to spend more time sampling regions of high probability, i.e sampling proportional to the posterior. The sequence of parameter values on the walk depends only on the previous set making the resulting sequence of parameter values a Markov Chain [100]. The Metropolis-Hastings algorithm [101, 98] is used to generate the Markov Chains used in BAT. This procedure is detailed below and an illustration of the process for only two parameters θ_1 and θ_2 is shown in Figure 11.2.

1. The chain begins at position $\boldsymbol{x_1}$ in the parameter space to be sampled.

2. The next position, $\boldsymbol{x_2}$, is proposed by selecting each new parameter from a Breit-Wigner distribution centered on the value of the corresponding parameter for the

current position $\boldsymbol{x_1}$.

3. A random number r between 0 and 1 is selected from a uniform distribution.

4. The value of the posterior $p(\nu, \boldsymbol{\theta}|\text{data})$, given in Equation (11.2), is calculated for both $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$ resulting in $p(\nu, \boldsymbol{\theta}|\text{data})_1$ and $p(\nu, \boldsymbol{\theta}|\text{data})_2$.

5. If $r < \frac{p(\nu,\boldsymbol{\theta}|\text{data})_2}{p(\nu,\boldsymbol{\theta}|\text{data})_1}$, the algorithm transitions to the new position $\boldsymbol{x_2}$ and it is added to the chain. Otherwise the algorithm remains at $\boldsymbol{x_1}$ and is added to the chain.

6. This process is then repeated starting from the chosen position defined as position $\boldsymbol{x_1}$.

This random walk in parameter space generates a chain which preferentially transitions to positions corresponding to high probability regions of the posterior, effectively sampling the important regions of the posterior distribution. By plotting the frequency of occurrence for each parameter along the chain, and then normalizing the distribution to unity, the desired marginalized posterior for the signal parameter $p(\nu|\text{data})$ is found. The mode of this marginalized posterior, known as the maximum a posteriori probability (MAP), is used as the measured number of signal events (ν) and the standard deviation (σ) is used as the uncertainty on the measurement. Dividing $\nu \pm \sigma$ by $(\mathcal{L}_D \times \sigma_{\text{H}} \times \Gamma_{H\to b\bar{b}} \times \text{A}_s)$ [1] the signal strength $\mu_s \pm \sigma_s$ is found.

[1] $\mathcal{L}_D$ is the luminosity of the dataset. σ_{H} is the cross section for the boosted Higgs boson production production mechanism. $\Gamma_{H\to b\bar{b}}$ is the branching fraction of the Higgs boson to $b\bar{b}$. A_s is the signal acceptance of event selection.

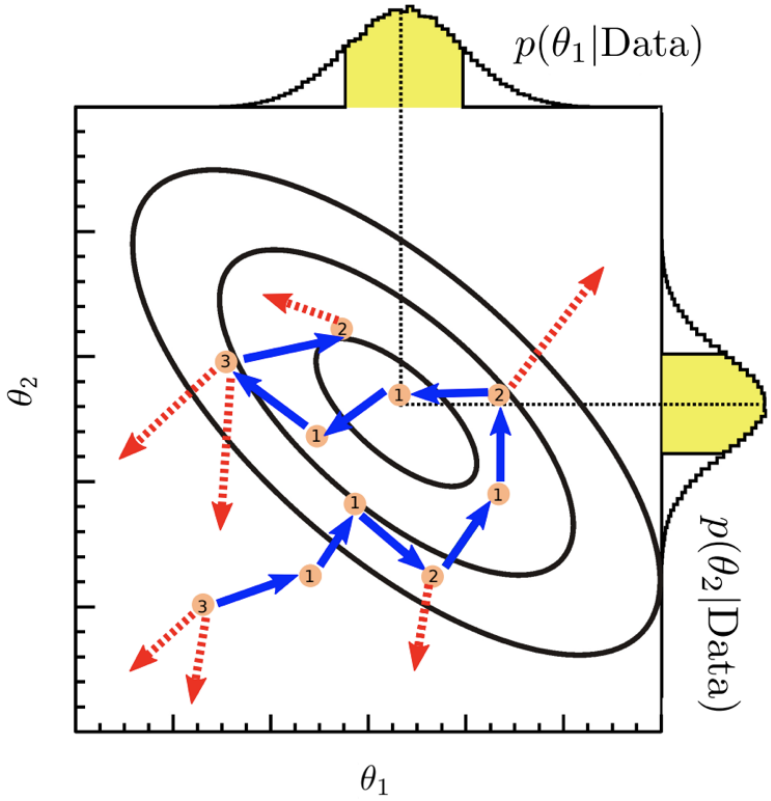


Figure 11.2: This figure illustrates a random walk in parameter space (θ_1, θ_2). The numbers indicate the number of iterations the chain remained at this point in parameter space, the blue arrows indicate accepted transitions, and the red arrows indicate rejected transitions. The marginalized posterior distributions obtained for the two parameters $p(\theta_1|\text{data})$ and $p(\theta_2|\text{data})$ are also shown, and the yellow bands correspond to the central 68% of the distributions [98].

