

**INAUGURAL ATLAS SEARCHES FOR RESONANT DI-HIGGS AND SH
SIGNALS IN THE BOOSTED, FULLY-HADRONIC $b\bar{b}VV^*$ FINAL STATE
AT ATLAS USING $\sqrt{S} = 13$ TEV DATA AND NOVEL MACHINE
LEARNING TECHNIQUES.**

Blake Christopher Forland

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements

for the degree

Doctor of Philosophy

in the Department of Physics,

Indiana University

December 2021



Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the
requirements for the degree of Doctor of Philosophy.

Doctoral Committee

Rick Van Kooten, Ph.D. (Chair)

Harold Evans, Ph.D.

Radovan Dermisiek, Ph.D.

Rex Tayloe, Ph.D.

Date of Defense: 12/03/2021

Copyright © 2021
Blake Christopher Forland

Dedicated to my nephews

Oliver & Oslo

along with any subsequent higher-order Forlands.

May you be the best of us.

ACKNOWLEDGEMENTS

My graduate experience has been an unconventional one and I would not have made it through were it not for the support and confidence given to me by an ever increasing number of remarkable people.

First of all I'd like to thank my advisors and mentors who were intimately involved with this work. Thanks you to Rick for your support, particularly with my transition into experimental particle physics. Gabriel for his patience, input, and trust without whom I cannot imagine having been successful in this endeavor. I often wavered in my confidence with my work and time and again you were there to give me the push I needed. Thank you Hal Evans for taking it upon yourself to give me much needed guidance and stability throughout my studies. I didn't make it easy with my coming to the show late but in the end we made it. Also thank you to Jason, Kira, and the other members of the *bbVV* and di-Higgs analysis groups.

To my undergraduate advisor, mentor, and friend Jim who gave me a shot and started this long road I hope now you feel you made the right choice. Jeff, we started on Fridays with Griffiths and now look at me, thanks for for appealing to my early *randonnées* into particle physics. Thank you Sarah for all my early research successes, the sun seems to be calling me back.

There are many more people that mean so much to me and have played a role in my having gotten this far that I need to mention. Thank you to...

My basecamp at IU, Andrew and Braden, and Matthew for your support in coursework and training. Graham for being an inspiration to me and such a supportive friend. Neil for so many long, wonderful conversations. Josh for entertaining my political discussions and have the motivation to start the PGSC together. Tim and Sarah for being the most wonderful family I could have hoped for in Indiana and Cole for keeping IU interesting. James my closest friend and study partner throughout my bachelor work. My whole CERN crew Dylan, Francesco, Tom, and Valentina, *Ciao!*. And Manon, my strongest source of support while living France, thank you for all you did for me.

To my dearest friends Eric and Kelley, thank you for sticking with me all these years,

I can never express to you just how much it has meant to me.

Finally to my family, thank you for everything you've done for me over my life that has helped lead me to this goal.

And thank you to myself, for continuing to endure.

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DE-SC-0010120.

PREFACE

The Standard Model (SM) of particle physics is the current best theory to explain the known fundamental particles of nature and the ways in which they interact. The SM is the result of centuries of relentless effort to better understand and explain the natural world. Continually tested and refined, it is an intellectual marvel with unrivaled measurement capabilities. These achievements of the SM are due in part to the tireless efforts to search for new physical phenomena since, in spite of all its successes, there is still much about the physical world that remains explained.

One such effort by the ATLAS collaboration at the Large Hadron Collider is the ongoing exploration of the nature of the higgs boson. Responsible for the SM's explanation of the mass for most fundamental particles, the higgs bosons is a significant component of the theory as well as a relatively new area of experimental investigation. Within the SM, the production of two Higgs bosons is expected to occur; however, at levels of detection much lower than the current expected sensitivity. This thesis presents a Beyond the Standard Model search for a heavy resonance X producing two higgs bosons that subsequently decay into two b -quarks and two vector (V) bosons. The search for the $b\bar{b}VV$ decay channel is performed in the high energy region of the final state in which both bosons decay exclusively to quarks. This is a complex final state that has few easily identifiable signatures motivating the development and implementation of a novel machine learning algorithm in order to identify, or tag, the vector bosons. This is the first time this search has been performed and limits set by the ATLAS collaboration and is done so using full Run 2 data. The analysis used to perform the di-Higgs search is reinterpreted using a different model where one higgs boson is instead replaced by an

additional heavy resonance S .

The work in this thesis is structured in the following way: Chapter 1 presents the SM and the theory underlying the searches of this thesis. Chapter 2 summarizes the experimental setup, both the ATLAS experiment as well as the method for simulating its data. Chapter 3 explains the definitions and methods for reconstructing and calibrating the physics objects used in ATLAS analyses. In Chapter 4 relevant machine learning concepts are discussed followed by a detailed explanation of the development, design, and performance of the tagger. The calibration and uncertainty estimation strategy for the tagger follow in Chapter 5. Finally the resonant di-Higgs and SH searches and results are presented in Chapters 6 and 7 respectively, with concluding remarks in Chapter 8.

Blake Christopher Forland

**INAUGURAL ATLAS SEARCHES FOR RESONANT DI-HIGGS AND SH SIGNALS IN
THE BOOSTED, FULLY-HADRONIC $b\bar{b}VV^*$ FINAL STATE AT ATLAS USING $\sqrt{S} = 13$
TEV DATA AND NOVEL MACHINE LEARNING TECHNIQUES.**

In this work two distinct yet intimately related efforts are presented. First, the development of a jet tagger consisting of a hybrid, parameterized convolutional neural network optimized for discriminating boosted, four prong jets against a majority QCD background. Uncertainties in tagging four prong jet are estimated, calibrating the tagger for use on such objects in data for the first time in ATLAS. Second, the aforementioned tagger is used as a core component in an analysis to set 95% CL limits for resonant di-Higgs production into the fully hadronic $b\bar{b}VV$ final state. This analysis is reinterpreted under the $X \rightarrow SH$ model into the same final state.

CONTENTS

Acknowledgements	v
Preface	vi
Abstract	ix
List of Figures	xv
List of Tables	xxiii
1 Theory	1
1.1 Underlying Fundamentals	1
1.1.1 Principle of Least Action	2
1.1.2 Weyl and Fock's Insight and Noether's Theorem	4
1.2 Quantum Electrodynamics	5
1.3 Additional symmetries	6
1.4 Electroweak Sector	6
1.5 Electroweak Spontaneous Symmetry Breaking	8
1.5.1 Spontaneous Symmetry Breaking	8
1.5.2 Higgs Mechanism	9
1.6 Strong Sector	11
1.7 Fermion Mass and Yukawa Matrix	12
1.8 Summary of Standard Model Particles	12
1.9 Beyond the Standard model	14
1.9.1 Neutrinos	14
1.9.2 Matter anti-matter asymmetry	14
1.9.3 Dark matter and dark energy	14

1.9.4 Gravity	15
1.10 Higgs Physics	15
1.11 Di-Higgs Production	17
1.11.1 Non-resonant di-Higgs Production	17
1.11.2 Resonant di-Higgs and SH production	18
1.12 The $HH \rightarrow bbVV$ analysis	20
1.12.1 Interpreting New Scalars	22
1.13 Other di-Higgs channels	22
1.13.1 Future of Di-Higgs Physics	23
2 Experimental Setup	24
2.1 Overview of Collider Physics	24
2.1.1 Parton Distribution Functions	26
2.2 CERN	26
2.3 The Large Hadron Collider	27
2.4 A Toroidal LHC ApparatuS	29
2.4.1 Coordinate System	30
2.4.2 The inner detector	31
2.4.3 The Liquid Argon and Tile Calorimeters	33
2.4.4 Muon Spectrometer	36
2.4.5 Magnet System	36
2.4.6 The ATLAS Forward Detector project	38
2.5 Trigger	38
2.6 Performance	39
2.7 Future of ATLAS	40
2.8 Simulating Collision Events	41
2.8.1 Event Generation	42
2.8.2 Detector Simulation	42
2.8.3 Digitization	43
2.8.4 Reconstruction	43
3 Object Reconstruction	44
3.1 Tracks	44
3.2 Topological Calorimeter Clusters	47
3.3 Jets	48

3.3.1	Reconstructing Jets	49
3.3.2	Trimming	51
3.3.3	Jet Calibration	52
3.3.4	ATLAS Jet Definitions	55
3.3.5	Track Assisted Reclustered Jets	57
3.3.6	b Jets	59
3.4	Muons	60
3.5	Electrons and Photons	61
3.6	Taus	61
3.7	Missing Transverse Momentum	62
4	Identifying Boosted Four-Prong Objects	63
4.1	WW jet as four-prong object	63
4.1.1	Prongs	64
4.1.2	TAR jets for boosted $HH \rightarrow bbWW$	65
4.2	Machine Learning in HEP	65
4.2.1	Machine learning philosophy	66
4.3	Artificial Neural Networks	66
4.3.1	Feed Forward Neural Networks	67
4.3.2	Deep Neural Networks	69
4.3.3	Supervised Learning	69
4.3.4	Optimization	70
4.3.5	Convolution	71
4.4	A Boosted Four-Prong Tagger	72
4.4.1	Features	73
4.4.2	The Lund Jet Plane	76
4.4.3	Parameterization	80
4.4.4	Algorithm Specifics	83
4.4.5	Feature Scaling	85
4.4.6	Network Training	87
4.5	Performance	87
4.5.1	Working point definitions	89
4.6	Limitations and Extensions	90
4.6.1	Higgs mass	90

5	Four-Prong Jet Calibration	92
5.1	Calibration strategy	92
5.2	Scale Factor uncertainty	93
5.3	MC uncertainty	94
5.4	Four-Prong Tagger Efficiency Uncertainty	96
6	Search for boosted resonant di-Higgs production	98
6.1	Overview	98
6.1.1	Previous resonant di-Higgs Results	98
6.2	Data and simulation samples	99
6.2.1	Data samples	99
6.3	Jet Definitions	100
6.3.1	b -tagging	101
6.4	Event Selection	101
6.4.1	Triggers	101
6.4.2	Event criteria	102
6.4.3	Signal Selection	102
6.5	Analysis regions	104
6.5.1	Analysis for each scalar mass	105
6.5.2	Composition	106
6.6	Background estimation	106
6.6.1	The ABCD Method	112
6.6.2	Estimate uncertainty	112
6.7	Uncertainties	114
6.7.1	Experimental Uncertainties	114
6.7.2	Background estimation uncertainty	115
6.7.3	Four-Prong Tagger Uncertainty	116
6.7.4	b -tagging Uncertainty	116
6.7.5	Uncertainties in theory	116
6.8	Statistics for Setting Upper Limit	117
6.8.1	Likelihood-based method for physics analysis	118
6.8.2	Setting upper limit	119
6.9	Statistical Analysis	122
6.9.1	m_{HH} Fit Binning	122

6.9.2	Fit Results	124
6.9.3	Cross section Normalization	124
6.10	Results	129
7	$X \rightarrow SH$ production sensitivity	131
7.1	MC and Data samples	132
7.2	$X \rightarrow HH$ vs. $X \rightarrow SH$ Topologies	132
7.3	Statistical Fit	133
7.4	Systematics	133
7.5	Results	136
8	Conclusions	139
A	Network Dissection	142
A.1	Introduction	142
A.2	Understanding Convolutional Filters	142
A.2.1	Filters of Handwriting	143
A.3	Analysing Tagger Filters	144
A.4	Physical Connection	145
A.5	Feature Correlation	145
B	Four Prong Tagger Materials	154
C	Signal sample variations	158
D	Tiggers	159
E	Experimental Uncertainties	160
	Acronyms	162
	References	165
	Curriculum Vitae	

LIST OF FIGURES

1.1	The Lagrangian of the Standard Model of Particle physics [5].	2
1.2	(a) Shape of the elementary Higgs potential. (b) Cartoon illustrating the loss of symmetry from the choice of a VEV.	9
1.3	A summary of all the particles in the Standard model [18]	13
1.4	Feynman diagrams of main SM production processes for the Higgs particle at the LHC.	16
1.5	(a) The relative sizes of the cross sections for the different production methods [24]. (b) The relative sizes of the branching ratio for Higgs particle decays [25] with (c) extending the BR to a Higgs mass of up to 1000 GeV.	16
1.6	Feynman diagrams of main SM production processes of di-Higgs events at the LHC [28].	17
1.7	(a) Upper-limit measurements of the di-Higgs cross section for many channels at $27.5 - 36.1 \text{ fb}^{-1}$ as well as their combination [29]. (b) Updated upper-limit measurements for $27.5 - 139.1 \text{ fb}^{-1}$ for available channels [30].	18
1.8	Di-Higgs production ggF cross section as a function of the tri-linear coupling parameter κ_λ [32].	19
1.9	Feynman Diagrams for the production of a resonance via Gluon Gluon Fusion. The resonance then subsequently decays into an HH or SH pair.	20
1.10	Cartoon depicting the change in the topology of the $bbVV$ hadronic final state as the p_T of the Higgs boson increases.[33]	20
1.11	Cartoon of the the resonant di-Higgs (SH) boosted all hadronic final state [33].	21
2.1	An overview of the CERN complex and subsystems, the LHC ring, and detectors and experiments [40].	27

2.2	A overview cutaway of the ATLAS experiment showing the relative location of each of the sub-components as well as the global size [51].	29
2.3	This cartoon shows the difference coordinates used by the ATLAS experiment, from the side on the left and a straight on view on the right. Both representations are centered on the Interaction Point (IP) [52].	30
2.4	The ATLAS Inner Detector, on the left the barrel section of the three ID subsystems are shown and on the right the end-caps [54].	32
2.5	Cut-away of the calorimeter system showing the Liquid Argon and tile calorimeters [59].	34
2.6	(a) Cut-away view of the muon system in the x-y plane. (b) Cut-away view of the muon system in the z-y plane [60].	35
2.7	Depiction of the path that different particles take through the ATLAS detector [65].	37
2.8	(a) ATLAS online luminosity for each data tacking period through 2018 [71]. (b) Total integrated luminosity at ATLAS categorized by the total amount delivered to ATLAS, total recorded, and the portion of the luminosity that is available for physics use [72].	39
2.9	The pileup of each data taking period as well as the total for the full data range [72].	40
2.10	Long-term upgrade and run schedule for ATLAS through HL-LHC as well as the anticipated run luminosity and E_{COM} [74]	41
2.11	A brief flowchart of the steps taken to produce simulations of ATLAS collision data.	41
3.1	This diagramatic representation of a track shows the relationship the track has with the coordinate system as well as its characteristic parameters d_0 and z_0 [89].	47
3.2	ATLAS simulation of the steps taken in topocluster formation for energy deposits from a dijet event into the FCal. Seed cells passing the initial significance threshold are shown in (a). Seeds that then pass the neighbor cell threshold that are shown in (b). Finally, (c) shows the final topoclusters outlining their associated cells [91].	48
3.3	An example of jets reclustered with (a) anti- k_T and (b) C/A for the sample event. The jets are shown shown on the $y - \phi$ plain with the p_T of the jets along z. The total p_T of each jet is represented by the tallest bin in each reclustering [95].	50

3.4	Cartoon depicting the steps in the trimming method of jet grooming [97]. ...	51
3.5	A flowchart and brief description of the states of JES calibration, adapted from [98].	52
3.6	Flowchart summarizing the sequential steps of the TAR algorithm [108].	59
4.1	Cartoon illustrating the collimation of jets as the p_T of the jets increases and the diagrammatic relationship between TAR jets and their underlying small- R jets.	63
4.2	The number of quarks found in $H \rightarrow WW$, $R = 1.00$ jets as a fraction of the total up to 2000 GeV. The fraction of jets with 4 quarks become overwhelming at around 400 GeV and increases to near 100% by 2000 GeV. [116].	64
4.3	(a) The minimum ΔR of the quarks in TAR jets from the $W \rightarrow qq$ decay and (b) the maximum separation of these quarks and the ability to fully contain them in a TAR jets with $R = 1.0$ [116].	65
4.4	(a) A simple artificial neural network with one input layer, a single hidden layer, and an output layer. This is commonly referred to as a MLP. (b) A detailed example of the relationship between the inputs x_n , the weights w and biases b of the nodes, and the output y of a single layer of a network [52].	67
4.5	(a) Linear, (b) sigmoid, and (c) ReLU activation functions. The behavior of these functions at different values of inputs is readily apparent from the shape of the functions.	68
4.6	(a) Cartoon illustrating the first three steps in the filter convolution of a simple image. (b) A single step in an example of the application of maximum pooling. A pool size of 2x2 can be seen in yellow, with the maximum values that have been selected seen in the second array.	72
4.7	(a) The p_T of jets generated from a 5 TeV scalar compared to that of QCD jets. (b) p_T distribution of the $H \rightarrow WW$ jets used for training.	74
4.8	MC examples of calo images for W jets [127] and $H \rightarrow WW$ jets. The additional two subjets in the WW topology obscure the due prong features of the W calo images.	77
4.9	(a) An average Lund jet plane for a large collection of $H \rightarrow WW$ jets. (b) Explanation of the physical interpretations of the different regions of the Lund plane [128].	79
4.10	Distribution of the number of nodes per jet for each training process.	80

4.11	Average truncated Lund jet planes for the five training processes (a) $H \rightarrow WW$, (b) QCD, (c) $t\bar{t}$, (d) W +jets, and (e) $b\bar{b}$	81
4.12	(a) The distribution for the signal training samples as a function of μ . (b) A comparison of the trained and untrained μ network response values as a function of μ , demonstrating the untrained samples have the same response characteristics.	82
4.13	n_2 , m_2 , τ_{42} , and c_4 substructure variables scaled to the range $[0,1]$ for the five processes used to train the four-prong tagger: $H \rightarrow WW$, QCD, $t\bar{t}$, W +jets, and $b\bar{b}$ jets	84
4.14	The architecture of the neural network used for the four-prong tagger. The Lund plane is processed by a convolutional branch shown here on the left, and the substructure features processed by the feed-forward branch on the right. These two branches are combined into a single feed forward to produce the final multi-class output.	86
4.15	(a) The network response R_{WW} for each of the training classes $H \rightarrow WW$, QCD dijet, $t\bar{t}$ jets, W +jets, and $b\bar{b}$ jets. (b) The $1/(\text{QCD dijet rejection})$ vs. signal efficiency is shown comparing the performance to the single variable discriminate τ_{42} and again with a mass windows applied that is consistent with the Higgs mass.	88
4.16	Comparison of the performance for each individual branches and the full four-prong tagger neural network.	88
4.17	(a) A distribution of the network response, R_{WW} , for the $H \rightarrow WW$ process illustrating the global WP. The selection of jets that pass the <i>Loose</i> and <i>Tight</i> WP's are shown in shaded regions of the distribution. (b) The same response as (a) but shown for the p_T bin $1800 < \text{Jet}_{p_T} < 3000$ GeV. All six distributions are show in B.2.	90
4.18	The sculpting of the m_{TAR} for QCD jets. Before training is shown in blue, jets incorrectly tagged as signal shown in black demonstrating a more signal-like, red, mass distribution.	91
5.1	(a) Efficiency, ϵ , for data and MC QCD jets are shown in the top histogram for the <i>Loose</i> WP. The bottom ratio in the efficiencies $\epsilon_{\text{bin}}^{\text{data}}/\epsilon_{\text{bin}}^{\text{MC}}$ with the difference from 1.0 representing scale factor uncertainty. In (b) the sample relationship is shown for the <i>Tight</i> WP.	93
5.2	The p_T of the $H \rightarrow WW$ jets are shown in the upper histograms with the same binning that was applied in the QCD scale factor uncertainty calculations. In the lower histograms the normalized efficiency per p_T bin of the tagger is shown. In red are the uncertainties in these efficiency values derived from the efficiencies in the variations in Table 5.2.	95

5.3	The total uncertainty σ_{TOT} for each p_T bin as well as each contributing source of uncertainty of efficiency for relative size comparison.	97
6.1	(a) The $27.5 - 36.1 \text{ fb}^{-1}$ limits for the resonant di-Higgs production in each individual channel as well as the combination. (b) The most recent results for resonant di-Higgs production for $126 - 139 \text{ fb}^{-1}$	99
6.2	Efficiency curve for two large-R jet triggers demonstrating the efficiencies as well as the effect of jet mass cuts on the efficiency plateau.	101
6.3	p_T distribution for the first four jets in MC for dijet and three signal masses. Jet multiplicity after the first 2 TAR jet cut is also shown in (e).	103
6.4	TAR jet mass for the first four jets in the events in MC for dijet and three signal mass samples. The Higgs mass peak is clearly present in the leading and sub-leading jets in the signal MC.	103
6.5	Four-prong tagger features for dijet and three signal mass samples MC jets.	104
6.6	(a) Diagram of the analysis regions. The y-axis is defined by the number of BTags of the bbCandidate. The x-axis is defined by the <i>Loose</i> and <i>Tight</i> Working Points of the four-prong tagger. (b) An illustrative example of the $\mathcal{R}_{WW}$ response plot showing the regions of the tagger response for each WP.	105
6.7	The kinematic distributions of (a) p_T , (b) η , (c) jet mass, (d) number of b tags, (e) ΔY , and (f) the HH candidate mass for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region.	107
6.8	The distributions for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}	107
6.9	The kinematic distributions of (a) p_T , (b) η , (c) jet mass, (d) number of b -tags, (e) ΔY , and (f) the HH candidate mass for the $b\bar{b}$ and WW jet candidates from the 1200 GeV signal mass samples in the 2 bTag signal region.	108
6.10	The distributions for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}	108
6.11	The kinematic distributions of (a) p_T , (b) η , (c) jet mass, and (d) number of b tags for the $b\bar{b}$ and WW jet candidates from the the 3000 TeV signal mass samples in the 2 bTag signal region.	109
6.12	The distributions for the $b\bar{b}$ and WW jet candidates from the the 3000 TeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}	109

6.13	The kinematic distributions of (a) p_T , (b) η , (c) jet mass, and (d) number of b tags for the $b\bar{b}$ and WW jet candidates from the the 3000 GeV signal mass samples in the 2 bTag signal region.	110
6.14	The distributions for the $b\bar{b}$ and WW jet candidates from the the 3000 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}	110
6.15	MC composition of the background processes considered for the analysis. ..	111
6.16	The distributions of m_{hh} for the data, background estimation, and signal of the (a) 2 and (b) 5 TeV mass points in both validation regions.	113
6.17	Cartoon depicting the relationship between the PDF $f(q_\mu \mu)$, Z, and the p -value shaped in red [152]	120
6.18	The m_{HH} discriminant distribution using (a) an optimized fixed width definition and (b) the binning transform from Equation 6.18 with $z_b = z_s = 4$	122
6.19	Plots showing the pre- and post-fit data in both signal regions for 3000 GeV scalar mass.	123
6.20	Post-fit data in both signal regions for the (a) 800 and (b) 1200 GeV scalar masses.	125
6.21	Post-fit data in both signal regions for the (a) 3000 and (b) 4000 GeV scalar masses.	126
6.22	Systematic pull and impact plots for the 800, 1200, 3000, and 4000 GeV mass points.	127
6.23	Normalisation pull and impact plots for the 800, 1200, 3000, and 4000 GeV mass points. $\Delta\theta$ here is the uncertainty in the fit.	128
6.24	Expected upper limits at 95% C.L. on the cross-section of resonant di-Higgs production as a function of the signal scalar resonance mass.	129
7.1	Feynman Diagrams for the production of a resonance via ggF. The resonance then subsequently decays into a SM Higgs and BSM Higgs-like scalar S	131
7.2	Comparison of 2000 GeV scalar mass point for di-Higgs and $m_S = 170$ SH sample. The similarities between the channels can be seen given the low mass of the S scalar.	133
7.3	p_T distribution for the first four jets in MC for dijet and three signal masses. Jet multiplicity after the first 2 TAR jet cut is also shown in (e).	134

7.4	TAR jet mass for the first 4 jets in the events in MC for Di-Jet and 3 signal mass samples. The expected higgs mass is present in the leading and sub-leading jets in signal.	134
7.5	m_{TAR} for the bb and VV candidates for the $m_X = 2000$ SH sample and a range of m_S masses.	135
7.6	m_S in the 2 b -tag signal region. The boosted S scalar is much more successfully reconstructed.	135
7.7	The NP pulls and limit impacts for the systematic for (a) $m_X = 3000, m_S = 170$ GeV and (c) $m_X = 2000, m_S = 170$. (b) and (d) so the pulls and the rankings for the normalization factors for the same mass pairs.	136
7.8	Comparison of the expected upper limits for the $X \rightarrow SH$ signal model for $m_S = 750, 1000, 2000$, and 3000 GeV samples.	137
7.9	Expected upper limits for the $X \rightarrow SH$ signal model for $m_S =$ (a) 750, (b) 1000, (c) 2000, and (d) 3000 GeV samples.	137
A.1	A single hand written 6 taken from the mnist dataset.	143
A.2	Simple filter definitions corresponding to (a) vertical and (b) horizontal edge detection.	143
A.3	The convolution of Figure A.1 with (a) Figure A.2a and (b) Figure A.2b visually demonstrating the types of features the filters detect.(a) Strong vertical edge detection and (b) strong horizontal edge detection.	144
A.4	Lund jet plane for a large collection of (a) $H \rightarrow WW$ jets and (b) QCD jets. (c) shows the normalized difference between the two images.	145
A.5	Linear correlation of between the labeled variation and the the Lund planes for WW and QCD jets combined.	146
A.6	Correlation matrix for the input features of the feed-forward branch of the network as well as the jet mass and the network response $\mathcal{R}_{WW}$	147
A.7	Columns 1 and 2 show the max-pooled output for each convolutional layer in the network for an averaged sample of WW and dijet lund planes. The last column shows the difference in the output puts highlighting the difference in the strength of the response between the two processes.	147
A.8	All filters from the first layer of the tagger.	148
A.9	The difference between the convolution of the first layer's filter with of QCD and WW average Lund jets planes.	149

A.10	All filters from the second layer of the tagger.	150
A.11	The difference between the convolution of the filters in the second layer with QCD and WW average Lund planes.	151
A.12	All filters from the third layer of the tagger.	152
A.13	The difference between the convolution of the filters in the third layer with QCD and WW average Lund planes.	153
B.1	Correlation matrix of the testing features. The features found in Table 4.1 are a subset of features in the matrix that have minimal correlations with other features.	155
B.2	The performance of the tagger rejecting QCD dijet MC versus signal efficiency. Compared here are two of the more powerful input parameters of the network and conventionally considered for signal variable discrimination.	156
B.3	The performance of the tagger rejecting QCD dijet MC versus signal efficiency. Compared here are two of the more powerful input parameters of the network and conventionally considered for signal variable discrimination.	157

LIST OF TABLES

1.1	BR percentages for the different decay channels of di-Higgs events [28].	17
1.2	Branching ratios for the WW decay channels.	21
4.1	Table containing the list of variables tested as possible substructure inputs. The AUC score and percent difference in the AUC score are listed. In the last column the variables are ranked in order of their impact on the performance of the network from highest to lowest.	77
4.2	The network response values for each p_T dependent WP.	89
4.3	The efficiencies for the Tight WP per p_T bin for all signal jets using a global or individual working points for each p_T bin.	90
5.1	Details of the nominal di-Higgs MC samples used for the MC uncertainty derivation.	95
5.2	Efficiency of each signal sample variation studied and used to generate the uncertainty in four prong tagging.	96
5.3	Summary of the uncertainty values per p_T bin for each of the sources considered, as well as the total uncertainty coming from the combination of the two.	96
6.1	Definitions for generating the resonant signal MC samples.	100
6.2	Definitions for generating the MC for the background processes considered.	100
6.3	Breakdown of the naming scheme used and the description of each pieces in the identifier chain used for triggers.	102
6.4	MC background yields and relative yields in signal regions SR1 and SR2 for 1000, 2000, 3000, and 5000 GeV mass points.	106

6.5	Signal and estimate yields in the validations regions for the 2000 GeV scalar mass point.	114
6.6	Data and estimate yields the estimation validation regions VR1 and VR2. ...	114
6.7	An breakdown of the naming scheme used and the description of each pieces in the identifier chain used for triggers.	116
6.8	Expected upper limits at 95% CL for all signal mass points.	130
7.1	m_X and m_S scalar mass ranges for the SH signal MC samples.	132
7.2	Definitions for generating the signal MC samples.	132
7.3	Expected limits for 750 GeV X scalar mass sample.	137
7.4	Expected limits for 1000 GeV X scalar mass sample.	138
7.5	Expected limits for 2000 GeV X scalar mass sample.	138
7.6	Expected limits for 3000 GeV X scalar mass sample.	138
B.1	Loose WP efficiency for each signal sample variation studied and used to generate the uncertainty in four prong tagging.	154
C.1	A brief description of each of the physics list variations, model inputs, as well as the detector differences summarized from [154]..	158
D.1	Complete list of the unrescaled single large-R jet triggers.	159
E.1	Complete list of systematic uncertainties included in the analysis.	162

Chapter 1

THEORY

One of the most important discoveries in the last half-century in particle physics was the detection of a Higgs boson by the ATLAS [1] and CMS [2] experiments at the Large Hadron Collider (LHC) [3]. It is hard to overstate the significance of this detection. The Higgs particle had long been [4] theorized as a necessary piece of the Standard Model (SM) making its discovery a crucial source of validation for the current model of particle physics. The SM is one of the most successful scientific theories to date both in the breadth of phenomena that it describes as well as the accuracy to which it describes them. Perhaps the most iconic representation of the SM for both physicists and layman alike is the formalism of the Lagrangian pictured in its infamous form in Figure 1.1. In the following chapter the mathematical fundamentals underlying the SM are outlined. These foundations contextualize the subsequent discussion of SM di-Higgs production as well as the resonant di-Higgs and SH models that are the subject of this thesis. Finally, examples of Beyond the Standard Model (BSM) extensions that offer interpretations of the new resonances are briefly summarized.

1.1 Underlying Fundamentals

The purpose of this section is to explain the underlying mathematics of the SM and connect it to the physics objects it describes. The SM is often described as elegant and perhaps the most clear example of this is seen in the physical symmetries and the mathematical groups that described them. The SM is a gauge quantum field theory constructed on a small foundational set of ideas: the Principle of Least Action, Noether's

$$\begin{aligned}
\mathcal{L} = & -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \\
& + i \bar{\psi} \not{D} \psi + h.c. \\
& + \chi_i^\dagger Y_{ij} \chi_j \phi + h.c. \\
& + |D_\mu \phi|^2 - V(\phi)
\end{aligned}$$

Figure 1.1: The Lagrangian of the Standard Model of Particle physics [5].

theorem, and the Principle of Locality. The SM is a quantum field theory that describes fields in spacetime, three spatial dimensions and one time dimension. The SM has a $SU(3)_c \otimes SU(2)_L \otimes U(1)_Y$ internal symmetry above the Electroweak (EW) energy scale and $SU(3)_c \otimes U(1)_{EM}$ below. The transition between the high and low energy regimes is manifested by the Spontaneous Symmetry Breaking of the $SU(2)_L \otimes U(1)_Y$ symmetry facilitated by the Englert–Brout–Higgs (EBH) mechanism.

1.1.1 Principle of Least Action

A Lagrangian is a mathematical expression that describes the trajectory of the particles in a given system and in general is expressed as the difference between the kinetic and potential energy of that system. They are used to determine the action of the system

$$S = \int d^4x \mathcal{L} \ , \quad (1.1)$$

where S is the action and $\mathcal{L}$ is the lagrangian density. The lagrangian density here has three requirements: it must only be a function of the fields, it must be a real-valued scalar with mass dimension four in natural units, and it must be Lorentz invariant. The lagrangian of the SM describes fields that are a generalization of the path of a particle to

not only depend on time, but also on the position in space. This is known as spacetime coordinates. The lagrangian for a free, real valued scalar field is

$$\mathcal{L} = \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 \quad , \quad (1.2)$$

which when used to minimize the action from Equation 1.1 yields the equation of motion of the system. This equation

$$\square \phi + m^2 \phi = \left(\frac{\partial^2}{\partial^2} - \nabla^2 + m^2 \right) \phi = 0 \quad (1.3)$$

is known as the Klein-Gordon equation and is a second-order differential equation representing the dynamics of a particle ϕ with a mass m the solutions to which are plane waves. To take this to a quantum field the dynamic variables are promoted to operators. This change brings about issues with the solutions in that they seem to require negative energies. This issue is resolved through the introduction of the Dirac spinor Ψ . The end result is the Dirac lagrangian

$$\mathcal{L}_D = i \bar{\Psi} \gamma^\mu \partial_\mu \Psi - m \bar{\Psi} \Psi \quad , \quad (1.4)$$

with the first term describing the dynamics of the spinors, the second term giving the mass, and γ^μ are the gamma matrices. The spinors, Ψ , represent a bispinor field

$$\Psi = \begin{pmatrix} \Psi_L \\ \Psi_R \end{pmatrix} \quad (1.5)$$

with left and right spinor components. The left and right components are named for their helicities that result from the direction of motion of spin-1/2 particles such as electrons. Electrons are electrically charged particles and conservation of this charge in observations leads to the introduction of symmetry groups.

1.1.2 Weyl and Fock's Insight and Noether's Theorem

An extraordinary insight that helped to shape the world of modern theoretical physics and was crucial to the development of the SM is Emmy Noether's theorem [6]. Noether's theorem identified a fundamental relationship between differentiable symmetries and conserved quantities. It shows that for each symmetry of the system there exists a conservation law characterized by the group charge and conserved through the Noether's current. An instructive demonstration of this is to look at a coordinate transformation. An underlying assumption of physics is that any theory must be independent of the reference frame. This means that any acceptable theory must produce the same results regardless of the location in space (or spacetime generally). For instance, a transformation in a single dimension $x \rightarrow x + \epsilon$ for a field $\phi(x)$ is

$$\phi(x + \epsilon) = \phi(x) + \epsilon \frac{\partial \phi}{\partial x} , \quad (1.6)$$

All possible such transformations can be described by a group M where all the members of the group can be expressed as $M(\theta) = e^{-i\theta} [7]$ parameterized by θ . The transformations of the group are carried out by its generator $\mathbf{X}$ defined as $M(\theta) = e^{-i\mathbf{X}\theta}$. An expansion in the parameter θ yields

$$M(\theta) = 1 - i\mathbf{X}\theta , \quad (1.7)$$

an expression of the generator of infinitesimal translations in one dimension. Noether's theorem relates the generators of a group through a current. Inspection of the two above equations shows that the generator of translations in x is

$$\mathbf{X} = -i \frac{\partial}{\partial x} , \quad (1.8)$$

yielding the Noether current for 1D translation and revealing the conserved quantity, momentum. This one-dimensional translation is merely a part of a larger group, the Poincaré group. The Poincaré group is responsible for 3D translations, 3D rotations, 3D Center-of-Mass (COM) velocities, and time. These 10 transformations lead to conser-

vation of momentum, angular momentum from rotations, and COM frame invariance, and energy. This means that any theory that works with special relativity would have to be invariant under Poincaré group transformation. It also demonstrates how group symmetries and Lorentz invariance are fundamental to the physics of the SM. The connection between the above relationship and fields comes from Fock and Weyl [8]. In order for the fields in a quantum field theory to be invariant under the same sort of local transformation it must remain unchanged when multiplied by a space-time dependant phase such as

$$\Psi \rightarrow \Psi = \Psi' e^{-i\chi\theta} \quad (1.9)$$

which yields an interesting consequence. Given that the generator of the $\mathcal{U}(1)$ symmetry group is the identity $X = 1$, imposing local gauge invariance on a quantum field theory immediately implies the existence of a $\mathcal{U}(1)$ symmetry. Applying this to the Dirac Lagrangian uncovers the first electromagnetic symmetry group $U(1)_{EM}$.

1.2 Quantum Electrodynamics

While the conservation of electric charge was suggested as far back as 1746 and proof was supplied by Faraday in 1846 [9], the introduction of gauge invariance to the particle description explicitly described the conserved quantity and current. The $U(1)_{EM}$ gauge transformation applied to Equation 1.11 yields an additional field. To account for this new field and to maintain local gauge invariance under the $\mathcal{U}(1)$ symmetry, a new vector field A_μ which obeys the transformation $A_\mu \rightarrow A_\mu + \partial_\mu\theta$ is introduced. With this new field the dynamics of the Lagrangian changes and the covariant derivative can be defined as

$$D_\mu = \partial_\mu + iQA_\mu \quad (1.10)$$

where Q is the charge of the $\mathcal{U}(1)$ symmetry group associated with the electric charge and A_μ is the gauge field that mediates the EM interactions, the photon. This transforms the

Dirac Lagrangian into the Lagrangian for Quantum Electrodynamics (QED)

$$\mathcal{L}_{\text{QED}} = i\bar{\Psi}\gamma^\mu D_\mu\Psi - m\bar{\Psi}\Psi - A_\mu J^\mu - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad , \quad (1.11)$$

where J^μ is the electromagnetic current and $F_{\mu\nu}$ is the field strength tensor for the electromagnetic interaction called the Faraday tensor whose components are the electric and magnetic fields. This success of expressing local gauge invariance and symmetry groups motivates the desire to describe the interactions of other observed symmetries with higher order groups.

1.3 Additional symmetries

Before moving on to the groups of the EW and Strong forces, there are additional symmetries resulting from the particle/anti-particle relationship of the SM particles. While the theory correctly describes these pairs, there is an ambiguity in the assignment of the fields to either the particle or the anti-particle. The solution to this is the introduction of three additional symmetries Charge Conjugation (C), Parity Transformation (P), and Time Reversal (T) whose invariance imply the choice is arbitrary. Hinted at by the names, C swaps the charges of the fields, P inverts the spatial coordinates, and T inverts the time coordinate. The SM is invariant under the successive application of all three of these symmetries, known as CPT symmetry. However, while logically motivated, nature is not as simple with both P and CP violations observed in the weak force.

1.4 Electroweak Sector

Before discussing the EW interaction it is first necessary to discuss the weak interaction alone. The phenomena of β decay suggested a new force that was very weak and short ranged. The first theory to explain this phenomena was put forth by Enrico Fermi in 1934 [10]. His proposal was a four particle point interaction theory to describe the weak force. This model was highly successful, however the discovery of the weak P [11] and CP [12] violations lead to the development of a gauge theory for the weak interaction.

In β decay an up quark decays into a down quark and supposing a $SU(2)$ symmetry

for the weak interaction, these fields are then treated as different representations of the same particle, differing by the charge of the group called isospin. While successful at describing weak processes, it is immediately clear that a solely $SU(2)_L$ theory for the weak interaction is not a correct description of nature. This can be seen in the observed difference in the mass of the up and down quark or in the gauge fields associated with the generators of a weak interaction which are massless under $SU(2)_L$ but, even before detecting the W and Z bosons, the weak interaction mediators were known to be massive. The solution to this comes from the Glashow-Weinberg-Salam model of EW unification [13]. This model proposes a $SU(2)_L \otimes U(1)$ symmetry group that couples to left-handed fields to describe both the EM and weak interactions as a combined phenomena called Electroweak. This gauge symmetry results in a covariant derivative

$$D_\mu = \partial_\mu - ig_B Y B_\mu - ig_W T_j W_\mu^j \quad , \quad (1.12)$$

where the gauge fields B_μ and W_μ^i have been introduced to maintain local symmetry for the $U(1)_Y$ and $SU(2)_L$ groups respectively. The charge of the $U(1)_Y$ group is Y known as hypercharge and T^j are the 3 weak-isospin charges of the $SU(2)_L$ group. The Lagrangian then for the EW interaction is

$$\mathcal{L}_{EW} = i\bar{\Psi}\gamma^\mu D_\mu\Psi - m\bar{\Psi}\Psi - \frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a,\mu\nu} \quad , \quad (1.13)$$

where the field strength tensor are $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ and $W_{\mu\nu} = \partial_\mu W_\nu - \partial_\nu W_\mu - gf^{abc}W_\mu^b W_\nu^c$. Curiously, there is a new phenomena that has been introduced to the SM, now and that is the non-Abelian nature of the $SU(2)$ group. Non-Abelian in QM in general means that the operators do not commute, that is $AB - BA \neq 0$. For a gauge symmetry this means that the generators of the groups do not commute. The $(-gf^{abc}W_\mu^b W_\nu^c)$ term is a result of the lack of commutation. The implications here are that the gauge fields W_μ^i are capable of interacting with each other.

The left-handed doublets of the EW interaction describes particles that differ from each other by the inversion of the third component of the weak isospin T_3 , for instance

the up and down quarks or the electron neutrino and the electron have $1/2$ and $-1/2$ T_3 respectively. While this inversion is completely acceptable within the EW group, it doesn't however conserve the EM charge Q . Anticipating this conserved quantity of the $U(1)_{EM}$ symmetry group it is advantageous to express the gauge fields of the EW group in terms of their EM charge;

$$W^\pm = (W^1 \pm iW^2)/\sqrt{2} \quad (1.14)$$

$$Z = \cos \theta_w W^3 - \sin \theta_w B \quad (1.15)$$

$$A = \sin \theta_w W^3 + \cos \theta_w B \quad , \quad (1.16)$$

where Z and A are uncharged fields. This redefinition results in the relationship

$$Q = (T_3 + \frac{1}{2}Y) \quad (1.17)$$

between the charges of the EW and EM groups. These fields mediating the EW interaction are massless, but as has already been stated, the weak bosons are observed to be massive in nature. Something more is needed to connect these gauge fields and the mediators of the separated the weak and EM forces. The Higgs mechanism described below facilitates the symmetry breaking.

1.5 Electroweak Spontaneous Symmetry Breaking

Electroweak (EW) Spontaneous Symmetry Breaking (SSB) breaks the symmetry group that described the unified EW interaction into the individual weak and EM interactions.

1.5.1 Spontaneous Symmetry Breaking

It's important to note that SSB is not a phenomena unique to the standard model or particle physics. There are many other examples in nature where symmetries are spontaneously broken, for instance the loss of ferromagnetism above the Curie temperature [14] in ferromagnetic materials. In the SM the EW symmetry is broken at low energies which means the $SU(2)_L \otimes U(1)_Y$ symmetry groups is broken. Curiously, symmetry

breaking is associated with a phase transition, analogous to ferromagnetism, but in the SM the transitions is a continuous crossover.

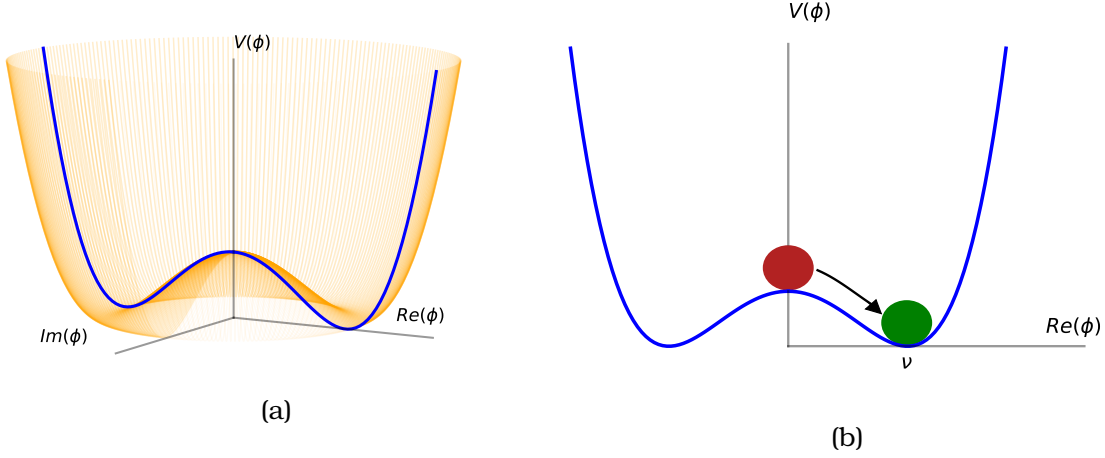


Figure 1.2: (a) Shape of the elementary Higgs potential. (b) Cartoon illustrating the loss of symmetry from the choice of a VEV.

1.5.2 Higgs Mechanism

The Higgs Mechanism begins with the introduction of a complex scalar (spin-0) field with an associated potential. This potential has a Vacuum Expectation Value (VEV) that breaks the symmetry of the theory. Since the symmetry to be broken is $SU(2) \otimes U(1)$ the introduced field must be a $SU(2)$ doublet. The functional from of the field therefore is

$$\phi(x) = \frac{v}{2} e^{iW^a \tau^a} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \quad (1.18)$$

where ϕ_1 and ϕ_2 are complex components of the doublet, $v = \frac{\mu}{\sqrt{\lambda}}$, T^a are generator of $SU(2)$, and W_i is the gauge field. The Higgs potential is

$$V(\phi) = -\mu^2 |\phi|^2 + \lambda^2 |\phi|^4, \quad (1.19)$$

which is illustrated in Figure 1.2. The choice of the functional form 1.19 isn't determined but is the simplest potential that can facilitate the symmetry breaking and still respect the underlying unbroken symmetries [15]. Clearly for non-zero values of v this potential

has two minima on the real axis, $\pm v$. The lowest energy state of the Higgs field then is at one of these two VEV's. Taking $+v$ as the value for the real component of ϕ_2 Equation 1.21 becomes

$$\langle \phi(x) \rangle \approx \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix}. \quad (1.20)$$

Since only one component of the Higgs field is non-zero after a VEV is chosen, the symmetry of the $SU(2) \otimes U(1)$ is broken for non-zero values of v , illustrated by the green circle in Figure 1.2 moving from the symmetric state to a point in the potential "well". This means the internal $SU(2)_L \otimes U(1)_Y$ symmetry is *spontaneously* broken. This new ground state of the field can be expanded around by some excitation H

$$\phi(x) \approx \begin{pmatrix} 0 \\ \frac{v+H}{\sqrt{2}} \end{pmatrix}. \quad (1.21)$$

The Lagrangian for this new scalar field is

$$\mathcal{L}_\phi = D_\mu \phi^\dagger D^\mu \phi - \mu^2 |\phi|^2 + \lambda^2 |\phi|^4, \quad (1.22)$$

where D_μ is Equation 1.12. Expanding this equation with Equation 1.12 and the relationships in Equation 1.16 gives

$$\mathcal{L}_\phi = \frac{1}{2} \partial H \partial H - \frac{1}{2} M_H^2 H^2 - \frac{1}{2} M_Z Z^2 + \frac{1}{2} M_W W^+ W^- - \lambda v^2 H^2 - \lambda v H^3 - \frac{1}{4} \lambda H^4 + \dots \quad (1.23)$$

where mass terms for the Higgs field and the W and Z bosons have emerged, though no mass term appears for the second uncharged field A . This remaining massless boson corresponds to the photon of the $U(1)$. The relationship between these bosons and the Higgs mass are

$$M_\gamma = 0 \quad (1.24)$$

$$M_W = \frac{vg_W}{2} = \frac{M_H}{\sqrt{2}\lambda} \quad (1.25)$$

$$M_Z = \frac{vg_W}{2\cos\theta_W} = \frac{M_W}{\cos\theta_W} \quad , \quad (1.26)$$

and the H^3 term corresponds to the Higgs tri-linear coupling interaction illustrated by κ_λ in Figure 1.6.