Chapter 12

Results

This chapter summarizes the results and interpretation of the boosted $H \to b\bar{b}$ search described in this dissertation and published in Reference [88]. To arrive at this point; the theory was presented (Chapters 2 and 3), the data were collected (Chapters 4 and 5), a simulation of the data from theory was constructed (Chapter 6), physics objects were built in data and MC (Chapter 7), events were then selected for analysis (Chapter 8), the backgrounds were modelled (Chapter 9), sources of error were discussed (Chapter 10), and finally all of these considerations were combined in statistical fit (Chapter 11). This procedure gives an estimate of the observed signal significance and signal strength (μ_s) for the boosted $V \to b\bar{b}$ which is used as a standard candle to validate the analysis, and the boosted $H \to b\bar{b}$ which is the main signal of interest. The following sections briefly outline the extraction of the results and present the measured quantities.

12.1 Measurement Procedure

To measure the Standard Model signal of interest a model comprised of V + jets, $H \to b\bar{b}$ and $t\bar{t}$ templates along with a QCD multijet model function is fit to the data. The $t\bar{t}$ template is constructed from a MC sample with its normalization corrected with a k-factor derived from a dedicated $t\bar{t}$ Control Region. This fit simultaneously extracts the signal strengths of the V + jets and $H \to b\bar{b}$ process (μ_V and μ_H respectively) which are parameterized with a flat prior. The comparison of data to the maximum a posteriori probability (post-fit) model is seen in Figure 12.1 where the nuisance parameters were constrained as shown in Figure 12.2.

12.2 Observation of Boosted $V \to b\bar{b}$

The observed signal strength for the V + jets process is

$$\mu_V = 1.5 \pm 0.22\ (\text{stat.})^{+0.29}_{-0.25}\ (\text{syst.}) \pm 0.18\ (\text{theo.})\,,$$

corresponding to an observed significance of 5σ with an expected significance of 4.8σ [102]. This measurement represents the first direct observation of boosted vector bosons decaying to bottom quark pairs in ATLAS for $\sqrt{s} = 13$ TeV.

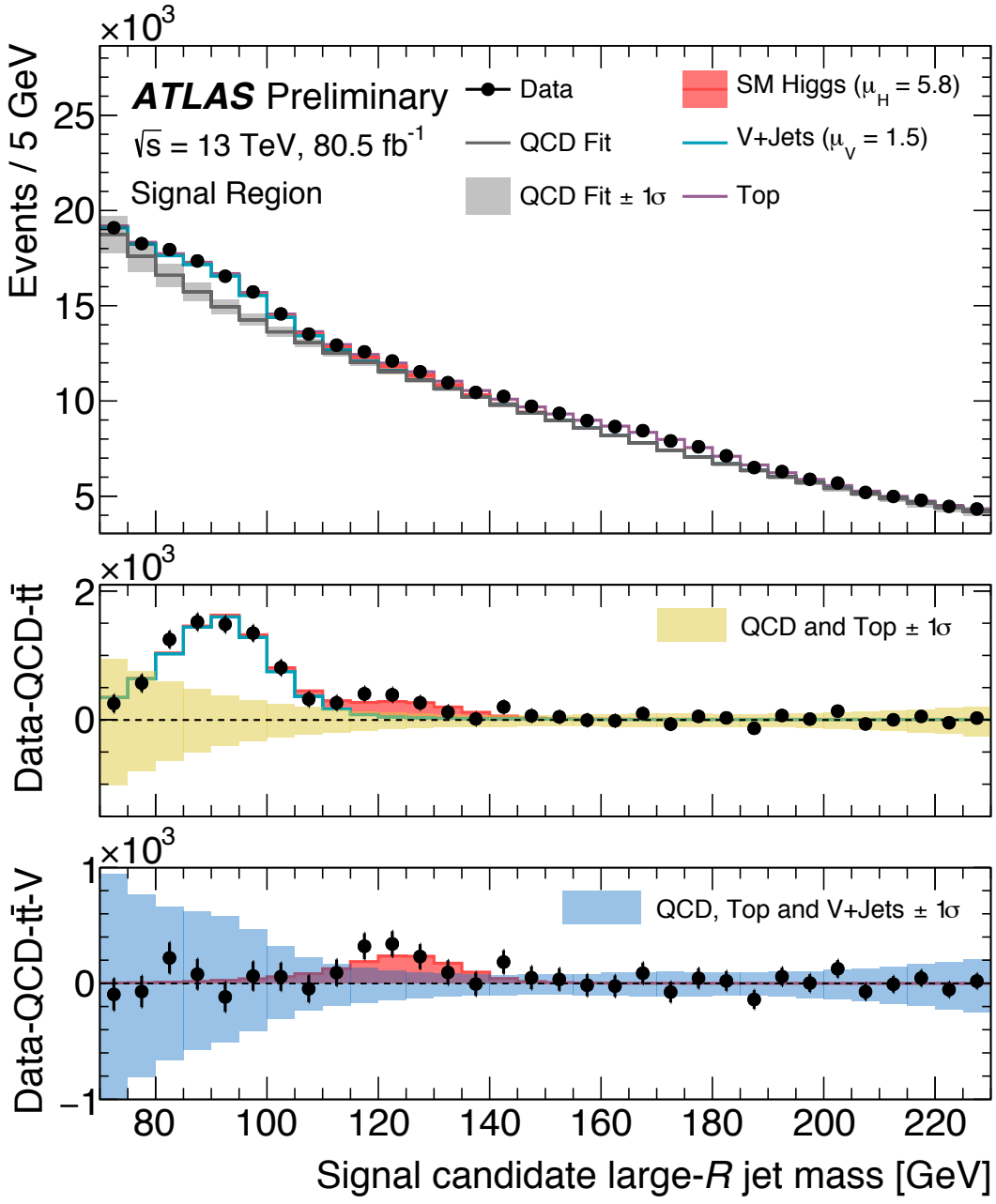


Figure 12.1: The top panel shows the post-fit comparison of the signal candidate large-R jet mass distribution for the combined SM Higgs boson, V + jets, $t\bar{t}$ and QCD model to the observed data [88]. The middle panel gives the ratio of the post-fit model and the data with the QCD and $t\bar{t}$ components subtracted, highlighting the large resonance from V + jets. The bottom panel gives the ratio of the post-fit model and the data with the QCD, V + jets, and $t\bar{t}$ components subtracted, highlighting a slight excess of events near $m_J = 125$ GeV.

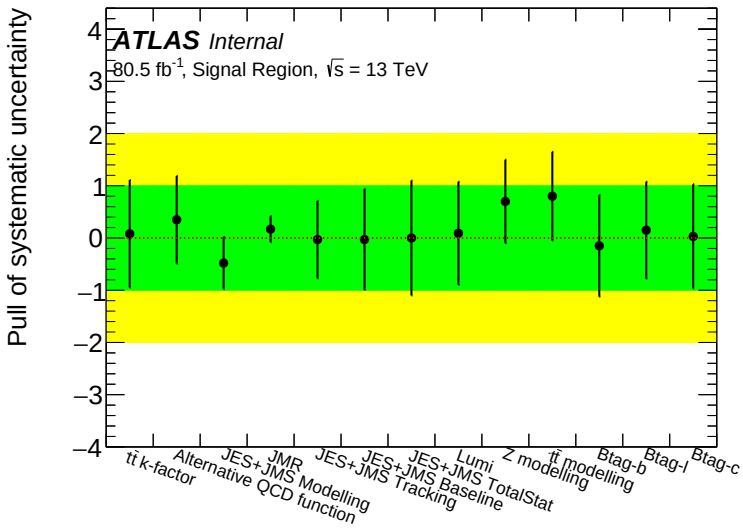


Figure 12.2: Pulls of nuisance parameters in the final combined fit [63].

12.3 Measurement of Boosted $H \to b\bar{b}$

For the $H \to b\bar{b}$ process, the observed signal strength is:

$$\mu_H = 5.8 \pm 3.1 \text{ (stat.)} \pm 1.9 \text{ (syst.)} \pm 1.7 \text{ (theo.)}\,.$$

Using the generator level cross sections ($\sigma_{\text{generated}}$) from Table 6.1 and the efficiency (ϵ) times acceptance (A) from Table 8.2, a restricted cross section ($\sigma_{\text{restricted}}$) for the phase space represented by the presented analysis selections can be calculated for $H \to b\bar{b}$ inclusive of ggF, VBF and VH production mechanisms. Using the values from Table 12.1 and multiplying by the observed μ_H gives the observed Higgs boson cross section:

$$\sigma_H = 15 \pm 8.2 \text{ (stat.)} \pm 5.0 \text{ (syst.)} \pm 4.5 \text{ (theo.) fb}\,.$$

	$\sigma_{\text{generated}}$(fb)	N_{SR}	$\epsilon \times A$	$\sigma_{\text{restricted}}$(fb)
ggF	476.3	115	0.002997	1.428
VBF	100.24	53	0.006319	0.6334
WH	8.2914	26	0.03895	0.3230
ZH	4.5178	21	0.05774	0.2609
Total	589.3	215		2.645

Table 12.1: Summary of pertinent boosted Higgs boson cross section and selection values in simulation. N_{SR} is the number of simulated Higgs bosons that pass into the SR.

Given the uncertainties, this result is consistent with the background-only hypothesis at 1.6σ with an expected sensitivity of 0.28σ [102]. This constitutes a measurement of the boosted Higgs decaying to a bottom quark pair but not a direct observation which, would require 5σ significance. The result of the combined fit of V + jets and $H \to b\bar{b}$ agrees with the Standard Model prediction of $\mu_H = \mu_V = 1$, as seen in Figure 12.3.

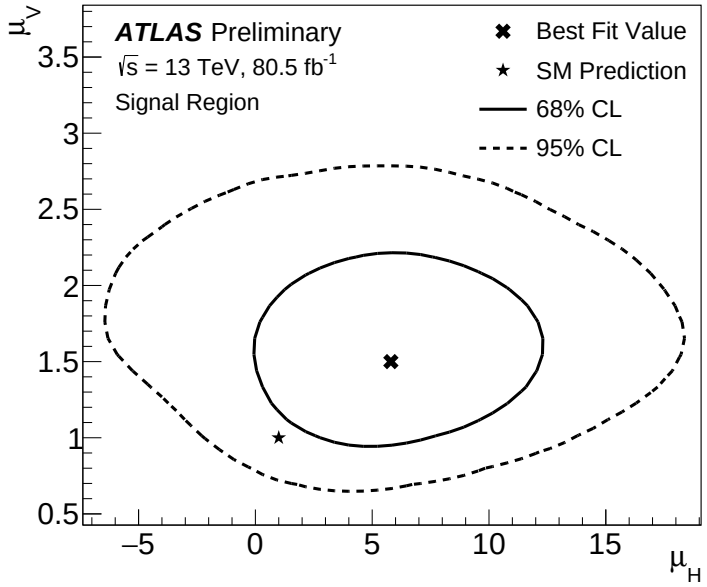


Figure 12.3: Combined marginalized posterior distributions of μ_H and μ_V in the Signal Region [88]. It is seen that the best-fit values for the signal processes lie within the 68% Credibility Level (2σ) of the Standard Model prediction.