1.6 Strong Sector

The theory used to describe the interactions of the strong force is known as Quantum Chromodynamics (QCD). In much the same way that the EW for is described by $SU(2)_L \otimes U(1)_Y$, the strong interaction is described by the $SU(3)_c$ symmetry group. The charge of this group is known as *color charge* of which there are three: red, green, and blue. All quarks are charged under the $SU(3)_C$ group. The covariant derivative describing the dynamics of the theory is

$$D_\mu = \partial_\mu - i\frac{g_c}{2}\lambda_\alpha G_\mu^\alpha \quad , \quad (1.27)$$

with G_μ^α representing the gauge field for the gluons, g_c the coupling strength, and λ_α are the Gell-Mann matrices. The Lagrangian then for the strong interaction can be compactly written as

$$L_{\text{QCD}} = i \sum_q \bar{\Psi}_q^k \gamma^\mu D_{\mu,jk} \Psi_q^k - \sum_q m_q \bar{\Psi}_q^j \Psi_q^j - \frac{1}{4} G^{k,\mu\nu} G_{k,\mu\nu} \quad , \quad (1.28)$$

where $G_{\mu\nu}$ is the strong field strength tensor and the index q sums over the quarks with the indices i, j, k corresponding to the color [16]. As with the non-abelian group of the EW interaction, the $SU(3)_C$ group of the strong force also has self-interacting gauge fields.

1.7 Fermion Mass and Yukawa Matrix

The process described in 1.5.2 not only describes the origin of the weak gauge boson masses, but also accounts for the masses of the fermions in the SM. The introduction of the Higgs field to the SM produces mass terms such as

$$M_\Psi = \dots - \frac{\lambda(v + H)}{\sqrt{2}} \bar{\Psi}\Psi + \dots \quad (1.29)$$

The 3 flavors of quarks (e.g. u, c, t) are a non-conservative quantity. Flavor mixing, or the process of a quarks mixing with another flavor is described by the Cabibo-Kobayashi-Maskawa (CKM) matrix. The CKM matrix

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{-i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}, \quad (1.30)$$

where the c/s_{nm} are cos and sin of the CKM mixing angles θ_{12}, θ_{13} and θ_{23} relates the different flavors through it's mixing angles. In addition to the three mixing angles is the phase of the CKM δ . This phase along with the requirement that the CKM matrix be unitary sets the limit on the minimal number of flavors in the SM to be three. Experimental measurements have also greatly constrained the possibility of additional flavors [17].

1.8 Summary of Standard Model Particles

Having now discussed the underlying mathematical framework for the SM a brief summary of the particles is warranted. The two types of particles in the SM are fermions and bosons. Fermions, the quarks and leptons, are particles that follow the Pauli exclusion principle and have half integer spin ($\frac{1}{2}, \frac{3}{2}, \dots$). There are six flavors of quarks separated into three generations: up (u) and down (d), charm (c) and strange (s), and bottom (b) and top (t). As was stated, quarks are charged under all symmetry groups of the standard model and therefore interact through all known forces. There are also three generations of leptons, each generation containing a charged lepton and its neutrino counterpart:

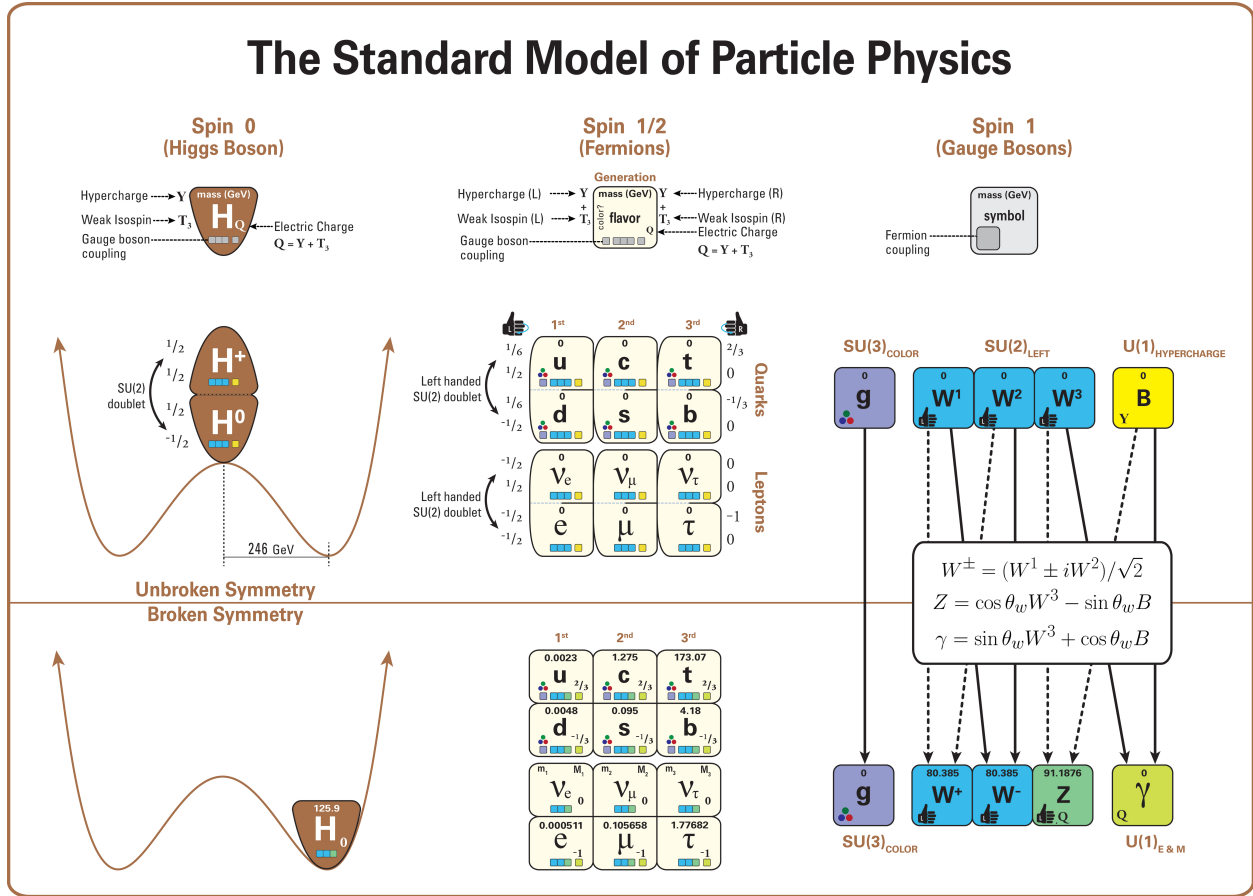


Figure 1.3: A summary of all the particles in the Standard model [18]

electron (e) and electron neutrino (ν_e), muon (μ) and muon neutrino (ν_μ), and the tau particle (τ) and tau neutrino (ν_τ). The strong force interaction coupling is described by the color charge (R, G, B) of each quark with composite particles either being made up of even numbers of color/anti-color pairs (mesons) or odd numbers of r,g,b (baryons). The Yukawa interaction gives mass to all EM-charged fermions; however, there isn't a description within the SM to account for the massive nature of neutrinos (Section 1.9.1).

The bosons are integer spin particles that follow Bose-Einstein statistics. As illustrated in the previous sections the force-mediating particles are all bosons and the Higgs particle is as well. There are 13 bosons in the SM in total, 8 strong force mediating gluons, 4 EW bosons which after SSB become combine as the 3 massive gauge bosons of the weak force and the massless photon that mediates the EM interaction. The final boson is the Higgs particle associated with the Higgs mechanism responsible for the

symmetry breaking of the SM. The force carriers bosons, W, Z, g , and γ , are spin-1 and the Higgs is spin-0. A summary of all of the particles is found in Figure 1.3.

1.9 Beyond the Standard model

Despite the overwhelming successes of the SM there are many observations of the natural world that remains unexplained.

1.9.1 Neutrinos

Unlike the mass of the charged fermions, the nature of the origin of neutrino mass is not explained within the SM. Once thought of as massless, through experimental measurements it is known that neutrinos do indeed have mass, though the values and the mechanism for mass generations are unknown [19, 20]. There are simple extensions that can account for the peculiar nature of neutrinos. One of these such explanations is the assertion that neutrinos may in fact be their own anti-particle. This type of particle is known as a Majorana particles and stand in stark contrast to the particle/anti-particle pairing of the charged fermions in the SM.

1.9.2 Matter anti-matter asymmetry

The proportion of matter to anti-matter in the universe is wildly disproportionate with there being billions of times more matter than antimatter. While the SM does allow for difference in the production rates of matter and anti-matter through CP violation in the CKM matrix it doesn't come close to accounting for the observed differences. Experiments are ongoing, including experiments at CERN, to further understand this problem however to date there is not a clear candidate for a solution.

1.9.3 Dark matter and dark energy

All of the particles described by the SM are often referred to as *ordinary* matter. Ordinary matter comprises around 4.9% of the universe. The remainder is made up of 26.6% of so called *Dark Matter* and 68.5% *Dark Energy* [21]. These two phenomena are entirely

absent in the SM. As with many of the phenomena presented in this section there exist extensions to explain them. The origins of dark matter can be explained for instance by a larger mass scalar field describing a dark matter particle. As for dark energy, the vacuum energy density of General Relativity behaves similarly mathematically to that of the SM, however the observed values differ by a factor of $\sim 10^{120}$ [22].

1.9.4 Gravity

The physics of the SM describes the nature of particles interacting in a spacetime that is itself described by the theory of General Relativity and to date there has not been any observations that would suggest a connection between the two theories. There is also the additional complication that the mathematics of the two theories are not compatible. This is not to say that there are not *many* efforts to connect them. For instance, if gravity is quantized in a way similar to that of the particles in the SM, the force of the interactions then requires a mediating particle, a so called Graviton. There are many theories of the nature and source of a graviton explained through extensions to the SM. Related to the work presented are Kaluza-Klein (KK) gravitons which could be produced as new massive spin-2 scalars [23] in an analogous way to the spin-0 model which is the subject of this thesis.

1.10 Higgs Physics

In the SM there are four first-order methods for producing Higgs bosons in proton-proton collisions which are: gluon-gluon Fusion (ggF), Vector Boson Fusion (VBF), Vector Boson Associated (VH) production, and top-associated production (ttH). At the LHC the dominant mechanism for the production of higgses is ggF. In ggF a Higgs is produced from a top triangle at Leading Order (LO), seen in Figure 1.6. The second most probable method is VBF where the Higgs is produced with two forward jets followed by VH in which the Higgs is emitted from a Vector boson in so called *Higgs-strahlung* radiation. Lastly the most rare method is a Higgs produced with an associated top/anti-top pair. The branching ratios for these different production methods can be seen in Figure

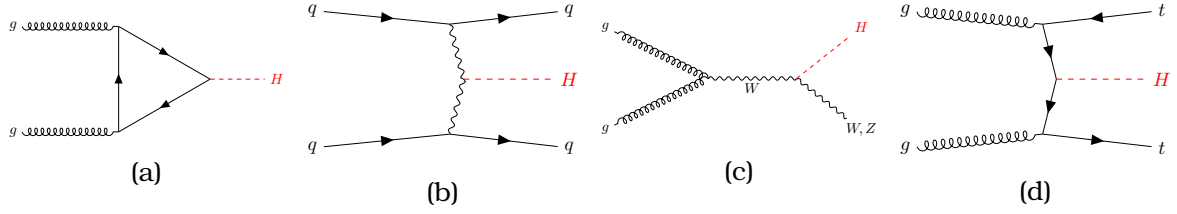


Figure 1.4: Feynman diagrams of main SM production processes for the Higgs particle at the LHC.

1.5.

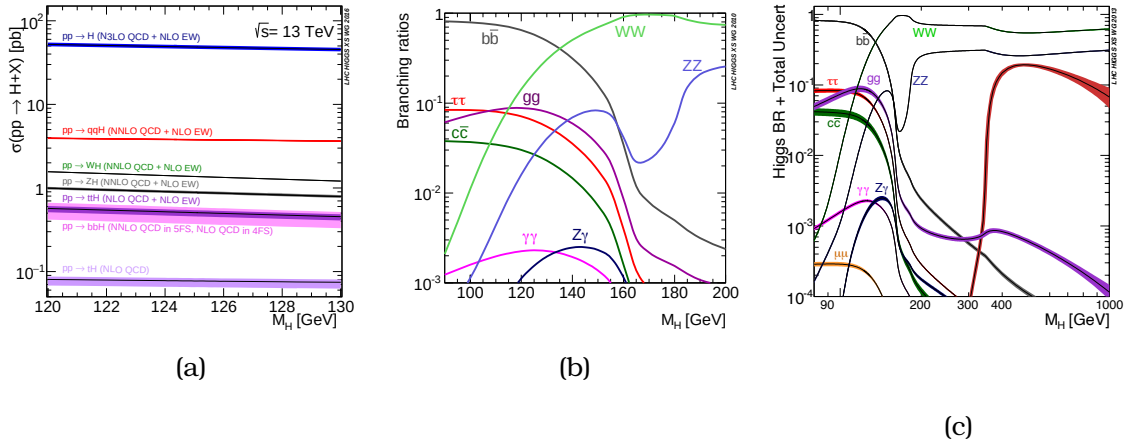


Figure 1.5: (a) The relative sizes of the cross sections for the different production methods [24]. (b) The relative sizes of the branching ratio for Higgs particle decays [25] with (c) extending the BR to a Higgs mass of up to 1000 GeV.

The branching ratios of Higgs decay processes are depicted in Figure 1.5 as a function of the mass of the Higgs. The ratios are a function of the mass of the particle that before the Higgs boson discovery meant that searching in a particular decay channel was more complicated; however, with the mass of the SM Higgs being fixed to 125.09 GeV [26] these BRs are also settled for the SM Higgs boson. The primary decay channel of a Higgs boson is to two b -quarks with the second most probable, and the focus of the analyses in this thesis, being the decay into two W bosons [27].

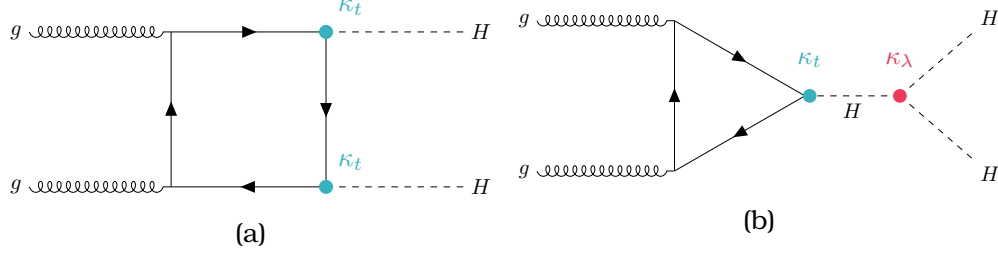


Figure 1.6: Feynman diagrams of main SM production processes of di-Higgs events at the LHC [28].

1.11 Di-Higgs Production

Searching for di-Higgs events at ATLAS is a significant and complex endeavor growing in interest within the ATLAS collaboration. Searches are performed for both standard model di-Higgs production as well as for Higgs pairs produced through more exotic BSM methods.

1.11.1 Non-resonant di-Higgs Production

The two main SM methods of producing Higgs pairs are shown in Figure 1.6 and their BR's in Figure 1.1.

	bb	WW	$\tau\tau$	ZZ	$\gamma\gamma$
bb	33%				
WW	25%	4.6%			
$\tau\tau$	7.4%	2.5%	0.39%		
ZZ	3.1%	1.2%	0.34%	0.076%	
$\gamma\gamma$	0.26%	0.10%	0.029%	0.013%	0.0005%

Table 1.1: BR percentages for the different decay channels of di-Higgs events [28].

Previous Results

The increased luminosity of the Run 2 dataset has lead to a significant increase in the sensitivity to di-Higgs events at ATLAS. Each channel produces limits for their own final state but eventually these limits are combined into limit for di-Higgs production itself. A

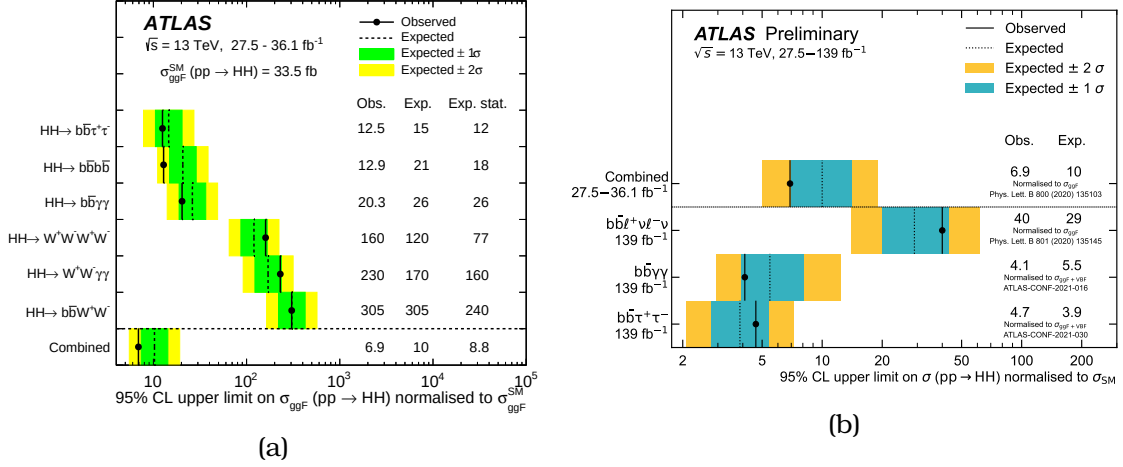


Figure 1.7: (a) Upper-limit measurements of the di-Higgs cross section for many channels at $27.5 - 36.1 \text{ fb}^{-1}$ as well as their combination [29]. (b) Updated upper-limit measurements for $27.5 - 139.1 \text{ fb}^{-1}$ for available channels [30].

summary of the 95% confidence interval limits for previous results are shown in Figure 1.7a and more recent results for channels using the full Run 2 dataset are in Figure 1.7b. The relative limits from the previous combination of all channels are compared to current Run 2 data in Figure 1.7b as well.

In addition to setting limits on di-Higgs sensitivity, ATLAS also sets limits on the Higgs self-coupling. The so called triangle diagram, referencing the quark loop, shown in Figure 1.6b offers the ability to more precisely measure the Higgs tri-linear coupling by means of the constant κ_λ . The coupling $\lambda_{\text{SM}} v H^3$ from Equation 1.23 is parameterized with $\kappa_\lambda = \lambda_{HHH}/\lambda_{\text{SM}}$ allowing searches for deviations of the coupling from the SM value of $\kappa_\lambda = 1.0$ [31]. Figure 1.8 shows previous results for the upper limits on ggF di-Higgs production as a function of κ_λ . Direct measurements of Higgs coupling will help to better understand the nature of the potential that is responsible for spontaneous symmetry breaking.

1.11.2 Resonant di-Higgs and SH production

The 125 GeV SM Higgs boson responsible for SSB may not be the only Higgs or Higgs-like particle in nature. In fact there is no imposed limit known on the number of scalar particles in the SM and illustrated by the initial inclusion of the new scalar field in the

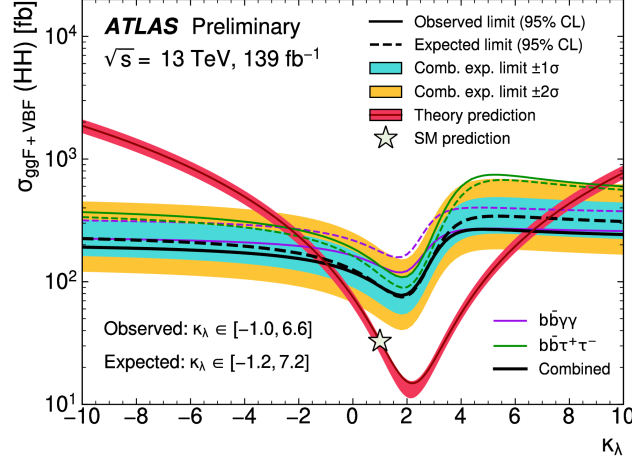


Figure 1.8: Di-Higgs production ggF cross section as a function of the tri-linear coupling parameter κ_λ [32].

Higgs mechanism. In fact, there are numerous theories that exist to explain phenomena unresolved by the SM that predict additional scalar particles. Resonant di-Higgs production hypothesizes a scalar resonance X which decays to a pair of Higgses. Searches for these heavy scalar particles can then be conducted in the final states of SM di-Higgs production. The Feynman diagram for the ggF production of a resonant scalar X decaying into two SM Higgs particles is shown in Figure 1.9. There is no limitation on the mass of the scalar particle X in Figure 1.9 therefore X can take on any value $m_X > 125$ GeV that is not already eliminated by Higgs mass searches. Comparing Figure 1.9 and Figure 1.6b it's clear that when the mass is simply 125 GeV the SM production channel is recovered. This non-SM scalar particle X would be ground breaking in and of itself; however, it could also explain BSM phenomena discussed in Section 1.9.

In addition to the X resonance from the previous section, the scalar sector can be further expanded to include an additional massive scalar particle S where X decays to a SM Higgs and the additional heavy scalar S . This is a relatively new area of study for the ATLAS collaboration as the results in this thesis are the first SH results for this final state.

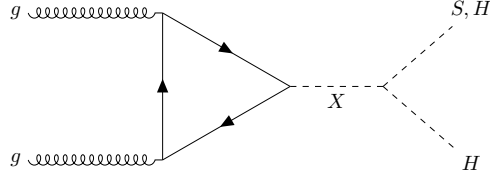


Figure 1.9: Feynman Diagrams for the production of a resonance via Gluon Gluon Fusion. The resonance then subsequently decays into an HH or SH pair.

1.12 The $HH \rightarrow bbVV$ analysis

The $HH \rightarrow b\bar{b}VV$ analysis searches for di-Higgs events in the $bbVV$ decay channel with two b -quarks and two V bosons which could be either W or Z bosons. Depending on the final state these vector bosons can either be differentiated or are both treated as vector bosons V from a kinematic perspective. The decays of these b -quarks and vector bosons results in a wide range of final states to be analyzed. The topologies of these final states can vary significantly, so that it makes sense to further sub-divide the channel in order to maximize the sensitivity to $bbVV$.

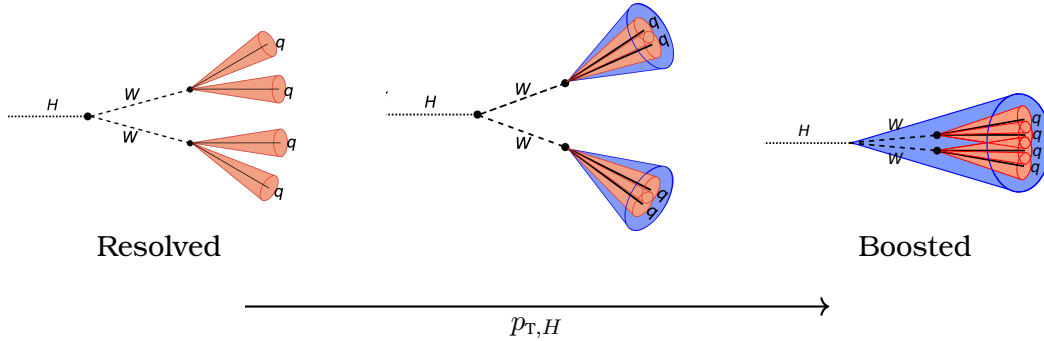


Figure 1.10: Cartoon depicting the change in the topology of the $bbVV$ hadronic final state as the p_T of the Higgs boson increases.[33]

The analysis is broken down by the transverse momentum of the b -quarks and vector bosons as well as the lepton content of the final state. When considering the decay of a W boson, the capability for analyses to resolve its decay products, or the W 's themselves as individual objects is dependant upon the p_T of the Higgs particle. This change in topology is illustrated Figure 1.10. As the p_T of the Higgs increases the $bbVV$ topology transitions from one where the decay products of the W 's are fully resolvable to the boosted regime where both W 's overlap too much to be individually discerned. The intermediary topology

	qq	$l\nu_l$
qq	47%	
$l\nu_l$	21%	32%

Table 1.2: Branching ratios for the WW decay channels.

in Figure 1.10 isn't as consequential to $X \rightarrow HH$ phenomena but as will be discussed in Section 1.12, when the mass of one of the Higgs particles is allowed to increase, the $bbVV$ topology because much more complex.

The lepton content is a significant delineator for $bbVV$ analyses final states. The BR for the W bosons are shown in Table 1.2. All of these branching ratios are investigated by ATLAS however the most probable final state for the W 's $qqqq$ is the subject of this thesis. Resolved and low- p_T resonant final states for this topology have so far not been explored.

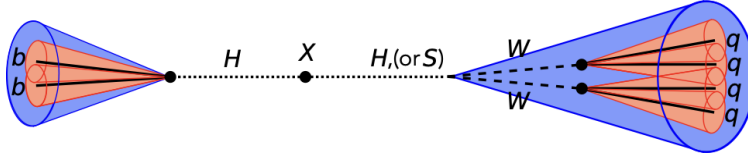


Figure 1.11: Cartoon of the the resonant di-Higgs (SH) boosted all hadronic final state [33].

$$X \rightarrow HH$$

Resonant production often results in the two Higgs particles being *boosted* due to the increased mass of X . The boosted nature of this decay results in b -quarks and V bosons that have very little spatial separation. This in turn leads treating them as unified bb and VV objects. This overlap can be seen in the cartoon of the decay process in Figure 1.11. Tagging of this unified VV object and the overall process is subject of this thesis covered in detail in Chapters 4 and 6.

$$X \rightarrow SH$$

The additional scalar particle S further complicated things since the topology depends heavily on the relative mass of X and now S . As the mass of S approaches that of X the V bosons become much less boosted and more resolved as shown in Figure 1.10. This

point and further implications are elaborated on thoroughly in Chapter 7.

1.12.1 Interpreting New Scalars

Resonant di-Higgs analyses are performed with models that govern the generation of the signal model MC. Two attractive theories that are used broadly throughout di-Higgs resonance searches to explain the the resonant scalar particle described above are the Two-Higgs-doublet model (2HDM) and Kaluza-Klein (KK) gravitons.

Two-Higgs Doublet Model

The scalar section of the SM, i.e. the Higgs sector, is relatively unexplored compared to the fermionic sector. These models introduce new Higgs doublets and thereby new scalars to the SM. A resonant spin-0 scalar predicted by such 2HDM's is used for the signal model analyzed in Chapters 6 and 7. These models have many potential goals and since they introduce new degrees of freedom to the SM, they have possible explanations for dark matter and neutrino masses [34].

Kaluza-Klein Gravitons

The original Kaluza [35] and Klein [36] theory added an additional, cyclic, compact, spatial dimension to the SM therefore expanding the SM to five dimensions. With this additional dimension the results of working through the general relativistic machinery is a series of spin-2 gravitons dubbed *KK gravitons*. These KK gravitons couple to the SM Higgs and are detectable in resonant di-Higgs production.

1.13 Other di-Higgs channels

In addition to the $b\bar{b}VV$ channel many other di-Higgs final states are investigated by the ATLAS collaboration. As with the work in this thesis, the objective currently of these analyses is to set an upper limit of the sensitivity required for detection.

$b\bar{b}\gamma\gamma$

The $b\bar{b}\gamma\gamma$ analysis covers a broad range of phenomena. Principally SM and resonant ggF production [37] whose combined limits are shown in Figure 1.7b.

$b\bar{b}\tau\tau$

Thoroughly investigated, the $b\bar{b}\tau\tau$ final states has analysis for SM and resonant production as well as highly boosted SM τ 's. The combined limits are also shown in 1.7b.

$b\bar{b}b\bar{b}$

The di-Higgs final state with the largest BR, " $4b$ " takes advantage of the highly refined efforts in identifying b -quark physics in the analysis. Previous ggF results are shown in Figure 1.7b.

Multilepton

Di-Higgs events eventually decaying into final states with any number of leptons are a large part of the ATLAS di-Higgs analysis. The di-Higgs decay channel would obviously include $b\bar{b}VV$; however, the fully hadronic nature of the search in this thesis means there is no overlap with the final state.

1.13.1 Future of Di-Higgs Physics

One of the clear goals for the future of di-Higgs physics at the LHC and beyond is the continued constraint on the scalar sector and the Higgs self coupling. As can be seen in Figure 1.8 the constraint on the possible values of the coupling as observed by ATLAS are still quite loose. With the Run 2 full dataset as well as the more than double expected luminosity of Run 3, there is expected to be continued improvements on the limits of di-Higgs detection as well as a tightening of κ_λ constraints. It may be a while until the first di-Higgs event is recorded though, as extrapolations from current results indicate that first measurements of di-Higgs production will require the luminosity gains anticipated from Run 4 [38].

Chapter 2

EXPERIMENTAL SETUP

Buried between 45 m and 170 m under the ground just outside of Geneva, Switzerland at the Conseil Européen pour la Recherche Nucléaire (CERN) is the Large Hadron Collider (LHC). The goal of the LHC is to provide increasing numbers of collision data in order for the detectors there to continue to validate the Standard Model and search for possible signs of new physics. The primary physics program at the LHC is proton proton collisions, though heavy ions are also collided [3]. One of the flagship experiments designed to examine the protons collisions at the LHC is the experiment known as ATLAS (Toroidal LHC ApparatuS).

2.1 Overview of Collider Physics

Before presenting the particulars of the LHC and the detectors operating there, it is first important to briefly discuss the physics that determines the design of colliders. The primary characteristic of any collider is the luminosity. The luminosity is the number of collisions that can be produced per cm^2/s , with the instantaneous luminosity being defined as [3]

$$L = \frac{N_1 N_2 n_b f_{\text{rev}} \gamma}{4\pi \sigma_x \sigma_y} F. \quad (2.1)$$

Here N_1 and N_2 are the number of protons per bunch with n_b being the number of bunches. Bunches are groups of tightly packed protons that have been gathered to increase the probability of collisions to occur within the experiments. f_{rev} is the frequency of revolution of the collider. σ_x and σ_y represent the root mean square of the beam

width in the x and y directions at the collision point [3]. A Lorentz boost factor of γ is applied along with a geometric factor F that accounts of the crossing angle of the two beams at the interaction point. It is particularly useful to talk about colliders in terms of luminosity because it is physics-process independent being only a result of the particular characteristics of the collider itself allowing for the direct comparison of performance across different designs. The total number of events produced by a collider is referred to as the *integrated luminosity*

$$\mathcal{L} = \int L dt \quad , \quad (2.2)$$

and is calculated by integrating Equation 2.1 over time. The integrated luminosity is typically expressed in inverse femtobarns (fb^{-1}) or picobarns (pb^{-1}). The unit of inverse barn is the same units as one over the cross section for a given process $1/\sigma$. This means

$$N = \mathcal{L} \sigma \quad , \quad (2.3)$$

therefore the unit fb^{-1} is describing the number of events per fb expected for a given process at a given scale e.g. femto, pico, etc.

In addition to the luminosity, colliders are also defined by their Center-of-Mass energy. Typically when calculating the COM energy of a particle collision one would sum the four momentum ($p^\mu = (E, \vec{p})$) of the two particles; however, at the LHC this is simplified since the two particles are both protons resulting in equal and opposite momentum. This results in the simple expression

$$\sqrt{s} = \sqrt{(p_1 + p_2)^2} = \sqrt{4E^2} \quad (2.4)$$

where E a single proton energy and s is the COM energy. Unfortunately, since protons are composite particles p_1 and p_2 should not be thought of as simple the momentum two protons. Hadrons, such as the protons collided at the LHC, aren't made up of a set number of constituent quarks but are in fact composed of colliding and interacting quarks and gluons collectively termed *partons*. When protons collide in the LHC it is

in reality the partons that interact and produce the subsequent collision event. The momentum of the two colliding protons is defined by partons' momentum distribution.

2.1.1 Parton Distribution Functions

The Parton Distribution Function (PDF) encapsulates the reality of these collisions by modeling the internal structure of a hadron in term of the partons. The sum of the total parton makeup must equal the momentum of the proton therefore functions are defined that expresses the density of the hadron in terms of the fractional momentum of the partons [39]. These PDF's are formally parameterized by the momentum transfer of the colliding particles. The differential cross section for the desired interaction is acquired through the calculation of the incoming and outgoing particles initial and final state and the total cross section is achieved through integrating over the PDF, parton fractional momenta of both hadrons, as well as all initial and final states of the colliding particles. Different types of PDF's are characterized by the order in which the partons' interactions are considered when integrating performing the integration's.

2.2 CERN

The four main experiments at the CERN are A Toroidal LHC ApparatuS (ATLAS) [41], Compact Muon Solenoid (CMS) [42], A Large Ion Collider Experiment (ALICE) [43], and Large Hadron Collider beauty (LHCb) [44]. Each is located at one of the four crossing points positioned around the ring. There are an additional three experiments that use this collision data, TOTEM [45], LHC Forward (LHCf) [46], and MoEDAL [47] and a forthcoming experiment FASER [48]. Each of these experiments has a specific physics objective. ATLAS and CMS are both general purpose detectors designed to study a large range of physics phenomena. LHCf is designed to simulate cosmic rays in lab conditions by measuring particles boosted forward by the collisions occurring at the ATLAS experiment. The physics of the protons themselves is examined through precise measurements carried out by the TOTEM experiment. In much the same way LHCb is dedicated to the study of the b -quark. MoEDAL is searching for magnetic monopoles and FASER will

CERN's Accelerator Complex

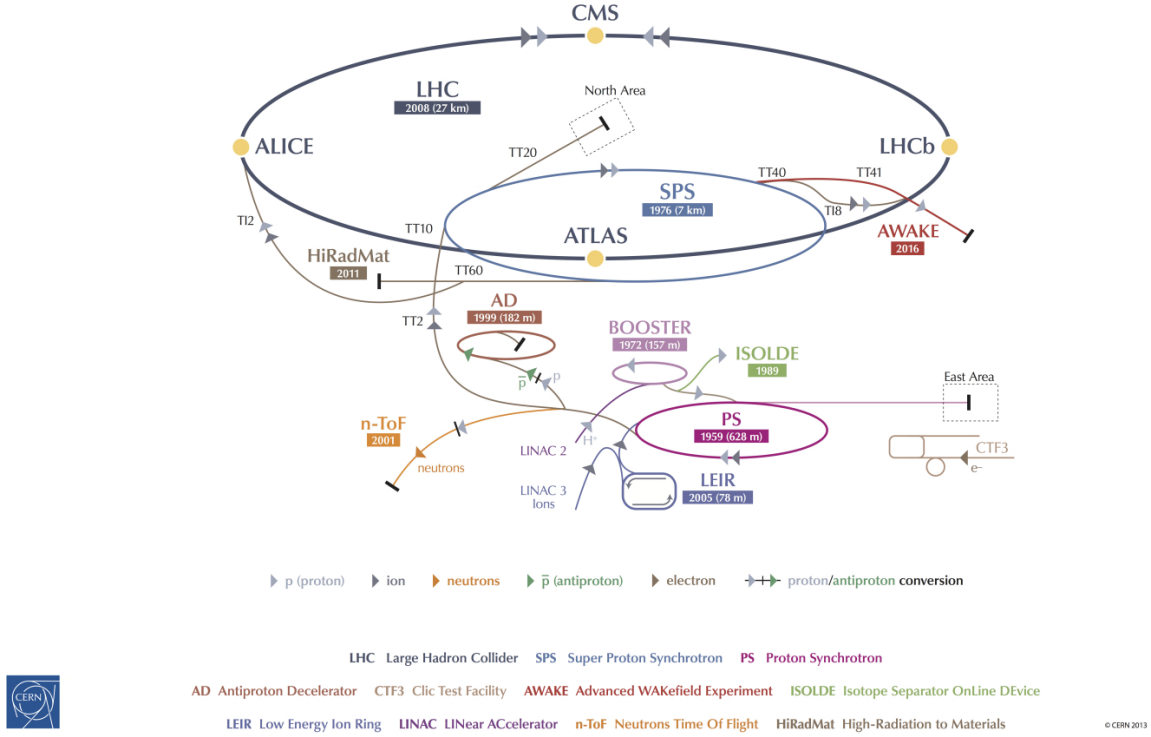


Figure 2.1: An overview of the CERN complex and subsystems, the LHC ring, and detectors and experiments [40].

search for new light and weakly interacting particles. Finally, unlike the others, ALICE is designed to examine the collisions of heavy ions like Pb (lead) to study the physics of strongly interacting matter at extreme energy densities. The CERN complex as well as all detectors and sub-systems can be seen in Figure 2.1.

2.3 The Large Hadron Collider

The LHC is designed to deliver an instantaneous luminosity of $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$ with a COM energy for the proton-proton collisions of $\sqrt{s} = 14 \text{ TeV}$. At the time of this thesis, this COM energy has not been achieved. There have been two physics operational runs at the LHC as of 2021: Run I (2010-2013) and Run II (2015-2018). Run 1 was at 7 TeV and 8 TeV collision energies and Run 2 had 13 TeV. There have already been two major

upgrade periods at the LHC. Long shutdown (LS1) from 2013-2015 took place after Run 1 to allow for upgrades that would bring each beam up to an energy of 7 TeV; however, due to issues with the superconducting magnets, Run 2 was run at a beam energy of 6.5 TeV. The second major upgrade period LS2 began in June 2018 and continued until the Fall of 2021. Run 3 is scheduled to begin after LS2 in the first half of 2022. Following the conclusion of Run 3, a new era for the LHC known as the High Luminosity LHC will begin. This era will start with the third long shutdown LS3 where upgrades to the equipment and experiments will take place that will allow the HL-LHC to operate with luminosities up to 7.5×10^{34} [49].

Before protons can be collided at the LHC to provide the data that is used in physics analyses they must be accelerated. All protons used at the LHC get their humble beginnings in a small container of H_2 gas tucked away in a corner in the caverns of CERN near the Linear accelerator (Linac). The first step for these protons is to be first accelerated to 450 GeV. In order to accelerate the hydrogen atoms they are ionized. The Linac4 was designed to accelerate H^- ions, hydrogen atoms with an additional electron, to 160 MeV [50]. The two additional electrons are stripped from the proton when they are injected into the Proton Synchrotron Booster (PSB) that in turn accelerates them to 1.4 GeV. After the PSB, the Proton Synchrotron (PS) accelerates them to 25 GeV and from there the protons are accelerated one final time in the Super Proton Synchrotron (SPS) before being injected into the two beam pipes of the LHC. The protons in the beams of these two pipes circulate in opposite directions, and in order to reach the design COM energy, they must be accelerated to 7 TeV each. This is accomplished using sixteen 400 MHz Radiofrequency (RF) superconducting cavity systems. The field in these RF cavities oscillate, accelerating or reducing the energy of the protons until they reach the desired energy.

This oscillation additionally performs something of a sorting, resulting in the protons being collected into bunches. In order to keep these protons moving in a circle as they are accelerated around the LHC powerful dipole magnets continuously bend the path of the protons. Since the goal is to maximize the number of collisions quadrupole magnets throughout the ring tighten the beam since the denser the particles, the more likely

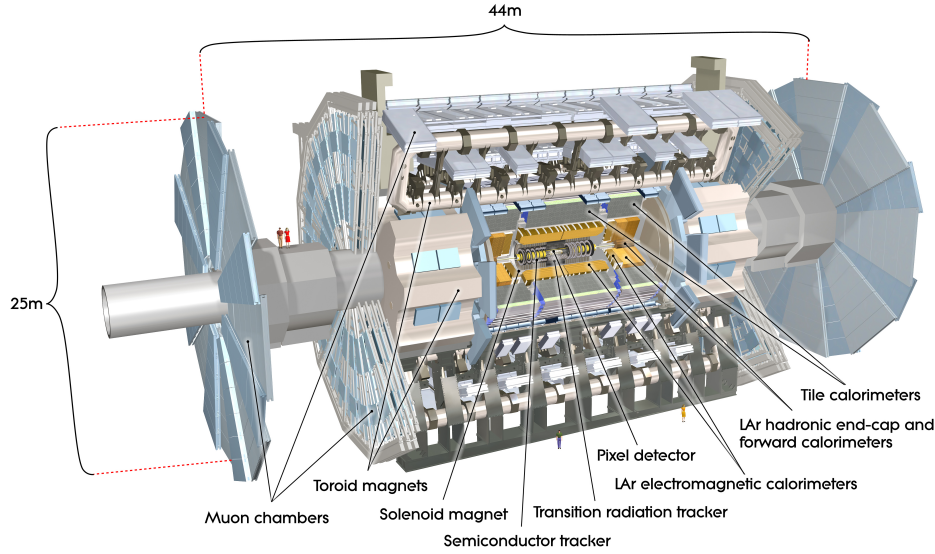


Figure 2.2: A overview cutaway of the ATLAS experiment showing the relative location of each of the sub-components as well as the global size [51].

they are to collide. The last component of the LHC before the beams cross inside a detector such as ATLAS is the so called inner triplet. The inner triplet is a grouping of three quadrupole magnets whose purpose is to squeeze the proton bunches as closely together as possible in the plane transverse to their motion in order to minimize $\sigma_{x/y}$ in Equation 2.1. There are 8 of these inner triplets located on either side of the four main detectors at the LHC. There can be a maximum total number of 2556 proton bunches per beam each containing 1.15×10^{11} protons. These bunches are crossed to allow for crossing to occur at a rate of one every 25 ns. Despite the number of protons per bunch being there is still only ~ 20 collisions per crossing. At a frequency of 25 ns though, there are still 800 million collisions per second [3].

2.4 A Toroidal LHC ApparatuS

The ATLAS detector has an effective detection region that nearly covers the entire solid angle around the collisions that occur in the center of the detector. The ATLAS detector is comprised of three distinct elements, the ID, calorimeter system, and the muon spectrometer [41] as well as the four forward detectors.

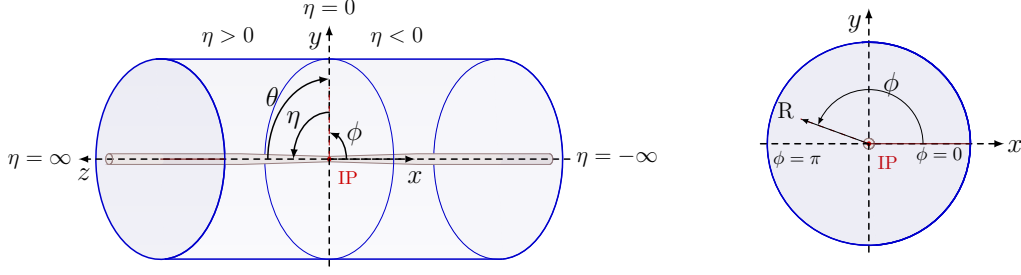


Figure 2.3: This cartoon shows the difference coordinates used by the ATLAS experiment, from the side on the left and a straight on view on the right. Both representations are centered on the Interaction Point (IP) [52].

2.4.1 Coordinate System

The ATLAS detector is designed to have the collisions take place at the center of the detector and the coordinate system used at ATLAS reflects this orientation. In Figure 2.3 the cylindrical coordinates for the ATLAS detector are illustrated. R gives the distance from the origin to a location in the $x - y$ plane, θ then is the angle from the central axis, and ϕ is the azimuthal angle. It is advantageous to use these coordinates to defined Lorentz invariant variables, or variables that have the same value in the lab frame as they do in the COM frame along the z -axis. First, the azimuthal angle can be defined as

$$\phi = \tan^{-1} \frac{p_y}{p_x} , \quad (2.5)$$

where p_x and p_y are the x and y component of the momentum. Next, rapidity, which is defined as

$$y = \frac{1}{2} \ln \frac{E + p_z}{E - p_z} . \quad (2.6)$$

While the the rapidity is not Lorentz invariant, difference in rapidity are. When the mass of the particle is negligible with respect to it's energy, as if often true for particles at the LHC, the rapidity reduces to the pseudorapidity

$$\eta = - \ln \tan \frac{\theta}{2} , \quad (2.7)$$

which describes an angle relative to the beam axis shown in Figure 2.3. The distance R

separating two particles can then be defined as

$$R = \sqrt{(\Delta y)^2 + (\Delta \eta)^2}. \quad (2.8)$$

The momentum of physics objects used in analyses is expressed in terms of the transverse momentum,

$$p_T = p |\sin \theta| \quad (2.9)$$

or the component of the momentum that is perpendicular to the beam axis.

2.4.2 The inner detector

The inner-most component of the ATLAS detector is the Inner Detector (ID). The highly energetic particles moving through this region of the detector are not absorbed there for the goal is instead to precisely measure the path of charged particles described by an object known as a track. Depictions of tracks originating from various fundamental particles can be seen in Figure 2.7. The precise tracking of particles is required for momentum measurements, particle and vertex identification, fundamental physics object reconstruction, and ultimately event reconstruction. The range of the ID as a whole is $|\eta| < 2.5$. The ID has three sub-components: the Pixel Detector (PD), the semi-conductor tracker (SCT), and Transition Radiation Tracker (TRT). These three sub-components deliver precise measurements and tracking over a broad range of $R - \phi$ and z .

The Silicon Pixel Detector plays a valuable role in the success of the ATLAS experiment. By providing tracking information so close to the interaction point, accurate particle recognition can be achieved via lifetime determination. The PD is the primary component responsible for the ID's ability to find secondary vertices [53]. As is true for all major components of the ATLAS detector the PD has two different geometries, a barrel tracker and two end-cap trackers. The barrel tracker is made of three layers, a B-layer and the layers 1 and 2. The sensors in the barrel tracker have a minimum size in $(R - \phi) \times z$ of $50 \times 400 \mu\text{m}^2$ per pixel. The PD can achieve accuracies of $10 \mu\text{m}$ in $R - \phi$ and $155 \mu\text{m}$ in z . There are six PD end cap layers, three on either side of the interaction

points. The end caps have the same pixel size as the barrel with accuracies of $10 \mu\text{m}$ in $R - \phi$ and $155 \mu\text{m}$ in R . The end-caps are arranged in rings around the beam axis. In total the PD has approximately 80.4M readout channels throughout the 9 layers. The pixel detector barrel and end-caps in Figure 2.4. The silicon pixel detector works through the creation of electron-hole pairs when a particle pass through them and these pairs are read out to determine where the particle pass through each layer of the PD.

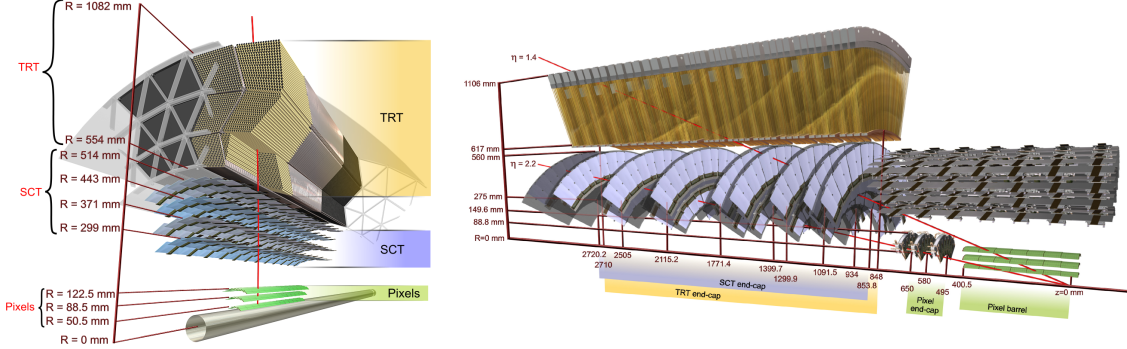


Figure 2.4: The ATLAS Inner Detector, on the left the barrel section of the three ID subsystems are shown and on the right the end-caps [54].

The next component, the semi-conductor tracker (SCT), works in a similar way to the PD but is made up of silicon strip instead of pixels. The SCT has barrel and end-cap modules as well. The barrel of the SCT has four layers comprised of 12 cm strips with a pitch of $80 \mu\text{m}$ and has an accuracy in $(R - \phi)$ of $17 \mu\text{m}$ and $580 \mu\text{m}$ in z . There are three layers per end cap for the SCT with strips half the length of the barrel at 6 cm. The accuracy of the end cap layers is the same in $(R - \phi)$ and $580 \mu\text{m}$ in R . The SCT provides four measurements per track in the intermediate radial range [55] through its 6.3M readout channels.

The last and outermost component of the ID is the Transition Radiation Tracker which has an effective range of $|\eta| < 2.0$. The TRT is a gaseous detector comprised of straw tubes 4 mm in diameter and with a maximum length of 150 cm. The straws' drift tubes contain gold-plated tungsten wires. The TRT only used a predominantly xenon (70%) gas mixture but due to leaks, sections have been replaced with an argon mixture. The use of Xenon is necessary for the transition radiation particle identification [56]. Again the TRT is separated into the barrel and end-cap regions. The barrel has 50000

straws running parallel to the beam pipe and the end-caps contain 320000 radial straws. The nature of the straw detector means that the TRT only provides measurements in $R-\phi$. Both the end-cap and barrel regions have a granularity of $130\text{ }\mu\text{m}$ per straw. The tubes detect particles through the ionization that occurs during a particle's crossing. The size of the TRT active area allows for many measurements throughout charged particle's path improving measurements of momentum. Particles that pass through the straws of the TRT ionize the gas within the tube resulting in a current along the wires leading to a detection. Electrons can be identified by their increase in ionization deposited in each straw in comparison to non-electron particles.

2.4.3 The Liquid Argon and Tile Calorimeters

The ATLAS calorimetry system has two distinct subsystems, the Liquid Argon calorimeter and tile hadronic calorimeter, with $|\eta| < 4.9$ and full ϕ coverage around the beam axis [41]. Electrons and photons are measured with high precision by the Electromagnetic calorimeter. Jets and Missing Transverse Energy (MET) of collision events are reconstructed from measurements taken throughout the whole of the calorimetry system. An addition design consideration that is important to the calorimeter system is ensuring containment of EM and hadronic showers in order to limit punchthrough into the muon system. The base goal of a calorimeter is total capture of the particles under consideration, that is, hadrons as well as photons and electrons produced in the collisions within the ATLAS detector are absorbed by the calorimetry system. This relative travel distance through the detector for different particles is illustrated in Figure 2.7.

As with the sub-components of the ID, the calorimetry system is made up of barrel and end-caps sections. The EM calorimeter has two half-barrel sections separated at the center of the detector and two end cap wheels. The barrel sections are effective out to $\eta < 1.475$ and the end-caps extend from $1.375 < \eta < 3.2$. Both sections are made up of lead and stainless-steel absorbers with gaps in between that house "accordion-shaped" electrodes [57]. The peculiar shape of the electrodes ensures full coverage in ϕ . The electrodes consist of copper layers separated by plastic sheets to insulate them. The end cap wheels sit one on either side of the barrel halves and themselves are made up

of an inner and outer coaxial wheel. The barrel halves have a combined total number of 109696 calorimetry cells and the end caps have 62208. The EM calorimeter barrel section has a granularity in $\Delta\eta$ ranging from 0.003125 to 0.075 and in $\Delta\phi$ from 0.025 to 0.1 rad. The end-caps have similar ranges, 0.03125 to 0.1 and 0.025 to 0.1 rad in $\Delta\eta$ and $\Delta\phi$ respectively. Both the barrel and endcaps are made up of four layers including a presampler layer. The presamplers cover a small range in η , $\eta < 1.8$, and recover and correct for energy that is not accounted for by the EM calorimeter [58].

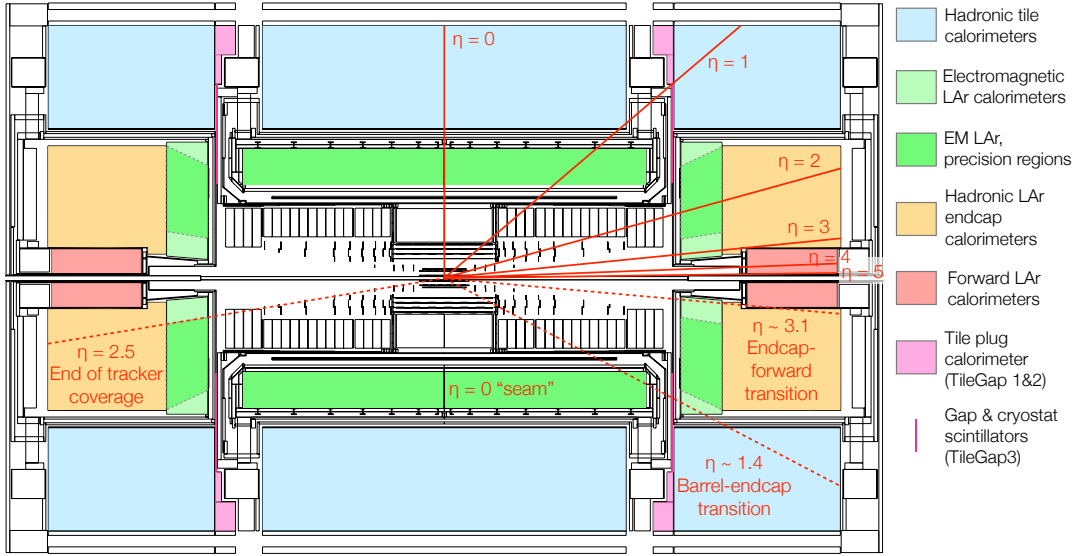


Figure 2.5: Cut-away of the calorimeter system showing the Liquid Argon and tile calorimeters [59].

Outside the EM calorimeter end caps are the Liquid Argon Hadronic End Cap Calorimeter end caps both consisting of two independent wheels ranging from $1.5 < |\eta| < 3.2$. The absorbers of the wheels are 25 mm copper plates near to the interaction point and 50 mm plates further way. Liquid argon gaps in between these plates provide the active medium for the sampling. The granularity in $\Delta\eta \times \Delta\phi$ of the HEC is 0.1×0.1 for $1.5 < |\eta| < 2.5$ and 0.2×0.2 for $2.5 < |\eta| < 3.2$.

Finally the FCal finishes out the LAr system effective in the range $3.1 < |\eta| < 4.9$. The FCal has three modules, one made of copper designed for EM measurements and two made of tungsten designed for hadronic measurements. The modules consist of

longitudinal channels parallel to the beam axis that contain concentric rods and tubes with a LAr gap between themselves and the metal of the modules acts as the sensitive medium. The minimum granularity of the FCal is 5.4×4.7 .

Directly outside the LAr system is the tile calorimeter. The tile calorimeter extends out to $|\eta| < 1.7$. It is made up of three sections, the central of which has a range from $|\eta| < 1.0$ and two outer sections from $0.8 < |\eta| < 1.7$. The tile calorimeter consists of layers of tiles. These tiles are made up of steel and plastic and measure the energy deposition of charged and neutral hadrons. The two sides of the tiles have wavelength shifting fibers that direct the photons from the scintillator into photomultiplier tubes. The end-caps of the tile calorimeter are located behind the LAr end-caps and also have inner and outer wheels like the LAr [57]. The physics requirements of the the HEC and tile calorimeters are different than the higher precision LAr calorimeters. The tile calorimeter layers' granularity of 0.1×0.1 and 0.2×0.1 in $\Delta\eta \times \Delta\phi$ for the inner and outermost layer respectively provides measurements sufficient for jet and MET reconstruction. The calorimeter system in its entirety can be seen in Figure 2.5.

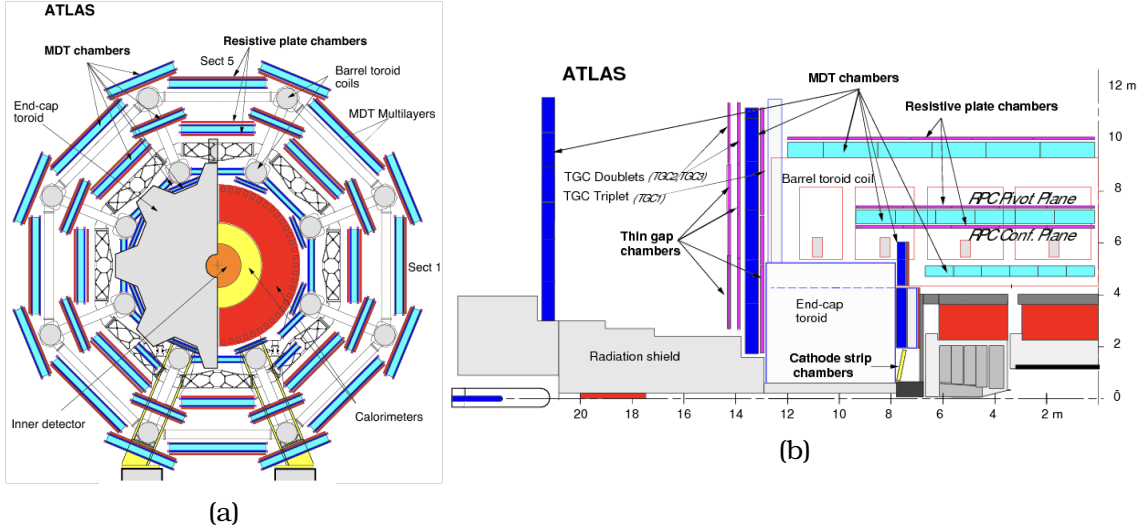


Figure 2.6: (a) Cut-away view of the muon system in the x-y plane. (b) Cut-away view of the muon system in the z-y plane [60].

2.4.4 Muon Spectrometer

The final component of the ATLAS detector that shrouds calorimeter system is the Muon Spectrometer (MS) shown in Figure 2.6. The MS is made up of precision tracking concentric cylindrical shell chambers in the barrel region at radii of 5 m, 7.5 m, and 10 m. In the two end-cap regions the muon chambers are made up of large wheels at $|z| \simeq 7.5$ m, 10.8 m, 14 m, and 21.5 m all perpendicular to the z-axis [61]. The powerful toroidal magnetic field produced by the magnetic system covered in the next Section 2.4.5 deflects muons exiting the calorimetry system and the MS tracks these particles providing precision measurements of their momentum [62]. The MS is made up of four different chambers that serve either precision tracking or triggering functionality. The two high precision chambers are the Monitored Drift Tubes (MDT) and the Cathode Strip Chambers (CSC). The MDT has a range of $|\eta| < 2.7$ and is comprised of 1088 chambers per layer with 1150 in the inner most layer each with a single sense wire delivering a resolution of $35\mu\text{m}$ in z . The other tracking chamber, the CSC, operates from in $2.0 < |\eta| < 2.7$ with 32 chambers containing multiple wires each and a resolution of $40 \times 5 \times 7$ in $\mu\text{m} \times \phi \times \text{ns}$. The two trigger systems work together to cover the range $|\eta| < 2.4$ with the Resistive Plate Chambers (RPC) operating at $|\eta| < 1.05$ and the Thin Gap Chambers (TGC) covering the remainder of the range. Both trigger chambers primarily provide bunch crossing identification and tracking information at a high rate for use by the Level-1 trigger and in second coordinate and transverse momentum measurements.

2.4.5 Magnet System

The magnet system of the ATLAS detector is designed to generate a magnetic field to allow for the precise measurement of the momentum of charged particles [63]. The system consists of four super conducting magnets: the central solenoid, the barrel toroid, and two end-cap toroids. The magnetic field strength of the components of the system differ with the central solenoid having a peak field strength of 2.6 T and the barrel and end-cap toroids having 3.9 T and 4.1 T strengths respectively [64]. The central solenoid is aligned on the beam axis in order to provide an axial magnetic field for the ID.

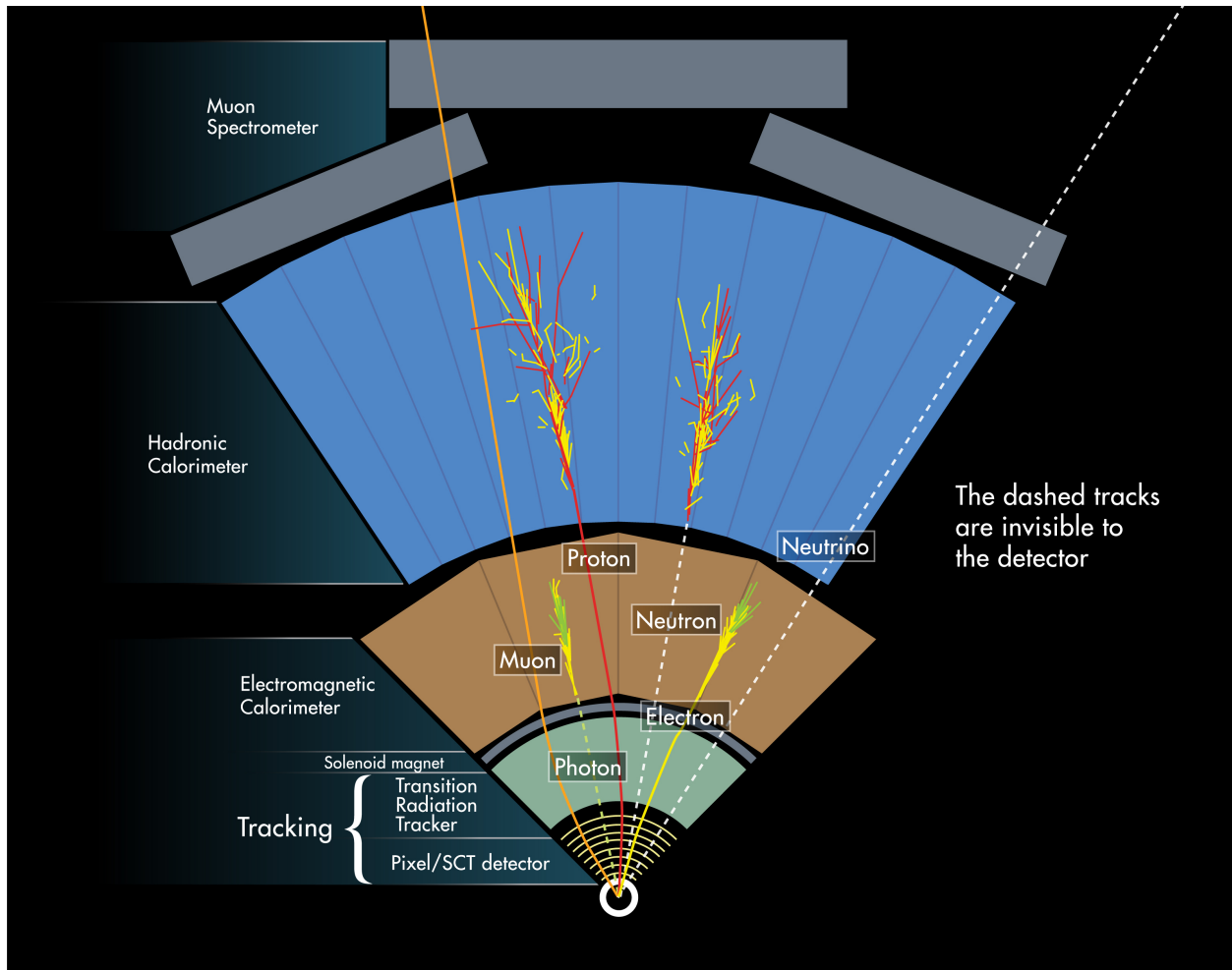


Figure 2.7: Depiction of the path that different particles take through the ATLAS detector [65].

2.4.6 The ATLAS Forward Detector project

The final parts of the ATLAS detector are the suite of additional detectors located along the beam pipe with a varying array of physics goals. The four components are mirrored pairs with detectors on either side of the ATLAS experiment.

LUCID The Luminosity measurement using Cerenkov Integrating Detector (LUCID) is a dedicated luminosity measurement experiment located 17 m from the center of ATLAS. The purpose of LUCID is to provide relative, independent measurements of the luminosity for ATLAS [66]. The detector uses measurements of Cerenkov radiation produced by particles in the ATLAS detector to measure the luminosity.

ZDC The Zero Degree Calorimeter (ZDC) pair is located where the beam pipes of the LHC merge on either side of the ATLAS detector around 140 m along the beampipe. The purpose of the ZDC is to aid in the detection and measurement of neutral particles, particularly those at with trajectories very near to the beam axis as the ADC has a functional range of $|\eta| > 8.3$ [67].

ALFA The Absolute Luminosity For ATLAS (ALFA) detectors are located 240 m away from ATLAS and like the LUCID detectors the purpose of ALFA is to provide a measurement of the absolute luminosity [68].

AFP The ATLAS Forward Proton (AFP) project is to measure characteristics of extremely forward protons. This offers new capabilities for the ATLAS collaborations to study phenomena associated with protons that remain intact after crossing [69].

2.5 Trigger

While the ATLAS detector is designed to detect billions of proton-proton collisions per second, the vast majority of these events contain mundane processes in which physicists are not interested. At a size of ~ 1.5 MB per event there is no feasible way to store all this data for later analysis anyway. To solve this, ATLAS uses hardware and software triggers to filter the data and decide which events to keep. The first level of filtering is

accomplished with a system of custom hardware electronics called the Level-1 hardware trigger. The Level 1 trigger uses information from the calorimeter system and the MS to decide whether or not to keep an event [70]. The events that are kept are passed on to the high level trigger (HLT). The HLT is a software based trigger running on a computing farm where analysis are run on either full event readout data, or by looking at particular isolated regions of the detector.

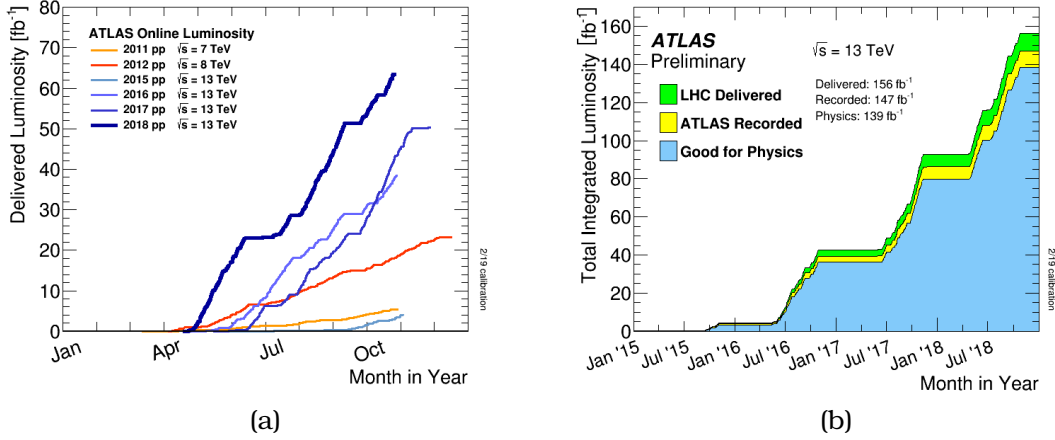


Figure 2.8: (a) ATLAS online luminosity for each data taking period through 2018 [71]. (b) Total integrated luminosity at ATLAS categorized by the total amount delivered to ATLAS, total recorded, and the portion of the luminosity that is available for physics use [72].

2.6 Performance

By the end of Run II the LHC delivered 156 fb^{-1} of proton-proton collisions to the ATLAS experiment. Of these, the ATLAS detector recorded 147 fb^{-1} . This is an efficiency of 94%. Of the events recorded, again 94% were deemed good for physics usage. Figure 2.8b shows the delivered luminosity versus recorded for ATLAS. Since the increase of collision energy in 2015, ATLAS has continued to record increasing luminosities every year. A consequence of this increase in luminosity is an increase in pileup ($\langle\mu\rangle$), defined as the average number of interactions per bunch crossing. There are two primary types of pile-up, in- and out-of-time pileup [73]. In-time pileup refers to additional proton-proton collisions that occur in addition to the collision of interest. Out-of-time pile-up refers to collisions that occur in bunch crossings immediately before or after the bunch crossing

of the collisions of interest. This increase leads to more complex events since the more particles that are involved lead to more interactions. This increase in interactions has consequences not only for the trigger system to maintain efficiency but also for offline analyses trying to correctly reconstruct events. Both hardware and software approaches are active areas of research looking for ways to reduce the impact of pileup. The relative ($\langle\mu\rangle$) per year at the LHC can be seen in Figure 2.9.

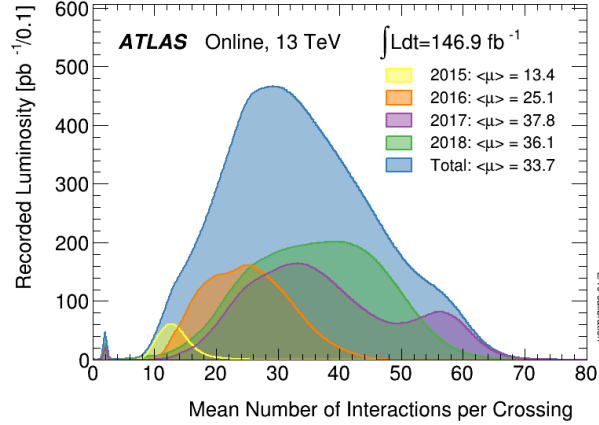


Figure 2.9: The pileup of each data taking period as well as the total for the full data range [72].

2.7 Future of ATLAS

Throughout its operating period the LHC hasn't always remained in operation. On top of unscheduled hardware issues that required shutdowns early on, the LHC has periodic and end-of-year technical stops for maintenance and scheduled periods of long shutdowns for enhancements and upgrades to be implemented. A timeline of this plan can be seen in Figure 2.10. The most recent of these long shutdowns lasted from 2018 to 2021. This shutdown was able to bring the collision energy up to 13.5 TeV. With this improvement the LHC plans to collect 300 fb^{-1} in Run 3 which will nearly double the current integrated luminosity. During this shutdown ATLAS also underwent its Phase I upgrades. The purpose of these upgrades were to prepare the detector for luminosities of $(2 - 3) \times 10^{34}$ which will, among other things, increase pileup and data rates. After Run 3, the LHC will again shut down and enter the next real phase of its operation,

known as the High Luminosity LHC (HL-LHC). This improvement will aim to increase the luminosity by a factor of 7.5 above the baseline LHC luminosity greatly expanding the possibilities of BSM physics searches such as the ones considered in this work.

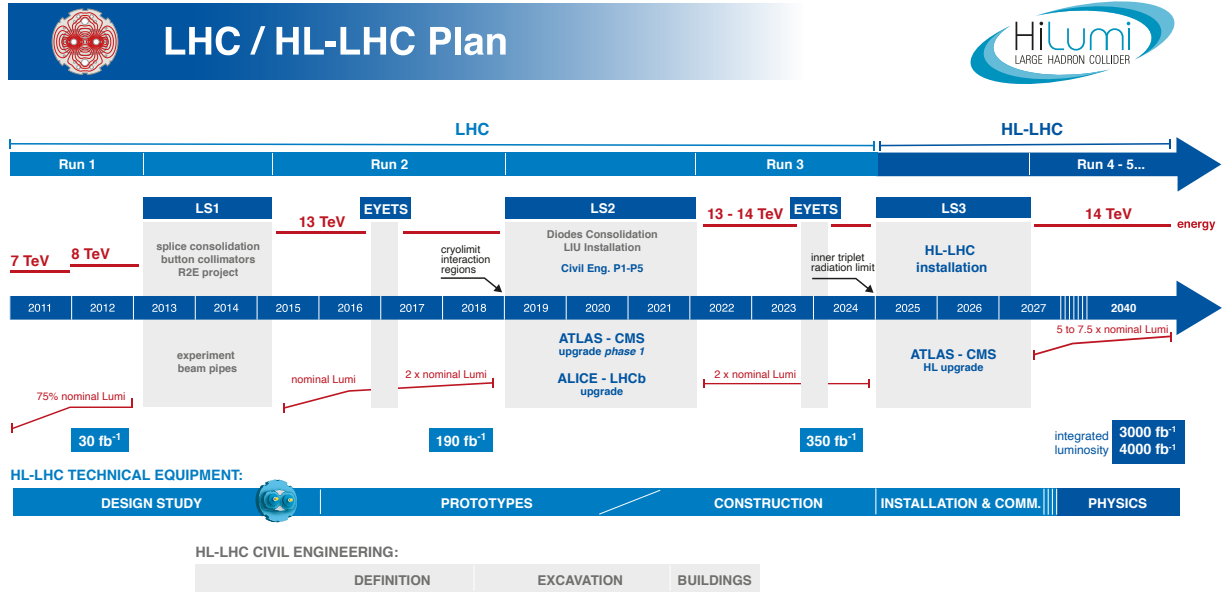


Figure 2.10: Long-term upgrade and run schedule for ATLAS through HL-LHC as well as the anticipated run luminosity and E_{COM} [74]

2.8 Simulating Collision Events

The connection between the theory of the SM laid out in Chapter 1 and the data collected by the ATLAS experiment hinges on the ability to accurately predict the rate and properties of physics events in order to use them for deeper analysis.

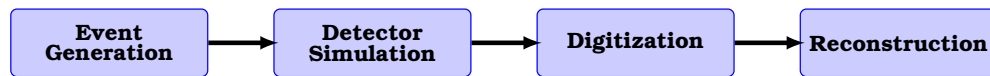


Figure 2.11: A brief flowchart of the steps taken to produce simulations of ATLAS collision data.

2.8.1 Event Generation

The process for generating simulated events begins with generating what is known as event-level MC. This is accomplished by using a generator to calculate the matrix element. There are a number of different MC generators that exist, and while the technical details of how they are used differs, in practice they do very much the same thing. The event-level data for this thesis were created using MadGraph [75], POWHEG [76], Pythia [77], and Sherpa [78]. Once the matrix element is calculated the hard scatter events are generated using a chosen PDF. These represent the initial and final state particles for the desired event. This event is then evolved in a similar way to what happens within the ATLAS detector, carrying out the fragmentation and hadronization for particles as they travel through the detector components. These secondary interactions can not be calculated in the same way as the initial interaction, therefore models are created with tune-able parameters that allow the users to fit the simulation to known data processes in order to ensure the simulation is properly recreating actual events. Many of the above generators are capable of simulating the parton shower as well. The general purpose generator Herwig7 [79] was used in some cases only for the parton showering.

2.8.2 Detector Simulation

The simulated MC event, while representative of the physics that takes place within collision, does not reflect the real signal that is produced by the ATLAS detector. The simulation of the detector is performed using the GEometry ANd Tracking (Geant4) toolkit. Geant4 simulates an event by propagating the particles from the event through the detector and evaluating the interactions, energy depositions, and momentum transfers that occur as the particles interact with the different materials [80]. The software also respects the particulars of the detector's geometry. The underlying physics of the interactions between the particles and the detector aren't universal but are defined by different physics models. The nominal simulation model for ATLAS is the Fritiof string model (FTF) [81] and Bertini Nucleon-nucleon Scattering model (BERT) [82] for hadronic interactions [83]. A different set of models and detector geometries are used in the un-

certainty estimation in Chapter 5 and summarized in Appendix C.

2.8.3 Digitization

When collision data is collected, the detector electronics interact and respond to the physical objects producing electric signals known as digitization. This process must be recreated in MC since there is no actual electronic detector responsible for the simulated data. Digitization is accomplished by converting the simulated energy deposits from the Geant4 toolkit into electrical responses. Noise is introduced to better reflect the hardware setup and finally the data readout structure of the ATLAS electronics is recreated.

2.8.4 Reconstruction

The final step in generating simulated events is to reconstruct the physics objects using the methods described in the following Chapter 3. Once reconstructed this simulated data is used as truth information to correctly relate the collision reconstruction objects to the particle level objects described by the MC. It's also used to calibrate and refined the methods of object reconstruction to derived systematic uncertainties in physics these processes. Additionally simulated data is crucial to successfully conducting physics analyses since it allows for the prototyping and development of new methods and analysis tools. MC is also used to estimate the efficiencies and rates of different physics processes in data.

Chapter 3

OBJECT RECONSTRUCTION

The collisions that occur in center of the ATLAS experiment produce showers of outgoing particles. In order to understand and study the events that produced them, the particles have to be correctly reconstructed and identified from the detector measurements. The objects that are used by ATLAS analyses are called reconstructed objects because they're reconstructed from the different components of the detector. These objects are jets, muons, photons, electrons, taus, and Missing Transverse Energy. By far the most complicated of these objects are MET and jets. Unlike the charged leptons and photons which are fundamental particles that are directly detected, jets and MET are composite objects reconstructed from many different detections to create a single reconstructed object. In this chapter the method for how each object is reconstructed by the ATLAS collaboration is presented. The Track Assisted Reclustered algorithm (TAR) is thoroughly detailed as well as the JER and JES calibrations that are critical to accurate jet measurements and uncertainties. In addition, the methods for identifying muons, electrons and photons, tau particles, and MET are covered.

3.1 Tracks

The first of two fundamental objects in the ATLAS reconstruction process are tracks. Tracks are charged particles that travel through the detector and that have multiple hits throughout the Inner Detector and are used in the reconstruction of all physics objects. The method for reconstructing these tracks has three steps. First, candidates for track reconstruction have to be identified and once they have been an ambiguity

solver is used to score the tracks in order to determine the best candidate [84]. This *track score* is the most important component in track reconstruction because its definition wholly determines the success of the track reconstruction. Finally, the hits on the track candidate are used in a fit to estimate parameters associated with the track like its transverse momentum.

The iterative process to identify track candidates begins with clustering of the raw measurements within the PD and SCT. Measurements are clustered by energy deposits that pass a given threshold as well as share a common edge or corner in the detector cells [85]. These clusters are used to create so called *space-points*. These space-points correspond to each point where a charged particle crossed the ID. Tracks are then seeded to a combinatorial Kalman filter [86]. Seed tracks are comprised of a minimum of three space-points in order to maximize the possible combination while insuring a legitimate first approximation for the track. An estimation of the trajectory of the track is used to further constrain the seed tracks by requiring that additional space-points correspond with the estimated path. The remaining space-points that are compatible with the track's estimated trajectory are filtered and used to define seed tracks [87]. The filter creates candidates from each seed track with the possibility of multiple candidate tracks if there exists compatible space-points within the same detector layer. Since all possible combinations of seed tracks and remaining space-points are made, there is a possibility of overlapping or incorrectly assigned space-points. This is sorted out by the ambiguity solver.

Each candidate track is assigned a track score and the ambiguity solver assesses each track according to this score in descending order. By prioritizing tracks with a high score the best match is ensured between the reconstructed track and the original trajectory of the charged particle. The track score is based on a combination of simple measurements that represent track quality. The track score is increased if calorimetry clusters exist that have been assigned to the track with the magnitude of the increase coming from preassigned weights. Holes, intersection of a track trajectory with detector planes that don't contain a matching cluster, reduce the score. A track fit is applied and χ^2 is evaluated, with poor track fits reducing the score. Lastly, the momentum of

the track is considered, favoring high- p_T tracks that tend to higher accuracy of correctly assigned clusters. All of these considerations lead to the track score of a given candidate. The ambiguity solver addresses each track, starting with tracks that have shared clusters. Each track candidate must have fewer than three shared clusters and each of the shared cluster can only be shared by at most two tracks. If a track fails these requirements, it is removed and all tracks that shared the cluster have their scores recalculated and are examined again. Track candidates fail the solver if they don't meet the minimum requirements.

After the ambiguity solver, neural networks are used to increase the reconstruction performance with networks trained to identify merged pixel clusters. This increases the precision of the cluster assignment. The last step is to perform a high-resolution fit of the tracks. This is done using networks [88] trained to estimate the trajectory of the track and compared it to the candidate track resulting in more precise tracking parameters.

Track candidates are defined by five kinematical variables: impact parameter $d0$, impact parameter $z0$, η , ϕ and the track's curvature. $d0$ is the distance of closest approach between the track and the transverse plane. $z0$ is the z -coordinate of the track at the point-of-closest approach to the global z -axis. η and ϕ are the η and ϕ values of the particle associated to the track. Finally the curvature of the track is related to charge over transverse momentum, q/p_T . The relationship of these parameters to a given track and the coordinate system is diagrammed in Figure 3.1.

Once the tracks have been successfully reconstructed, the vertices can be as well. The approach to vertex reconstruction is carried out in a very similar way to that of track reconstruction. Here there are two stages, vertex finding and vertex fitting. In the first stage the seed position for the vertex is calculated. This is accomplished using the z -coordinate of the seed track along with the x and y coordinates defined by the beam. An iterative process is then applied, minimizing the χ^2 fit to the vertex of the track, removing the lowest and starting over until the vertex position is determined. This process is repeated on all tracks [90].

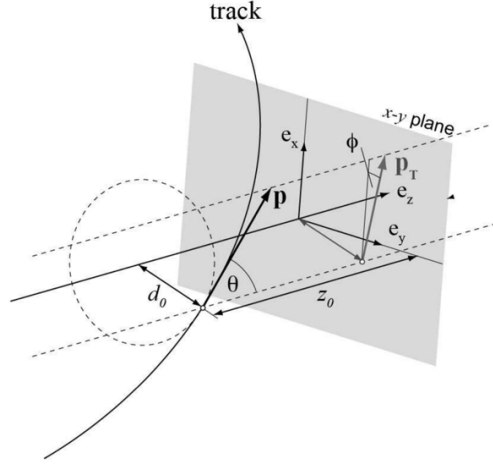


Figure 3.1: This diagrammatic representation of a track shows the relationship the track has with the coordinate system as well as its characteristic parameters d_0 and z_0 [89].

3.2 Topological Calorimeter Clusters

The second fundamental objects topological clusters, or topoclusters, are constructed from the energy deposited into the calorimetry cells of the ATLAS detector. These clusters are used as inputs in the reconstruction of jets, electrons and photons, and taus. The specific algorithms for reconstructing these physics objects are in Sections 3.3, 3.5, and 3.6 but before the cells' measured energy can be used in object reconstruction they must first be processed into collections of cells for improved performance. Using groupings of clusters instead of the individual cells themselves works to increase the quality of the desired signal and reduce unwanted noise. The cells are collected into three dimensional topoclusters. The clustering algorithm makes use of the cell signal significance [91]

$$\varsigma_{cell}^{EM} = \frac{E_{cell}^{EM}}{\sigma_{cell}^{EM}} \quad (3.1)$$

as the primary observable. This variable described the cell signal as a ratio to the cell's average noise. The formation of topoclusters is defined by two stages, collecting and then splitting.

The first step in the collection state is to define the seed cells for the clustering shown in Figure 3.2. These are cell that pass a predefined significance threshold. Seed cells are then ordered in terms of cell significance ς_{cell}^{EM} and the neighboring cells are collected

together with the seed cell to create proto-clusters. Proto-clusters are compared and if the neighboring cells of a pair of clusters pass a growth threshold, the proto-clusters are merged.

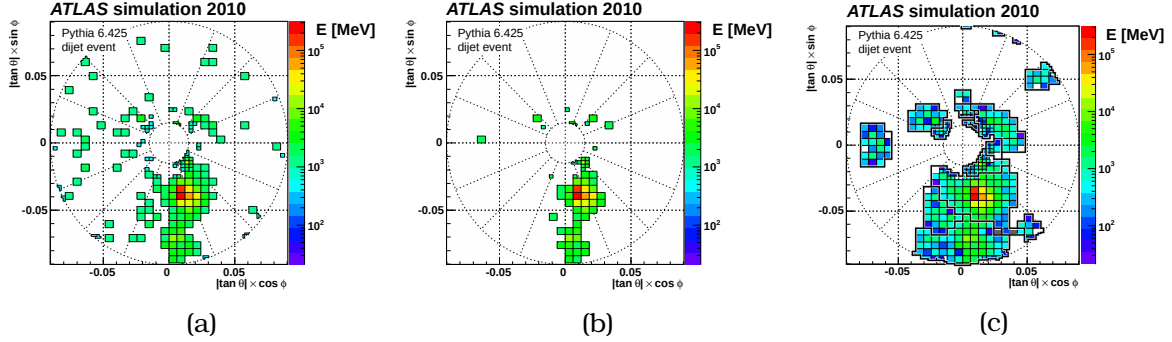


Figure 3.2: ATLAS simulation of the steps taken in topocluster formation for energy deposits from a dijet event into the FCal. Seed cells passing the initial significance threshold are shown in (a). Seeds that then pass the neighbor cell threshold that are shown in (b). Finally, (c) shows the final topoclusters outlining their associated cells [91].

The next stage is to evaluate the proto-clusters and to split them when necessary. The algorithm described above can result in clusters of cells that are too large to provide good measurements for the particles with which their energy is associated. Proto-clusters that are observed to have more than one local maximum above an energy threshold of 500 MeV are split. The proto-clusters that remain after the splitting stage are the topo-clusters used in object reconstruction and an example is shown in Figure 3.2c.

3.3 Jets

Quarks and gluons are the dominant products coming out of hard scatterings in the ATLAS experiment. These strongly interacting particles can not be directly measured, in contrast with leptons which can. Fortunately there are characteristics of the strong force that allow these particles to be observed, however indirectly. At very small distances the strong force coupling is weak which allows this particles to be treated as quasi-free partons. These partons cascade from the initial interaction point and quickly decay into hadrons at distances on the order of a fermi (10^{-13} m). It's not really a decay however, but more the spontaneous creation of quark pairs since lone quarks can't exist. This means that long before any of these quarks or gluons could be observed, additional quarks are

created combining with and forming composite hadrons. This process from quasi-free partons to hadrons is known as hadronization. These hadrons decay and interact with the detector resulting in a shower of particles. Long-lived hadrons continue to travel and decay until eventually they are detected through energy depositions in the cells of the hadronic calorimeter system. The detection of these initial particles in the decay shower necessitates a bookkeeping method known as jets. Jets combine the shower products in order to correctly associate them with the initial particle. Algorithms discussed in the next section describe the method of associating energy deposits in the detector to a given jet [92, 93].

3.3.1 Reconstructing Jets

Within ATLAS, the primary jet recombination algorithm is the anti- k_T algorithm. This is a sequential reclustering algorithm, one of three algorithms that are a result of the "general" k_T algorithm

$$d_{ij} = \min(k_{Ti}^{2p}, k_{Tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} \quad (3.2a)$$

$$d_{iB} = k_{Ti}^{2p} \quad , \quad (3.2b)$$

where the angular separation of the two pseudojets is $\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$ with y and ϕ being the rapidity and azimuthal angle of the two pseudojets i, j . k_T is simply the transverse momentum. d_{iB} is the distance between a input constituent and the beam and d_{ij} is the distance between the inputs themselves. R is a parameter that sets the spatial extent of the reconstructed jets. The value for the power p here determines the nature of the algorithm's reclustering with $p = 1$ corresponding to the k_T algorithm, $p = 0$ the Cambridge/Aachen (C/A), and $p = -1$ the anti- k_T algorithm. Regardless of the value of p chosen, they all follow a similar sequence. A single ordered list composed of d_{ij} and d_{iB} values is created. If a d_{ij} value is the minimum of this list, then clusters i and j are merged and the list is remade. If d_{iB} is the minimum, then cluster i is defined as a jet and removed from the list. The process is then iterated until no clusters remain in the list. All three algorithms use the k_T and the spatial separation between the constituents

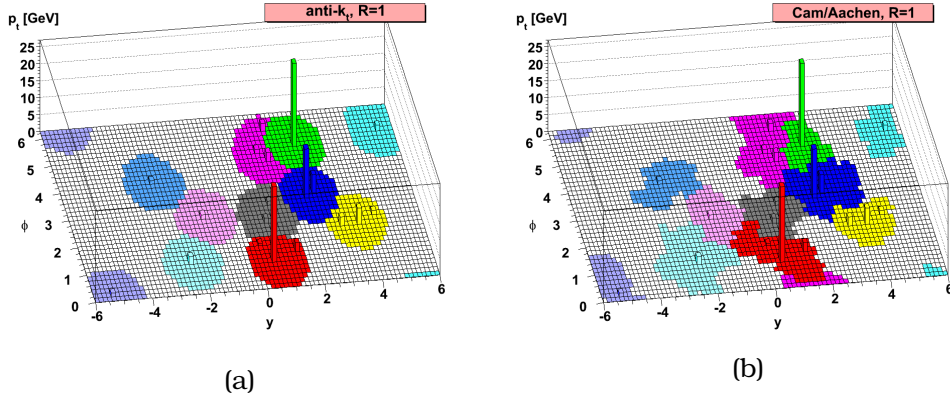


Figure 3.3: An example of jets reclustered with (a) anti- k_T and (b) C/A for the sample event. The jets are shown on the $y - \phi$ plane with the p_T of the jets along z . The total p_T of each jet is represented by the tallest bin in each reclustering [95].

i and j to cluster the jets although only two are relevant in this work and are covered in detail here.

$p = -1$ defines the anti- k_t algorithm [94] which is the dominant reclustering algorithm used by ATLAS. The separation distance d_{ij} of two constituents in the recombination is determined by Δ_{ij} and more importantly $1/k_t^2$. Due to this factor of $1/k_t$ the clustering is weighted towards higher k_T , or harder, particles. This is advantageous because it results in a small number of well defined recombination constituents called subjects. There is one thing to watch out for with this algorithm as the clustering sequence isn't necessarily physically meaningful. This is due to the fact that the recombination sequence starts with hardest particles.

$p = 0$ defines the Cambridge/Aachen (C/A) [96]. In stark contrast to the recombination nature of the anti- k_t algorithm, the C/A algorithm recombination is based solely on the ratio of the separation Δ_{ij} and the radius parameter R . This purely spatial recombination scheme results in a recombined jet that most successfully separates the large angle constituents when they split. While not preferable in creating jets to identify based on the energy density and sub-structure of the jet, this spatial separation has significant effect on the orientation of the subjects, or prongs, of the jets as well as the hard and soft constituents of the jets. Most relevant here is the impacts this recombination has on the Lund planes of jets (detailed later), informing much of the success in the tagger

presented in Chapter 4.

3.3.2 Trimming

Jet grooming techniques help to minimize the impact that pile-up or soft radiation has on the reclustering. By minimizing these effects, the resulting reclustered jet more correctly represents the original hard particle leading towards improved identification and tagging.

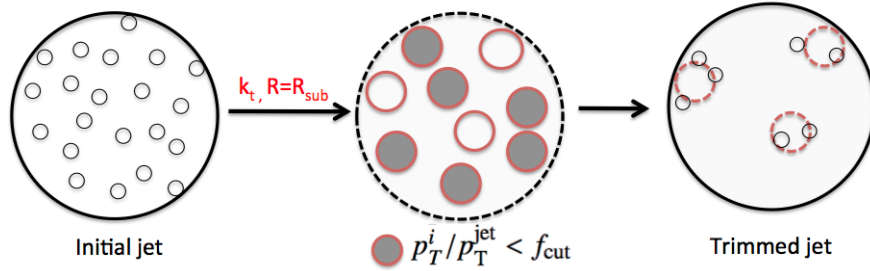


Figure 3.4: Cartoon depicting the steps in the trimming method of jet grooming [97].

While there are several types of grooming techniques within ATLAS, trimming has the widest application. The trimming method is carried out by first reclustering a jet using the anti- k_t algorithm into smaller subjets defined by a subjet radius R_{sub} . A p_T fraction f_{cut} is defined as the ratio of the p_T of the subjet to the jet and all subjets are required to have f_{cut} above a certain value. In ATLAS, $R_{sub} = 0.2$ and $f_{cut} = 0.05$ or 5%. The trimmed jet energy is then calculated from the four-momentum sum of all its subjets, with each subjet assumed to be massless. Figure 3.4 graphically outlines the process of trimming. The advantage of using the trimming can be in the resolution of mass distributions of a number of processes, particularly vector bosons.

Before the energy scale and resolution are calibrated, the origin of the reconstructed jets must be corrected. The origin of jets produced in the ATLAS detector are initially set to the center. The four momentum of the jet is recalculated in order for the origin of the jets to be corrected to the hard scatter primary vertex that is associated with the jet from reconstruction. This calibration improves the resolution in η .

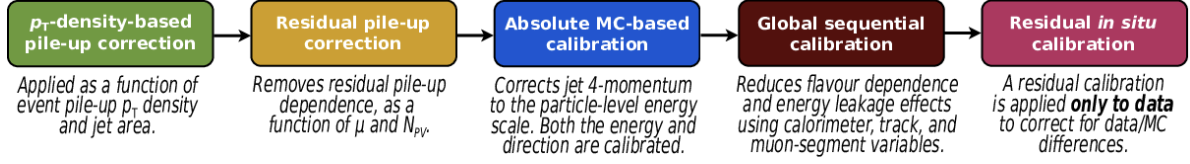


Figure 3.5: A flowchart and brief description of the states of JES calibration, adapted from [98].

3.3.3 Jet Calibration

In order for the jets that are constructed to be useful for precision physics measurements and sensitive BSM searches, their energy scale and resolution must be measured. Since analyses use MC samples to design and optimise their strategies that are used to make measurement of collision data, it is necessary for the measured values of the jets to be calibrated so that there is agreement between data and MC. This is accomplished through Jet Energy Scale (JES) and Jet Energy Resolution (JER) calibrations. These measurements and the calibrations methods also include the associated systematic uncertainties.

Jet Energy Scale Calibration

As the name implies, JES calibration restores the energy scale of reconstructed jets to simulated jets [59]. The calibration is accomplished through a number of steps whose end result are jet objects that have energies consistent with the underlying particle level jets. A flowchart outlining the process is shown in Figure 3.5.