Chapter 13

Conclusion

A search for boosted $H \to b\bar{b}$ was performed using an integrated luminosity of 80.5fb^{-1} of LHC proton-proton collision data recorded at the ATLAS experiment with a center-of-mass energy of $\sqrt{s} = 13$ TeV. The analysis of this data measured a signal strength of $\mu_H = 5.8 \pm 3.1$ (stat.) $\pm$ 1.9 (syst.) $\pm$ 1.7 (theo.), which is 1.6 standard deviations higher than the background-only hypothesis with an expected sensitivity of 0.28σ. The CMS collaboration performed a similar analysis in 2017 with an integrated luminosity of 39.5 fb^{-1} and found a signal strength for boosted $H \to b\bar{b}$ of $\mu_H = 2.3 \pm 1.5$ (stat.)$^{+1.0}_{-0.4}$ (syst.) which is consistent with this analysis' measurement within 2 standard deviations [103].

This is the first time this search has been performed in ATLAS and thus it represents important advancements in the use of boosted jet techniques and the implementation of new techniques such as the variable radius jets. Future work in this novel channel

will be necessary to optimize these new techniques and to reduce the major systematic uncertainties described in Chapter 10. The ATLAS calorimeter design gives an excellent energy resolution, making software improvements to jet technologies an important avenue of research for future versions of this analysis. Within ATLAS there is active work to include a particle flow algorithm into the large-R jet reconstruction. This algorithm combines calorimeter measurements of neutral-particle energies with information from charged hadron tracks in the ID to give better jet resolution. Furthermore, studies of new substructure-based triggers offer future improvements to the signal event selection.

In conclusion, this dissertation provides a measurement in the Higgs sector representing a new contribution to the field, serves as a validation of the predictions of the Standard Model, and presents new techniques for the analysis of boosted signatures in the ATLAS detector.

Bibliography

[1] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9 (cit. on pp. 2, 16).

[2] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508 (cit. on pp. 2, 21).

[3] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156 (cit. on pp. 2, 16).

[4] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321 (cit. on p. 2).

[5] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585 (cit. on p. 2).

[6] 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] (cit. on pp. 2, 37, 47).

[7] CMS Collaboration. "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex] (cit. on pp. 2, 37, 47).

[8] ATLAS Collaboration. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett. B* 786 (2018), p. 59. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex] (cit. on p. 2).

[9] CMS Collaboration. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex] (cit. on p. 2).

[10] Matthias Schlaffer et al. "Boosted Higgs Shapes". In: *Eur. Phys. J.* C74.10 (2014), p. 3120. DOI: 10.1140/epjc/s10052-014-3120-z. arXiv: 1405.4295 [hep-ph] (cit. on pp. 2, 33).

[11] Christophe Grojean et al. "Very boosted Higgs in gluon fusion". In: *JHEP* 05 (2014), p. 022. DOI: 10.1007/JHEP05(2014)022. arXiv: 1312.3317 [hep-ph] (cit. on pp. 2, 33).

[12] Massimiliano Grazzini et al. "Modeling BSM effects on the Higgs transverse-momentum spectrum in an EFT approach". In: *JHEP* 03 (2017), p. 115. DOI: 10.1007/JHEP03(2017)115. arXiv: 1612.00283 [hep-ph] (cit. on pp. 2, 33).

[13] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2 (cit. on pp. 4, 10).

[14] A. Salam. "Weak and Electromagnetic Interactions". In: *Elementary Particle Theory*. Ed. by N. Svartholm. Almqvist and Wiksell, 1968, p. 367 (cit. on p. 4).

[15] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264 (cit. on pp. 4, 10).

[16] John Campbell, Joey Huston, and Frank Krauss. *The Black Book of Quantum Chromodynamics*. Oxford University Press, 2017. ISBN: 9780199652747. URL: https://global.oup.com/academic/product/the-black-book-of-quantum-chromodynamics-9780199652747 (cit. on pp. 4, 13).

[17] E. Rutherford. "The scattering of alpha and beta particles by matter and the structure of the atom". In: *Phil. Mag. Ser.6* 21 (1911), pp. 669–688. DOI: 10.1080/14786440508637080 (cit. on p. 5).

[18] Arthur H. Compton. "A Quantum Theory of the Scattering of X-rays by Light Elements". In: *Phys. Rev.* 21 (5 May 1923), pp. 483–502. DOI: 10.1103/PhysRev.

21.483. URL: https://link.aps.org/doi/10.1103/PhysRev.21.483 (cit. on p. 6).

[19] Harald Fritzsch and Murray Gell-Mann. "Current algebra: Quarks and what else?" In: *eConf* C720906V2 (1972), pp. 135–165. arXiv: hep-ph/0208010 [hep-ph] (cit. on p. 6).

[20] ATLAS Collaboration. *Summary plots from the ATLAS Standard Model physics group*. https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/. [Online; accessed 24-October-2019]. 2015 (cit. on p. 7).

[21] Andrew Purcell. "Go on a particle quest at the first CERN webfest." In: BULNA-2012-269. 35/2012 (Aug. 2012), p. 10. URL: https://cds.cern.ch/record/1473657 (cit. on p. 8).

[22] Mark Thomson. *Modern particle physics*. New York: Cambridge University Press, 2013. ISBN: 9781107034266. URL: http://www-spires.fnal.gov/spires/find/books/www?cl=QC793.2.T46::2013 (cit. on pp. 9, 16).

[23] Particle Data Group. "Review of Particle Physics, Chapter 11: Status of Higgs Boson Physics". In: *Phys. Rev. D* 98 (3 2018), p. 030001. DOI: 10.1103/PhysRevD.98.030001. URL: https://link.aps.org/doi/10.1103/PhysRevD.98.030001 (cit. on pp. 23, 25, 29).

[24] D. de Florian et al. "Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector". In: (2016). DOI: 10.2172/1345634,10.23731/CYRM-2017-002. arXiv: 1610.07922 [hep-ph] (cit. on p. 25).

[25] L. A. Harland-Lang et al. “Parton distributions in the LHC era: MMHT 2014 PDFs”. In: *The European Physical Journal C* 75.5 (May 2015), p. 204. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-015-3397-6. URL: https://doi.org/10.1140/epjc/s10052-015-3397-6 (cit. on p. 27).

[26] ATLAS Collaboration. “Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV”. In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex] (cit. on pp. 30–32).

[27] S. Dawson, I. M. Lewis, and Mao Zeng. “Usefulness of effective field theory for boosted Higgs production”. In: *Phys. Rev.* D91 (2015), p. 074012. DOI: 10.1103/PhysRevD.91.074012. arXiv: 1501.04103 [hep-ph] (cit. on p. 33).

[28] ATLAS Collaboration. “Search for light resonances decaying to boosted quark pairs and produced in association with a photon or a jet in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Lett.* B788 (2019), pp. 316–335. DOI: 10.1016/j.physletb.2018.09.062. arXiv: 1801.08769 [hep-ex] (cit. on pp. 33, 72).

[29] Matthew Carl Feickert. “A Search for Boosted Low Mass Resonances Decaying to the $b\bar{b}$ Final State and Produced in Association with a Jet at $\sqrt{s} = 13$ TeV with the ATLAS Detector”. Presented 24 Jul 2019. Aug. 2019. URL: https://cds.cern.ch/record/2690521 (cit. on pp. 34, 52, 91, 99, 111).

[30] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: https://cds.cern.ch/record/782076 (cit. on p. 36).

[31] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001 (cit. on pp. 36, 43, 44).

[32] Chris Llewellyn Smith. "Genesis of the Large Hadron Collider". In: *Phil. Trans. Roy. Soc. Lond.* A373.2032 (2014), p. 20140037. DOI: 10.1098/rsta.2014.0037 (cit. on p. 36).

[33] CERN/AT/PhL. *Interim Summary Report on the Analysis of the 19 September 2008 Incident at the LHC*. EDMS 973073. CERN, 2008 (cit. on p. 37).

[34] Heather Gray and Bruno Mansoulie. *The Higgs boson: the hunt, the discovery, the study and some future perspectives*. https://atlas.cern/updates/atlas-feature/higgs-boson. [Online; accessed 26-October-2019]. 2018 (cit. on pp. 37, 47).

[35] Julie Haffner. "The CERN accelerator complex. Complexe des accélérateurs du CERN". In: (Oct. 2013). General Photo. URL: https://cds.cern.ch/record/1621894 (cit. on p. 38).

[36] R Scrivens et al. "Overview of the status and developments on primary ion sources at CERN*". In: CERN-ATS-2011-172 (Sept. 2011), 4 p. URL: https://cds.cern.ch/record/1382102 (cit. on p. 37).

[37] Ramon Cid Manzano Xabier Cid Vidal. *Proton Source: Taking a closer look at LHC*. https://www.lhc-closer.es/taking_a_closer_look_at_lhc/0.proton_source. [Online; accessed 26-October-2019] (cit. on p. 37).

[38] Giuseppe Battistoni et al. "The Application of the Monte Carlo Code FLUKA in Radiation Protection Studies for the Large Hadron Collider". In: *Prog. Nucl. Sci. Tech.* 2 (2011), pp. 358–364. DOI: 10.15669/pnst.2.358 (cit. on p. 40).

[39] Serge Dailler. "Cross section of LHC dipole.. Dipole LHC: coupe transversale." AC Collection. Legacy of AC. Pictures from 1992 to 2002. Apr. 1999. URL: https://cds.cern.ch/record/842530 (cit. on p. 41).

[40] J. Barranco García et al. "Long term dynamics of the high luminosity Large Hadron Collider with crab cavities". In: *Phys. Rev. Accel. Beams* 19 (10 Oct. 2016), p. 101003. DOI: 10.1103/PhysRevAccelBeams.19.101003. URL: https://link.aps.org/doi/10.1103/PhysRevAccelBeams.19.101003 (cit. on p. 44).