The objective of the first two corrections is to remove the contributions of pile-up to the p_T of the reconstructed jets. The first of these corrections is a p_T -density based correction. In general the p_T density of a jet is $\rho = \langle p_T/A \rangle$ where A is the area of the jet. This area is determined through the process of ghost association. Particles with infinitesimally small p_T are added uniformly to the event before reconstruction. After reconstruction the area of the jet can be determined through the number of associated particles. A pile-up p_T density $\rho_{\text{pile-up}} = p_T/A$ is derived from the median p_T of jets in low occupancy regions of the detector clusters [59] in the range $|\eta| < 2.0$. The median

value is used to reduce the influence of high p_T impacts. The density correction however leaves the $|\eta| > 2.0$ and high-occupancy regions of the detector with a residual pile-up dependencies.

This residual pileup correction is defined as the difference between the p_T of truth and reconstruction jets. This difference is then expressed as a function of both the number of primary vertices N_{PV} and the mean number of interactions per bunch crossing $\langle\mu\rangle$ with fitting parameters α and β respectively. N_{PV} and $\langle\mu\rangle$ are used because they are sensitive to in- and out-of-time pile-up respectively. These fit parameters are determined through a series of linear and logarithmic fits. A systematic uncertainty in the residual correction is derived from fluctuations in the logarithmic fit due to changes in jet p_T . The end result of the two pileup corrections is

$$p_T^{\text{corr}} = p_T^{\text{reco}} - \rho_{\text{pile-up}} \times A^{\text{reco}} - \alpha \times (N_{PV} - 1) - \beta \times \mu. \quad (3.3)$$

After the pileup corrections, the absolute JES and η of the jets are corrected. Truth MC jets with the pile-up corrections applied are used to derive a calibration for the JES and η . This correction is applied to the four momentum of the jets in order to correct them to truth values. This calibration corrects for biases in the reconstruction of η that arise from differences in detector geometry and material. Isolated jets are used to remove any ambiguity when associating the truth and reconstructed jets.

The previous calibrations applied corrections for pile-up, energy, and direction of the jets; however, these don't correct for fluctuations observed in the response of jet. These fluctuations are observed due to differences in jet makeup and constituent distribution as well as fluctuations from the jet's traversal of the calorimeter system. This is corrected by the Global Sequential Calibration (GSC) [99]. The GSC corrects the four momentum of the jet through the sequential application of six independent global observables derived from measurements in specific layers of the calorimeters, tracking information, or muon segments. No corrections are applied for correlations between the variables since none have been observed. The result is an improvement in the jet energy resolution while keeping the average jet energy response unchanged.

The final step in the JES calibration corrects for differences between data and MC jets known as residual *In situ* calibration. This step is only applied to data to correct for differences that are due to imperfect detector description and response to the jets. It also helps correct for imperfections in modeling of the jets e.g. hard scattering, showering, and hadronization. The *in situ* calibration is done using η -intercalibration, Z/γ +jets, and multijet balance. Corrections are defined for both MC and data and the ratio estimates the difference in JES between the two. The η -intercalibration compares forward jets to central jets to correct for poor modeling of forward jets to the much more well modeled central jets. A combination of the other two techniques corrects for response differences to jets in the central region, with Z/γ +jets applied to jets up to 950 GeV and the multijet balance extending the calibration to 2000 GeV.

The JES uncertainty is calculated by propagating the uncertainty from each of the calibration terms through to the JES, resulting in 80 separate uncertainties [99]. To simplify the utilization of the JES for analyses, as accounting for the full set would be cumbersome, many of the uncertainties are combined in order to significantly reduce the total number while maintaining the same correction to within percent differences.

Jet Energy Resolution

The *in situ* methods described above are also used to calibrate the *resolution* of the measurements. Whereas the the response was determined through measurement of the means of different observables, the Jet Energy Resolution (JER) is a measurement of the width of the detector response [100]. The fractional JER, $\frac{\sigma(p_T)}{p_T}$, is measured in the central region using the γ +jets and Z +jets with dijet topologies used in non-central regions. In addition to the described *in situ* calibrations, an additional calibration for the JER is determined by looking at the symmetry of the uncertainty in the p_T of a jet. Using dijet events the asymmetry observable is defined as

$$\sigma_{\mathcal{A}}^{\text{probe}} = 2 \frac{\sigma_{p_T}^{\text{probe}} - \sigma_{p_T}^{\text{ref}}}{p_T^{\text{probe}} + p_T^{\text{ref}}} , \quad (3.4)$$

where *probe* is a reference to a probe jet in the area of interest and *ref* is a reference jet. This provides standard deviation corrections for any asymmetries in the width of the response to a jet and is used to calibrate the overall resolution resulting in the final JER measurement of

$$\frac{\sigma_{p_T}^{\text{probe}}}{p_T^{\text{probe}}} = \sigma_{\mathcal{A}^{\text{probe}}} \frac{\sigma_{p_T}^{\text{ref}}}{p_T^{\text{ref}}}. \quad (3.5)$$

Very low p_T jets suffer relatively more from increased pileup and the following parameterization is used to improve the resolution [101]

$$\frac{\sigma_{p_T}}{p_T} = \frac{N}{p_T} \oplus \frac{S}{\sqrt{p_T}} \oplus C, \quad (3.6)$$

where N is a noise term accounting for pileup and electronic effects, S relates to the sampling of the calorimeter, and C is a constant that is independent of p_T . The symbol $\oplus$ implies the quadrature sum of the three parameters. A scale factor is derived by measuring both N and S in MC and data and used to further calibrate the JER [102].

3.3.4 ATLAS Jet Definitions

Analyses in ATLAS use jet objects reconstructed and calibrated using the methods discussed in the previous sections. The jet definitions in ATLAS attempt to provide jet objects that encompass the majority of physics typologies.

Small R

The term Small-R jet is predominately used in ATLAS to refer to jets that have an $R = 0.4$ and they are used in the vast majority of ATLAS analyses. These jets use the anti- k_T algorithm for reclustering because of its high level of confidence in relating jet constituents to the jet itself. Their small size means that they are not particularly suited for boosted studies. This also means that they are typically normalized to the electromagnetic scale.

Large R

In ATLAS large- R jets normally have a $R = 1.0$. The standard large- R jet definition in ATLAS uses local cluster weighting (LCW or LC) to weight the clusters to the appropriate

energy scale, either electromagnetic or hadronic.

***R* Scan**

R-Scan jets are the result of the ongoing effort to offer more specialized jets for analyses that require more highly tuned object definitions. These jets use non-standard (not 0.4 or 1.0) radius parameters in the reclustering. The *R*-Scan effort has created definitions for jets with $R = 0.2$ and $R = 0.6$. Relevant for this thesis are the $R = 0.2$ jets since these, by the nature of their very small radius, can be used in anticipation of energetic, highly boosted objects. These small radius jets are also used to create larger jet such as the TAR jets covered in Section 3.3.5

Variable Radius Track Jets

Instead of fixing the radius of a jet for use in the reclustering, Variable Radius (VR) track jets parameterize the anti- k_T reclustering algorithm in Equation 3.2 in term of the transverse momentum and a fixed scale ρ as [103, 104]

$$R \rightarrow R_{\text{eff}}(p_T) = \frac{\rho}{p_T}. \quad (3.7)$$

This parameterization leads to a more precise reclustering of the constituents of a jet resulting in performance gains particularly for boosted b -jets.

PFlow, TCC, UFO Jets

A technique that has become more common and led to the creation of improved jet definitions is the usage of tracking information along with calo clusters. Two new newer jet definitions that began to be used within ATLAS at the end of Run 2 that apply this technique are Particle Flow (PFlow) jets [105] and Track-Calo-Cluster (TCC) jets. The benefit to using tracking information along with the calo-clusters is the added angular resolution that the tracks provide. PFlow jets are $R=0.4$ reclustered jets while TCC jets are an alternative to large- R jets. A more recent advancement in jet definitions are so called Unified Flow Objects (UFO) jets that seek to create a combined definition that

combines TCC and PFlow jets.

3.3.5 Track Assisted Reclustered Jets

Briefly discussed in 3.3.4 were various types of jet definitions used in ATLAS. The analysis covered in Chapters 6, 7 uses a type of jet known as a Track Assisted Reclustered algorithm (TAR) jet. A primary benefit gained from the use of TAR jets is an increase in the resolution of certain jet substructure variables, particularly the mass of the jet. They also provide increase performance in boosted regimes over other jet definitions. As the name implies, TAR jets are a combination of two techniques, track assistance and reclustering.

Track Assisted Jets

The structure and kinematics of jets are derived from the four-momentum of the constituents, for instance the jet mass

$$m = \sqrt{\left(\sum_{i \in J} E_i\right)^2 - \left(\sum_{i \in J} \vec{p}\right)^2}, \quad (3.8)$$

which can either be a sum over caloclusters for m^{calo} or tracks for m^{trk} . These labelings should not be confused with the mass of the tracks and clusters which is zero. As jets become more boosted, they become more collimated when they interact with the detector. This collimation of the jets leads to the energy being deposited in fewer cell resulting in a more dense merging of clusters and a resolution reduction. This reduction in the angular resolution makes m^{trk} superior to m^{calo} for boosted jets, however tracks only exist for charged particles. Track assistance recovers some of this angular resolution lost in the calorimeter by correcting m^{trk} with the momentum difference measured between the tracker and the calorimeter. This correction is simply a ratio of the two momenta which gives a new track assisted mass m^{TA} shown in Figure 3.9.

$$m^{\text{TA}} = \frac{p_T^{\text{calo}}}{p_T^{\text{trk}}} \times m^{\text{trk}}. \quad (3.9)$$

Since both m^{TA} and m^{calo} are better in their respective p_{T} regions, high and low, it makes sense to combine the two. A combined mass definition m^{comb} is constructed for jet mass combining the two as a single variable using an inverse-resolution weighted mean. m^{comb} is the standard mass for Large-R jets in ATLAS. To further improve the resolution of the substructure of highly boosted jets, particularly those with complex final state topologies like $H \rightarrow VV$, reclustering is used. Within ATLAS a method of reclustering known as Track-assisted subjet algorithm (TAS) [106] is often used. TAS matches p_{T} -scaled tracks to the subjets of a trimmed Large-R jet through ghost association [107, 95]. Tracks that can not be ghost associated are matched to the nearest subjet requiring $\Delta R_{\text{track,subj}} < 0.3$. Since scaled tracks are used the neutral hadronic components from the calorimeter are accounted for by scaling the old tracks to the p_{T} of the subjet:

$$p_{\text{T}}^{\text{track,new}} = p_{\text{T}}^{\text{track,old}} \times \frac{p_{\text{T},j}^{\text{subjet}}}{\sum p_{\text{T},j}^{\text{track,old}}}. \quad (3.10)$$

TAS demonstrates an improvement in track association because it associates the tracks using the subjets which have tighter matching constraints. However, although TAS uses rescaled tracks, it still depends on the original reclustering structure of the large-R jet since the association is to the original subjets. TAR jets removes this dependency by reclustering the Large-R jet from the track directly.

Track Assisted Reclustered Jets

The TAR algorithm improves upon TAS by making use of the scaled tracks directly to recluster a large radius jet instead of rematching [108]. This is done by first clustering a large radius jet from calibrated anti- k_t $R = 0.2$ jets. A trimming fraction is applied to the subjets of $f_{\text{cut}} = 0.05$. Tracks are matched to the $R = 0.2$ jets after the trimming. Tracks that can be associated to the jets through ghost association are done first, with the remaining tracks associated to their nearest $R = 0.2$ jet as long as the $\Delta R < 0.3$. While this value of ΔR can be adjusted, it was found in initial studies that a value of 0.3 was optimal. As with TAS jets, tracks are scaled by the p_{T} of their associated jet using 3.10. The final jet is built from these $R = 0.2$ jets and is referred to as a TAR jet, with the

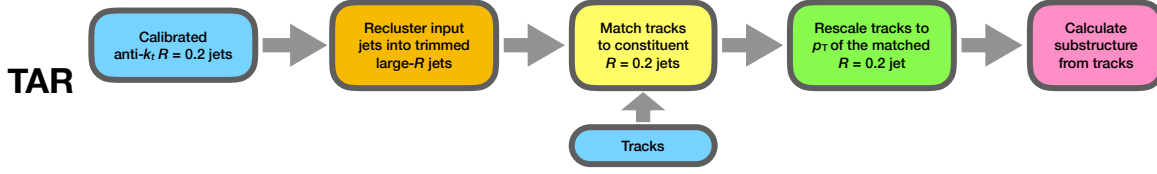


Figure 3.6: Flowchart summarizing the sequential steps of the TAR algorithm [108].

the $R = 0.2$ jets being called the constituent jets and the tracks the constituent tracks. All substructure variables including m^{TAR} are calculated from the scaled tracks. A brief summary of the algorithm is depicted in Figure 3.6.

There are added benefits on top of the improvements to substructure resolution. Since TAR jets are clustered using calibrated $R = 0.2$ jets this means the TAR jets themselves do not have to be calibrated and the uncertainties of the jets can be taken directly from the reclustering constituents. This so called bottom-up technique derives the TAR jet by propagating the uncertainties from the more precisely defined uncertainties of the tracks and $R = 0.2$ jets. This technique also means that the TAR jets are more flexible in their definition. The calibration of the large- R jets from Sections 3.3.4 is specific to the radius parameter of the jet meaning that a redefinition of the jet would require a new calibration to be derived. TAR jets on the other hand can redefine the radius and recluster, with the calibration remaining the same, propagated from the constituents associated to the jet under the new radius value.

3.3.6 b Jets

b quarks that are produced in the LHC quickly hadronize into a jet object, a so-called b jet. These jets are of extreme interest at the LHC as they allow for the study of many different aspects of the SM as well as being a tool for exploring BSM theories. More specifically, the topology of the $HH \rightarrow bbVV$ channel has two b -quarks in it and the precise identifications of the jets that come from these two quarks is fundamental to the analyses presented in Chapters 6 and 7.

***b*-tagging Jets**

Jets produced from b quarks are uniquely identifiable because b -hadrons have a mean lifetime of 1.6 ps which, while incredibly short, is significant resulting a significant and measurable distance between the primary vertex and the decay of the b hadron. The location of the decay is called the secondary vertex. b -jet tagging in ATLAS is a busy pursuit, always trying to improve upon the existing techniques. In this thesis a high-level b -jet tagger based off of a deep neural network known as DL1r [109] is used. The DL1r tagger uses working points defined by the efficiency to correctly tag b jets (and the associated c - and light-jet rejection factor) using the discriminant:

$$D_{\text{DL1r}} = \ln \frac{p_b}{f_c \cdot p_c + (1 - f_c) \cdot p_{\text{light}}} \quad , \quad (3.11)$$

where p_b , p_c , and p_{light} are the probability for a jet to originate from a b -quark, c -quark or light-quark/gluon respectively. The f_c fraction is a tunable parameter allowing the discriminant to be set based on the desired performance. The tagger has multiple Working Point (WP)'s though the 77% WP is recommended. This corresponds to performance in the b -tagging efficiency of 77% when applied to jets.

3.4 Muons

While the fully hadronic final state of the analysis in the thesis contains no leptons, the lepton veto is still dependent on proper particle identification. Muons are an incredibly important physics object that is used in a wide variety of studies including new physics searches and SM measurements. There are four different types of reconstructed muons. Combined muons are muons that use tracks from both the ID and MS. Segment-tagged muons defines muons that do not reach all the layers of the MS. The third type of muon are the calorimeter-tagged muons that are used in regions of the detector that do not have complete coverage. The final muons are extrapolated muons which are defined for regions beyond the reach of the ID.

Four quality requirements are defined in order to select muons with high efficiency

while suppressing background processes. Muon identification within an analysis is performed by selecting one of these requirements informed by particular need of the analysis. The definitions for each of the quality requirements comes from specific requirements for each of the sub-systems responsible for muon classification [110]. Medium selection is the default selection for the ATLAS collaboration. It only uses combined muons and tracks from the MS. The Loose selection is used to maximize reconstruction efficiency and also in tau particle ID. Tight muon selection is used when muon purity is prioritized and needs to be maximal. The last selection criteria is high- p_T which results in muons with the best momentum resolution of the selections.

3.5 Electrons and Photons

Electron and photon reconstruction begins in the EM LAr calorimeter. Energy depositions are matched with tracks within the ID. Energy depositions with a well constructed track are classified as electrons and those without an associated track are classified as photons. A third classification for photon conversion is used when a clearly constructed conversion vertex is associated to the energy deposit. As with muons, electrons and photons have identification and isolation criteria. A multivariate technique is applied to a variety of discriminating variables [111] and selections Loose, Medium, and Tight are defined by cutting on the discriminate of a likelihood-based method with increasing background rejection for each selection [112, 113].

3.6 Taus

The tau particle does not feature in the analysis for this thesis but it still warrants a brief discussion. The tau particle is a particularly unique objects. Unlike the lighter leptons, tau leptons have a dominant hadronic BR of around 60%. Their massiveness also means that they often do not interact with the ID, decaying too quickly. This makes tau lepton identification a particularly difficult task, hybridizing in a way the techniques described for jets in Section 3.3 and the lighter leptons in Section 3.4 and Section 3.5. Hadronic tau decays are identified through the selection of a lepton using the loosest

identification criteria for muons and electrons. These leptons are seed an $R = 0.4$ anti- k_T jet [114] which becomes a tau candidates. The tau vertex is selected the associated track with the largest fraction p_T to the jet. Tau identification is performed using MVA techniques and is optimized for light QCD jet rejection.

3.7 Missing Transverse Momentum

The second composite and last physics object to discuss is Missing Transverse Energy (MET). All the objects in an event have associated four momenta and as described in the previous sections, these are used to reconstruct all the physics objects used in analysis. This reconstruction isn't perfect, and in some cases is significant. Since the momenta of the incoming protons are essentially longitudinal, the momentum in the transverse plane should sum to zero. This sum is reconstructed as the MET object. An expression to reconstruct the MET from the other objects is [115]

$$E_T^{\text{miss}} = \sqrt{(E_x^{\text{miss}})^2 + (E_y^{\text{miss}})^2}, \quad (3.12)$$

$$E_{x(y)}^{\text{miss}} = -(E_{x(y)}^{\text{jets}} + E_{x(y)}^{\text{electrons}} + E_{x(y)}^{\text{photons}} + E_{x(y)}^{\text{taus}} + E_{x(y)}^{\text{muons}} + E_{x(y)}^{\text{ST}}) \quad ,$$

where $E_{x(y)}^{\text{object}}$ is the x and y components of the sum of the different reconstruction variables discussed previously. $E_{x(y)}^{\text{ST}}$ is the sum of the tracks and clusters not associated to another objects in the event. The most common example of the relation between the MET and a physical particle is represented by the neutrino. Undetected by ATLAS these neutrinos carry momentum away that is associated to the MET in events.

Chapter 4

IDENTIFYING BOOSTED FOUR-PRONG OBJECTS

4.1 WW jet as four-prong object

Briefly discussed in Section 1.12, the topology of $H \rightarrow WW \rightarrow qq\bar{q}\bar{q}$ is heavily dependent on the p_T of the Higgs. At low transverse momentum, the four quarks can be resolved and treated as four independent jets. As the p_T of the Higgs increases, the topology changes to one where the W 's are still separated and finally at sufficiently high p_T values, these W jets become so collimated so that they are indistinguishable, and treating the two W jets as a single jet becomes advantageous, even necessary.

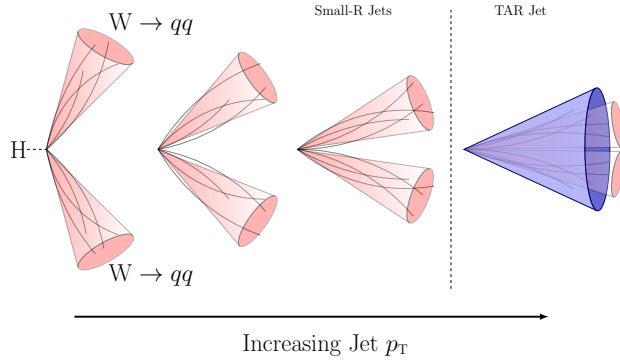


Figure 4.1: Cartoon illustrating the collimation of jets as the p_T of the jets increases and the diagrammatic relationship between TAR jets and their underlying small- R jets.

These boosted WW jets are a relatively unexplored physics objects due to the difficulty in identifying them. One of the primary reasons for this difficulty is due to the complexity of the environment where these WW jets are common. The transition of the topology of these WW jets with p_T can be seen in Figure 4.1. The transition happens at around 300 – 400 GeV. The transition p_T range is illustrated in Figure 4.2 by the fraction

of quarks contained within a single $R = 1.0$ jet for a hadronically decaying $H \rightarrow WW$ process. At 400 GeV 75% of jets contain all four quarks and this fraction grows with p_T . These four quarks from the WW decay make for a highly complex internal substructure of the jet, particularly when contrasted with the smooth energy distribution found in QCD jets. This contrast motivates characterizing jets by the number of energetic decay products found within highly boosted jets where the partons are close together. A common feature this number of decay products is a *Prong*.

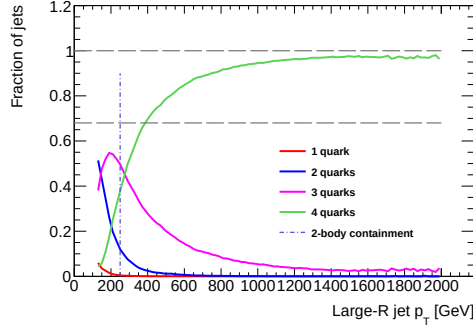


Figure 4.2: The number of quarks found in $H \rightarrow WW$, $R = 1.00$ jets as a fraction of the total up to 2000 GeV. The fraction of jets with 4 quarks become overwhelming at around 400 GeV and increases to near 100% by 2000 GeV. [116].

4.1.1 Prongs

The hadronic decay of pair of boosted W s is a clear example of a four-prong object, the prong characterization however applies to any jet. For instance, the hadronic decay of a single W boson, $W \rightarrow q\bar{q}$, produces two highly energetic quarks, resulting in a two prong jet. As previously stated, the number of prongs a jet has tells a great deal about the complexity of the substructure of a jet. Other relevant physical processes and their number of prongs are QCD jets (1 prong), W jets (2 prongs), $t \rightarrow Wb \rightarrow qqq$ jets (3 prongs), and $H \rightarrow WW \rightarrow qqqq$ jets (4 prongs). These processes feature heavily in di-Higgs physics motivating the use of the number of prongs in a jet for object identification. In order to take advantage of the link between particular processes and its prong count, kinematic variables have been defined to better identify the prongs. In addition to prongs, other substructure variables have been constructed that successfully characterize the underlying structure of a jet and the structure's relation to energy density and geometry of a

jet.

4.1.2 TAR jets for boosted $HH \rightarrow bbWW$

It is rather straight forward to motivate the use of large radius TAR jets for the complex $HH \rightarrow bbWW \rightarrow qq\bar{q}\bar{q}$ final state. Figure 4.3a shows that the maximum separation of the quarks in the $W \rightarrow qq$ decay is around 1.0 meaning that the hard prongs for a $H \rightarrow WW$ jet can be contained in a single $R = 1.0$ jet. The decrease in the minimum separation of the quarks in figure 4.3b necessitates maximizing the granularity of the jet.

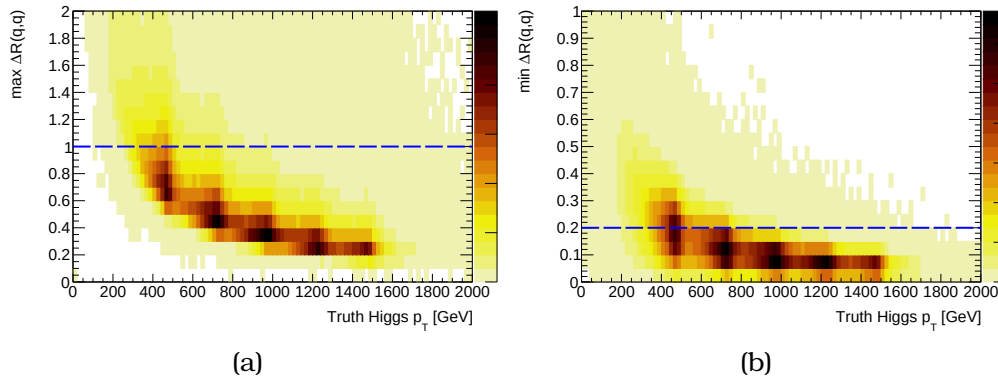


Figure 4.3: (a) The minimum ΔR of the quarks in TAR jets from the $W \rightarrow qq$ decay and (b) the maximum separation of these quarks and the ability to fully contain them in a TAR jets with $R = 1.0$ [116].

4.2 Machine Learning in HEP

A constant task in high energy physics is to develop techniques that can discriminate between different physical processes, most often a single signal process from a variety of backgrounds. However, in particle physics the signal is often many times smaller than the background. This necessitates the development of increasingly complex tools for discrimination with the goal of higher levels of precision or simply success where other methods have failed. The underlying complexity of the task of tagging a four-prong jet as well as the desire to explore sophisticated approaches easily motivates the use of machine learning. The advent of new machine learning techniques as well as the impressive, and readily available increase in raw computing power has catalyzed a increased use of machine learning in HEP. There are many different types of machine

learning algorithms that are employed, boosted decision trees, Markov chains, support vector machines, but recently one has begun to dominate as the preferential machine learning solution and that is the artificial neural network.

4.2.1 Machine learning philosophy

One of the most oft quoted "definitions" for machine learning, and one that guides much of the way that machine learning algorithms are designed comes from Tom M. Mitchell. He said that "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E [117]." This definition is however quite broad and can cover a large array of different machine learning techniques, some that have been in usage for decades. The focus of the machine learning discussion in this thesis and the application that is utilized in the analyses to follow focus on artificial neural networks. This particular type of machine learning and its many variants have been increasing in utilization in HEP in the last decade or so due to the increase in computing power that has facilitated the computational workload that they often require.

While machine learning and Multivariate analysis techniques are often described as a "black box" in high energy particles physics, and it can certainly feel that way at times, the fundamentals of how machine learning works is rather straight forward and understanding this can help to better understand the inner workings of a particular algorithm. At its most basic level, what a neural network accomplishes is learning a generalized function to solve a given problem. It learns this function through the repeated exposure to an predefined experience, "learning" with increased exposure until the task achieves a desired level of success.

4.3 Artificial Neural Networks

Artificial neural networks are inspired by nature, taking their language and structure from that of real neurons and synapses in biological brains. As with real neurological pathways, artificial neural networks are made up of collections of artificial neurons,

or nodes. Nodes are connected to other nodes in a similar way to how neurons are connected through synapses. Typically the nodes of an artificial neuron network are arranged into layers with a specified method of connection to subsequent layers.

4.3.1 Feed Forward Neural Networks

The simplest type of artificial neural network is known as a Multilayer Perception (MLP). While the term MLP can be ambiguous as it is periodically used to refer to any artificial neural network, here it will refer to an artificial neural network comprised of three layers shown in Figure 4.4a. The first layer contains the inputs to the network and the last layer produces the output of the network. Layers between the input and the output layers are known as hidden layers and in a MLP there is only one hidden layer. Each link between nodes contains a weight and the "learning" of the network is accomplished through the repeated updating of these weights in response to the information that passes through them. In Figure 4.4b the relationship between the input data represented here as the vector x_m and the weights of a network w_m are shown through the sum that is used to generate the output $\hat{y}$ of a given network.

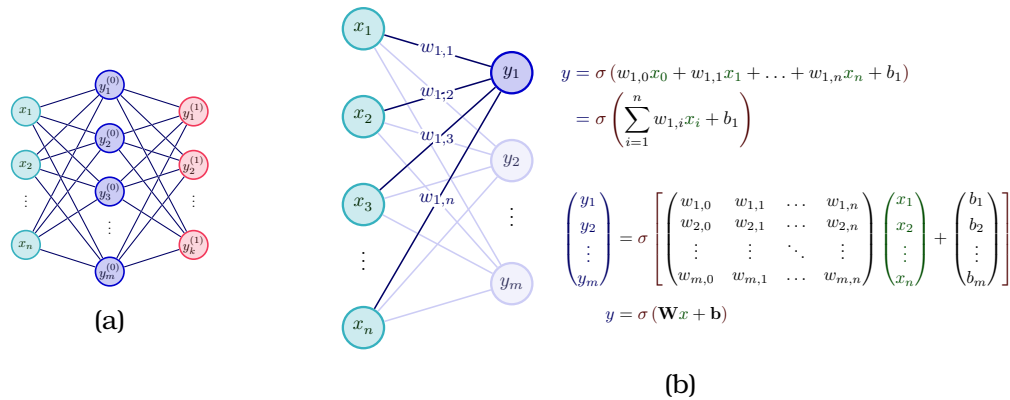


Figure 4.4: (a) A simple artificial neural network with one input layer, a single hidden layer, and an output layer. This is commonly referred to as a MLP. (b) A detailed example of the relationship between the inputs x_n , the weights w and biases b of the nodes, and the output y of a single layer of a network [52].

The values for the weights between the layers are initialized before training with the distribution of weights chosen having an impact on the performance of the network. The most common initialized weights for a feed forward neural network are derived from the

Xavier distribution [118]. Defined as a uniform distribution with the range $[-\text{limit}, \text{limit}]$ where $\text{limit} = \sqrt{2/(\text{fan}_{in} + \text{fan}_{out})}$. Here fan_{in} and fan_{out} are the input and output number of weight units. This distribution helped to solve issues with poor performance that arose during the optimization of deep networks. The uniformity in the distribution insured that the non-linearity of the activation functions combined with the weights didn't saturate the network signal stopping network from meaningfully updating the weight.

The layers of a neural network are made up of any number of nodes. Nodes turn the weighted sum of inputs into outputs through an activation function. An activation function is a function that defines how the output of a node is transformed and depending on the particular task, any number of activation functions can be deployed. It's helpful to look a few functions in order to better understand their usage as well as the way in which they transform a network.

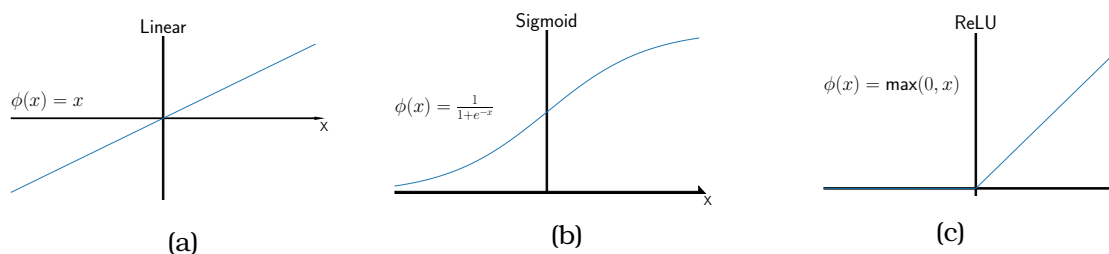


Figure 4.5: (a) Linear, (b) sigmoid, and (c) ReLU activation functions. The behavior of these functions at different values of inputs is readily apparent from the shape of the functions.

The most basic function is simply $\phi(x) = x$, or a linear relationship between the inputs and the outputs. This function was one of the first to be used and is able to solve basic tasks with neural networks. However, as the tasks become more complex, and the relationships between the inputs and the outputs become more non-linear, non-linear activation functions are applied. The sigmoid function is a non-linear function that improves on linear functions, but suffers from a tendency to "snap" to 0 and 1 for large and small input values. This can be mitigated somewhat through proper feature normalization, however can still cause a network to fail to minimize, or to stop prematurely. An improvement to the sigmoid and one of the most common used in neural networks is the Rectifier activation function (ReLU) activation function. The ReLU function results

in improved optimization both in the time to convergence as well as helping to ensure that the updates to the weights don't become sufficiently small as to be meaningless through an issue in machine learning known as the vanishing gradient problem. Figure 4.5 shows a representation of the functional form of these activation functions.

For the output layer of a neural network the desire is to produce a probability $\hat{y}$ associated with each class. This is generally accomplished using the sigmoid function. The "snap" tendency of the sigmoid characterized above is not necessarily desirable within the layers of a network however as an output layer this characteristic produces the best confidence for a given network response. The sigmoid function only works on tasks making binary decisions between two choices. For networks trained on more than two classes, the softmax function is used. The softmax function [119]

$$\phi(\mathbf{x})_i = \frac{e^{x_i}}{\sum_i e^{x_i}} \quad , \quad (4.1)$$

is simply a generalization of the sigmoid function where x is a vector of the network signal as opposed to a single value for the sigmoid.

4.3.2 Deep Neural Networks

Deep neural networks broadly refers to any network that has more than one hidden layer. This description has become less helpful in present day applications because network depth can range from a few layers to hundreds but both networks would fall under that definition of "deep". A descriptor like number of parameters on the other hand is a very simple way of characterising a networks complexity or "depth" in a more concrete fashion.

4.3.3 Supervised Learning

While there are many different algorithms that are used under the umbrella term of machine learning such as supervised learning, unsupervised learning, reinforcement learning, and anomaly detection [120] [121], supervised learning is the primary focus of this thesis because it is the underlying algorithm used to train the Four-Prong Tagger.

Supervised learning is the process of exposing a neural network to a task repeatedly and correcting its performance. Supervised learning is limited to situations where the processes in question is well enough understood to be able to construct training sets with truth information to supervise the learning.

4.3.4 Optimization

The heart of machine learning and where a neural network learns from the data is through the minimization process. Nearly all of deep learning is powered by single principle, Stochastic Gradient Descent (SGD) [117]. In general, the gradient descent approach is used to minimize a function. In machine learning this function is often referred to as the loss, or cost, function. The data that is fed into a network works it ways through the nodes and eventually produces an output. The type of output is defined by the goal, for instance for a classifier the output would be the probabilities of each classes, and this output is converted into the loss of the network through the loss function. For classification tasks such as jet tagging, the function typically used is the log-loss function. For a binary classifier discriminating between only two classes it is written as

$$L = -\frac{1}{N} \sum_i^N (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) , \quad (4.2)$$

where y are the binary labels (0,1) for each sample and $\hat{y}$ are the probabilistic responses from the network and N represents the total number of inputs samples. It's clear that this function minimizes when the "guesses" $\hat{y}$ are equal to the labels. SGD propagates the guess back through the network in order to use the loss to update the weights of the network. This process through which the weights of the network are updated in response to the loss function is where the "learning" of a network occurs. This process is repeated until the loss function is minimized. For multi-classifiers Equation 4.2 can be generalized for M classes

$$L = -\frac{1}{N} \frac{1}{M} \sum_i^N \sum_j^M y_{ij} \log(\hat{y}_{ij}). \quad (4.3)$$

4.3.5 Convolution

The concept of convolution is heavily associated with machine learning. In the recent past it has become more common to apply it to neural networks to reduce the parameter space of a task more quickly, decreasing computation time and opening up the doors for more complex features and tasks. A great example of this can be seen through the application of a convolutional neural network to item identification using images. Convolutional Neural Networks have been shown to learn edges of objects, shapes, sub-features (such as the nose on a human face) and much more. In general, convolution is a process that expresses how the shape of one function alters another and is expressed as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau)g(t - \tau)d\tau \quad , \quad (4.4)$$

where the produced convolution represented the strength of the alteration. In the case of machine learning, it is easier to think of the two functions as the inputs and weights of an algorithm. The structure of a convolutional neural network layer is analogous to that of the simpler fully connected layers. There's a defined number of nodes, in this case filters, that contain the weights of the layer. Unlike the weights of the fully connected layer, where each link has a single weight, the filters of a convolutional layer have themselves dimensionality. These filters convolve the input producing lower dimensional output. In Figure 4.6 the first three steps in the convolution of a single 5×5 "image" with a 3×3 filter of weights is shown. The element-wise product is taken between the filters and the corresponding entry in the image, the sum of which makes the output of the convolution. This is repeated until the filter reaches the end of the image. This process is often referred to as the dot product between the image and the filter but this is not, strictly speaking, correct.

This simple illustration outlines clearly how the filter weights interact with the values of an image showing how the size of the output changes with the strength of the response of the filter to the image. The weights of convolutional layers, particularly for deep networks are often initialized with the He distribution [122]. Absent from Figure 4.6 is

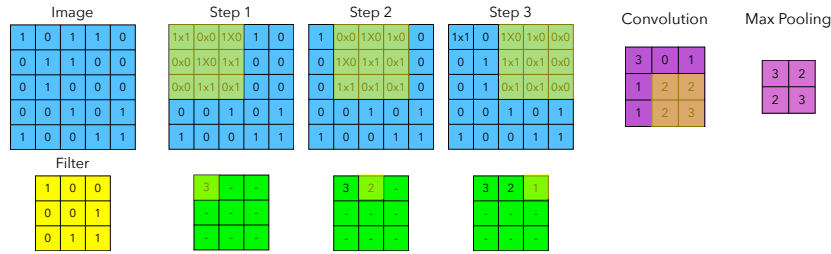


Figure 4.6: (a) Cartoon illustrating the first three steps in the filter convolution of a simple image. (b) A single step in an example of the application of maximum pooling. A pool size of 2x2 can be seen in yellow, with the maximum values that have been selected seen in the second array.

the activation function that is applied to the outputs of the convolutional layer before being passed to subsequent layers.

Desirable features of the convolution can be preferentially selected through the process of pooling. Pooling selects values out of a defined subset of the convolution based on the necessary criteria. In the case of the tagger and many other application of Convolutional Neural Networks pooling is used to select the strongest signal, thereby minimizing the convolution to only the prominent features. This is known as maximum pooling. Often in fact convolutional layers followed by a pooling layer are simply referred to as the convolution layer, or a convolution "block".

The output of the convolutional layers described above are a $x \times y \times (\text{filter number})$ tensor. This has to be reduced or restructured in some way in order to be used to produce the desired probabilistic response used in classification tasks. There are many ways to accomplish this but a common and straight forward way is to flatten the output of the final convolutional block into a 1D feature vector of length $x \times y \times (\text{filter number})$ which is typically reduced further using standard feed forward techniques.

4.4 A Boosted Four-Prong Tagger

Ian Goodfellow [117, 123] has said that it is often not immediately clear *a priori* in machine learning applications what algorithm will be successful therefore the best approach is to design and test an idea. This is particularly true with the application of

machine learning for the first time to the task of discriminating four-prong jet from QCD. A common approach to machine learning based taggers is to select a relatively maximal number of features, remove features that offer little to no additional discrimination power, then optimize and train a network architecture tailored to the high-impact features. The development of the four-prong tagger takes a slightly a different approach. In the following section each of the jet kinematics and substructure features used as input to the four-prong tagger are discussed as well as the physical reasoning for their inclusion. In Section 4.4.1 the method for reducing these features with similar physics properties or strong correlations presented. The Lund jet plane is presented as well as the reasoning reasoning behind using images for jet tagging. In the final sections the method and reasoning for parameterization are covered followed by the technical makeup of the network and its performance.

4.4.1 Features

In general with neural networks the number of features used for input has a trade off between computational complexity and performance. Put another way, typically the more features that you use in a network the better it performs but the more difficult and time consuming the computation. For High Energy Physics applications explainability plays a much more significant role in the design of a network. It is rarely enough to be able to design an algorithm that successfully accomplishes a task. For machine learning in HEP, in addition to high levels of performance, a successful implementation is accompanied by an explanation of the reasons for the performance observed and what physics can explains its successes. This desire for explainability motivates reducing number of inputs used neural network based tagger, simplifying preemptively the task of thoroughly understanding the network and its performance.

Jet p_T

The momentum of boosted four-prong jets can be powerful feature for discrimination since these jets typically have an much higher p_T than the anticipated background, in this case, QCD jets. In order for the tagger to be effective throughout the broad p_T range

the MC samples for the $H \rightarrow WW$ jets used was constructed in such a way that there is a relatively equal number of four-prong jets throughout the entire p_T range. This method avoids complications or increases in statistical uncertainty that arise through the process of reweighting. A similar logic was used with regards to the dijet QCD jets. Shown in Figure 4.7, the dijet samples were made to be heavily populated at high p_T ranges which is non-physical. This uncharacteristically high p_T distribution for QCD is to help improve dijet rejection in high p_T ranges where four-prong signal is anticipated.

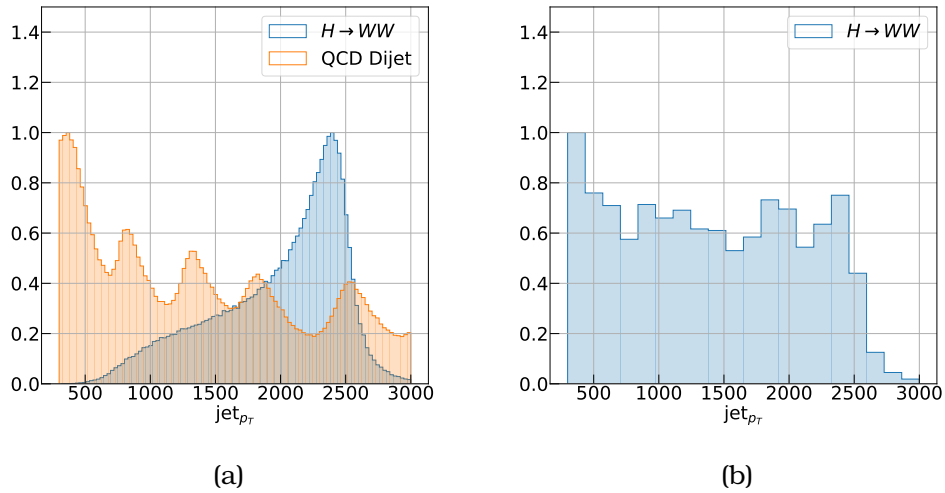


Figure 4.7: (a) The p_T of jets generated from a 5 TeV scalar compared to that of QCD jets. (b) p_T distribution of the $H \rightarrow WW$ jets used for training.

Jet Substructure Variables

Jet substructure variables are defined and used not only to tell something about the nature of the constituents inside a jet, but that they also can be used to help identify the process that produced a jet. The substructure variables covered here were investigated because of their potential strength in discriminating the most likely number of prongs a jet has.

N-subjettiness (τ_N) measures how well a jet can be described as having at least N

“prongs” [124]. Defined as

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \dots, \Delta R_{N,k}\} \quad , \quad (4.5)$$

where the sum is over the constituents k of the jet and $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ which gives the difference in $\eta - \phi$ between a given subjet axis of the jet, and $N = 4$ for a four-prong jet. The variable is normalized by the p_T of the constituents and the radius used in the reclustering with $d_0 = \sum_k p_{T,k} R_0$. In addition to n-subjetiness, ratios are also commonly used to differentiate between two numbers of subjetiness with

$$\tau_{N,N-1} = \frac{\tau_N}{\tau_{N-1}} \quad , \quad (4.6)$$

where τ_{42} would give the relative subjetiness of four versus two prong jets.

Energy-correlation functions [125] are used in the same manner as n-subjetiness, however they don't depend on the reclustering radius. The energy correlation function (ECF) series is

$$\text{ECF}(i, \beta) = \sum_{j_1 < j_2 < \dots < j_i \in J} \left(\prod_{a=1}^i p_{T,j_a} \right) \left(\prod_{b=1}^{i-1} \prod_{c=b+1}^i \Delta_{j_b j_c} \right)^\beta \quad , \quad (4.7)$$

where i is the number of energy point correlations desired used in the sum over the constituents j of the jet. The series defined below are constructed from e_i^β variables, ECF's from Equation 4.7 normalized by the single energy correlation $\text{ECF}(1, \beta)^i$. Three substructure variables are used as input for the tagger which are members of the c_i [125], n_i , and m_i [126] series. The c_i series is defined as

$$c_i^\beta = \frac{e_{i+1}^\beta e_{i-1}^\beta}{(e_i^\beta)^2} \quad , \quad (4.8)$$

and give a determination of the softness of higher order radiation from LO substructure of a jet. The m_i and n_i series use an updated definition for the energy correlation function in Equation 4.7 ${}_\nu e_i$ where i is still the number of particles to be correlated however the parameter ν truncates the product at a desired minimum number of pairwise angles

[126]. Using this new definition the m_i series is defined as

$$m_i^\beta = \frac{1e_{i+1}^\beta}{1e_i^\beta} \quad , \quad (4.9)$$

and is a good discriminator of jets with i prongs as long as the jets are sufficiently well groomed. Finally the n_i series is defined as

$$n_i^\beta = \frac{2e_{i+1}^\beta}{(1e_i^\beta)^2} \quad , \quad (4.10)$$

and offers similar information about n-subjetiness. Higher order members of each of these series were constructed and used in the initial development of the tagger to determine which were successful in discriminating boosted four-prong jets.

Reducing feature count

The individual input features for the tagger were selected as either being a primary kinematic feature of the jet or a substructure variable known to define four-prong objects well. The total set of substructure variables considered are listed on the axis of Figure B.1. This relatively small set of substructure variables was further reduced through an (n-1) performance comparison. Each substructure variable series is examined, e.g. c_i, n_i , etc., and one variable from each pair that has a linear correlation coefficients $> 70\%$ is eliminated. The resulting subset of features, listed in Table 4.1, are ranked according to their impact on the AUC score of the network once removed. The five features with the largest impact on the performance of the network as characterized by the AUC score were chosen.

4.4.2 The Lund Jet Plane

Before discussing the Lund jet plane and it is worth taking a moment to discuss the application of images and convolution to the task of jet tagging.

Variable	AUC	AUC $\Delta\%$	Ranking
All	0.9011	-	-
p_T	0.8407	6.70	1
c_4	0.8576	4.872	2
τ_{32}	0.8920	1.0010	3
m_2	0.8940	0.7879	4
n_2	0.8998	0.1442	5
τ_{21}	0.9000	0.1221	6
d_2	0.9008	0.0333	7
d_3	0.9008	0.0333	8
τ_{43}	0.9007	0.0444	9
τ_{42}	0.9008	0.0333	10
n_3	0.9009	0.0222	11

Table 4.1: Table containing the list of variables tested as possible substructure inputs. The AUC score and percent difference in the AUC score are listed. In the last column the variables are ranked in order of their impact on the performance of the network from highest to lowest.