[41] *ATLAS TWiki: Run 2 Luminosity Public Results, § Luminosity summary plots for 2018 pp data taking: Peak Luminosity per Fill in 2018*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2#Luminosity_summary_plots_for_201. Accessed: 2019-10-22 (cit. on p. 44).

[42] ATLAS Collaboration. *ATLAS Luminosity Public Results for Run 2*. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2. [Online; accessed 26-October-2019]. 2019 (cit. on pp. 44, 45).

[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 (cit. on pp. 48, 50, 55–64).

[44] *ATLAS: technical proposal for a general-purpose pp experiment at the Large Hadron Collider at CERN*. LHC Tech. Proposal. Geneva: CERN, 1994. URL: https://cds.cern.ch/record/290968 (cit. on p. 47).

[45] ATLAS Collaboration. *First beam and first events in ATLAS*. http://atlas.cern/updates/atlas-news/first-beam-and-first-events-atlas. [Online; accessed 27-October-2019]. 2008 (cit. on p. 47).

[46] Gabriella Pasztor. "The Upgrade of the ATLAS Electron and Photon Triggers towards LHC Run 2 and their Performance". In: (Sept. 2015). URL: https://cds.cern.ch/record/2048697 (cit. on p. 51).

[47] Giordon Holtsberg Stark. "The search for supersymmetry in hadronic final states using boosted object reconstruction". Presented 26 Apr 2018. May 2018. URL: https://cds.cern.ch/record/2317296 (cit. on p. 53).

[48] *ATLAS inner detector: Technical Design Report, 1*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: http://cds.cern.ch/record/331063 (cit. on p. 54).

[49] S Haywood et al. *ATLAS inner detector: Technical Design Report, 2*. Technical Design Report ATLAS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/331064 (cit. on p. 54).

[50] Karolos Potamianos. *The upgraded Pixel detector and the commissioning of the Inner Detector tracking of the ATLAS experiment for Run-2 at the Large Hadron Collider*. Tech. rep. ATL-PHYS-PROC-2016-104. 15 pages, EPS-HEP 2015 Proceedings. Geneva: CERN, Aug. 2016. URL: https://cds.cern.ch/record/2209070 (cit. on p. 54).

[51] B. Abbott et al. "Production and integration of the ATLAS Insertable B-Layer". In: *JINST* 13 (2018), T05008. DOI: 10.1088/1748-0221/13/05/T05008. arXiv: 1803.00844 [physics.ins-det] (cit. on p. 55).

[52] ATLAS Collaboration. "Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties". In: *JINST* 12.05 (2017), P05002. DOI: 10.1088/1748-0221/12/05/P05002. arXiv: 1702.06473 [hep-ex] (cit. on p. 57).

[53] Christian Wolfgang Fabjan and F Gianotti. "Calorimetry for Particle Physics". In: *Rev. Mod. Phys.* 75.CERN-EP-2003-075 (Oct. 2003), 1243–1286. 96 p. DOI: 10.1103/RevModPhys.75.1243. URL: https://cds.cern.ch/record/692252 (cit. on p. 58).

[54] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva: CERN, 1996. URL: https://cds.cern.ch/record/331061 (cit. on p. 62).

[55] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator". In: *Journal of High*

Energy Physics 2015.5 (May 2015), p. 140. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2015)140. URL: https://doi.org/10.1007/JHEP05(2015)140 (cit. on p. 70).

[56] John M. Campbell et al. "NLO Higgs boson production plus one and two jets using the POWHEG BOX, MadGraph4 and MCFM". In: *Journal of High Energy Physics* 2012.7 (July 2012), p. 92. ISSN: 1029-8479. DOI: 10.1007/JHEP07(2012)092. URL: https://doi.org/10.1007/JHEP07(2012)092 (cit. on pp. 70, 74).

[57] Keith Hamilton et al. "Merging H/W/Z + 0 and 1 jet at NLO with no merging scale: a path to parton shower + NNLO matching". In: *JHEP* 05 (2013), p. 082. DOI: 10.1007/JHEP05(2013)082. arXiv: 1212.4504 [hep-ph] (cit. on p. 70).

[58] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph] (cit. on pp. 70, 74).

[59] J. Pumplin et al. "New generation of parton distributions with uncertainties from global QCD analysis". In: *JHEP* 07 (2002), p. 012. arXiv: hep-ph/0201195 (cit. on pp. 70, 74).

[60] David J. Lange. "The EvtGen particle decay simulation package". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 462.1 (2001). BEAUTY2000, Proceedings of the 7th Int. Conf. on B-Physics at Hadron Machines, pp. 152–155. ISSN: 0168-9002. DOI: https://doi.org/10.1016/S0168-9002(01)00089-4. URL:

http://www.sciencedirect.com/science/article/pii/S0168900201000894 (cit. on pp. 70, 72, 74).

[61] Paolo Nason and Carlo Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". In: *JHEP* 02 (2010), p. 037. DOI: 10.1007/JHEP02(2010)037. arXiv: 0911.5299 [hep-ph] (cit. on p. 70).

[62] Bruce Mellado Garcia et al. "CERN Report 4: Part I Standard Model Predictions". In: (May 2016). URL: https://cds.cern.ch/record/2150771 (cit. on p. 70).