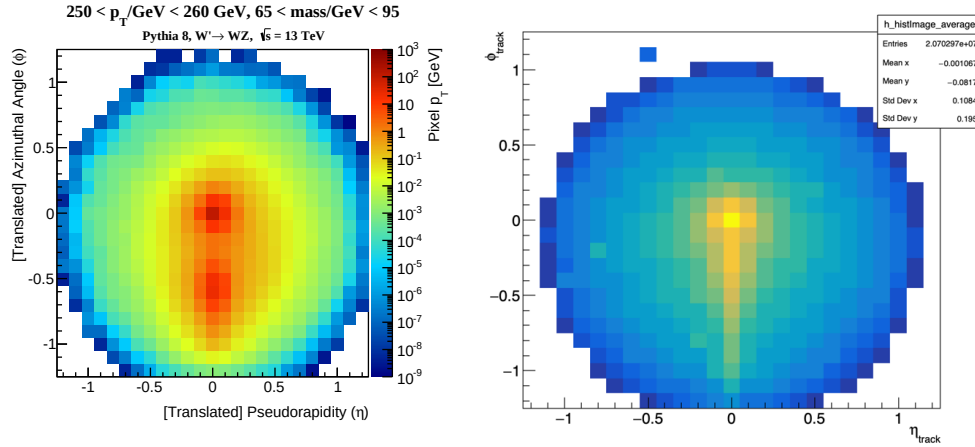


Figure 4.8: MC examples of calo images for W jets [127] and $H \rightarrow WW$ jets. The additional two subjects in the WW topology obscure the due prong features of the W calo images.

A comment Calo Jet Images

Jet images [127] are 2-D images created using the p_t , η , and ϕ of the subjects and tracks of a jet to create an "image" representation analogous to a jet's energy deposition in the calorimeter. Contemporaneous with this thesis was the emergence of solid performance in the use of calo jet images in the tagging of boosted objects. This in turn motivated a study to use calo jet images to tag four-prong jets. This study resulted in the realization that calo jet images are not a good representation of four-prong jet. Prototype calo jet image based four-prong taggers performed exceptionally poorly. Even more fundamental though is the structure of the jet. Figure 4.8 shows an example of boosted WZ and $H \rightarrow WW$ jets as calo images. Calo images clearly show a clear two prong structure for the WZ jets that is not present in the four-prong jets. The concept of jet images persisted however and led to the Lund jet images used in the four-prong tagger.

A Lund jet plane [128] is a representation of a jet based on Lund diagrams [129]. The Lund plane depicts a jet in terms of the transverse momentum and the angle of the emission with respect to the emitter of the constituents of the jet. [130]. This representation is of particular interest for four-prong objects, in this case $H \rightarrow WW$, because this first initial emission is quite hard due to the masses of the particles emitted. This insight is heavily reflected in the Lund planes for $H \rightarrow WW$ jets. In Figure 4.9a the jets have a defining characteristic located in the upper left had corner of the plane corresponding to splittings that are both large in the angle Δ and k_t .

The Lund planes of a jet are constructed using variables that are calculated during the declustering of a jet. Jets that have been clustered using the Cambridge-Aachen algorithm with a R parameter of 1.0 are used because of the C/A preferential spatial clustering outlined in Section 3.3.1. The Lund planes of a jet are created by working backwards through the jet-clustering constituents. By declustering a jet into two pseudojets a and b the clustering "tree" can be followed to construct a path of the emission. At each step in the declustering the combination of the two pseudojets is considered to be the emitter. This emitter can be thought of as a "node" in the jet clustering, and for each node clustering variables can be calculated. It is from these reclustering variables

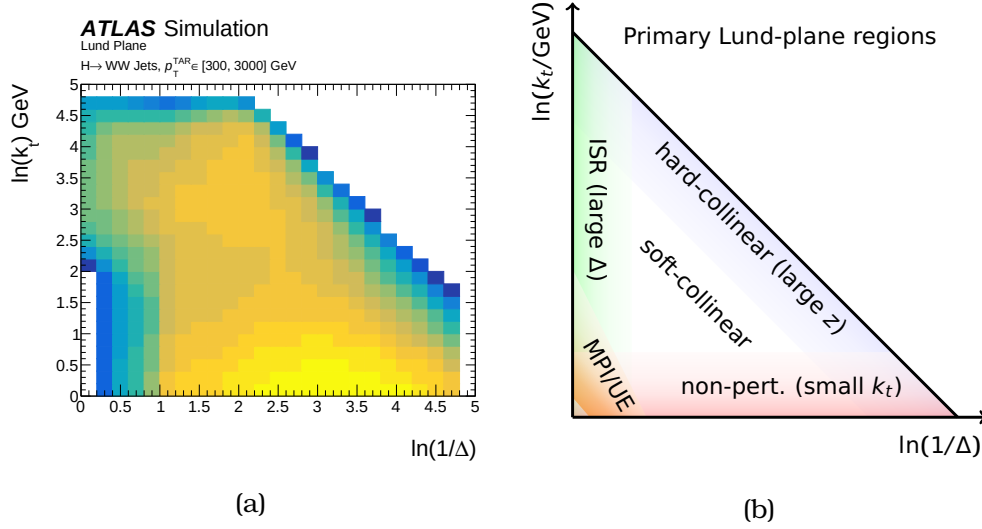


Figure 4.9: (a) An average Lund jet plane for a large collection of $H \rightarrow WW$ jets. (b) Explanation of the physical interpretations of the different regions of the Lund plane [128].

that Lund planes are generated. There are a broad number of reclustering variables that have been investigated; however, two

$$k_t = p_{Ta} p_{Tb} \Delta_{ab} \quad (4.11)$$

$$\Delta_{ab} = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2} \quad (4.12)$$

where $p_{Ta} > p_{Tb}$ are particularly relevant for W boson physics. Each of these variables is calculated for each node in the declustering, resulting in a number of pairs of k_t, Δ variables that represent the transverse momentum and emission angle of the nodes. Figure 4.9 shows a Lund jet plane for a large selection of boosted $H \rightarrow WW$ jets and the physical interpretation of the location in the plane $\ln 1/\Delta$ and $\ln k_T$ plane. Each one of these pairs can be represented by a point in the Lund plane. The number of reclustered nodes for each of the class of the tagger can be seen in Figure 4.10.

The Lund plane is not by any means the only way to use these reclustered nodes and variables, however given that the number of nodes varies from jet to jet, in order to use them as a feature for machine learning they need to be normalized in some way. A simple

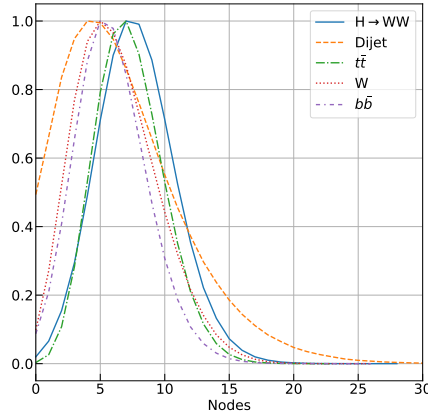


Figure 4.10: Distribution of the number of nodes per jet for each training process.

way to do this is by selecting a minimum number of nodes required per jet or padding all nodes to the maximum number with empty data. The Lund plane alternatively sets no limit on the number of nodes, instead projecting the nodes onto the 2d plane as an image. Each node has an associated Lund plane but combining all the planes provides a complete image of the emission recombination.

The Lund planes in Figure 4.11 contain all of the planes in the re-clustering, however they have been truncated to $\log k_t > 0$ since the region below this threshold is dominated by soft radiation emissions and dominates all other features. This focuses the plane to highlight the hard splitting features in larger k_t and Δ values that marks the difference between few and many-prong processes.

4.4.3 Parameterization

The final features in the network is a parameter that is introduced in order to generalize the performance of the network with respect to the resonance X in the model used to generate the the four-prong, $H \rightarrow WW$ jets used in training. The method of parameterizing a network uses a characteristic of the signal as an input. A simple example of network parameterization can be demonstrated using the Gaussian distribution.

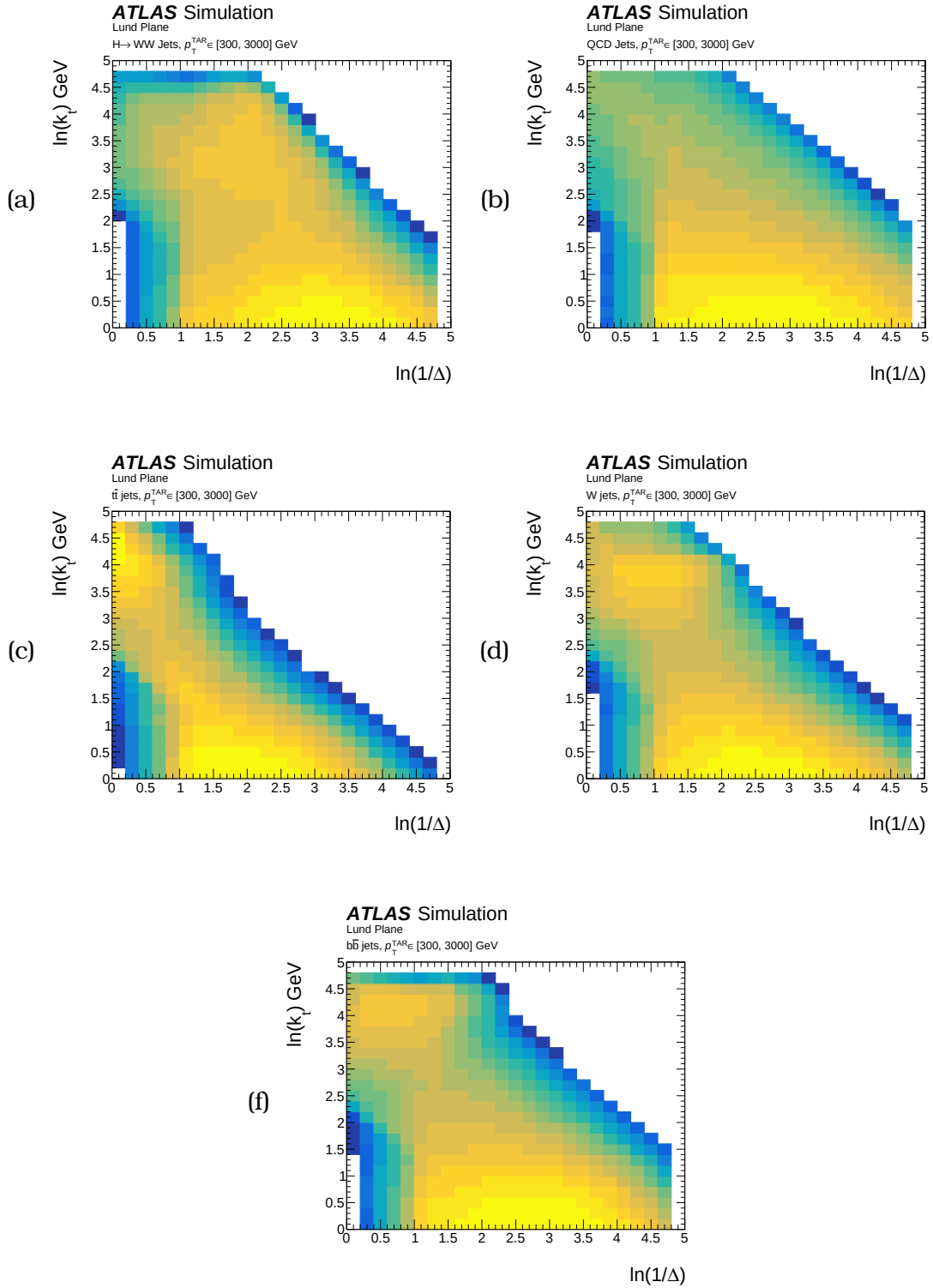


Figure 4.11: Average truncated Lund jet planes for the five training processes (a) $H \rightarrow WW$, (b) QCD, (c) $t\bar{t}$, (d) W +jets, and (e) $b\bar{b}$.

A simple parameterized task

Suppose a dataset that contains contributions from six types of samples: in five sample categories the variable in question is drawn from Gaussians with means $[-2, -1, 0, 1, 2]$. For the sixth (non-Gaussian) category of events, the input value of μ is chosen randomly for each event from the five possible values. The Gaussian drawn samples are illustrated in Figure 4.12a. A network can be parameterized in terms of μ by including the true value of μ in the training data. Once trained, a parameterized network has an improvement over standard network to extrapolate to examples it hasn't seen. In Figure 4.12b the response of the network vs the samples generated with $\mu = -1.0, 0.5, 2.0$, values that were not trained on, are shown. The network has clearly learned the Gaussian relationship to μ as the response as a function of μ has the same nature for the trained and untrained μ samples. For a network that had not been parameterized the distribution would be wider and the response from the network at the mean value would be significantly less strong.

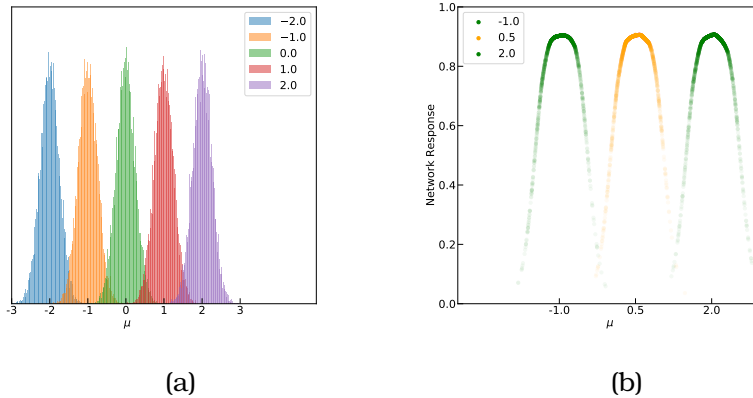


Figure 4.12: (a) The distribution for the signal training samples as a function of μ . (b) A comparison of the trained and untrained μ network response values as a function of μ , demonstrating the untrained samples have the same response characteristics.

The process that can give rise to boosted, four-prong objects can originate from different models. Many share the characteristic of a massive particle being generated from the initial collision $pp \rightarrow X$ as is the case for the analysis discussed in this thesis. The subsequent decay then gives rise to the four-prong objects. As was discussed in detail

in Section 1.11.2, the signal samples for these processes require the mass of the scalar X to be specified and eventually a range of these masses is determined that spans the scope of the desired search. This set of masses obviously isn't exhaustive and if it is later desirable to add a mass point it can be difficult and time consuming. The $X \rightarrow HH \rightarrow WWWW$ signal model is already a function of the mass of the scalar particle that is used to derive the signal jets; therefore, it was an obvious extension to the tagging algorithm to parameterize the inputs of the network in terms of this characteristic. The major benefit to parameterizing it in this way is it allows for a network to be trained once on all signal hypotheses while still maintaining the same performance as if the network was trained for each signal sample individually. There is also the freedom to test new signal hypothesis outside the training set without having to retrain the network to maintain the same level of performance. It reduces the computational overhead and can reduce the complexity of the application of the network. The primary drawback is specialization of the network task since it is now dependant on the particulars of a specific a models, i.e., those with a resonance mass.

4.4.4 Algorithm Specifics

The network for the four-prong tagger was developed in Python using the Keras [131] deep learning framework with the Tensorflow [132] backend. Shown in Figure 4.14, it is a hybrid neural net with two branches, a fully connected branch that takes the jet features as input, and a convolutional branch that processes the Lund Jet planes. These branches are flattened and combined into a final network before producing the output.

The inputs to the convolutional branch are 25×25 Lund jet plane images. The branch is comprised of three convolutional layers with 128 nodes, each followed by a maxpooling layer. The output of each convolutional and maxpooling layer pair is a feature map tensor, for example the output of the maxpooling is $15 \times 15 \times 128$. The activation function applied after each layer is the ReLU function. Overtraining is addressed on the convolutional branch through the application of L2 regularization [133] with a value of 0.01 for the regularization parameter. This method combats overtraining by adding a penalty to the loss function. Given the sparseness of the inputs images, a large filter of 10×10 is

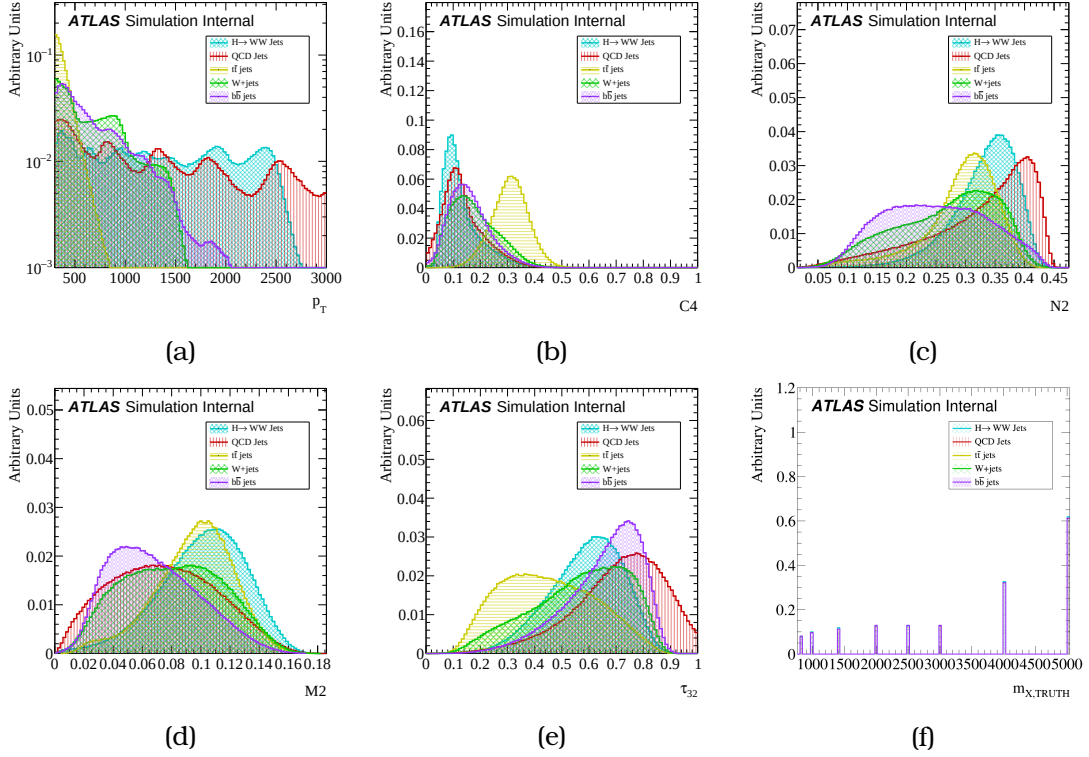


Figure 4.13: n_2 , m_2 , τ_{42} , and c_4 substructure variables scaled to the range $[0,1]$ for the five processes used to train the four-prong tagger: $H \rightarrow WW$, QCD, $t\bar{t}$, W +jets, and $b\bar{b}$ jets

defined for the first convolutional layer with the next two layers having a size of 4×4 . The pooling applied after each convolution is 2×2 . After the last layer in the convolutional branch the output is flattened to 1D in order to be combined the output of the feed forward branch.

The feed forward branch also has three fully connected layers, each 256 nodes and ReLU activation. A common technique known as dropout [134] is used to regularize the branch. The dropout technique helps to prevent the overtraining of the network by "dropping" a different randomized percentage of the nodes of the layers epoch. This percentage is set to 20%.

The output of the two branches are concatenated together and fed into two final layers to produce a combined network response. The last two fully connected layers have 64 and 32 nodes respectively, using the same dropout scheme as above, with a final output layer that uses the softmax activation function. The network produces a response for the five classes, dijet, W +jets, $b\bar{b}$, $t\bar{t}$, $H \rightarrow WW$. A diagram of the network can be seen in the Figure 4.14.

4.4.5 Feature Scaling

Each feature x_i was scaled to the range (0,1) using

$$X_{i,\text{scaled}} = \frac{X_i - X_{i,\min}}{X_{i,\max} - X_{i,\min}} \quad , \quad (4.13)$$

where X is the feature vector of i features and $X_{\min}$ and $X_{\max}$ are the minimum and maximum values for each input feature. Many other scaling methods were explored including standardization $\frac{x_i - \bar{x}_i}{\sigma_i}$, Gaussian normalization, and Scikit-learn [135] transformations such as RobustScaler. Different scaling technique had no significant impact on the results, though the simpler method used greatly reduced the complexity of integrating the tagger into the ATLAS software framework.

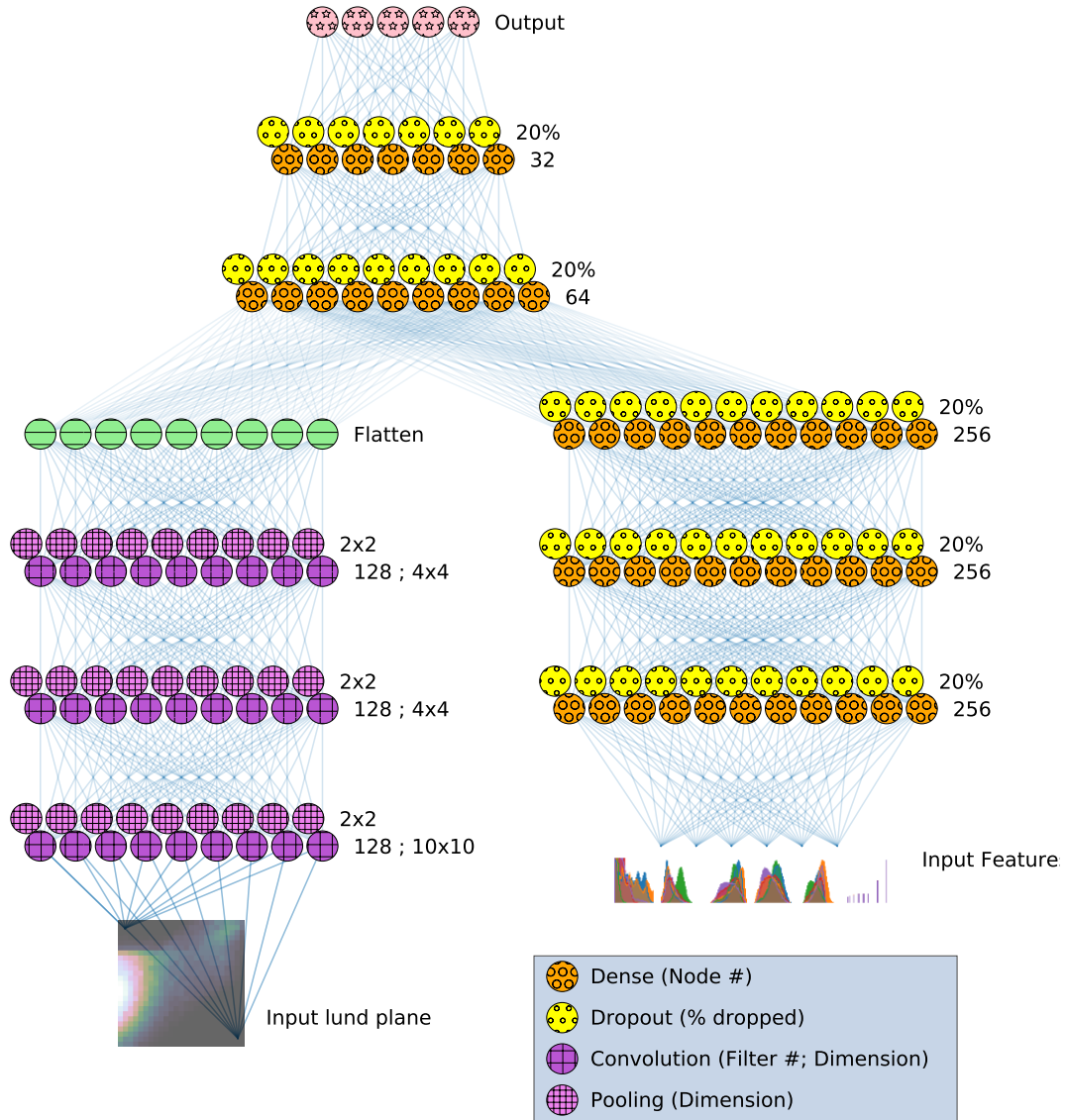


Figure 4.14: The architecture of the neural network used for the four-prong tagger. The Lund plane is processed by a convolutional branch shown here on the left, and the substructure features processed by the feed-forward branch on the right. These two branches are combined into a single feed forward to produce the final multi-class output.

4.4.6 Network Training

The network was optimized using the Adam [136] optimization method with a learning rate of 0.001 and optimization parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Initially the network was trained without a number of epochs defined but instead used early stopping. The early stopping points achieved during repeated trainings of the network for optimization determined that 25 epochs was sufficient for training. Training was performed on Tesla p100 GPUs on the Indiana University Carbonite Deep Learning cluster. It was trained using 5 million jets of equal population across the five classes. The MC samples generation information for the $HH \rightarrow bbWW$ samples are listed in Table training samples is listed in Table 5.1 with the other processes in Table 6.2.

4.5 Performance

The network produces probabilistic responses, y_i , for each of the five classes. These responses are combined into a single log response

$$\mathcal{R}_{WW} = \log \left(\frac{\hat{y}_{WW}}{\hat{y}_{di\,jet} + \hat{y}_{t\bar{t}} + \hat{y}_{V\,jets} + \hat{y}_{b\bar{b}}} \right), \quad (4.14)$$

where $\mathcal{R}_{WW}$ is the combined response of the network for WW jets. The distribution $\mathcal{R}_{WW}$ for each type of jet can be seen in Figure 4.15. A first-order comparison of the performance of the tagger can be done by examining approaches that conventionally have been used to discriminate sources of multi-prong jets. Single-variable taggers are often the first attempt at multi-prong tagging. A baseline comparison then is to look at the performance of the network compared to simple cuts on individual features often used for multi-prong discrimination shown Figure 4.15 (b). Not surprisingly, the network significantly outperforms these single variable taggers motivating not only this tagger, but the use in general of MVA techniques for these more complex jet objects.

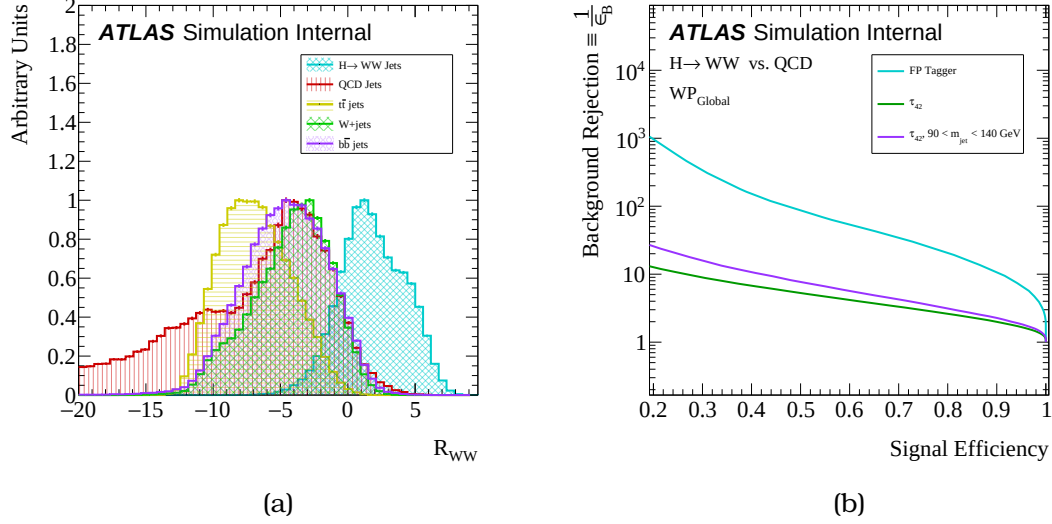


Figure 4.15: (a) The network response R_{WW} for each of the training classes $H \rightarrow WW$, QCD dijet, $t\bar{t}$ jets, W +jets, and $b\bar{b}$ jets. (b) The $1/(\text{QCD dijet rejection})$ vs. signal efficiency is shown comparing the performance to the single variable discriminate τ_{42} and again with a mass windows applied that is consistent with the Higgs mass.

Performance vs, baseline, each branch

The individual performance of each branch compared to the full network can be seen in Figure 4.16. Both branches have performance levels more than double that of the single variable tagger from Figure 4.15b. The convolutional branch demonstrated around 25% improvement over the feed-forward branch as well but much more striking is the almost order of magnitude improvement in performance from the combination of the two branches.

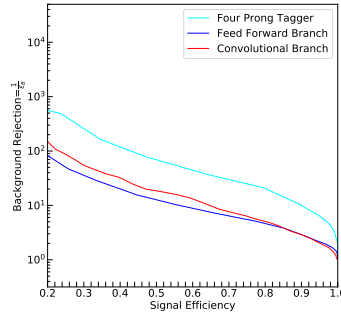


Figure 4.16: Comparison of the performance for each individual branches and the full four-prong tagger neural network.

4.5.1 Working point definitions

The working points for the tagger are defined by the efficiency for tagging $H \rightarrow WW$ jets. The actual value for the WP's are the $\mathcal{R}_{WW}$ values corresponding to the desired efficiency. Initially, a single working point for the entire p_T range was used. However it was observed that the p_T dependence of the network resulted in significant performance differences between the low and high p_T ranges. This is understandable since for boosted objects the p_T range is broad and the performance should be defined in such a way that it reflects the tagging of boosted four-prong jets, not simply jets with high transverse momentum. The steps described in 4.4.1 taken to reduce the jet p_T influence did not sufficiently normalize the performance across the broad p_T range. Two Working Points (WP) are defined for the tagger, *Loose* (80% signal efficiency) and *Tight* (50% signal efficiency). These two WP percentages were selected based on precedent set by other boosted taggers. To ensure constant performance in p_T the values for R_{WW} for each working points are determined in bins of p_T and are shown in Table 4.2. Figure 4.17 shows the response distribution for the global and one bin of the p_T binned WPs with all distributions shown in Appendix B.

p_T [GeV]	300 - 600	600 - 900	900 - 1200	1200 - 1500	1500 - 1800	1800 - 3000
R_{WW} (Global WP)	1.70					
R_{WW} (Binned WPs)	-0.311	0.089	0.758	1.583	2.724	4.154

Table 4.2: The network response values for each p_T dependent WP.

The binning that is used to determine these p_T -dependant working points is the same as the binning described used in the calibration of the tagger described in Section 5.2. This selection is a compromise between a global working point that had the concerns described above and the limitations of lower statistics in higher p_T bins. A comparison of the *Tight* WP efficiencies for a global and binned p_T approach on the nominal training sample shown in Figure 4.3.

Since the *Tight* WP has a 50% signal efficiency, in the p_T binned method, each bin has a resulting efficiency very close to 50%. This is particularly helpful in increasing the signal efficiency at the lower p_T values, though it is worth mentioning that this results in

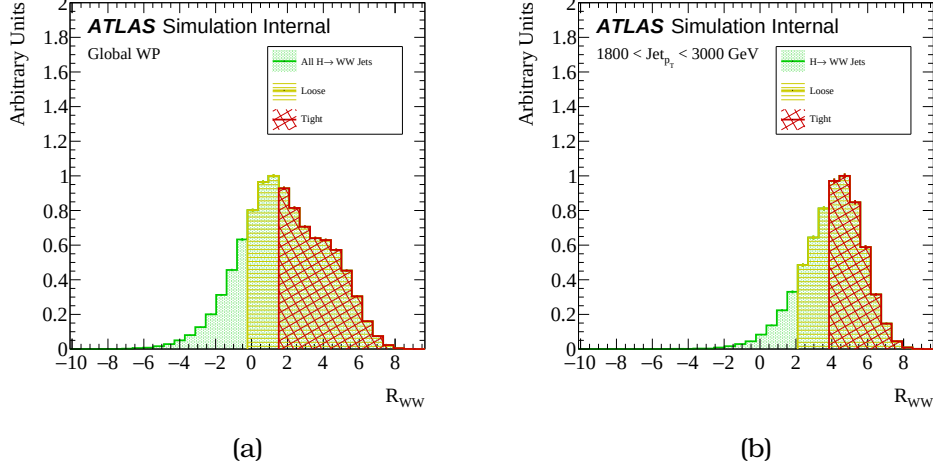


Figure 4.17: (a) A distribution of the network response, $\mathcal{R}_{WW}$, for the $H \rightarrow WW$ process illustrating the global WP. The selection of jets that pass the *Loose* and *Tight* WP's are shown in shaded regions of the distribution. (b) The same response as (a) but shown for the p_T bin $1800 < \text{Jet}_{p_T} < 3000$ GeV. All six distributions are show in B.2.

p_T [GeV]	300 - 600	600 - 900	900 - 1200	1200 - 1500	1500 - 1800	1800 - 3000
Eff. (Global WP)	0.033	0.117	0.244	0.470	0.745	0.904
Eff. (Binned WPs)	0.520	0.518	0.530	0.506	0.507	0.506

Table 4.3: The efficiencies for the Tight WP per p_T bin for all signal jets using a global or individual working points for each p_T bin.

a *reduction* in signal efficiency at higher p_T values since around 2000 GeV is the optimal p_T for the global working point.

4.6 Limitations and Extensions

Throughout the development of the tagger different biases were discovered and steps were taken, where permitted or possible, to reduce these biases. Even in the cases where reduction was not possible or practical, it was important to at least understand these biases or shortcomings in order to better explain the scope of its use as well as to explore possible methods of extending its applications.

4.6.1 Higgs mass

Since the tagger was trained on four-prong jets originating from the decay of a Higgs boson, the mass of the Higgs is a learnable piece of information and this is indeed case

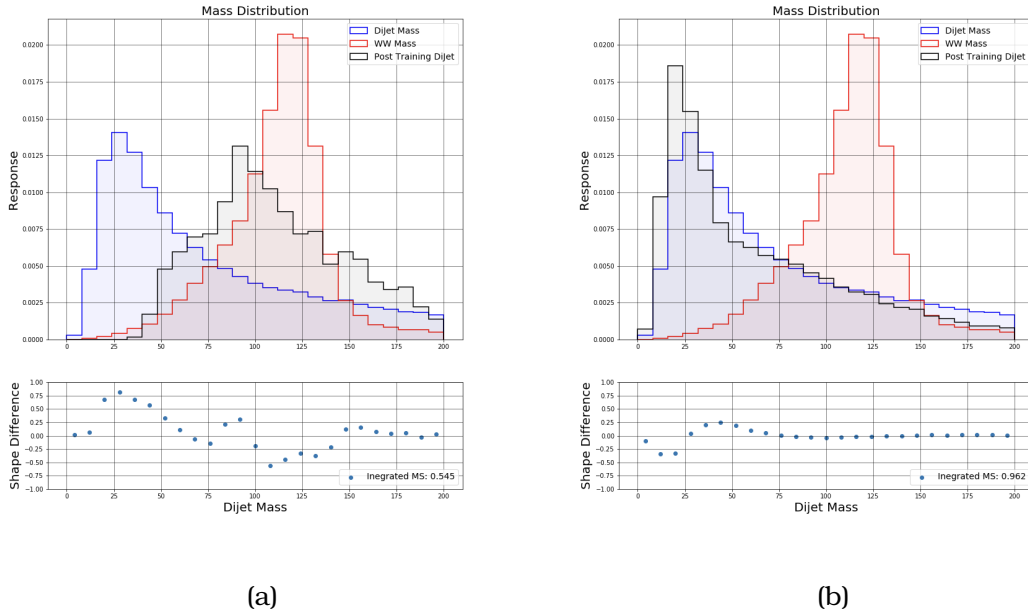


Figure 4.18: The sculpting of the m_{TAR} for QCD jets. Before training is shown in blue, jets incorrectly tagged as signal shown in black demonstrating a more signal-like, red, mass distribution.

for the four-prong tagger shown in Figure 4.18a. The reason this could be considered a limitation of the network is that, without retraining, it is biased to select jets with masses nearer the Higgs mass, regardless of their source. This is particularly interesting given that the jet mass is not a direct input to the network. For a more generally applicable four-prong tagger the performance of the network would need to be independent of the mass of the jet's decay parent. A method for removing, or reducing, the ability for the network to learn the mass of the Higgs boson, or in general to suppress the ability of a neural network to learn any undesirable piece of information, is through the use of adversarial learning. A method of adversarial learning developed and applied to physics studies [137] was adapted and used to train the tagger adversarially. Figure 4.18b shows the significant reduction in the sculpting of the Higgs mass in "signal like" QCD jets due to adversarial training. Though the tagger demonstrated the capability to suppress the learning of the Higgs mass, the performance suffered in the intended analysis use cases. In the end, the desire to optimize the tagger to maximized performance for $H \rightarrow WW$ tagging meant not training the tagger adversarially.

Chapter 5

FOUR-PRONG JET CALIBRATION

As was briefly covered in Section 3.3.3 it is necessary to calibrate objects and tools that are used on data to the MC samples used to develop them. In many instances this is a rather straight forward, if difficult, task that involves calculating scale factors between data and MC for either the objects themselves, or on efficiencies of selection. Unfortunately in the case of four-prong jets, there does not exist a sufficient number of data samples to perform this scale factor calculation directly. A new method for calibrating these objects was developed and is presented in the following sections.

5.1 Calibration strategy

The calibration strategy for the four-prong tagger uses differences in the efficiency of the tagger on QCD data as well as the uncertainties that come from different MC sample generators. The data to MC tagger performance uncertainty was achieved by calculating the efficiency of the tagger tagging a single prong object, in other words, a QCD jet. Since a four-prong jet is comprised at first order of four individual single prong objects, the primary source of modeling uncertainty for a hadronic four-prong object would come from the physics modeling of these single prong jets. Since this technique is not a perfect characterization of four-prong modeling difficulties, the four-prong MC/data scale factor was chosen to be set to one with an uncertainty equal in magnitude of the difference in efficiency between a QCD enriched data control sample (described below) and the QCD MC.

This addresses the need to calibrate the tagger performance to data, however there is

still the issue of tagger performance on the four-prong MC itself. For this, MC samples that are derived using different variations in the physics modeling assumptions and the geometric modeling of the ATLAS detector are derived to construct a broad, though in no way exhaustive, phase space that encompasses many of the assumptions made in the generation of ATLAS MC data. A list of these variations can be found in Table C.1.

5.2 Scale Factor uncertainty

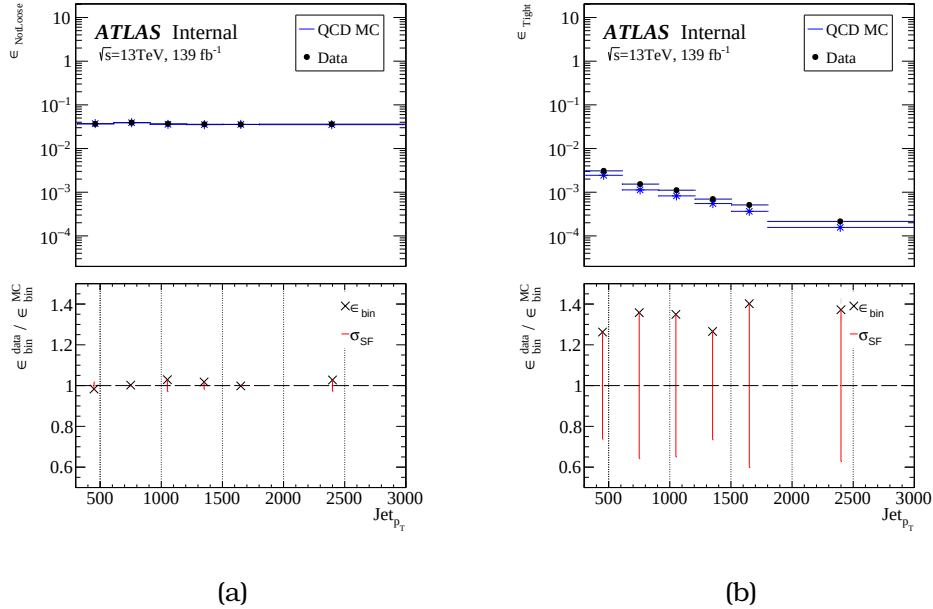


Figure 5.1: (a) Efficiency, ϵ , for data and MC QCD jets are shown in the top histogram for the *Loose* WP. The bottom ratio in the efficiencies $\epsilon_{\text{bin}}^{\text{data}} / \epsilon_{\text{bin}}^{\text{MC}}$ with the difference from 1.0 representing scale factor uncertainty. In (b) the sample relationship is shown for the *Tight* WP.

To reiterate for clarity, the scale factors for the four-prong signal is are assumed to be 1. The uncertainty in these scale factors are then derived from QCD dijet modeling discrepancy from data. The comparison of collision data to QCD jets requires as clean a selection of MC jets as possible. Since the tagger is designed and trained for boosted objects, all the jets that were selected for the scale factor uncertainty calculation were required to have a $p_T > 300$ GeV, however there is a significant drop in the efficiency for leading jet triggers at this p_T so leading jets p_T s are required to greater than 450 GeV. All leptons are vetoed, and since there is a bias of the tagger towards the mass

of the Higgs at 125 GeV, a cut on the mass of the jets is placed at < 90 GeV. What remains is a dijet selection with a very high purity. Both *Loose* and *Tight* working points are tested for their efficiency. The binning selected for the calibration region is such that there are a maximum number of bins while keeping the statistics per bin at a reasonable population. This is complicated by the fact that tagger background rejection is quite high, particularly so for regions where dijet and collision statistics are already low. Figure 5.1 shows the p_T distribution for MC and data in the upper histogram and the ratios of the two in the lower plot. The low efficiency is understandable as four-prong-like objects are exceedingly rare in QCD jets. The scale factor uncertainty is calculated by taking the difference between 1 and the ratio of dijet MC to data. The values for the uncertainties are summarized in Table 5.3. Given that these uncertainties are derived from a scale factor of data and MC, the larger uncertainties observed for the *Tight* working point reflects the quality of the modeling description of QCD jets. The more signal like the QCD jets are from a four-prong jet perspective the more illuminated the general uncertainty in modeling these QCD in this previously unexplored modeling phase space becomes.

5.3 MC uncertainty

As discussed in Section 2.8.1 the ATLAS collaboration uses Geant4 to generate simulated MC for use in analyses. MC samples are nominally generated using recommended parameters for general use throughout the collaboration and typically this is sufficient for a physics analysis. It could be though that the four-prong tagger is successful, even dependent, on assumptions used to generate the nominal MC samples. In order to estimate the uncertainty due to the MC generator choice, many difference MC sample alternatives were generated using ATLAS physics lists summarized in Appendix C. A nominal sample set based on the recommended inputs was generated for resonant signal with $m_X = 800, 1000, 1400, 2000, 2500, 3000, 4000, 5000$ GeV. The tagger was trained on this sample set to provide a baseline of its performance. The p_T Range of 300-3000 GeV was selected for the jets, with no other cuts applied. The efficiency for each of the vari-

Process	Generator	Parton Shower	PDF
ggF $X \rightarrow HH \rightarrow bbWW$	Madgraph5_aMCatNLO [138]	Herwig7 [79]	NNPDF23_lo_as_0130_qed [139]

Table 5.1: Details of the nominal di-Higgs MC samples used for the MC uncertainty derivation.

ations was determined using the nominal sample trained network and working points and these are recorded in Table 5.2. Throughout all p_T bins and variations, the tagger remains relatively consistent. The difference between the efficiencies of each variation and the nominal are calculated then summed in quadrature to generate the uncertainty in the four-prong signal efficiency. This uncertainty determination is performed for both working points with the *Tight* efficiencies shown in Table 5.2 and *Loose* in Table B.1.

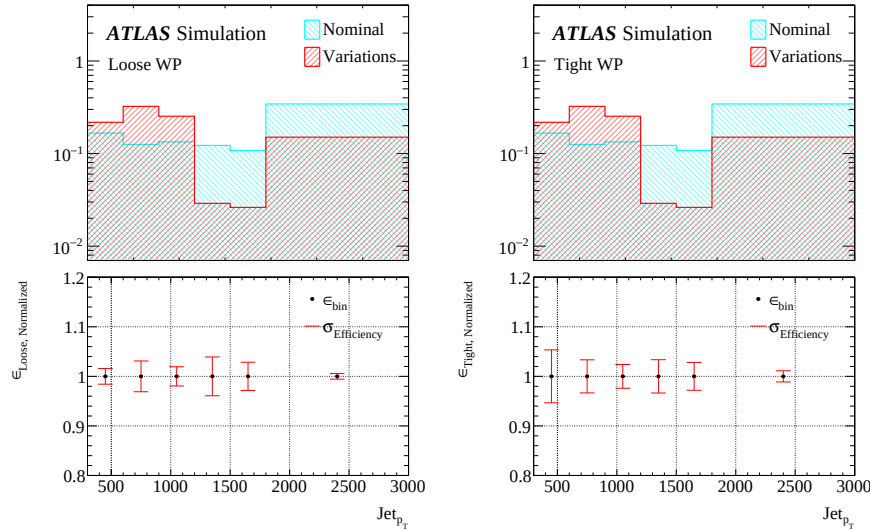


Figure 5.2: The p_T of the $H \rightarrow WW$ jets are shown in the upper histograms with the same binning that was applied in the QCD scale factor uncertainty calculations. In the lower histograms the normalized efficiency per p_T bin of the tagger is shown. In red are the uncertainties in these efficiency values derived from the efficiencies in the variations in Table 5.2.

The uncertainties for both working points were similar and relatively small with a maximum of 3.9% for *Loose* and 5.3% for *Tight*. The per p_T bin uncertainty for both working points along with the relative p_T distribution are shown in Figure 5.2.

Bin p_T [GeV]	300 - 600	600 - 900	900 - 1200	1200 - 1500	1500 - 1800	1800 - 3000
nominal	0.5195	0.5182	0.5297	0.5059	0.5074	0.5063
s3295	0.5305	0.534	0.5214	0.4911	0.5274	0.5126
s3296	0.5178	0.525	0.5137	0.4971	0.5211	0.52
s3297	0.5353	0.5135	0.5177	0.5197	0.528	0.5166
s3298	0.5225	0.5333	0.501	0.5171	0.5349	0.5153
s3299	0.5342	0.532	0.5001	0.5126	0.5349	0.517
s3373	0.5285	0.5209	0.5181	0.5156	0.5322	0.518
s3374	0.5332	0.5257	0.504	0.5103	0.5445	0.5212
s3375	0.5319	0.5176	0.5009	0.493	0.5296	0.5155

Table 5.2: Efficiency of each signal sample variation studied and used to generate the uncertainty in four prong tagging.

5.4 Four-Prong Tagger Efficiency Uncertainty

The SF and the MC uncertainties, σ_{SF} and σ_{MC} , are combined in quadrature to determined the overall uncertainty in the tagger efficiency σ_{TOT} . This uncertainty is incorporated in jets that are tagged by the tagger through an adjustment of the weights of each jet when tagged during analysis selection. The individual uncertainties that contribute as well as the total uncertainty for both working points are shown in Figure 5.3 as a function of p_T . The largest source of uncertainty for both working points is the same, originating from the SF uncertainties.

Bin p_T [GeV]	300 - 600	600 - 900	900 - 1200	1200 - 1500	1500 - 1800	1800 - 3000
Loose						
σ_{MC}	1.6	3.1	1.9	3.9	2.9	5.7
σ_{SF}	1.7	0.2	2.9	1.8	1.6	2.8
Total	2.5	3.2	3.6	5.0	4.0	3.0
Tight						
σ_{MC}	5.3	3.3	2.4	3.4	2.8	1.1
σ_{SF}	26.3	35.8	34.9	26.5	40.2	37.2
Total	26.8	36.0	35.0	26.9	40.4	37.3

Table 5.3: Summary of the uncertainty values per p_T bin for each of the sources considered, as well as the total uncertainty coming from the combination of the two.