[63] ATLAS Collaboration. *Search for boosted dijet resonances decaying to two b-quarks and produced in association with a jet*. Tech. rep. ATL-COM-PHYS-2018-300. Geneva: CERN, Mar. 2018. URL: https://cds.cern.ch/record/2310645 (cit. on pp. 71, 73, 75–77, 98, 101, 103, 108, 114, 134).

[64] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph] (cit. on p. 72).

[65] Stefano Carrazza, Stefano Forte, and Juan Rojo. "Parton Distributions and Event Generators". In: *Proceedings, 43rd International Symposium on Multiparticle Dynamics (ISMD 13)*. 2013, pp. 89–96. arXiv: 1311.5887 [hep-ph] (cit. on pp. 72, 74).

[66] T. Gleisberg et al. "Event generation with SHERPA 1.1". In: *JHEP* 02 (2009), p. 007. DOI: 10.1088/1126-6708/2009/02/007. arXiv: 0811.4622 [hep-ph] (cit. on pp. 72, 74).

[67] Hung-Liang Lai et al. "New parton distributions for collider physics". In: *Phys. Rev.* D82 (2010), p. 074024. DOI: 10.1103/PhysRevD.82.074024. arXiv: 1007.2241 [hep-ph] (cit. on pp. 72, 74).

[68] M. Bahr et al. "Herwig++ Physics and Manual". In: *Eur. Phys. J. C* 58 (2008), p. 639. DOI: 10.1140/epjc/s10052-008-0798-9. arXiv: 0803.0883 [hep-ph] (cit. on p. 74).

[69] Andy Buckley and Holger Schulz. "Tuning of MC generator MPI models". In: *Adv. Ser. Direct. High Energy Phys.* 29 (2018), pp. 281–301. DOI: 10.1142/9789813227767_0013. arXiv: 1806.11182 [hep-ph] (cit. on p. 74).

[70] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph] (cit. on p. 74).

[71] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph] (cit. on pp. 79, 81, 82).

[72] S. Catani et al. "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions". In: *Nucl. Phys.* B406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M (cit. on p. 81).

[73] Yuri L. Dokshitzer et al. "Better jet clustering algorithms". In: *JHEP* 08 (1997), p. 001. DOI: 10.1088/1126-6708/1997/08/001. arXiv: hep-ph/9707323 [hep-ph] (cit. on p. 81).

[74] ATLAS Collaboration. "In situ calibration of large-radius jet energy and mass in 13 TeV proton–proton collisions with the ATLAS detector". In: *Eur. Phys. J.* C79.2 (2019), p. 135. DOI: 10.1140/epjc/s10052-019-6632-8. arXiv: 1807.09477 [hep-ex] (cit. on pp. 82, 116).

[75] ATLAS Collaboration. "Search for new phenomena with large jet multiplicities and missing transverse momentum using large-radius jets and flavour-tagging at ATLAS in 13 TeV *pp* collisions". In: *JHEP* 12 (2017), p. 034. DOI: 10.1007/JHEP12(2017)034. arXiv: 1708.02794 [hep-ex] (cit. on p. 82).

[76] 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] (cit. on p. 82).

[77] ATLAS Collaboration. *Performance of shower deconstruction in ATLAS*. ATLAS-CONF-2014-003. 2014. URL: https://cds.cern.ch/record/1648661 (cit. on p. 83).

[78] 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] (cit. on p. 82).

[79] ATLAS Collaboration. *Variable Radius, Exclusive-k_T, and Center-of-Mass Subjet Reconstruction for Higgs($\to b\bar{b}$) Tagging in ATLAS*. ATL-PHYS-PUB-2017-010. 2017. URL: https://cds.cern.ch/record/2268678 (cit. on pp. 82, 84–87).

[80] ATLAS Collaboration. "Performance of b-Jet Identification in the ATLAS Experiment". In: *JINST* 11.04 (2016), P04008. DOI: 10.1088/1748-0221/11/04/P04008. arXiv: 1512.01094 [hep-ex] (cit. on p. 84).

[81] ATLAS Collaboration. *Expected performance of the ATLAS b-tagging algorithms in Run-2*. ATL-PHYS-PUB-2015-022. 2015. URL: https://cds.cern.ch/record/2037697 (cit. on p. 87).

[82] Andy Chisholm. *Introduction to Heavy Flavour Jet Tagging with ATLAS*. ATLAS Higgs to $b\bar{b}$ / Flavor Tagging Workshop. 2017. URL: https://indico.cern.ch/event/655628/contributions/2670400/ (cit. on pp. 88–91).

[83] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017. URL: https://cds.cern.ch/record/2273281 (cit. on p. 89).

[84] ATLAS Collaboration. *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*. ATL-PHYS-PUB-2017-003. 2017. URL: https://cds.cern.ch/record/2255226 (cit. on p. 90).

[85] Andrea Sciandra. *Development of a new Soft Muon Tagger for the identification of b-jets in ATLAS*. Tech. rep. ATL-PHYS-PROC-2017-190. Geneva: CERN,

Oct. 2017. DOI: 10.22323/1.314.0768. URL: https://cds.cern.ch/record/2287545 (cit. on p. 90).