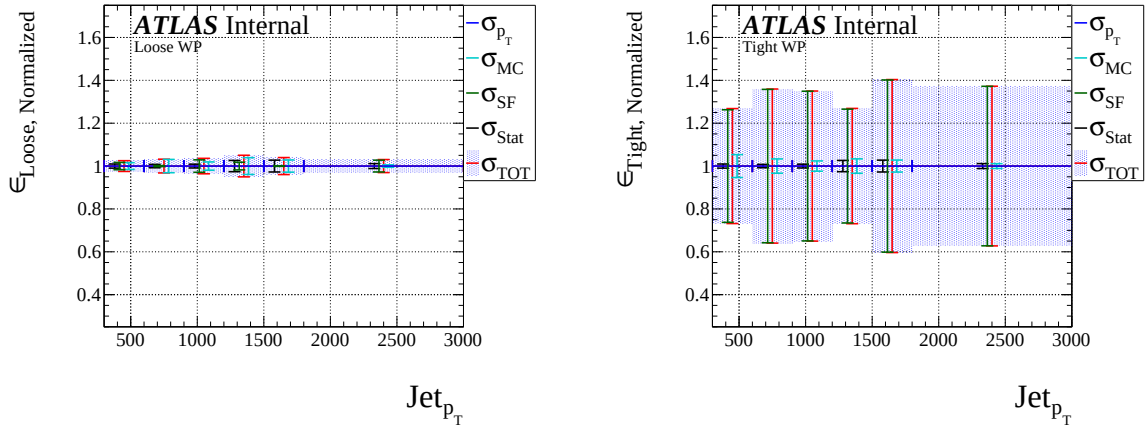


Figure 5.3: The total uncertainty σ_{TOT} for each p_T bin as well as each contributing source of uncertainty of efficiency for relative size comparison.

Chapter 6

SEARCH FOR BOOSTED RESONANT DI-HIGGS PRODUCTION

6.1 Overview

The boosted, all-hadronic $bbVV$ final state of $X \rightarrow HH \rightarrow bbVV$ has a complicated topology. As discussed in Sections 1.12 and 4.1, the two V bosons resulting from the Higgs boson decay are not individually resolvable. The reasoning for this is equally true for the two b -quarks resulting in a final state that consist of two TAR jets, one containing the decay products from the V 's and one the decay products from the b 's. The kinematic signatures of W and Z bosons are not distinguishable enough in this analysis; therefore W 's and Z 's are considered as vectors bosons instead of a more specific state of $bbWW$. The primary challenge for this analysis is distinguishing the final signal state from a large, signal-like background. The bb jet has a strong discriminant in the b -quarks as discussed in Chapter 3; however, the VV jet does not. In Section 4.1.1 the motivation for treating the $H \rightarrow VV$ process as a four-prong object and tagging it as such for identification is discussed. In this Chapter the previous results for resonant $bbVV$ limits are reviewed followed by a description of the analysis' methodology and the results of this search.

6.1.1 Previous resonant di-Higgs Results

The previous results for resonant di-Higgs production for the $bbWW$ channel for 27.5 fb^{-1} of integrated luminosity [140] are shown in Figure 6.1a. From this figure the best limit for the cross section of the $bbWW$ channel is $\sim 100 \text{ fb}$. More recent results for a variety of other channels at up to 139 fb^{-1} are shown in Figure 6.1b with the current combined

limit set at ~ 1 fb at $m_X = 3000$ GeV.

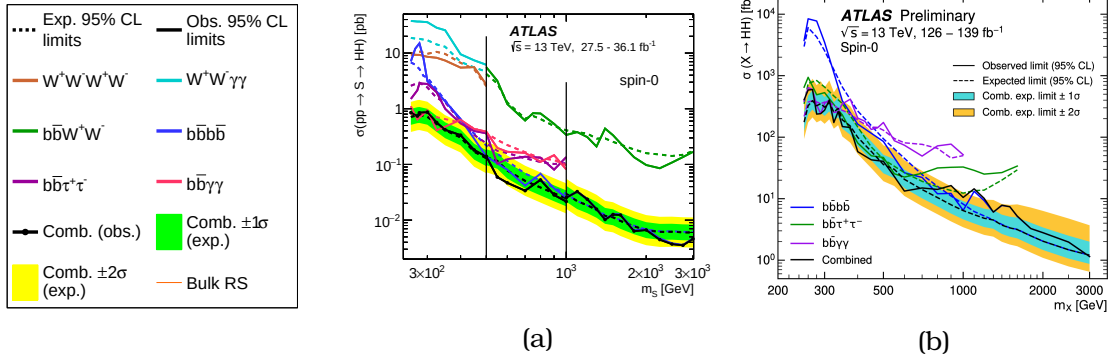


Figure 6.1: (a) The 27.5 – 36.1 fb $^{-1}$ limits for the resonant di-Higgs production in each individual channel as well as the combination. (b) The most recent results for resonant di-Higgs production for 126 – 139 fb $^{-1}$.

6.2 Data and simulation samples

The collisions of the protons within the LHC are ultimately analysed for the signal process; however, it is necessary to understand the makeup of the data to be able to successfully separate signal from background. With this reason in mind, simulated MC samples are generated for possible background processes in order to determine which require significant consideration.

6.2.1 Data samples

The data set used in this analysis contains all collision data collected by the ATLAS detector through Run 2. This included all data collected from 2015 - 2018. The total integrated luminosity for the data is 139.0 ± 2.4 fb $^{-1}$ [141].

Signal MC

The simulated MC samples for the process $X \rightarrow HH \rightarrow bbWW$ were created using Madgraph5_aMCatNLO [75] to generate the initial hard scatter event. The input specifications for the scalar particle were a spin = 0 particle with a resonant mass at the specified value with a width of 10% allowed for a natural mass width. The decay of the scalar to two Higgs particles was required. The signal sample generator details are listed in

Table 6.1. The signal MC is scaled according to the cross section of the physics process in question. The Two-Higgs-doublet model introduced in Section 1.12.1 is used as the theoretical model in the production of the signal MC. The cross section for this process is assumed to be 1 pb for all masses. This cross section is used to normalize the fit results in Section 6.9.

Process	Generator	Parton Shower	PDF	$\sigma \times \text{BR} [fb]$
$ggF \ X \rightarrow HH \rightarrow bbWW \rightarrow qq\bar{q}\bar{q}$	Madgraph5_aMCatNLO	Herwig7 [79] + EvtGen [142]	PDF23LO	116.19

Table 6.1: Definitions for generating the resonant signal MC samples.

The resonant masses for the signal samples are 800, 900, 1000, 1200, 1400, 1800, 2000, 2500, 3000, 4000, 5000 GeV. While not exhaustive, this selection provides a thorough scan of the mass phase space.

Background MC

The main background processes considered for the analysis final state phase space were QCD multijet, $t\bar{t}$, single-top, W/Z + jets, and di-boson (VV). These samples are estimated using MC in the same manner as the signal. A full description of the simulated background samples can be found in Table 6.2.

Process	Generator	Parton Shower	PDF
QCD DiJet	Pythia8 [77] + EvtGen	Pythia8 + EvtGen	NNPDF23LO [139]
$t\bar{t}$	Powheg [76]	Pythia8 + EvtGen	NNPDF23LO
Single-Top	Powheg	Pythia8 + EvtGen	NNPDF23LO
DiBoson	Sherpa [78]	Sherpa 2.2.1	NNPDF30NNLO [143]
W+jets	Sherpa	Sherpa	NNPDF30NNLO
Z+jets	Sherpa	Sherpa	NNPDF30NNLO

Table 6.2: Definitions for generating the MC for the background processes considered.

6.3 Jet Definitions

The jets used for the analysis are TAR jets with reclustering parameters matching the recommended values presented in Section 3.3.5 except for the case of the jet R parameter. A value of $R = 0.75$ was found to have the best sensitivity. The jets found in an event are ordered in p_T such that jet_1 is the highest p_T jet, etc. This aids in the event selection

as the WW and $b\bar{b}$ jets are most often the highest- p_T jets in an event, and p_T ordering allows for a more straight forward selection.

6.3.1 b -tagging

The jet collection used for b -tagging are VR track jets detailed in Section 3.3.4 with a minimum $R_{min} = 0.02$ and a maximum of $R_{max} = 0.4$. The scale factor ρ in Equation 3.7 is set to the recommended value of 30. The DL1r tagger discussed in Section 3.3.6 is used to tag each jet at the 77% b -tagging efficiency working point and the number of b Tags is associated to the TAR jet.

6.4 Event Selection

The events to be considered for this analysis were selected using a small set of triggers and selection criteria before the analysis selection is applied.

6.4.1 Triggers

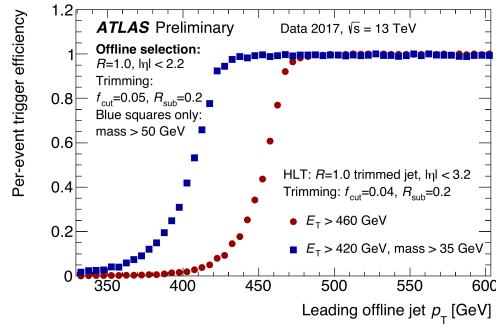


Figure 6.2: Efficiency curve for two large-R jet triggers demonstrating the efficiencies as well as the effect of jet mass cuts on the efficiency plateau.

The final state includes two boosted jets. Single large-R jet triggers are used to select events that have at least one large-R jet in the final state. A full list of the large-R jet triggers used to select events for this analysis is shown in D.1. An example trigger, showing how the naming convention describes its criteria is given in Table 6.3. The efficiency curve for two single large-R jet triggers can be seen in Figure 6.2. Most significant about

this efficiency is that it is essentially 100% for large-R jets with a greater than 35 GeV and $p_T > 450$ GeV.

HLT_j420_a10t_lcw_jes_35smcINF_L1SC111	
Chain piece	Description
HLT_j420	HLT single jet with $p_T > 420$ GeV
a10t	anti- k_T reclustered, trimmed jet with radius $R = 1.0$
lcw	Local calo weighted clusters
jes	JES calibration applied to jets
35smcINF	Jet mass cut > 30 GeV
L1SC111	Level 1 trigger; $E_T > 111$ GeV; implied $ \eta < 3.1$ since no explicit η reference

Table 6.3: Breakdown of the naming scheme used and the description of each pieces in the identifier chain used for triggers.

6.4.2 Event criteria

The criteria for the events is applied in order to thin the data. The final state requires two TAR jets. The following cuts are the applied for initial event selection: ≥ 2 TAR jet, $p_T > 100$ GeV, and $|\eta| < 2$. The $|\eta|$ cut is to ensure that the jets are within the barrel of the detector for maximum resolution.

6.4.3 Signal Selection

Finally, after the pre-selection, the signal event selection is applied to select events that match the desired signal process.

1. Leading TAR jets are required to have a $jet_{p_T,1} > 450$ GeV, and subleading $jet_{p_T,2} > 300$ GeV.
2. A mass cut is applied to both leading and subleading jets of $jet_m > 90$ GeV
3. Events with a third TAR jet with $jet_m > 50$ GeV are vetoed
4. One TAR jet must be DLR1 tagged as containing 1 or 2 b -quarks using the 77% working point and is labeled the $b\bar{b}$ candidate, while the other TAR jet must contain 0 b -quarks using the same 77% working point.

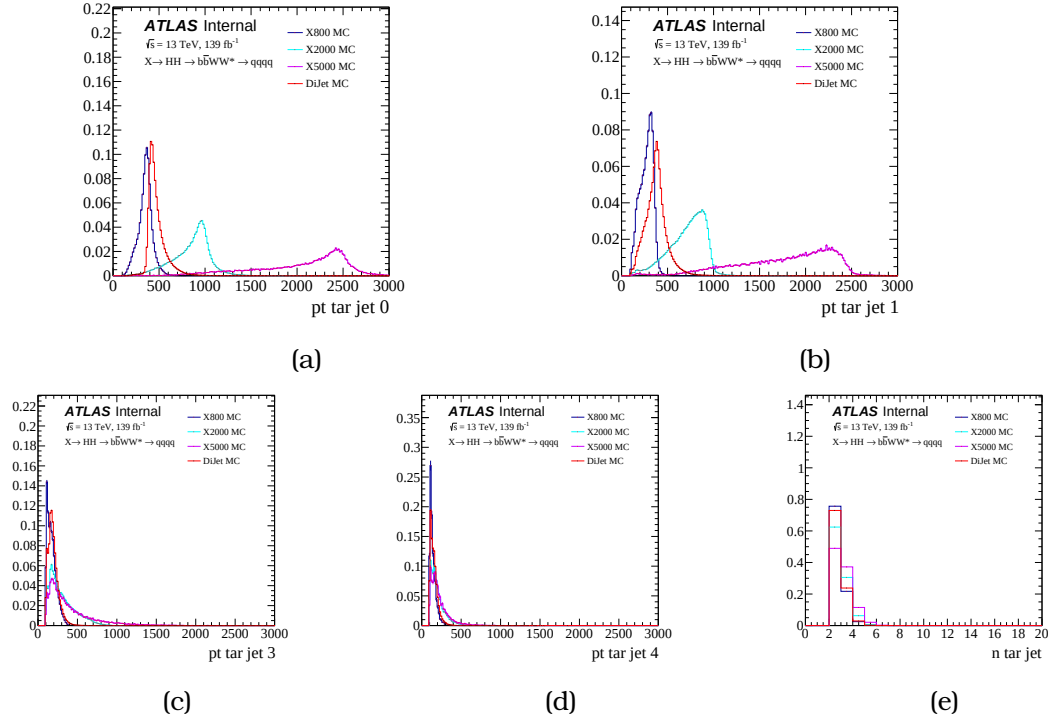


Figure 6.3: p_T distribution for the first four jets in MC for dijet and three signal masses. Jet multiplicity after the first 2 TAR jet cut is also shown in (e).

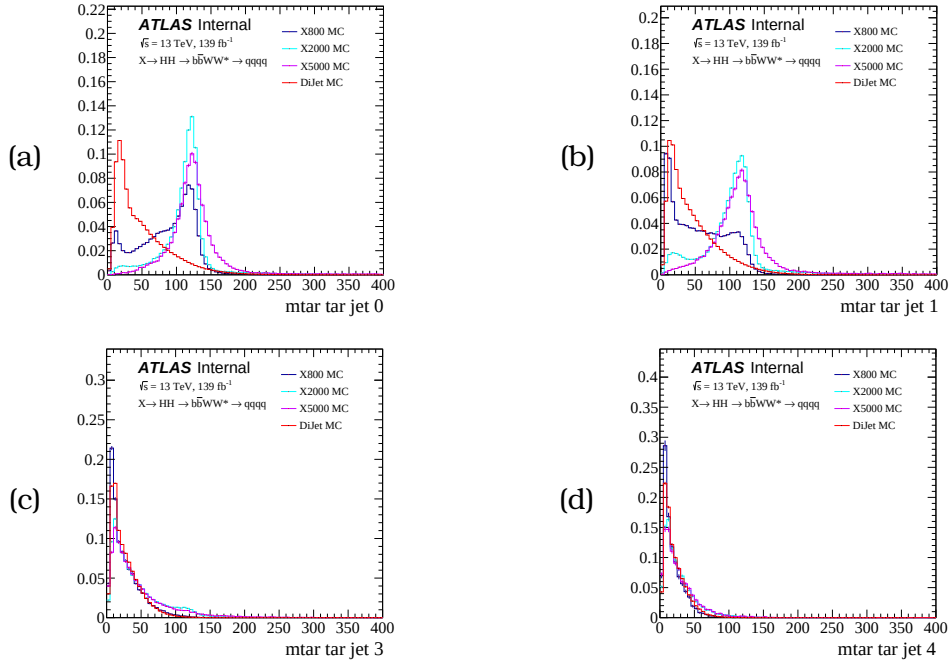


Figure 6.4: TAR jet mass for the first four jets in the events in MC for dijet and three signal mass samples. The Higgs mass peak is clearly present in the leading and sub-leading jets in the signal MC.

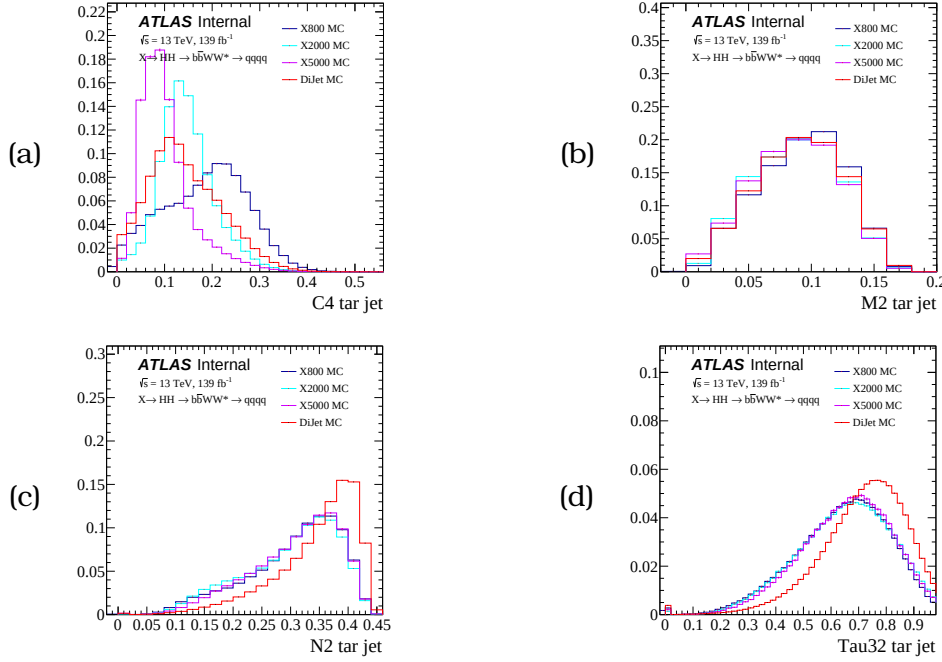


Figure 6.5: Four-prong tagger features for dijet and three signal mass samples MC jets.

5. The rapidity separation between the $b\bar{b}$ Candidate and the remaining jets, ΔY , must be less than 1.25
6. Finally, the VV candidate is the non- b tagged jet that passes the four-prong tagger *Tight Tight* working point

This selection results in two signal regions, both requiring a TIGHT tagged TAR jet and a b -tagged TAR jet; however, this b -tagged jet can have either 1 or 2 b -tags. As Table 6.4 shows, the dominant process contributing to the background in the analysis is QCD dijet events. The second-most populous background comes from top phenomena, $t\bar{t}$ and single top. The relatively small contribution for secondary processes motivates a QCD multijet-only background estimate to be fully data driven and the methodology for the background estimation covered in Section 6.6.

6.5 Analysis regions

The analysis signal regions are defined by two primary characteristics, the b and four prong tagging of the two leading jets. Jet's with both 1 or 2 b tags from the DL1r tagger

are used, each defining a different region. Both regions required the WW candidate to pass the *Tight* four-prong tagger selection. Three types of regions are defined. The control regions CR0-CR4 are used to independently validate the selection variables with minimal overlap. More importantly these control regions are used in the extrapolation of the data to the signal region in the background estimation method outlined in Section 6.6. The validation regions VR1 and VR2 are used to validate the estimation as well as for uncertainty extraction. The signal regions SR1 and SR2 are the regions in which the signal strength is maximized and the analysis is performed.

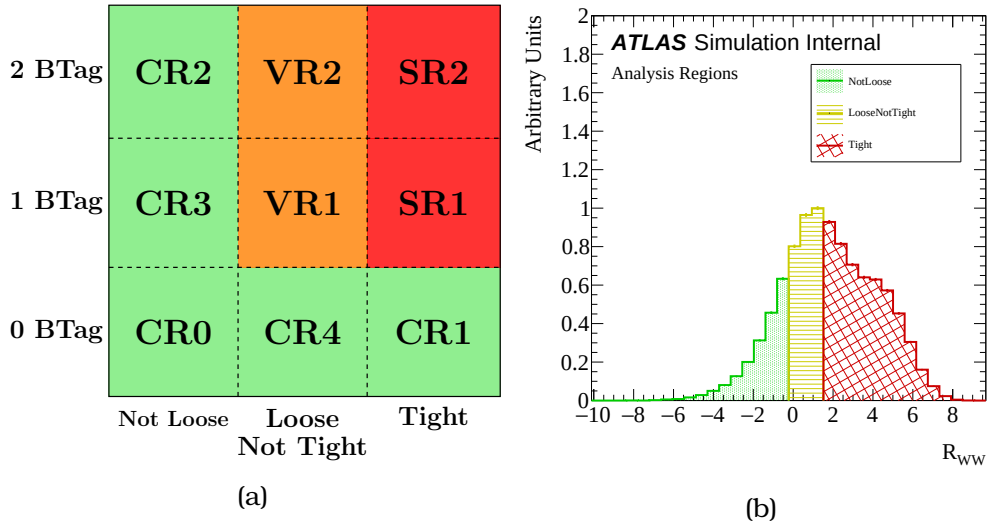


Figure 6.6: (a) Diagram of the analysis regions. The y-axis is defined by the number of BTags of the bbCandidate. The x-axis is defined by the *Loose* and *Tight* Working Points of the four-prong tagger. (b) An illustrative example of the R_{WW} response plot showing the regions of the tagger response for each WP.

6.5.1 Analysis for each scalar mass

Notice that in Figure 6.6b the horizontal axis is defined by the response from the four-prong tagger. This definition isn't as straight forward as the number of b -Tags due to the parameterized nature of the tagger's neural network. By design the response from the network is a function of the scalar mass value of the jet since this is the parameterizing feature of the network. Practically this means that the analysis must be run for each scalar mass considered. This only effects the values of the selection criteria for the *Loose* and *Tight* definitions of the four-prong tagger, but doesn't have any impact on the

overall analysis strategy. The presentation of the analysis is simplified, using subsets of the eleven scalar mass samples to illustrate the different steps in the analysis process. The details of the analysis of these mass points is presented in this chapter in their entirety. All other masses are documented in the Supplemental Files.

6.5.2 Composition

The background processes discussed in Section 6.2.1 are all the possible backgrounds that have to be considered when devising a strategy for background estimation within the analysis regions. Figure 6.15 shows the relative composition of the background in the signal regions and Table 6.4 shows the MC yields. Overwhelmingly the signal regions are dominated by QCD dijet. Due to the complexity of modeling QCD MC as well as the relatively low proportional contribution from secondary background sources, the background for the analysis is taken to be QCD multijet exclusively. Any bias introduced by the error in this assumption is included in the uncertainty in the background estimate outlined in Section 6.6.2.

Process	$m_X = 1000$		$m_X = 2000$	
	SR1(%)	SR2(%)	SR1(%)	SR2(%)
DiBoson	4 (0.08)	0.23 (0.03)	3.2 (0.15)	0.52 (0.16)
V+jets	67 (1.326)	7.5 (0.92)	51 (2.4)	6.00 (1.8)
Single Top	260 (5.217)	29 (3.6)	142 (6.7)	18 (5.3)
$t\bar{t}$	320 (6.389)	43 (5.3)	140 (6.7)	22 (6.7)
Dijet	4400 (87)	740 (90)	1800 (84)	290 (86)
	$m_X = 3000$		$m_X = 5000$	
	SR1(%)	SR2(%)	SR1(%)	SR2(%)
DiBoson	3.1 (0.2)	0.42 (0.19)	3.3 (0.23)	0.42 (0.22)
V+jets	40 (2.5)	4.6 (2.1)	40 (2.8)	4.5 (2.4)
Single To	100 (6.5)	14 (5.9)	90 (6.2)	11 (5.8)
$t\bar{t}$	96 (6.1)	18 (8.2)	90 (6.2)	17 (8.9)
Dijet	1300 (85)	180 (84)	1200 (85)	150 (83)

Table 6.4: MC background yields and relative yields in signal regions SR1 and SR2 for 1000, 2000, 3000, and 5000 GeV mass points.

6.6 Background estimation

A fully data driven background estimation method was chosen to estimate the background contribution in the signal regions of the analysis. This uses data in the place of dijet MC as is beneficial because of the difficulty in modeling QCD jets. In order to

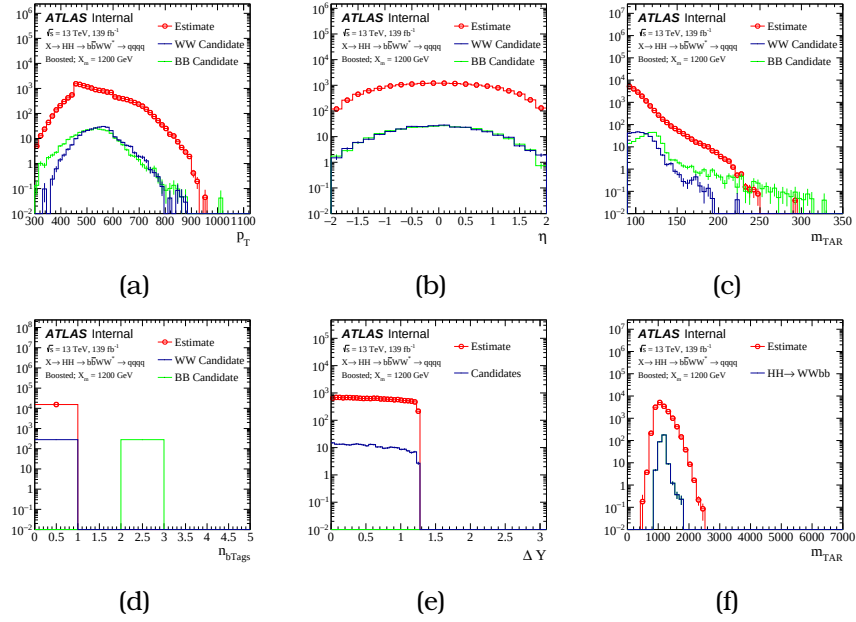


Figure 6.7: The kinematic distributions of (a) p_T , (b) η , (c) jet mass, (d) number of b tags, (e) ΔY , and (f) the HH candidate mass for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region.

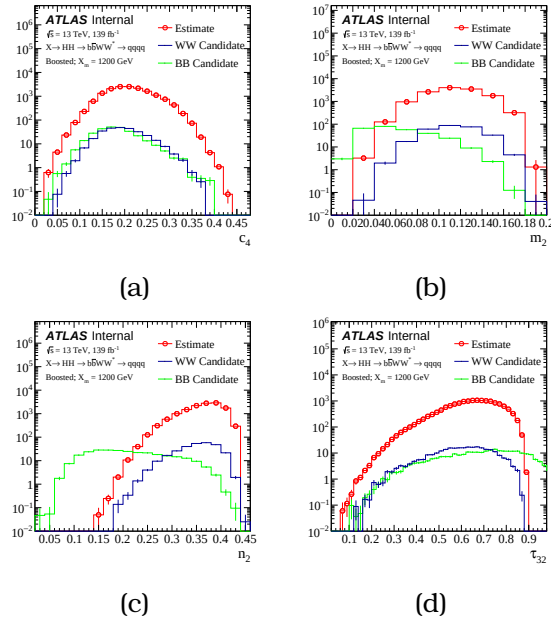


Figure 6.8: The distributions for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}

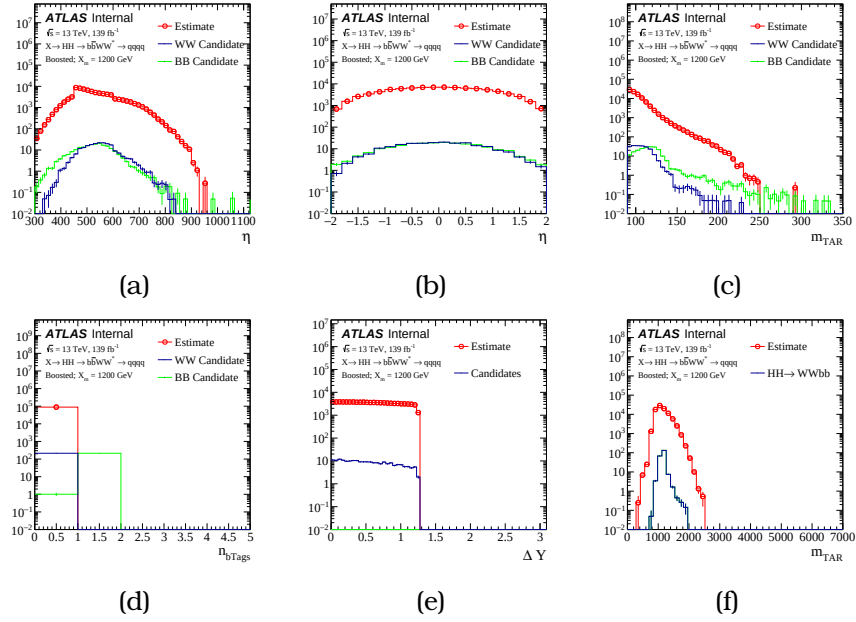


Figure 6.9: The kinematic distributions of (a) p_T , (b) η , (c) jet mass, (d) number of b -tags, (e) ΔY , and (f) the HH candidate mass for the $b\bar{b}$ and WW jet candidates from the 1200 GeV signal mass samples in the 2 bTag signal region.

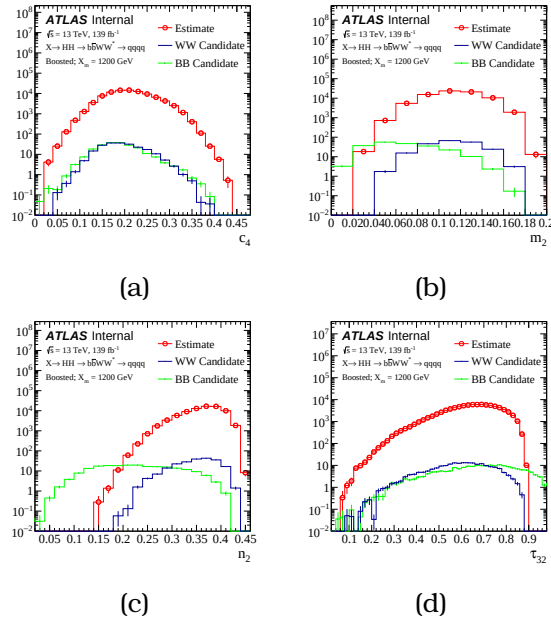


Figure 6.10: The distributions for the $b\bar{b}$ and WW jet candidates from the the 1200 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32} .

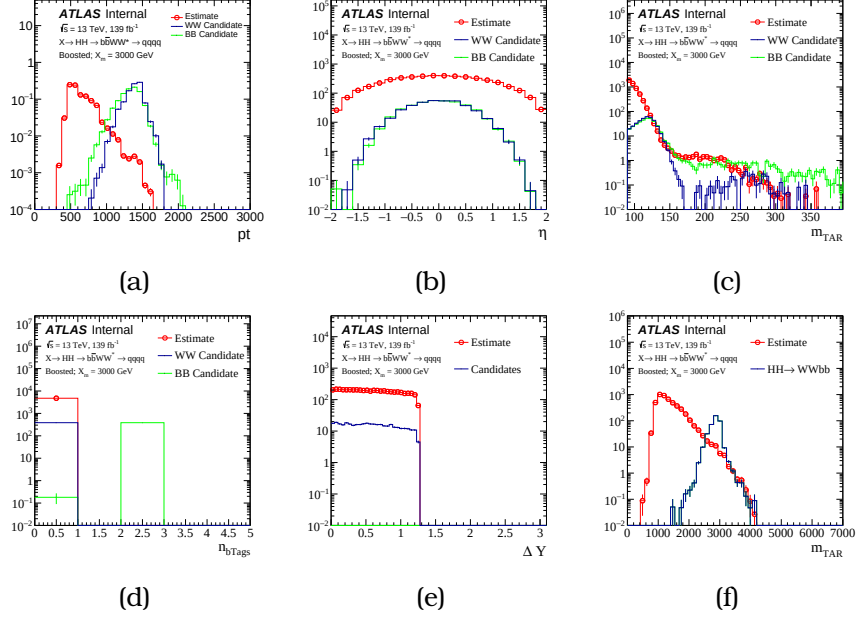


Figure 6.11: The kinematic distributions of (a) p_T , (b) η , (c) jet mass, and (d) number of b tags for the $b\bar{b}$ and WW jet candidates from the the 3000 TeV signal mass samples in the 2 bTag signal region.

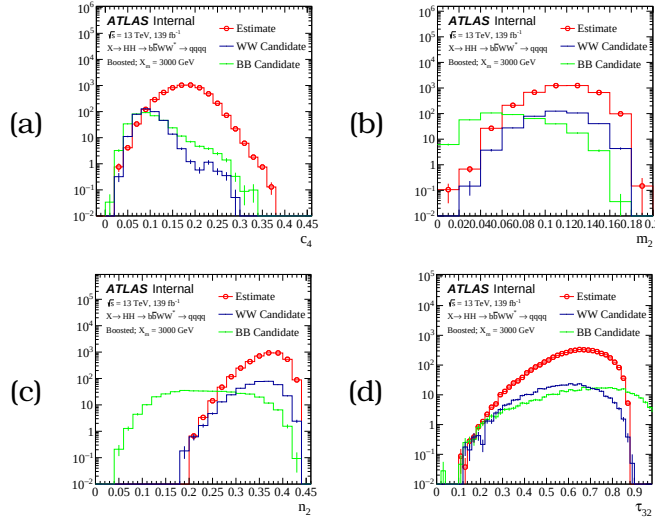


Figure 6.12: The distributions for the $b\bar{b}$ and WW jet candidates from the the 3000 TeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32} .

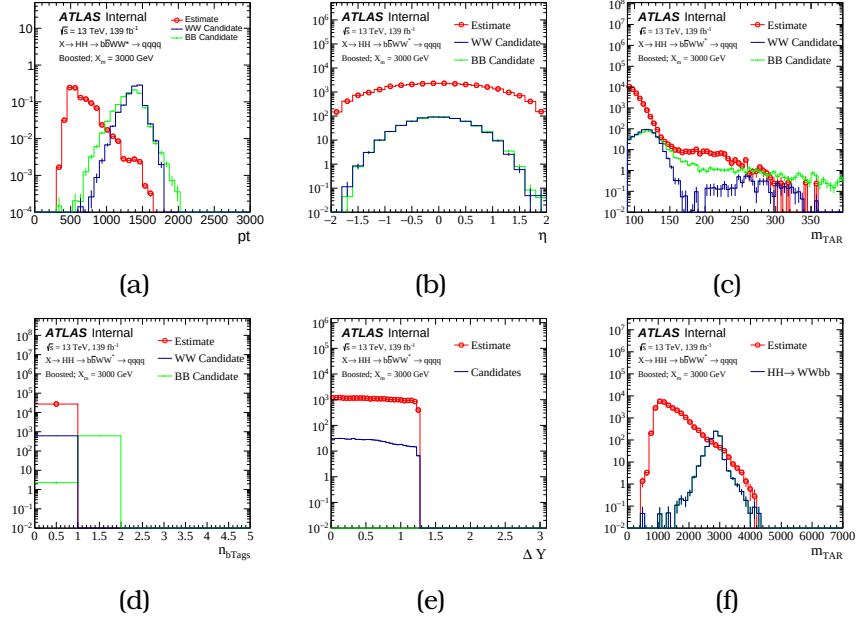


Figure 6.13: The kinematic distributions of (a) p_T , (b) η , (c) jet mass, and (d) number of $b\bar{b}$ and WW jet candidates from the the 3000 GeV signal mass samples in the 2 bTag signal region.

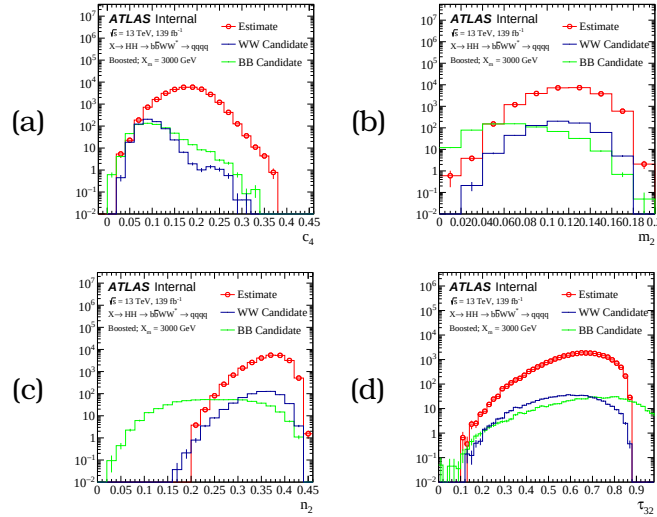


Figure 6.14: The distributions for the $b\bar{b}$ and WW jet candidates from the the 3000 GeV signal mass samples in the 2 bTag signal region of the four substructure features required for the tagger (a) c_4 , (b) m_2 , (c) n_2 , and (d) τ_{32}

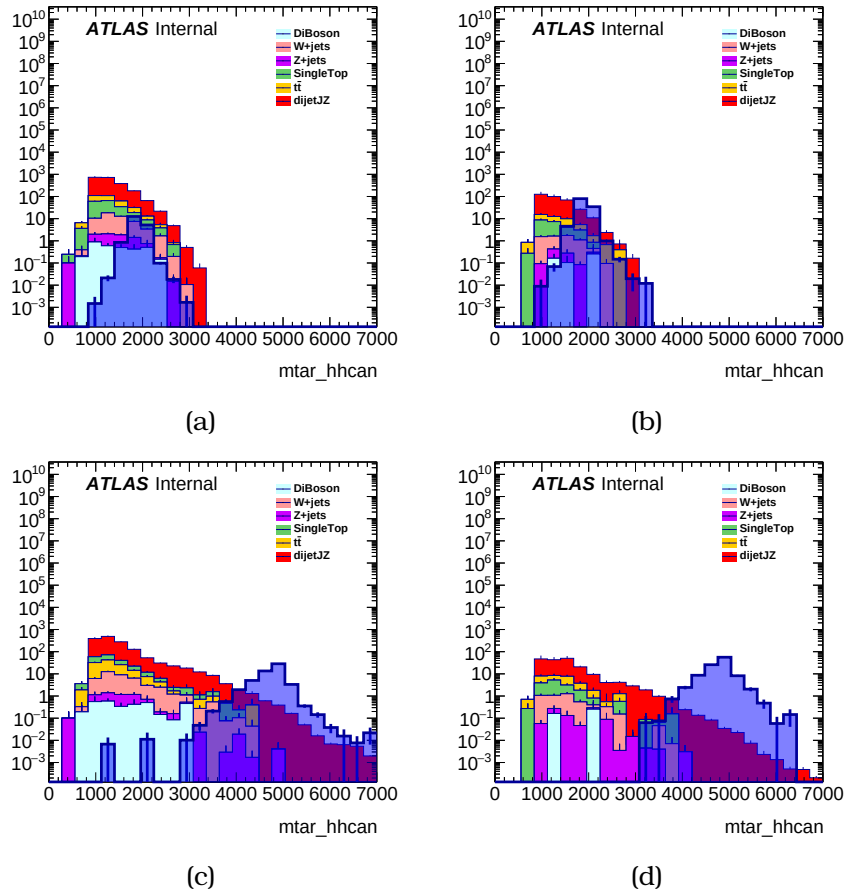


Figure 6.15: MC composition of the background processes considered for the analysis.

estimate the background within the two signal regions, the ABCD method is used to extrapolate from control regions where signal contamination is either low, or non-existent, into the signal regions with the goal of estimating the number of background events.

6.6.1 The ABCD Method

The ABCD method [144] requires two selections that form the control and signal regions. Here the number of b -tags and the four-prong tagger score are used as described by the regions outlined in Section 6.5. The arithmetic approach to determine the estimate assumes the following relationships to be true between the number of events in each analysis region.

$$\text{VR2} = \frac{\text{CR4} \cdot \text{CR2}}{\text{CR0}} ; \text{SR2} = \frac{\text{CR1} \cdot \text{CR2}}{\text{CR0}} \quad (6.1)$$

$$\text{VR1} = \frac{\text{CR4} \cdot \text{CR3}}{\text{CR0}} ; \text{SR1} = \frac{\text{CR1} \cdot \text{CR3}}{\text{CR0}} \quad (6.2)$$

These relationships can only be true if the correlation between the variables that define the ABCD plane are sufficiently uncorrelated. MC is used to verify that the events in the control region C1-C5 are free of signal. Regions VR1 and VR2 are used to validate the effectiveness of the method orthogonality to the signal regions. The validation allows the testing of the validity of the method by comparing the shape of the extrapolated distributions, as well as the extrapolated yields to the real values. The yields for the multijet background estimation and the signal samples for the 2000 and 5000 GeV samples are presented in Table 6.5.

6.6.2 Estimate uncertainty

It is necessary to quantify the uncertainty in the background estimation method outlined above. Table 6.6 shows the differences between the deviation and data yields in the background estimate validation regions VR1 and VR2. From the ratio of the yields it is clear that for smaller scalar mass samples the estimate uncertainty is quite small at $\sim 1\%$ and for larger mass samples it is around $\sim 4\%$. It is also clear from the ratio plots

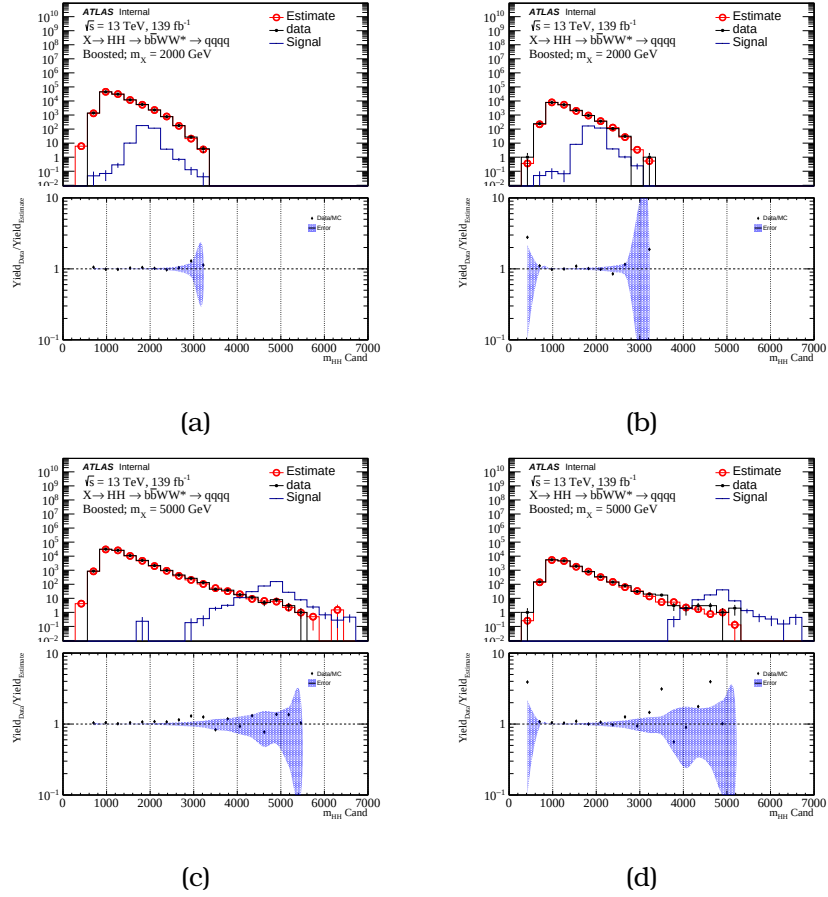


Figure 6.16: The distributions of m_{hh} for the data, background estimation, and signal of the (a) 2 and (b) 5 TeV mass points in both validation regions.

	Signal Region 1-btag	Signal Region 2-btag	Validation Region 1-btag	Validation Region 2-btag
Multijet	39000 ± 3900	6800 ± 680	98000 ± 9800	17000 ± 1700
Signal	460 ± 10	470 ± 10	308 ± 7	300 ± 7.0
Total	39460 ± 3900	7270 ± 680	98000 ± 9800	17300 ± 1700

Table 6.5: Signal and estimate yields in the validations regions for the 2000 GeV scalar mass point.

on the bottom of Figure 6.16 that the ratios are quite consistent with 1 within statistical uncertainties. Since the ratios are consistent and relatively small, the uncertainty in the background estimate is defined globally as 8% for the 4000 and 5000 GeV samples and 2% for all samples $m_X \leq 3000$ GeV. These uncertainties have been inflated from the maximum calculated value in order to account for statistical fluctuations.

Region	Scalar Mass (GeV)	Estimate	Data	Data/Estimate
VR1	2000	97000	98000	1.01
	5000	76000	78000	1.04
VR2	2000	17000	17000	1.00
	5000	13000	14000	1.08

Table 6.6: Data and estimate yields the estimation validation regions VR1 and VR2.

6.7 Uncertainties

The systematic uncertainties for the analysis come from the experimental uncertainties associated with the reconstructed objects, the theoretical uncertainties from the signal model, as well as analysis-specific uncertainties resulting from the methods. Examples of experimental uncertainties are the JER and JES uncertainties of the R-Scan jets from Section 3.3.3 or uncertainties arising from the efficiency of the triggers. A complete list of experimental uncertainties are compiled in Table E.1 and are described briefly below. Analysis-specific uncertainties are necessary for the background estimation discussed in the previous section as well as the efficiency of the four-prong tagger from Section 5.4. All sources of uncertainty and their sizes are summarized in Table 6.7.

6.7.1 Experimental Uncertainties

There are a large number of experimental uncertainties that are centrally produced by the ATLAS collaboration for use in analyses. These uncertainties are associated with the

reconstruction of the physics objects that are used as well as the measured characteristics of the LHC and ATLAS systems.

Luminosity

All cross sections measured at the LHC are subject to uncertainty in the measurement of the luminosity. The integrated luminosity has a combined uncertainty of 1.7% for the complete Run 1 and 2 data taking periods [145].

Trigger

Each trigger that is applied to the data has an associated uncertainty in its efficiency. This source of uncertainty isn't of particular concern given the analysis strategy employed. Figure 6.2 shows that the efficiency plateau for single large-R jet triggers is below the 450 GeV cut imposed on the lead jet leading to an efficiency of 100% for triggers in this analysis.

Leptons

The process for reconstructing leptons described in Sections 3.4 and 3.5 have uncertainties associated with the efficiency of reconstruction and identification. Fortunately, by not requiring a lepton trigger the only source of lepton uncertainty comes from the lepton veto. This has a very low impact on the overall uncertainty in the analysis.

TAR Jets

The TAR jets are effected by reconstruction uncertainties that are derived from the JES and JER calibration of the R-scan jets used to create the TAR jets as well as from the tracks that are associated to them. These uncertainties are calculated bottom up and therefore fully describe the uncertainty of the TAR jets.

6.7.2 Background estimation uncertainty

The background estimation method has a global uncertainty associated with the precision of the estimation yields in the validation regions VR1 and VR2. The value of the

uncertainty is dependent upon the scalar mass of the samples as described in Section 6.6.2.

6.7.3 Four-Prong Tagger Uncertainty

As detailed in Chapter 5 there is an uncertainty associated with the efficiency of the four-prong tagger. Figure 5.3 illustrates the p_T -based uncertainties for the *Loose* and *Tight* WP's.

6.7.4 b -tagging Uncertainty

The tagging of the bb candidate with the DLR1 tagger has an associated quark *flavor tagging* uncertainty. A conservative estimate of 10% is used to give an upper limit on the possible influence of this uncertainty motivated by results derived in similar boosted channels [146].

6.7.5 Uncertainties in theory

The largest source of uncertainty for resonant production comes from the MC generation and the choice of PS. In a similar way to the derivation of the MC for the four-prong tagger in Section 5.2, this uncertainty is estimated by generating MC samples using different generator for the PS and calculating the difference in the yields in the signal regions. This method was used in the analysis of the boosted $HH \rightarrow bbbb$ channel resulting in an estimate for the signal theory of 12%. Both channels use the same spin-0 scalar model therefore this can reasonably included as the signal theory uncertainty [146].

Source	$\pm\sigma$	Comment
Four-Prong Tagger	$\pm 22\%$	p_T based, post fit listed
Signal Theory	$\pm 12\%$	Estimated
Flavor Tagging	$\pm 7.5\%$	Estimated
Multijet Estimate	$\pm 2\%(8\%)$	≤ 3000 GeV (3000 GeV $>$)
Lepton Veto	$< \pm 1.0\%$	-
JES/JER	$< \pm 1.0\%$	-
Track	$< \pm 1.0\%$	-
Total	$\pm 4.70\%$	-

Table 6.7: An breakdown of the naming scheme used and the description of each pieces in the identifier chain used for triggers.

6.8 Statistics for Setting Upper Limit

The primary objective of this analysis and many BSM searches is to establish an upper limit on the sensitivity of the analysis to observe the signal process on top of background SM processes. When setting limits on a search the goal is to compare the compatibility of a background-only hypothesis and a background-plus-signal process hypothesis to what is observed in data. The agreement between a hypothesis and data is typically measured by a p -value that represents the probability of finding incompatibilities in the data under the test hypothesis. However the p -value is not the value that is often reported when particle physicists present a new discovery. The measure for discovery in particle physics is the significance:

$$Z = \frac{1}{2} (1 + \text{erf}(1 - p)) \quad , \quad (6.3)$$

where $\text{erf}()$ is the error function of $(1 - p)$. Z is defined such that a variable found Z standard deviations or more from the mean has a probability corresponding to p . Within the particle physics community values of Z that are accepted as a discovery are $Z = 5$ (obviously greater than 5 simply means more significant). This value of $Z = 5$ corresponds to a p -value of 2.87×10^{-7} which is an incredibly stringent threshold. For context, in some disciplines a p -value of 0.05 ($Z=1.64$) is considered to be a statistically significant level of confidence for a hypothesis test. The following section outline the statistical analysis approach used to estimate the upper limit of the signal strength with respect to its expected value μ_{up} . In brief, a test statistic based on a maximum likelihood ratio is defined. Using assumptions laid out by Wilks [147] and Wald [148] the Probability Density Function (PDF) and therefor its cumulant are estimated and used to determine μ_{up} for a desired p -value.

6.8.1 Likelihood-based method for physics analysis

As the name implies a likelihood based test centers around the likelihood function

$$L(x; \theta) = \prod_{i=1}^n f_x(x_i; \theta) \quad , \quad (6.4)$$

which is a product of the PDFs of individual observations. For purposes here, events aren't analysed individually, but are binned suggesting a Poisson distribution for the PDF. If for each event a kinematic variable x is measured, then the mean number of events found in bin i of the distribution of x is

$$s_i = s_{\text{tot}} \int_{\text{bin } i} f_s(x; \theta_s) dx \quad (6.5)$$

$$b_i = b_{\text{tot}} \int_{\text{bin } i} f_b(x; \theta_b) dx \quad , \quad (6.6)$$

where s_{tot} and b_{tot} are the mean number of signal and background events respectively, for the total distribution and the integrals represent the probability to be found in bin i . Since this likelihood method is being used to analyse the signal, the total number of events observed in bin i can be parameterized by a signal strength μ and expressed as

$$n_i = (\mu s_i + b_i) \quad , \quad (6.7)$$

The PDF's for the variable x of the signal and background are $f_s(x; \theta_s)$ and $f_b(x; \theta_b)$ are characterized by the parameters θ . For the nominal case θ would simply be the signal strength μ ; however, when considering systematical uncertainties, θ will also account for the Nuisance Parameters (NPs) associated with the uncertainties. The dependency of the PDF's on the NPs in θ means it is advantageous to have an independent constraint on the NPs that imposed by data in the signal region (where μ is also being determined). This can be done by constructing a new histogram of x with M bins in a control region where the expected signal is negligible such as CR1 in 6.6. The likelihood function for data is then a product of Poisson PDF's characterized by $(\mu s_j + b_j)$ for the signal region

and u_k for the control region is

$$L(\mu, \boldsymbol{\theta}) = \prod_{j=1}^N \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^M \frac{u_k^{m_k}}{m_k!} e^{-u_k} , \quad (6.8)$$

In order to test a hypothesis value of μ , test statistics are constructed from a ratio of profile likelihoods. In the case of upper limit setting where the value for μ is expected to be equal to zero or greater, the likelihood ratio used is expressed as

$$\tilde{\lambda}(\mu) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad \hat{\mu} \geq 0, \quad (6.9)$$

$$\tilde{\lambda}(0) = \frac{L(\mu, \hat{\hat{\boldsymbol{\theta}}}(\mu))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))} \quad \hat{\mu} < 0 , \quad (6.10)$$

where $\hat{\mu}$ and $\hat{\boldsymbol{\theta}}$ correspond to the values that maximize the likelihood. $\hat{\hat{\boldsymbol{\theta}}}$ are the values of $\boldsymbol{\theta}$ that maximize the likelihood for a fixed value of μ [149]. Notice the $\hat{\hat{\boldsymbol{\theta}}}$ in the denominator of the case with the $\hat{\mu} < 0$ conditional. This corresponds to the values for the NPs that maximize the likelihood at $\mu = 0$.

6.8.2 Setting upper limit

The corresponding test statistic as a function of the test value μ for the likelihood ratio in Equation 6.9 is

$$\tilde{q}_\mu = \begin{cases} 0 & \hat{\mu} > \mu, \\ -2\ln(\lambda(\hat{\mu})) & 0 < \hat{\mu} \leq \mu, \\ -2\ln(\lambda(0)) & \hat{\mu} < 0 \end{cases} , \quad (6.11)$$

, and tests the agreement between two likelihood functions with larger values of q_μ corresponding to increased incompatibility between the data the tested value of μ . How the above statistic is interpreted can either test for discovery or set an upper limit on the signal strength required to detect a new process. Regardless of the use case, the statistical analysis is based off the profile likelihood ratio 6.9 differentiated by different

assumptions concerning the signal strength.

For the purposes of this search the small size of the SM cross section of the process motivates setting an upper limit on the signal strength. To determine the value for the signal strength μ the definition for the p -value [150]

$$p_\mu = \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu) d\tilde{q}_\mu \quad (6.12)$$

is used. In this equation $f(\tilde{q}_\mu, \mu)$ represents the PDF of the test statistic $\tilde{q}_\mu$ under the assumption of signal strength μ . Despite the fact that the desired p -value is known (value corresponding to 95% confidence interval), determining $f(\tilde{q}_\mu, \mu)$ in practice can be very difficult. Fortunately Wald provides a way to estimate the PDF [151]. For suffi-

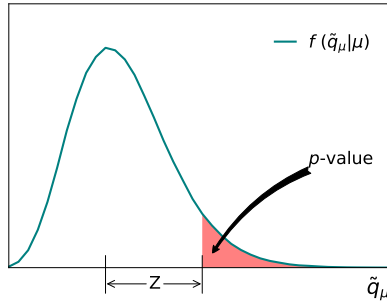


Figure 6.17: Cartoon depicting the relationship between the PDF $f(q_\mu|\mu)$, Z , and the p -value shaped in red [152]

ciently large numbers of background events, loglikelihood ratios can be approximated as $-2\ln(\lambda(\mu)) \approx \frac{(\mu - \hat{\mu})^2}{\sigma^2}$ where $\hat{\mu}$ is Gaussian distributed centered around $\bar{\mu}$ with a standard deviation σ . Higher-order terms in this approximation are suppressed by a factor of $1/\sqrt{b_{\text{tot}}}$. This leads to a simplified statistic

$$\tilde{q}_\mu = \begin{cases} 0 & \hat{\mu} > \mu, \\ \frac{(\mu - \hat{\mu})^2}{\sigma^2} & 0 < \hat{\mu} \leq \mu, \\ \frac{\mu^2}{\sigma^2} - \frac{2\mu\hat{\mu}}{\sigma^2} & \hat{\mu} < 0 \end{cases}, \quad (6.13)$$

Since the distribution for $\hat{\mu}$ is assumed to be Gaussian, the PDF in Equation 6.12 is

defined *a priori*. For the purposes of setting an upper limit on the signal significance $\bar{\mu} = 0$, but the standard deviation is still yet undetermined. It can be estimated through the use of an Asimov dataset. An Asimov dataset is a composite, representative dataset that is constructed from all MC samples. This dataset is used to estimate the variance of $\hat{\mu}$. Putting it all together yields the estimated PDFs

$$f(\tilde{q}_\mu|\mu) = \begin{cases} \frac{1}{2\sqrt{2\pi\tilde{q}_\mu}} e^{-\frac{q_\mu}{2}} & 0 < \tilde{q}_\mu \leq \mu^2/\sigma^2, \\ \frac{1}{\sqrt{2\pi}(2\mu/\sigma)} e^{-\frac{q_\mu + \mu^2/\sigma^2}{(8\mu/\sigma)^2}} & \tilde{q}_\mu > \mu^2/\sigma^2 \end{cases}. \quad (6.14)$$

The PDFs above are related to the p -value through their cumulant by $p = 1 - F(\tilde{q}_\mu|\mu)$ illustrated by the shaded region in Figure 6.17. The final step to determining an upper limit on μ is to set the p -value. If the p -value is found to be below a certain threshold, 95% here, then the signal is excluded with a confidence of 95%. The upper case in Equation 6.14 is relevant for upper limit setting therefore

$$F(\tilde{q}_\mu|\mu) = \frac{1}{2}(1 + \text{erf}(\sqrt{\tilde{q}_\mu})) , \quad (6.15)$$

where erf is the error function. The final expression for the significance is $Z_\mu = \sqrt{\tilde{q}_\mu}$. The upper limit μ_{up} is calculated by combining Equation 6.15 and 6.3 resulting in

$$\mu_{up} = \hat{\mu} + \frac{\sigma}{2}(1 + \text{erf}(1 - p_\mu)). \quad (6.16)$$

As stated in the beginning of this section upper limits are typically set at a 95% confidence interval that fixes the value for p_μ to 0.05 allowing for the μ_{up} to be determined by numerically scanning $\hat{\mu}$ since the cross section, σ , is a characteristic of the signal hypothesis. Once the value for the signal strength is determined, it is then expressed in terms of the cross section of the SM final state using

$$\mu = \frac{\sigma_{\text{TH}}}{\sigma_{sm}} \quad (6.17)$$

where σ_{sm} is the cross section of the SM process and σ_{TH} is the cross section of the new

physics theory being considered.

6.9 Statistical Analysis

The statistical analysis method described in the previous section is carried out individually for each mass sample using the regions defined in Figure 6.6, simultaneously fitting in both signal regions. The discriminating variable m_{HH} is used to carry out the fit. The binning for the input distributions are selected in order for the definition to constant across all mass samples for consistency.

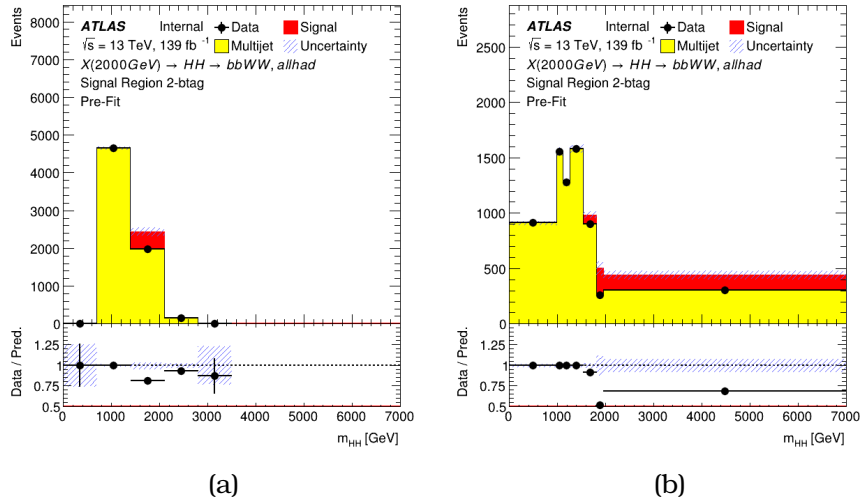


Figure 6.18: The m_{HH} discriminant distribution using (a) an optimized fixed width definition and (b) the binning transform from Equation 6.18 with $z_b = z_s = 4$.

6.9.1 m_{HH} Fit Binning

A fixed number of bins for all mass points doesn't result in an optimal distribution to fit, especially for the small and large values of m_X . Additionally, the statistical fluctuations in the extreme bins resulted in large statistical uncertainties. An investigation in alternative binning techniques led to the method described below [153]. Most importantly, this binning method resulted in an increase in signal sensitivity, particularly for the low mass samples. The binning transformation

$$Z = z_b \frac{n_b}{N_b} + z_s \frac{n_s}{N_s} \quad , \quad (6.18)$$

merges bins from the discriminating variable in order to produce a desired number of signal (background) bins z_s (z_b) which are scaled by the relative number of signal (background) events n_s (n_b). N_s (N_b) are the total number of signal (background) events. The optimal assignment of $z_b = z_s = 4$ resulted in the best sensitivity across all scalar mass samples. A comparison of a fixed-width bin definition and the automatic binning is shown in Figure 6.18. Pre- and post-fit data for the m_{HH} distribution can be seen for the $m_X = 3000$ GeV sample in Figure 6.19 and fit data in both signal regions for 800, 1200, 3000, and 4000 GeV samples are in Figures 6.20 and 6.21. All other fit plots can be found in the supplemental files.

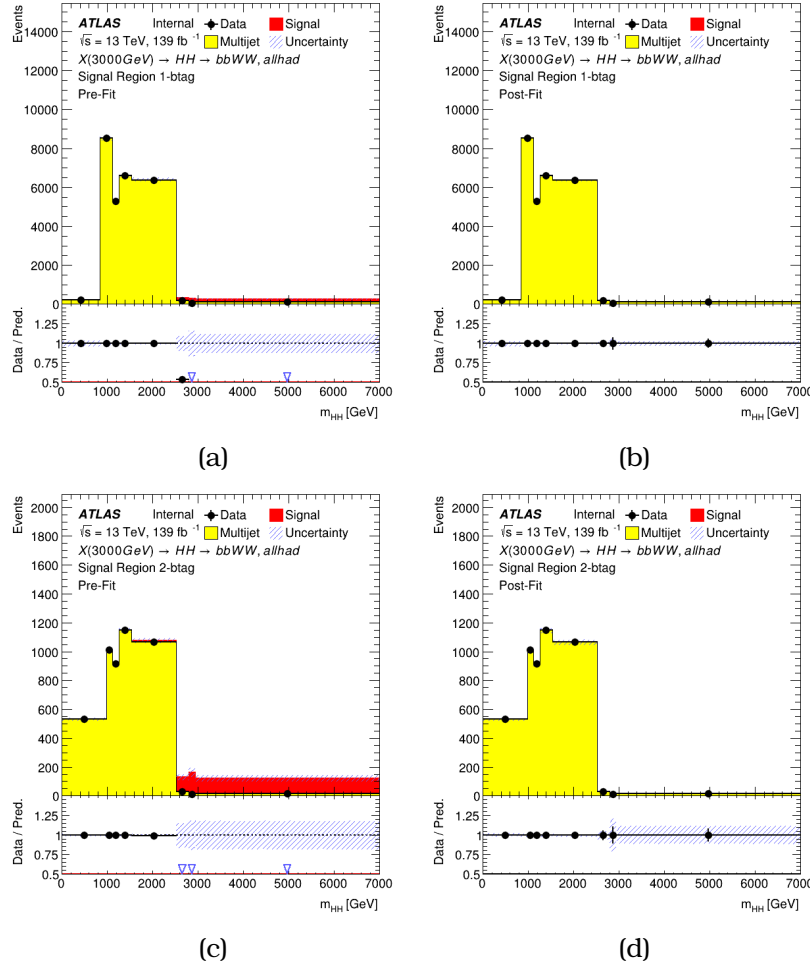


Figure 6.19: Plots showing the pre- and post-fit data in both signal regions for 3000 GeV scalar mass.

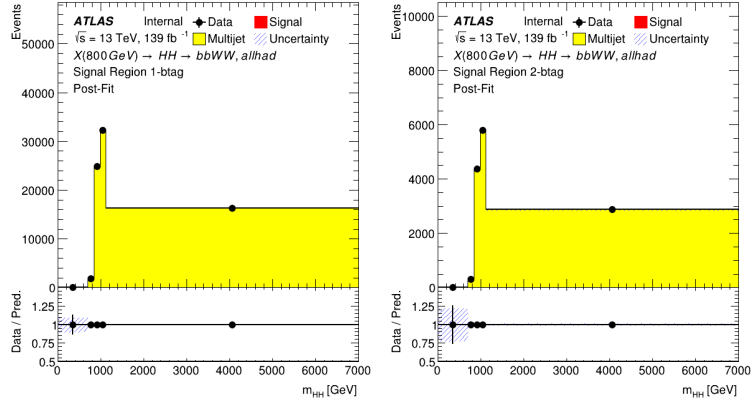
6.9.2 Fit Results

The systematic uncertainties from Section 6.7 are introduced as NPs into the statistical framework as parameters of the fit (see Equation 6.8). The fit parameters are varied individually to produce best-fit values for each. A method for testing the statistical framework as well as the fit is to analyse the "pulls" of each systematic when using Asimov data. A pull is defined as $(\hat{\theta} - \theta_0)/\Delta\theta$ where $\hat{\theta}$ is the best fit for a given NP, θ_0 is the expected value, and $\Delta\theta$ is the fit uncertainty for each NP. When testing the fit on Asimov data, the value for the pulls is expected to be very near, if not equal, to the expected value with an uncertainty of $\pm\sigma$ since the Asimov dataset is generated from the background with the assumption that all NPs are zero. Significant deviations from this expectation can highlight issues with the definitions and constraints of a given NP. The impact on the limit is determined by the uncertainty in μ over the best value $\hat{\mu}$ for each specific NP. In addition to the NP's associated with the systematics, there are normalization factors applied to the distributions of the multijet estimate in each analysis region.

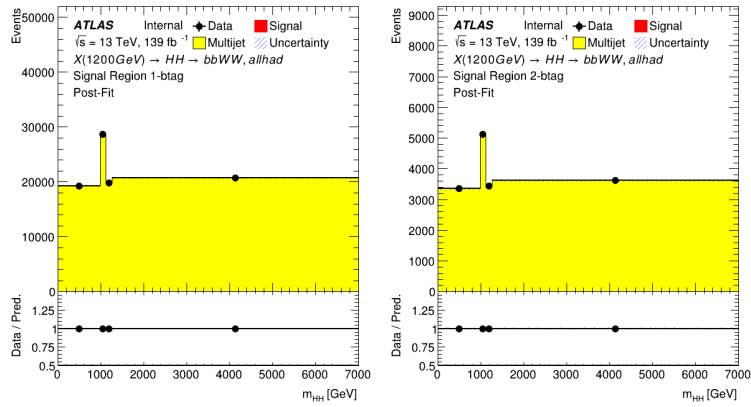
The limit impact and pulls for a subset of the scalar masses are shown in Figure 6.22. The ranking of the systematics varies with the m_X with the lower masses more effected by the background shape represented by the multijet estimate systematic. As the scalar mass increases, the limits become more impacted by the signal modeling uncertainty and the efficiency in both the b and four prong tagging. The pulls and impact for the normalization factors are shown in Figures 6.23, though the value for the pulls here is slightly different plotting the uncertainty in the fit in $\Delta\theta$.

6.9.3 Cross section Normalization

The results of the statistical fit is the parameter of interest μ_{up} from Equation 6.16. In order to take the fit parameter and express it as a physical upper limit, it must be normalized to the process under investigation. Using Equation 6.17 and the σ BR factor from the signal process in Table 6.1 the upper limits on the cross section for each scale

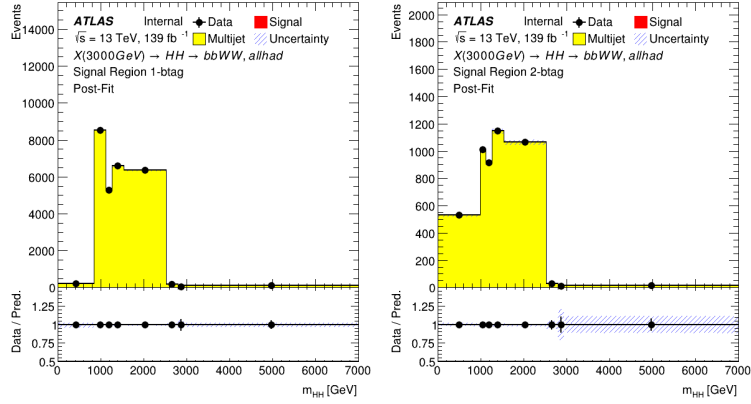


(a)

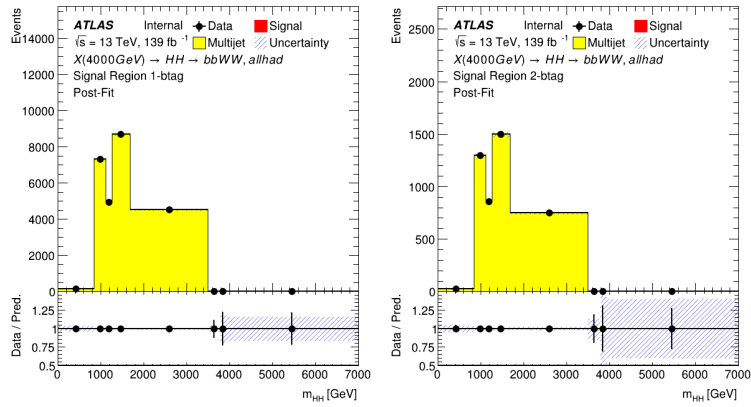


(b)

Figure 6.20: Post-fit data in both signal regions for the (a) 800 and (b) 1200 GeV scalar masses.



(a)



(b)

Figure 6.21: Post-fit data in both signal regions for the (a) 3000 and (b) 4000 GeV scalar masses.

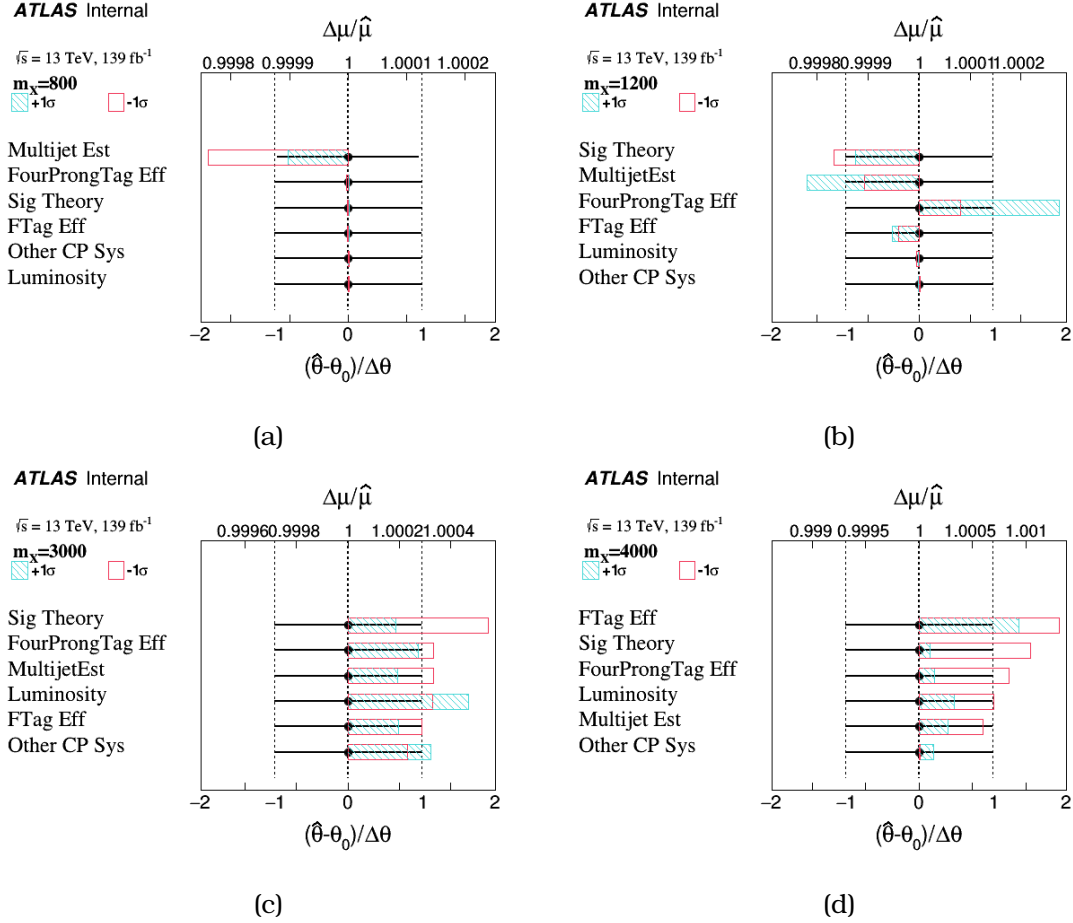


Figure 6.22: Systematic pull and impact plots for the 800, 1200, 3000, and 4000 GeV mass points.

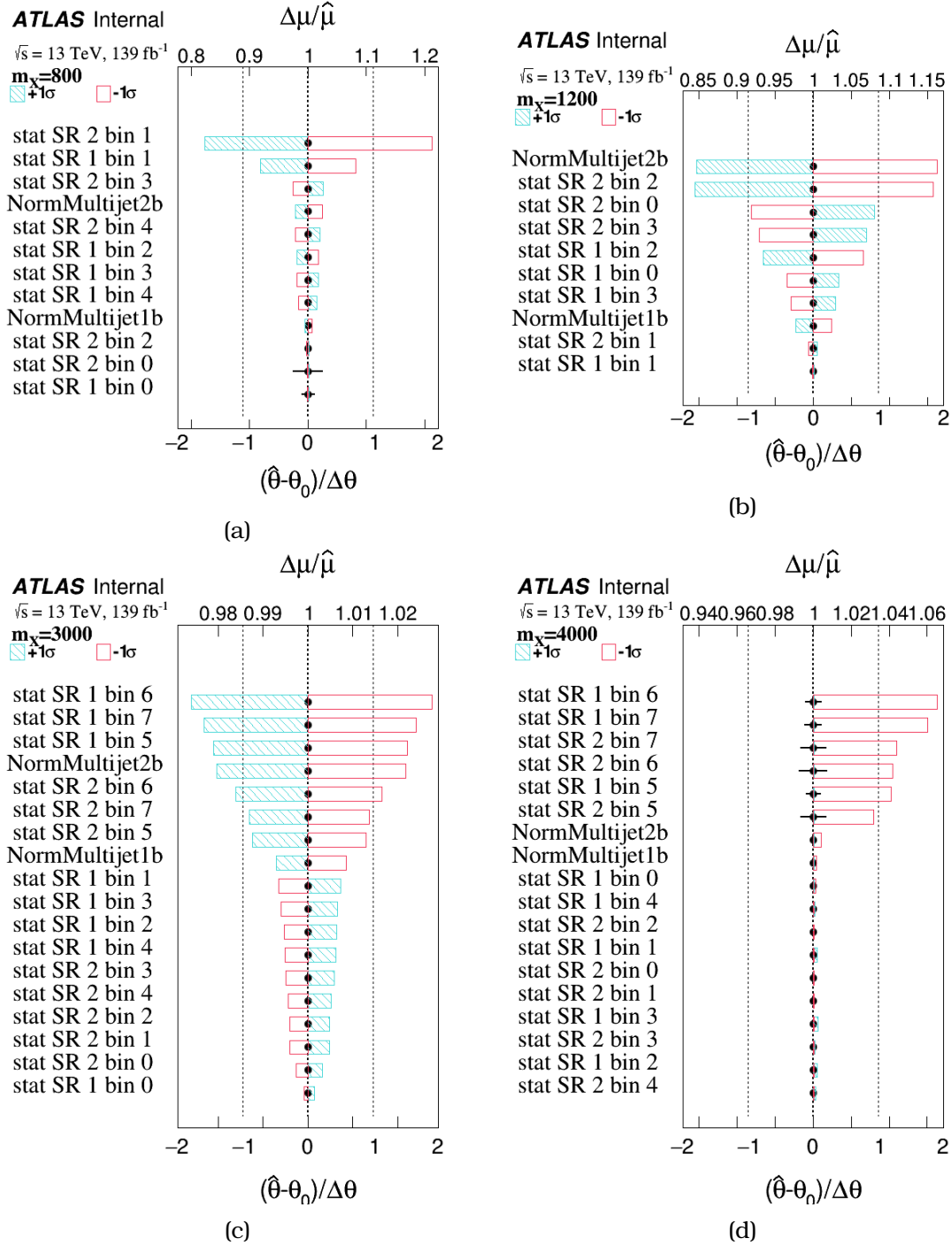


Figure 6.23: Normalisation pull and impact plots for the 800, 1200, 3000, and 4000 GeV mass points. $\Delta\theta$ here is the uncertainty in the fit.

mass in fb is

$$\sigma(pp \rightarrow X \rightarrow HH) = \mu_{UP, m_X} \times \sigma_{TH} \times BR_{bbWW} \quad . \quad (6.19)$$

6.10 Results

Expected 95% confidence interval upper limits are set for the cross section on the $X \rightarrow HH \rightarrow bbVV$ channel into the boosted, fully hadronic final state. These limits are presented as a function of the scalar masses in Figure 6.24. The analysis is most sensitive at higher scalar masses. The sensitivity of the analysis is relatively constant above 2000 GeV. The results are also of equal or greater sensitivity than those discussed in Section 6.1.1. The values associated with the limit plot are list in Table 6.8.

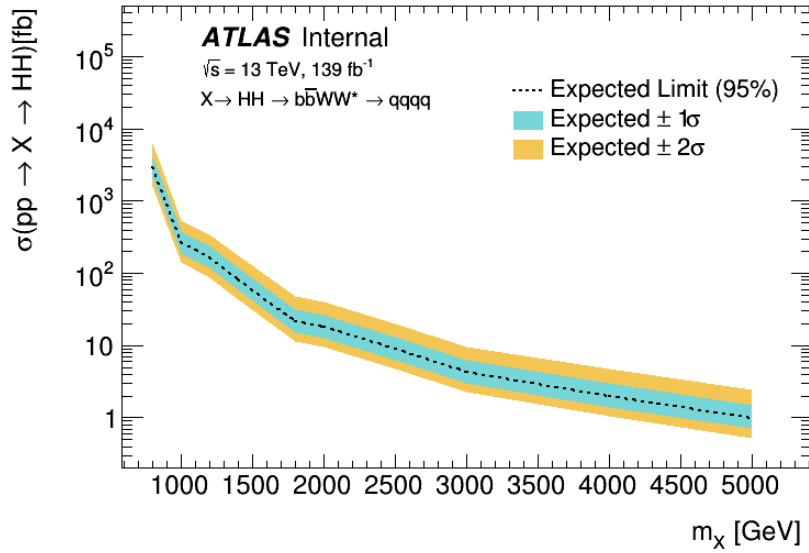


Figure 6.24: Expected upper limits at 95% C.L. on the cross-section of resonant di-Higgs production as a function of the signal scalar resonance mass.

Scalar Mass (GeV)	Estimated	$+1\sigma$	-1σ	$+2\sigma$	-2σ
800	3194.132	+ 1445.27	- 892.583	+ 3561.361	- 1479.759
900	532.092	+ 237.359	- 148.69	+ 582.416	- 246.504
1000	208.071	+ 94.843	- 58.144	+ 234.428	- 96.394
1200	108.913	+ 50.009	- 30.435	+ 124.539	- 50.457
1400	54.317	+ 25.481	- 15.178	+ 64.413	- 25.163
1800	18.074	+ 9.122	- 5.051	+ 24.157	- 8.373
2000	16.665	+ 8.446	- 4.657	+ 22.422	- 7.721
2500	8.139	+ 4.084	- 2.274	+ 10.762	- 3.77
3000	3.897	+ 1.928	- 1.089	+ 5.027	- 1.806
4000	1.728	+ 0.943	- 0.483	+ 2.58	- 0.801
5000	0.908	+ 0.502	- 0.254	+ 1.386	- 0.421

Table 6.8: Expected upper limits at 95% CL for all signal mass points.

Chapter 7

$X \rightarrow SH$ PRODUCTION SENSITIVITY

Presented in 1.11.2, the $X \rightarrow SH$ model is an extension of the resonant di-Higgs model where one of the Higgs particles is allowed to have a mass greater than SM Higgs of 125 GeV. The SH processes is a relatively un-examined process by the ATLAS collaboration. The limits presented in this chapter set a baseline sensitivity for the $bbVV$ final state. This study of the $X \rightarrow SH \rightarrow bbVV \rightarrow qqqq$ process reinterprets the analysis from Chapter 6. The decays for the H and S particles are $H \rightarrow b\bar{b}$ and $S \rightarrow VV$ respectively. Four m_X mass points and eight m_S mass points are studied and are summarized in Table 7.1.

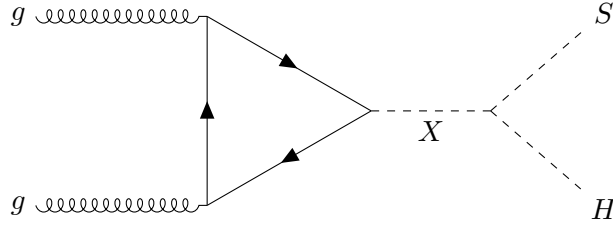


Figure 7.1: Feynman Diagrams for the production of a resonance via ggF. The resonance then subsequently decays into a SM Higgs and BSM Higgs-like scalar S .

m_X [GeV]	m_S [GeV]							
750	170	240	400	-	-	-	-	-
1000	↓	↓	↓	750	-	-	-	-
2000	↓	↓	↓	↓	1000	1500	-	-
3000	↓	↓	↓	↓	↓	↓	2000	2500

Table 7.1: m_X and m_S scalar mass ranges for the SH signal MC samples.

7.1 MC and Data samples

MC samples are generated for each of the twenty-one m_X, m_S combinations as listed in Table 7.1. The parameters for the SH MC are the same for each sample and are summarized in Table 7.2 .

m_X [GeV]	Generator	PDF	Parton Shower	$\sigma \times \text{BR}$ [pb]
750	Pythia8+EvtGen	NNPDF23LO	Pythia8+EvtGen	10.0
> 750	Pythia8+EvtGen	NNPDF23LO	Pythia8+EvtGen	1.0

Table 7.2: Definitions for generating the signal MC samples.

As with the Chapter 6 analysis, the full Run 2 $139.0 \pm 2.4 \text{ fb}^{-1}$ dataset is also used for the SH re-interpretation.

7.2 $X \rightarrow HH$ vs. $X \rightarrow SH$ Topologies

The increased mass of the S scalar has a considerable impact on the kinematics of the final state objects. This change results in a reduction of the efficiency of the selection outlined in Section 6.4.3. This is particularly true for the case when the S scalar is resolved, that is to say when $m_X - (m_H + m_S)$ is small (see Section 1.12 and Figure 1.10).

Since the analysis was designed and optimized for $X \rightarrow HH$ signal the event reconstruction is sub-optimal. In the di-Higgs analysis the mass of both Higgs particles is well reconstructed both in the event level reconstruction shown in Figures 7.4a and 7.4b as well as in the mass of the $b\bar{b}$ and WW candidates in Figures 6.7f, 6.9f, 6.11f, and 6.13f

As the new scalar particle becomes less boosted, the ability to construct the VV candidate and therefore the event final state is reduced. Figure 7.5 shows the m_S distribution for the $m_X = 2000 \text{ GeV}$ samples clearly showing m_H bias of the analysis. Given

the poor reconstruction of the S scalar at high masses, the m_X/m_{SH} distribution is necessarily affected. Figure 7.6 shows the reconstructed m_X for the di-Higgs case as well as for a range of m_S values. The HH reconstruction is clear, centered at 2000 GeV, however for the SH candidates, as the S becomes less boosted and more resolved, the peak becomes broader and more poorly formed. This is exemplified in the broken nature of the distribution when the two scalar masses are closest in Figure 7.6f.

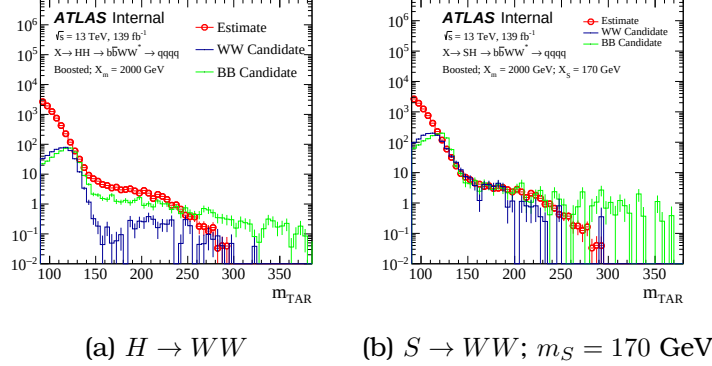


Figure 7.2: Comparison of 2000 GeV scalar mass point for di-Higgs and $m_S = 170$ SH sample. The similarities between the channels can be seen given the low mass of the S scalar.

7.3 Statistical Fit

The fitting process is the same as the HH analysis though the impact of the higher mass S scalar on the limit is significant. Analogous to the HH analysis, m_{SH} is used as the fitting discriminant. The methodology for the statistical analysis is identical to that of Section 6.9, repeated for each m_X, m_S signal sample pair.

7.4 Systematics

The systematic uncertainties are unchanged within the SH analysis as those from the HH analysis. The uncertainty estimate for the theory modelling uncertainty in Table E.1 is the only uncertainty that is subject to the new SH model however the uncertainty used is quite conservative and it is reasonable to assume that it is sufficient to characterize the uncertainty in the SH model as well. The NP parameter pulls and limit impacts are

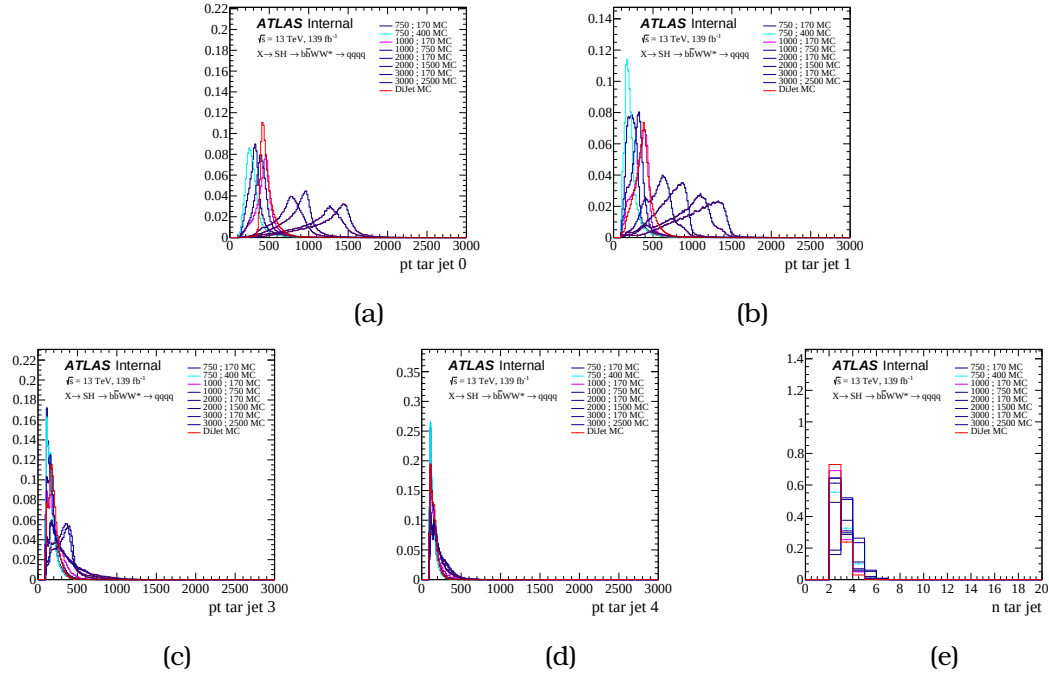


Figure 7.3: p_T distribution for the first four jets in MC for dijet and three signal masses. Jet multiplicity after the first 2 TAR jet cut is also shown in (e).

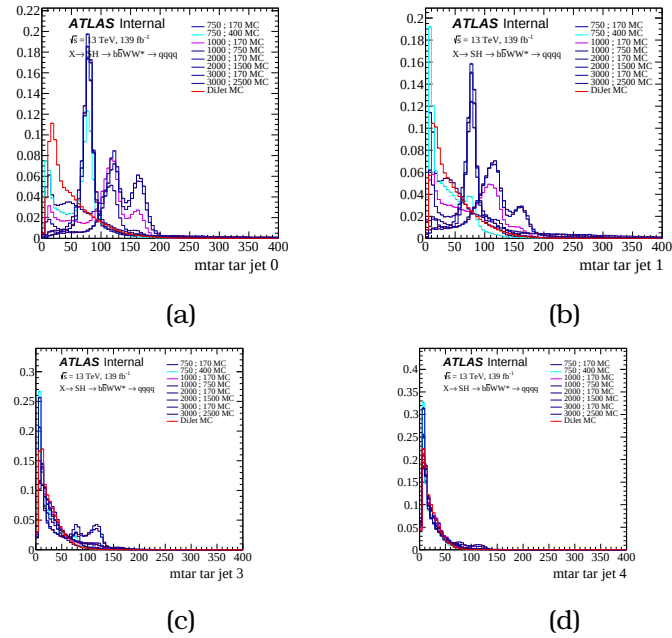


Figure 7.4: TAR jet mass for the first 4 jets in the events in MC for Di-Jet and 3 signal mass samples. The expected higgs mass is present in the leading and sub-leading jets in signal.

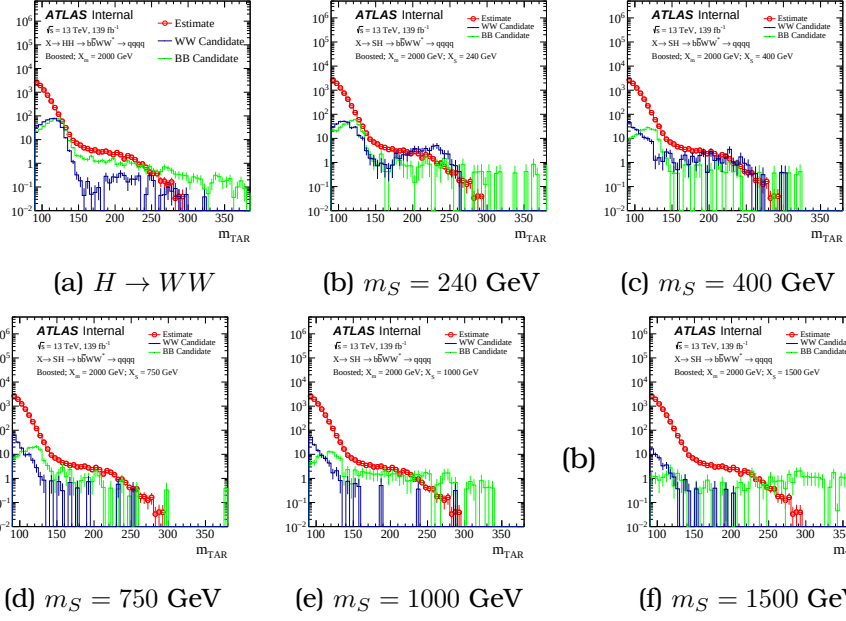


Figure 7.5: m_{TAR} for the bb and VV candidates for the $m_X = 2000$ SH sample and a range of m_S masses.

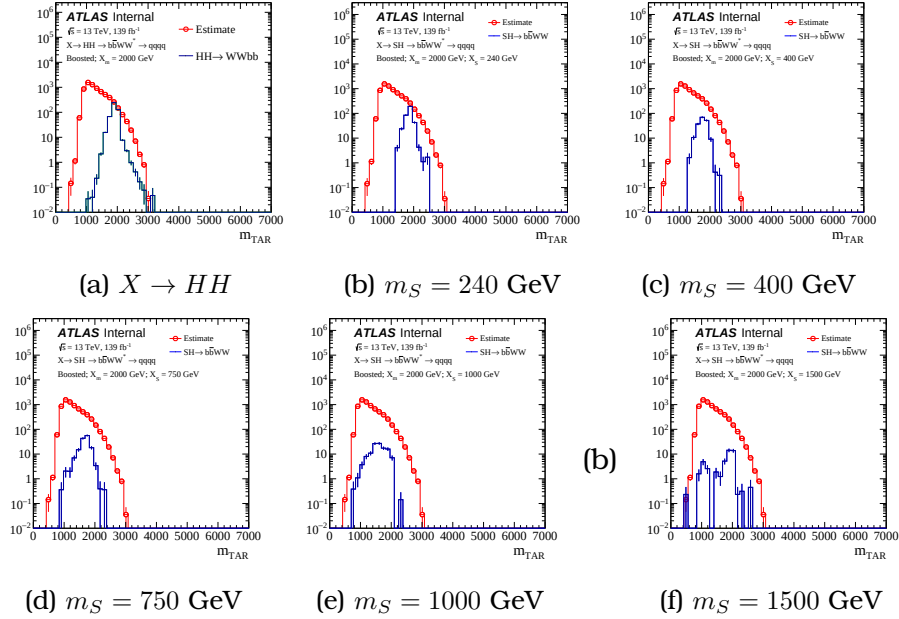


Figure 7.6: m_S in the 2 b -tag signal region. The boosted S scalar is much more successfully reconstructed.