[86] ATLAS Collaboration. "Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV". In: *JHEP* 08 (2018), p. 089. DOI: 10.1007/JHEP08(2018)089. arXiv: 1805.01845 [hep-ex] (cit. on pp. 90, 92, 116).

[87] ATLAS Collaboration. "Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s}$ =13 TeV". In: *Eur. Phys. J.* C76.5 (2016), p. 292. DOI: 10.1140/epjc/s10052-016-4120-y. arXiv: 1603.05598 [hep-ex] (cit. on pp. 92, 94, 95).

[88] ATLAS Collaboration. *Search for boosted resonances decaying to two b-quarks and produced in association with a jet at $\sqrt{s} = 13$ TeV with the ATLAS detector.* ATLAS-CONF-2018-052. 2018. URL: https://cds.cern.ch/record/2649081 (cit. on pp. 100, 120, 131, 133, 135).

[89] ATLAS Collaboration. *Improvements in $t\bar{t}$ modelling using NLO+PS Monte Carlo generators for Run 2.* ATL-PHYS-PUB-2018-009. 2018. URL: https://cds.cern.ch/record/2630327 (cit. on p. 106).

[90] John Erthal Gaiser. "Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'". PhD thesis. SLAC, 1982. URL: http://www-public.slac.stanford.edu/sciDoc/docMeta.aspx?slacPubNumber=slac-r-255.html (cit. on p. 106).

[91] Frederik Beaujean et al. “BAT: The Bayesian Analysis Toolkit”. In: *J. Phys. Conf. Ser.* 331 (2011), p. 072040. DOI: 10.1088/1742-6596/331/7/072040 (cit. on pp. 109, 121, 127).

[92] The ATLAS collaboration. “Measurements of $t\bar{t}$ differential cross-sections in the all-hadronic channel with the ATLAS detector using highly boosted top quarks in pp collisions at $\sqrt{s} = 13$ TeV”. In: (2016) (cit. on p. 109).

[93] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *Ann. Math. Statist.* 9.1 (Mar. 1938), pp. 60–62. DOI: 10.1214/aoms/1177732360. URL: https://doi.org/10.1214/aoms/1177732360 (cit. on p. 113).

[94] G.W. Snedecor and W.G. Cochran. *Statistical Methods*. Statistical Methods v. 276. Wiley, 1991. ISBN: 9780813815619 (cit. on p. 113).

[95] ATLAS Collaboration. “Improved luminosity determination in pp collisions at sqrt(s) = 7 TeV using the ATLAS detector at the LHC”. In: *Eur. Phys. J.* C73.8 (2013), p. 2518. DOI: 10.1140/epjc/s10052-013-2518-3. arXiv: 1302.4393 [hep-ex] (cit. on p. 117).

[96] J. M. Lindert et al. “Precise predictions for $V+$ jets dark matter backgrounds”. In: *Eur. Phys. J.* C77.12 (2017), p. 829. DOI: 10.1140/epjc/s10052-017-5389-1. arXiv: 1705.04664 [hep-ph] (cit. on p. 118).

[97] S. P. Jones, M. Kerner, and G. Luisoni. “Next-to-Leading-Order QCD Corrections to Higgs Boson Plus Jet Production with Full Top-Quark Mass Depen-

dence". In: *Phys. Rev. Lett.* 120.16 (2018), p. 162001. DOI: 10.1103/PhysRevLett.120.162001. arXiv: 1802.00349 [hep-ph] (cit. on p. 118).

[98] Lydia Audrey Beresford. "Searches for Dijet Resonances using $\sqrt{s} = 13$ TeV proton–proton collision data recorded by the ATLAS detector at the Large Hadron Collider. Dijet Resonance Searches with the ATLAS detector at 13 TeV and jet punch-through corrections". 2017. DOI: 10.1007/978-3-319-97520-7. URL: http://cds.cern.ch/record/2642397 (cit. on pp. 124, 127, 128, 130).

[99] Michael Betancourt. "A Conceptual Introduction to Hamiltonian Monte Carlo". In: *arXiv e-prints*, arXiv:1701.02434 (Jan. 2017), arXiv:1701.02434. arXiv: 1701.02434 [stat.ME] (cit. on p. 128).

[100] A.A. Markov. "Extension of the law of large numbers to quantities, depending on each other (1906). Reprint." eng. In: *Journal Électronique d'Histoire des Probabilités et de la Statistique [electronic only]* 2.1b (2006), Article 10, 12 p., electronic only–Article 10, 12 p., electronic only. URL: http://eudml.org/doc/128778 (cit. on p. 128).

[101] W. K. Hastings. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1 (1970), pp. 97–109. ISSN: 00063444. URL: http://www.jstor.org/stable/2334940 (cit. on p. 128).

[102] Matthew Feickert. *Searches for boosted low mass resonances decaying to b quarks with the ATLAS detector*. https://indico.cern.ch/event/732102/contributions/

3188460/. Higgs Couplings Conference in Tokyo, Japan. 2018 (cit. on pp. 132, 135).

[103] Albert M Sirunyan et al. "Inclusive search for a highly boosted Higgs boson decaying to a bottom quark-antiquark pair". In: *Phys. Rev. Lett.* 120.7 (2018), p. 071802. DOI: 10.1103/PhysRevLett.120.071802. arXiv: 1709.05543 [hep-ex] (cit. on p. 136).