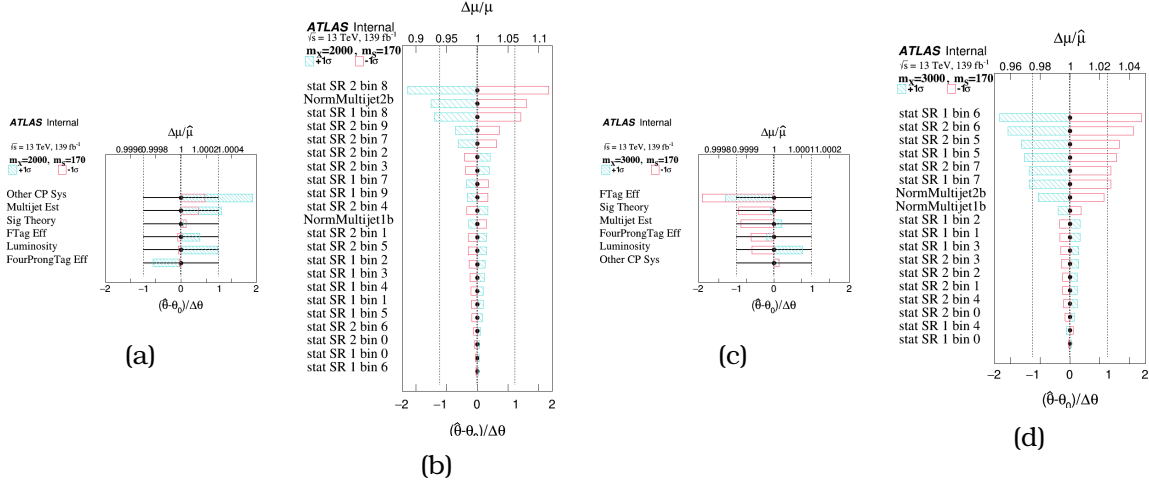


Figure 7.7: The NP pulls and limit impacts for the systematic for (a) $m_X = 3000, m_S = 170$ GeV and (c) $m_X = 2000, m_S = 170$. (b) and (d) so the pulls and the rankings for the normalization factors for the same mass pairs.

shown in Figure 7.7.

7.5 Results

Expected upper limits for the 95% confidence interval are set for the cross section of the $X \rightarrow SH \rightarrow bbVV$ channel into the boosted, fully hadronic $bbVV$ final state. The limits are presented for each m_X mass sample as a function of the scalar masses m_S in Figures 7.9. These are the first limits set for this channel, giving an initial estimate on the sensitivity for the SH process. As seen from the results in the di-Higgs limits, the sensitivity to SH is found to be larger than the di-Higgs limits, as expected, but are within an order of magnitude for the low m_S samples where the two analysis approach their overlap at 125 GeV. The relative expected limits are compared in Figure 7.8.

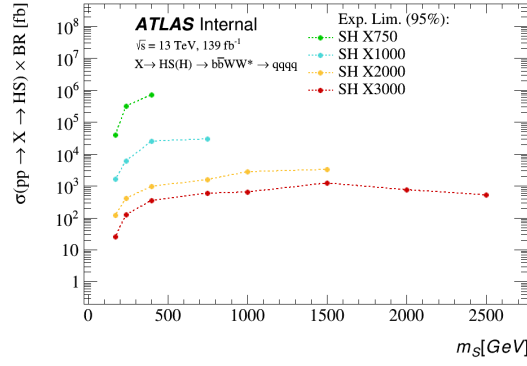


Figure 7.8: Comparison of the expected upper limits for the $X \rightarrow SH$ signal model for $m_S = 750, 1000, 2000$, and 3000 GeV samples.

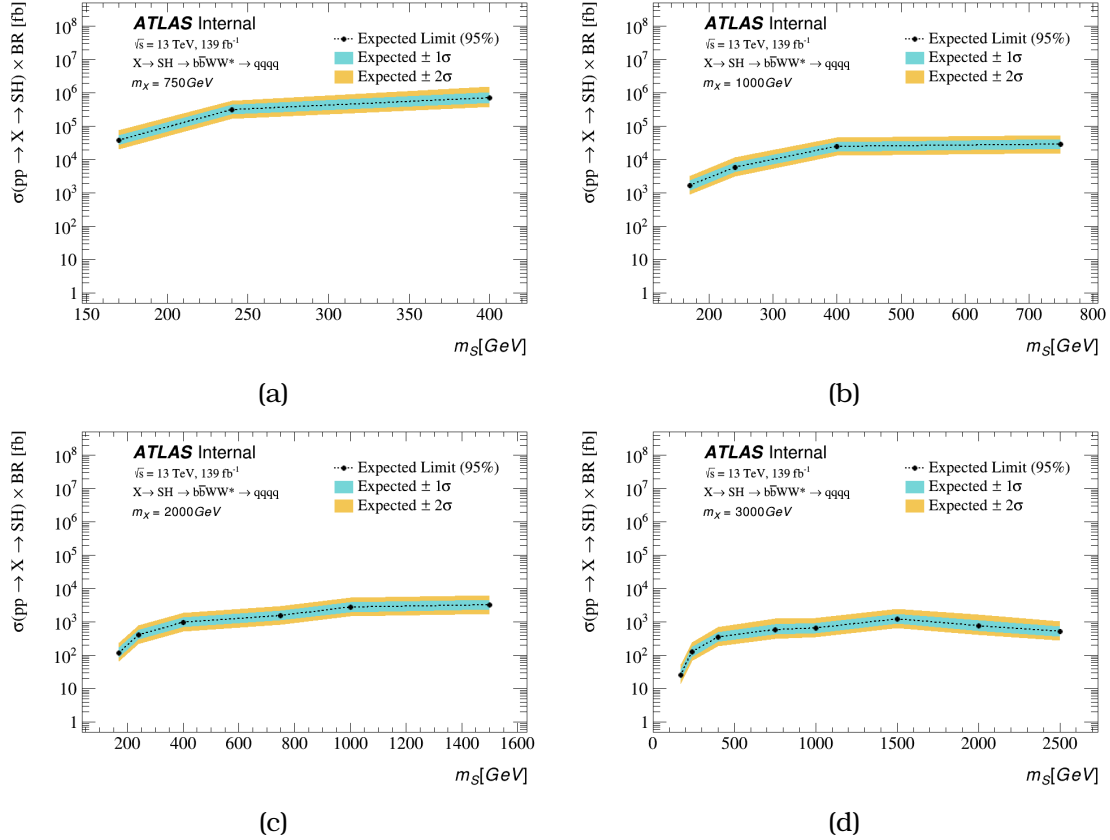


Figure 7.9: Expected upper limits for the $X \rightarrow SH$ signal model for $m_S =$ (a) 750, (b) 1000, (c) 2000, and (d) 3000 GeV samples.

Scalar Mass (GeV)	Estimated	+1σ	-1σ	+2σ	-2σ
170	33389.0	16886.0	9330.0	39429.0	15468.0
240	147867.0	58183.0	41321.0	129728.0	68503.0
400	319852.0	146427.0	89381.0	368468.0	148179.0

Table 7.3: Expected limits for 750 GeV X scalar mass sample.

Scalar Mass (GeV)	Esimated	$+1\sigma$	-1σ	$+2\sigma$	-2σ
170	1004.0	430.0	281.0	1022.0	465.0
240	3486.0	1482.0	974.0	3500.0	1615.0
400	18359.0	7366.0	5130.0	16526.0	8505.0
750	19811.0	7890.0	5536.0	17598.0	9178.0

Table 7.4: Expected limits for 1000 GeV X scalar mass sample.

Scalar Mass (GeV)	Esimated	$+1\sigma$	-1σ	$+2\sigma$	-2σ
170	102.0	48.0	29.0	121.0	47.0
240	356.0	146.0	99.0	332.0	165.0
400	949.0	386.0	265.0	875.0	440.0
750	1405.0	575.0	393.0	1312.0	651.0
1000	2747.0	1116.0	768.0	2547.0	1273.0
1500	3373.0	1365.0	942.0	3072.0	1562.0

Table 7.5: Expected limits for 2000 GeV X scalar mass sample.

Scalar Mass (GeV)	Esimated	$+1\sigma$	-1σ	$+2\sigma$	-2σ
170	28.0	14.0	8.0	35.0	13.0
240	124.0	53.0	35.0	125.0	57.0
400	348.0	152.0	97.0	361.0	161.0
750	594.0	277.0	166.0	680.0	275.0
1000	618.0	277.0	173.0	663.0	286.0
1500	1244.0	533.0	348.0	1236.0	576.0
2000	852.0	405.0	238.0	1000.0	395.0
2500	511.0	228.0	143.0	535.0	237.0

Table 7.6: Expected limits for 3000 GeV X scalar mass sample.

Chapter 8

CONCLUSIONS

The nature of the boosted, fully hadronic $b\bar{b}VV$ final state makes it a difficult topology in which to perform a di-Higgs search. This complexity motivated the assumption that the channel would not be sensitive enough to warrant investigation, highlighted by the fact that previous iterations of the resonant $b\bar{b}VV$ search did not include a fully hadronic channel. The exclusion limits presented in this work invalidate this previous assumption. The results, however, do not come without significant investment. The complex nature of the four-prong tagger implemented to tag the boosted VV pair is not a trivial, or particularly convenient, application of machine learning for a HEP search. This technically challenging yet sophisticated algorithm unquestionably led to the sensitivity observed justifying, however *post hoc*, the significant effort put into its development.

The sensitivity of the resonant di-Higgs limits set in Chapter 6 were found to increase, understandably, with the mass of the resonance. The best sensitivity of the channel was found to be at the 5000 GeV mass point. This result extends the sensitivity mass range from that of the previous $b\bar{b}VV$ search. The expected limits set are competitive with limits found for other channels as well as the current combined sensitivity for the $X \rightarrow HH$ search. Additionally these results have built a solid foundation of understanding to inform inevitable refinements to come and outlines the difficulties they will face, particularly the challenge of maintaining sensitivity at low resonance masses.

The SH model is explored for the first time in the $b\bar{b}VV$ channel. Confidence limits are also set and this sensitivity measurement defines a benchmark for future SH results. The search is most sensitive near the overlap with the di-Higgs topology as expected and

the significant dependence of the limit on the relative mass of X and S is demonstrated. This dependence motivates a deeper, and perhaps orthogonal, treatment of this intermediary boosted regime where the W are individually resolvable as the topology transitions from resolved to boosted.

As a whole the developed four-prong tagger combined with the results for both the di-Higgs and SH searches elucidate the phenomena of boosted vector boson tagging as well as $H \rightarrow VV$ physics. They also motivate the re-examination of previously unanalyzed final states especially considering the increased luminosity to be delivered during Run 3 at the LHC.

Appendices

Appendix A

NETWORK DISSECTION

A.1 Introduction

A common refrain that is stated often within the High Energy Physics community with regards to machine learning is that the networks that are created are "black boxes" and that it's not known what the network has learned or why the the network is successful. The novel network design of the boosted four-prong tagger discussed in Chapter 4 invites questions about the network and its learning.

A.2 Understanding Convolutional Filters

There are two different elements of a convolutional layer that can be analyzed to better understand the network. The the filters, or more accurately the weights, of the network can be examined directly for clues as the the features the network has identified. The outputs of each individual layer can also be produced in order to determine the strength of the network response at different stages of the network and for different processes. The way that filters work within a convolutional layer is outlined in Section 4.3.5. How these filters change and the features that are detected is more complicated. Before looking at the layers and filters of the tagger, it's helpful to examine a toy example in order to understand how the filters and the outputs of a convolutional later can be interpreted.

A.2.1 Filters of Handwriting

The Modified National Institute of Standards and Technology dataset is a subset of the NIST dataset of handwriting images and these images are often used to demonstrate image recognition and convolutional neural networks. For the purposes of understanding how the filters of convolution are analysed only a single image has to be used. Take for instance the number 6 in Figure A.1.



Figure A.1: A single hand written 6 taken from the mnist dataset.

In a typically neural network a filter of normalized weights would be passed over this image in order to learn to identify it. However, there are well known filters that illustrate what a network learns very clearly. Take for instance the filters in Figure A.2.

1	0	-1
1	0	-1
1	0	-1

(a)

1	1	1
0	0	0
-1	-1	-1

(b)

Figure A.2: Simple filter definitions corresponding to (a) vertical and (b) horizontal edge detection.

In Filter A.2a the vertical arrangements of the columns of values combined with the stark transition in weight size suggests that this filter should respond strongly to vertical edges. The same is true for Filter A.2b with horizontal edges. In Figure A.3 the mnist 6

character is convolved with both of these filters.



Figure A.3: The convolution of Figure A.1 with (a) Figure A.2a and (b) Figure A.2b visually demonstrating the types of features the filters detect.(a) Strong vertical edge detection and (b) strong horizontal edge detection.

The output images for the strength of the response is characterized by the brightness or darkness of an image. A more detailed way to think about it is in terms of large values corresponding to the sum of the weights dotted the image values. While it's not quite possible to use such clearly predefined filters in the tagger, and this wouldn't give the desired results, this last characteristic of the outputs of the convolution will guide the analysis of the tagger filters to better understand its learning.

A.3 Analysing Tagger Filters

The learned features of the tagger won't be able to be identified as clearly as for the more relatable task described above. The images of the Lund planes are less dense and less intuitive than images of numbers. This doesn't however mean that the same approach can't be applied in order to extract possible meaning from the filters as well as develop a more thorough understanding of what the network has learned. As described in Chapter 4 there are three sequential convolution layers. Each layer will be analysed in term, but first it is instructive to examine a single filter from the first layer. Recall that the three layers of the network have 128 filters and square 10, 4, and 4 filters respectively. All the filters for each layer are presented as images in Figures A.8,A.10, and A.12. Unlike in the simple case presented above, the weights of the filters in the tagger are defined

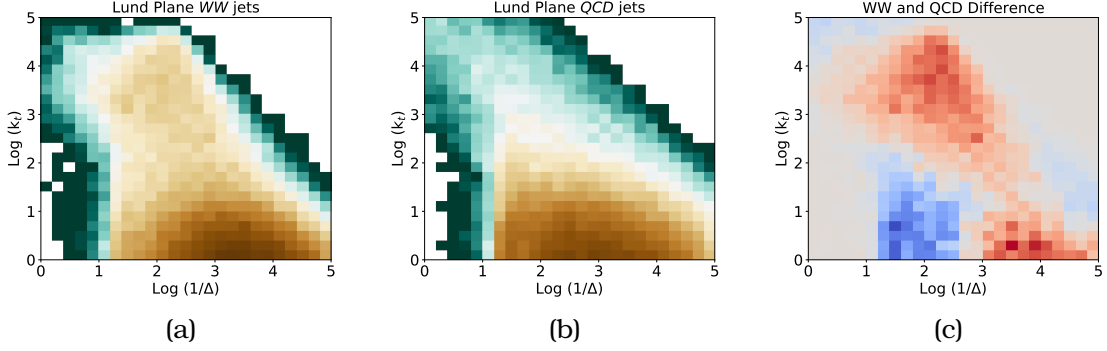


Figure A.4: Lund jet plane for a large collection of (a) $H \rightarrow WW$ jets and (b) QCD jets. (c) shows the normalized difference between the two images.

during the training of the network making interpreting their effect difficult on inspection. These weights for the three layers are presented as images in Figures A.8, A.10, and A.12 respectively. Initial visual inspection suggests that there are 4 filters of particular significance, 2 in the first and 2 in the third layer, however by looking the difference in the convolution of the filters in Figure A.9 with respect to $H \rightarrow WW$ and QCD a clear features is identified in the upper left corner of many filters. The difference average of all filters in the three layers in summarized in Figure A.7

A.4 Physical Connection

It is rather straight forward to interpret the feature that are identified by the first layer in the network. The large weight value in the upper left hand corner of Figure A.4a corresponds to the region of the Lund jet plane that is associated with large k_T and Δ values. Large k_T and Δ values of this nature are generated when a jet contains constituents with hard splittings. The second feature identified corresponds to low k_T for all Δ values which isn't particularly interesting as it is a common characteristic among all jets.

A.5 Feature Correlation

A second means of examining the Lund jet planes for information about the network is to look at the Pearson correlation coefficients between different inputs and features.

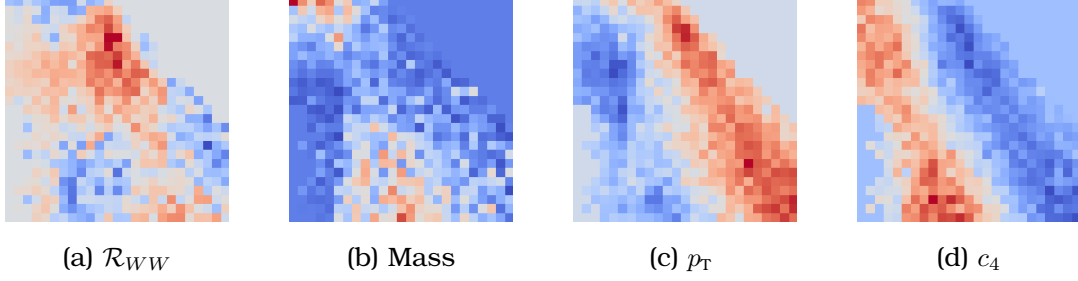


Figure A.5: Linear correlation of between the labeled variation and the the Lund planes for WW and QCD jets combined.

The correlations for the inputs to the feed forward branch as well as the jet mass and the network response are shown in Figure A.5. Nothing particularly new is learned here since the linear correlation was used in the selection process for these features and the p_T and m_X dependence of the network is a known characteristic.

However, the same analysis can be applied to the Lund jet planes. Figure A.5 shows the correlation between the "pixel" location for the Lund jet plane in Figure A.4a and different quantities. The correlation with the network response $\mathcal{R}_{WW}$ has interesting implications when compared to the other plots. First, examining it alone, the correlation says that the location most responded to in the Lund jet plane for $H \rightarrow WW$ jets is the region associated with the hard splitting. The relative density of nodes between the hard splitting region and the low k_t region is the same for the jets however the network response in these two regions is entirely different. This suggests a response to the splitting, not just the intensity in the image. When compared to Figure A.5b the region of the Lund jet plane most highly correlated with the jet mass has little to no overlap with the region correlated with the response implying that the strong feature in the Lund plane has little to no association with the learned mass from Figure 4.18a. Lastly, while there is overlap between the p_T of the jet and the Lund plane, it is not all encompassing, suggesting again that the feature associated with the hard splitting in the upper left corner of the Lund jet plane is being learned.

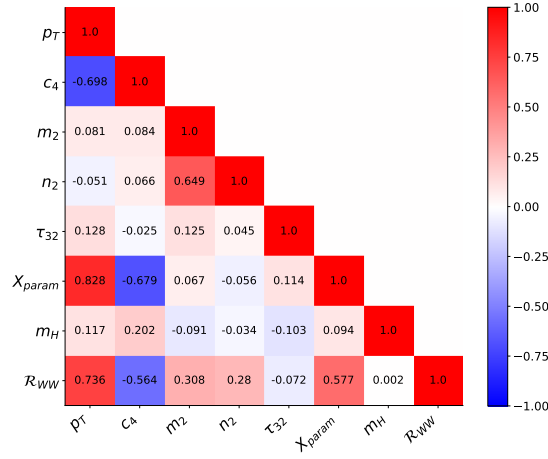


Figure A.6: Correlation matrix for the input features of the feed-forward branch of the network as well as the jet mass and the network response R_{WW}

Layer

Output

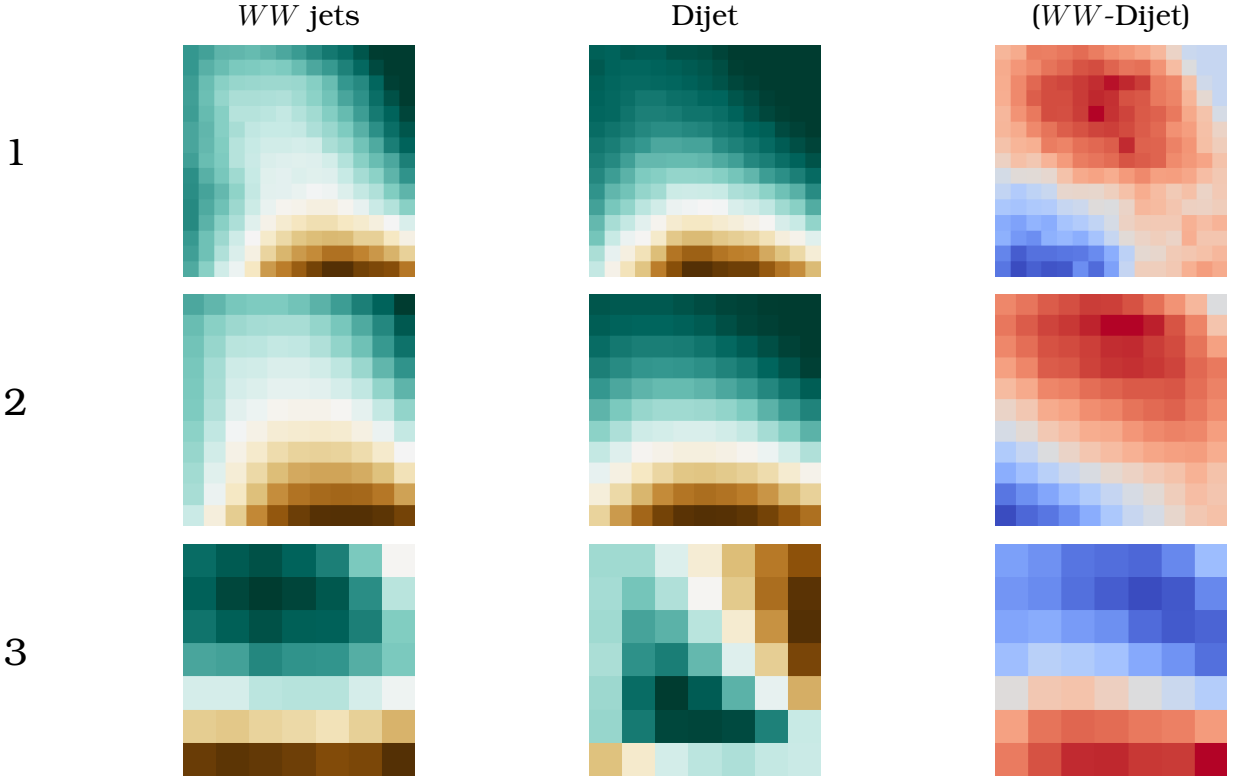


Figure A.7: Columns 1 and 2 show the max-pooled output for each convolutional layer in the network for an averaged sample of WW and dijet lund planes. The last column shows the difference in the output puts highlighting the difference in the strength of the response between the two processes.

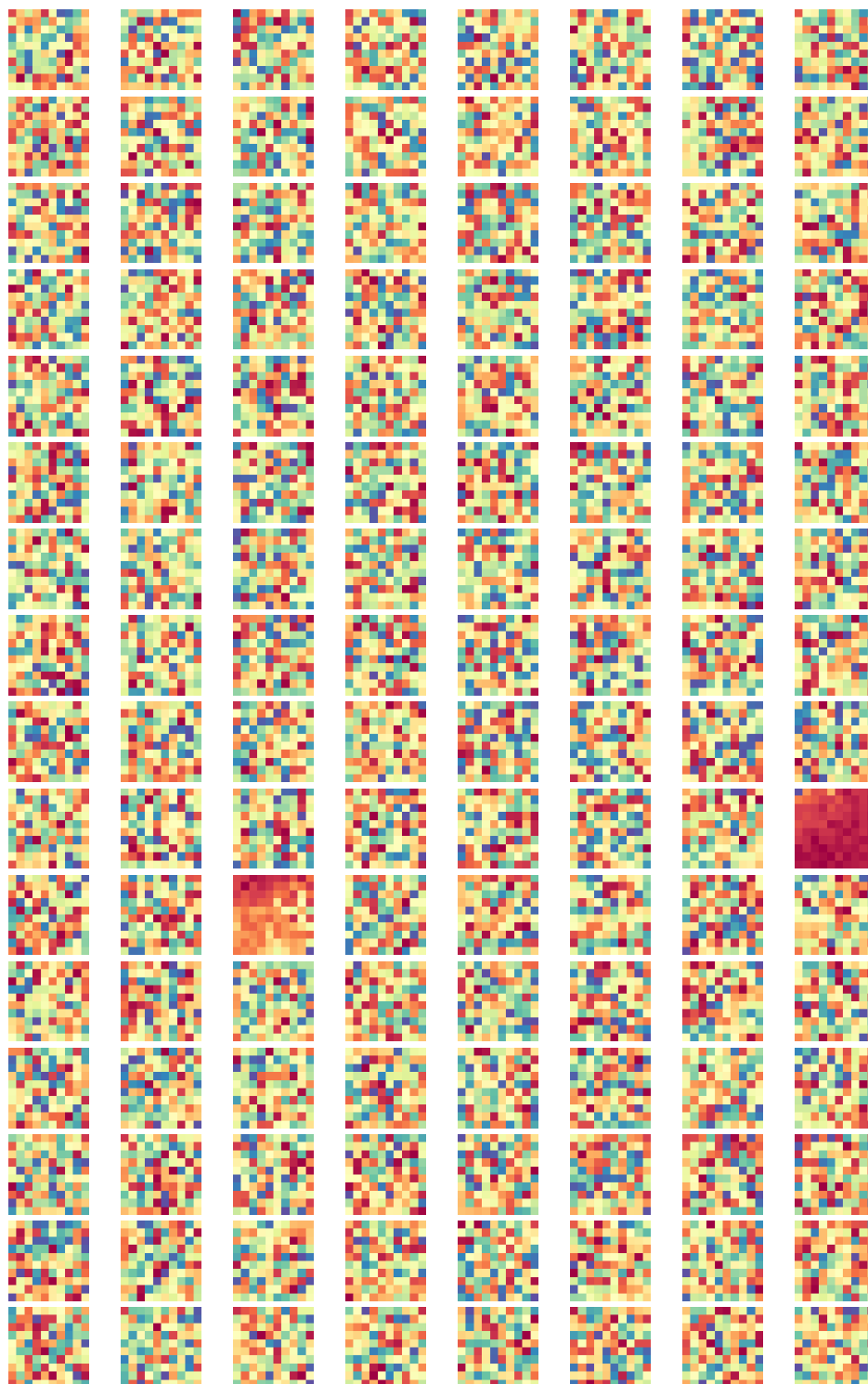


Figure A.8: All filters from the first layer of the tagger.

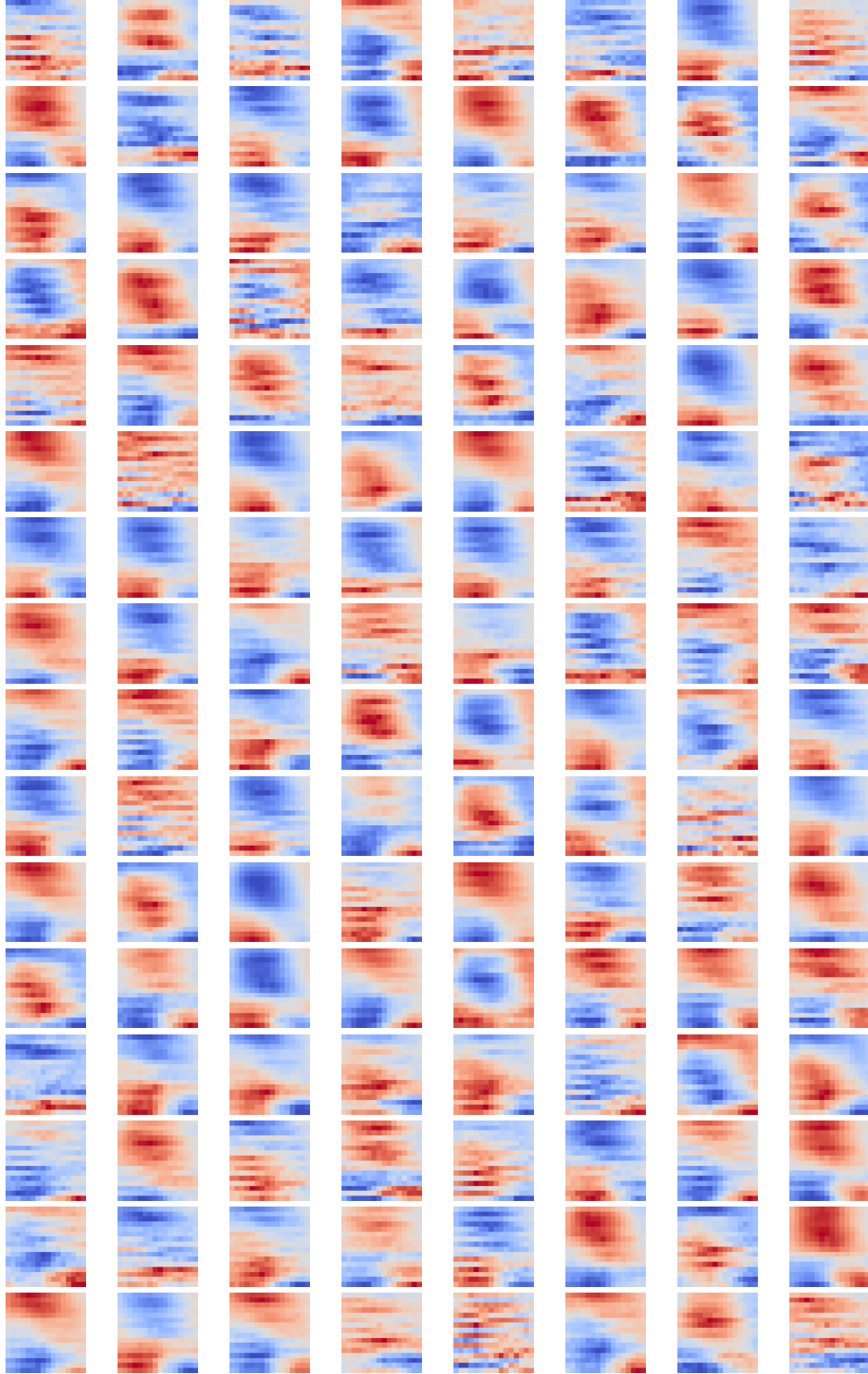


Figure A.9: The difference between the convolution of the first layer's filter with of QCD and WW average Lund jets planes.

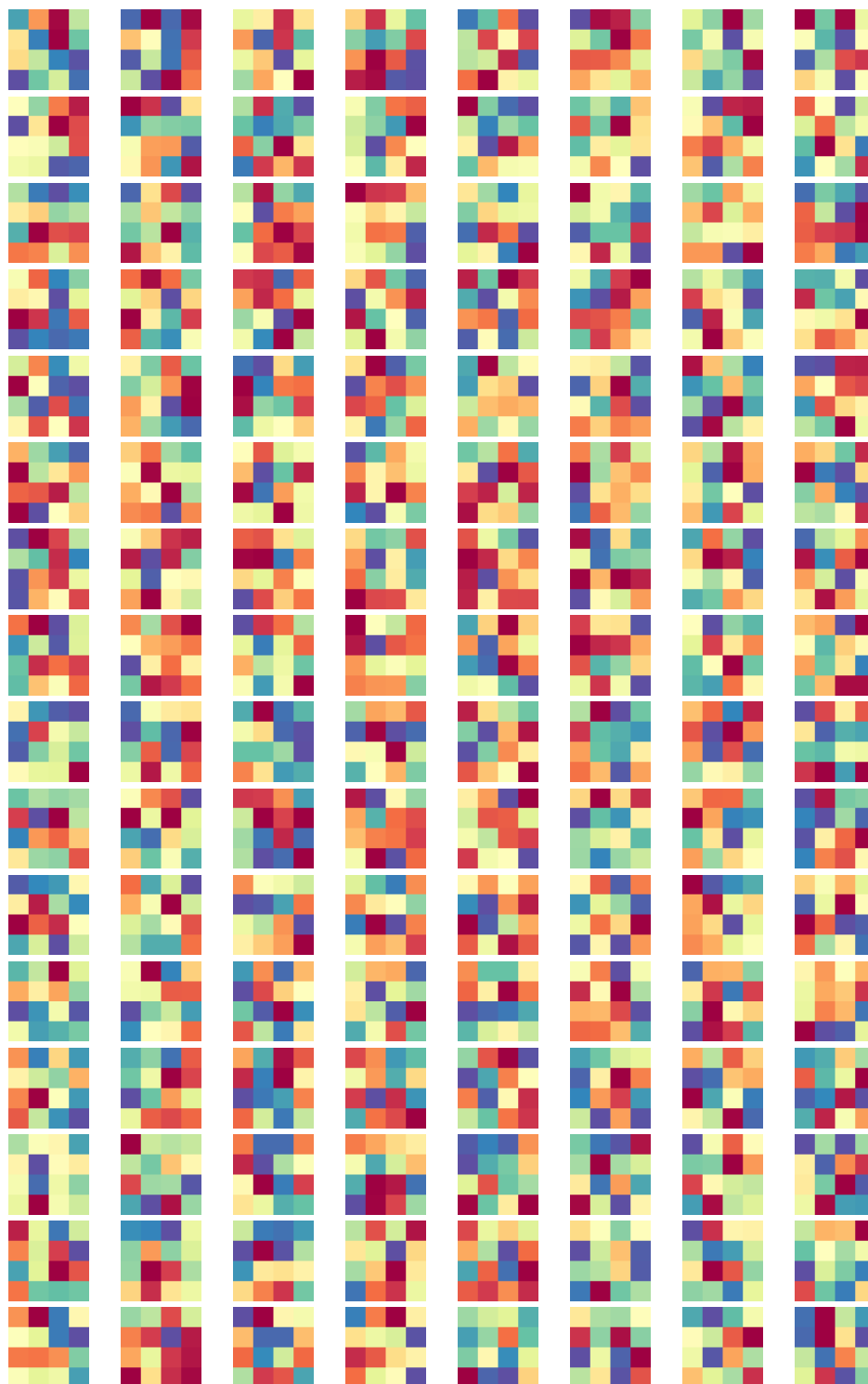


Figure A.10: All filters from the second layer of the tagger.

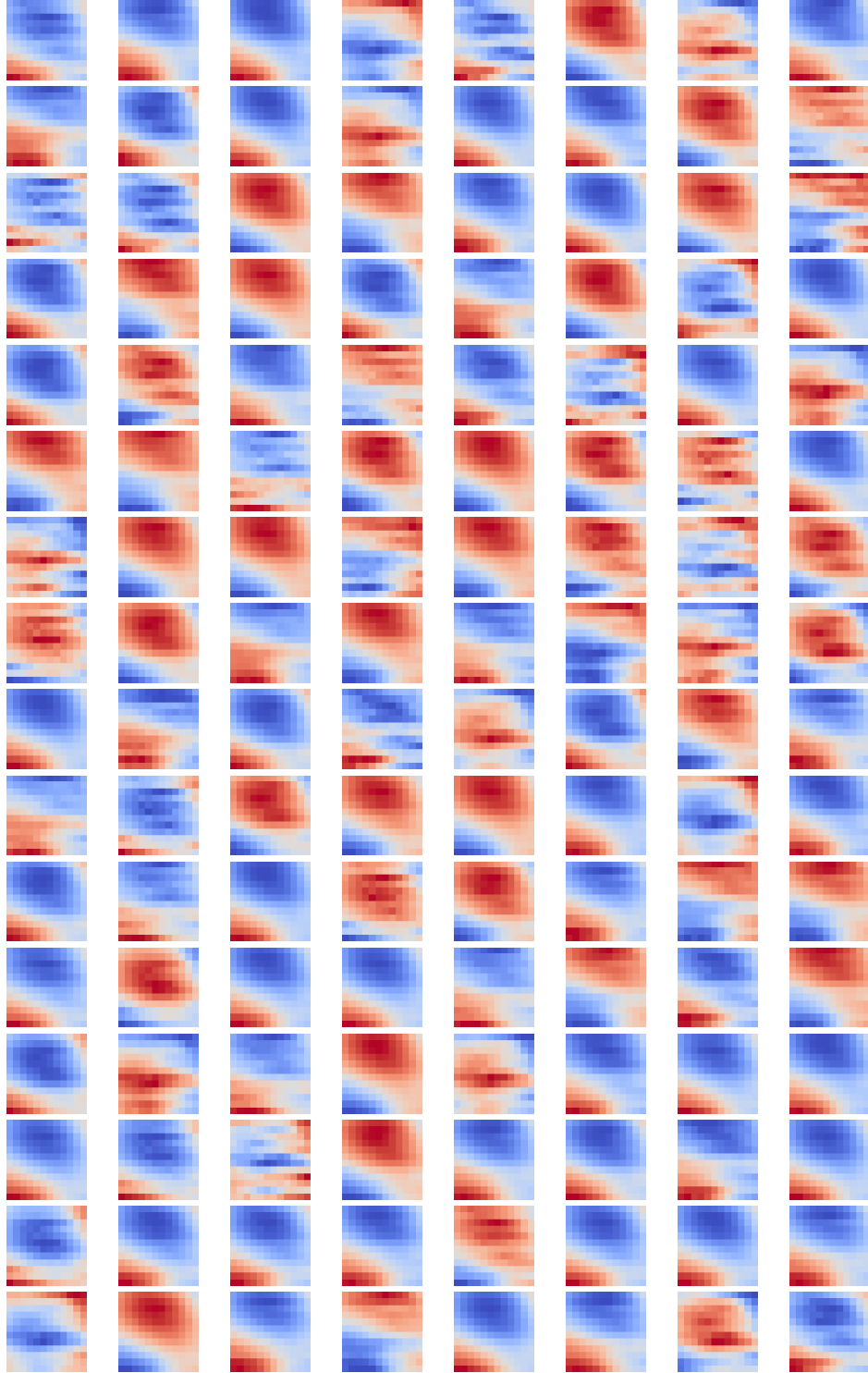


Figure A.11: The difference between the convolution of the filters in the second layer with QCD and WW average Lund planes.

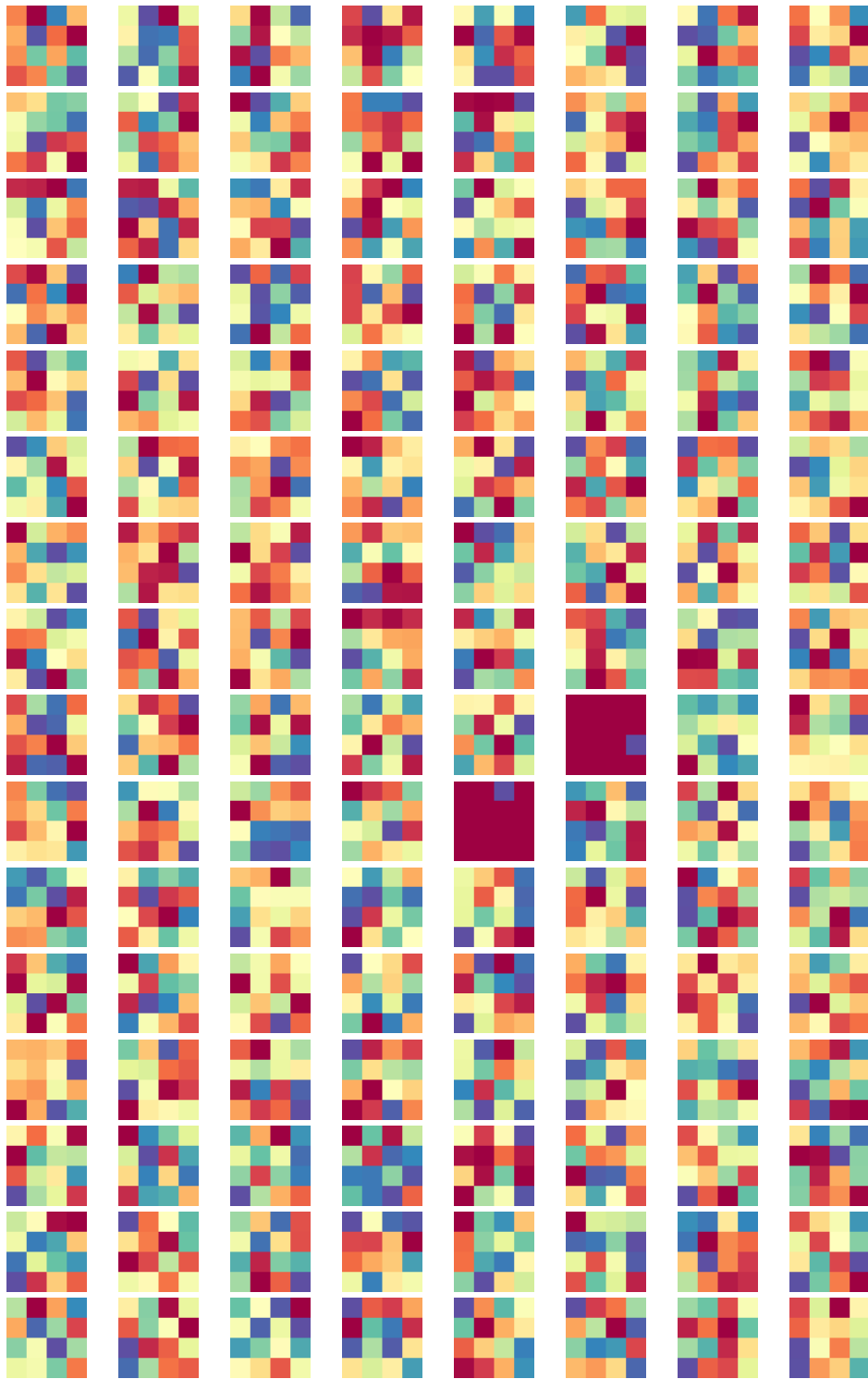


Figure A.12: All filters from the third layer of the tagger.

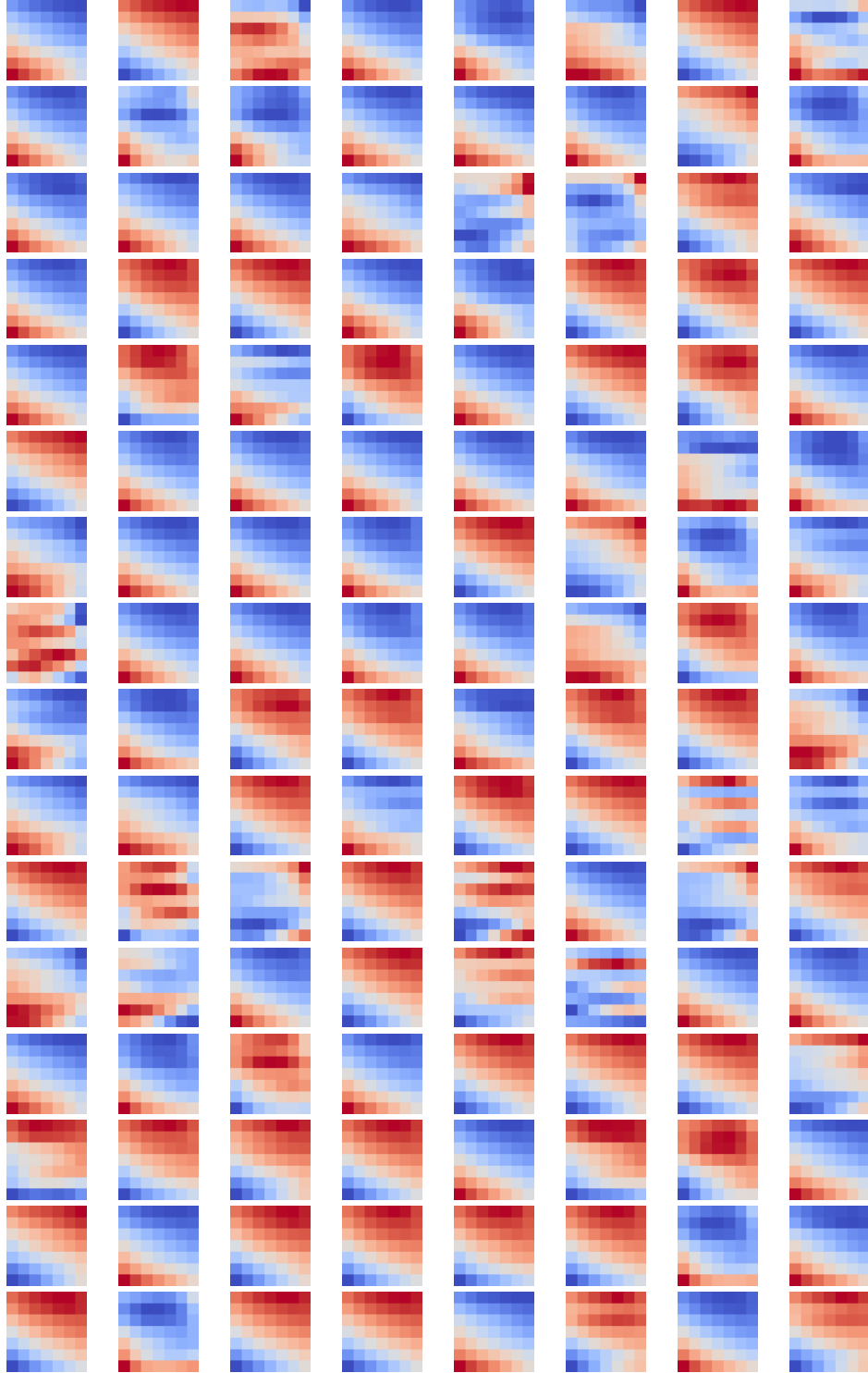


Figure A.13: The difference between the convolution of the filters in the third layer with QCD and WW average Lund planes.

Appendix B

FOUR PRONG TAGGER MATERIALS

p_T Bin	300 - 600	600 - 900	900 - 1200	1200 - 1500	1500 - 1800	1800 - 3000
nominal	0.7876	0.8202	0.8159	0.8052	0.8275	0.8425
s3295	0.8131	0.8168	0.8049	0.8203	0.8061	0.8052
s3296	0.8074	0.8184	0.7976	0.8145	0.7896	0.805
s3297	0.811	0.8116	0.8153	0.8281	0.8151	0.8032
s3298	0.8114	0.813	0.8022	0.8185	0.7978	0.8023
s3299	0.8161	0.8152	0.7957	0.8162	0.7981	0.8074
s3373	0.8094	0.8102	0.7967	0.8192	0.8046	0.8051
s3374	0.8066	0.8103	0.7925	0.8328	0.8106	0.8033
s3375	0.8108	0.8012	0.8014	0.8056	0.8028	0.8033

Table B.1: Loose WP efficiency for each signal sample variation studied and used to generate the uncertainty in four prong tagging.

Correlation Matrix (signal)

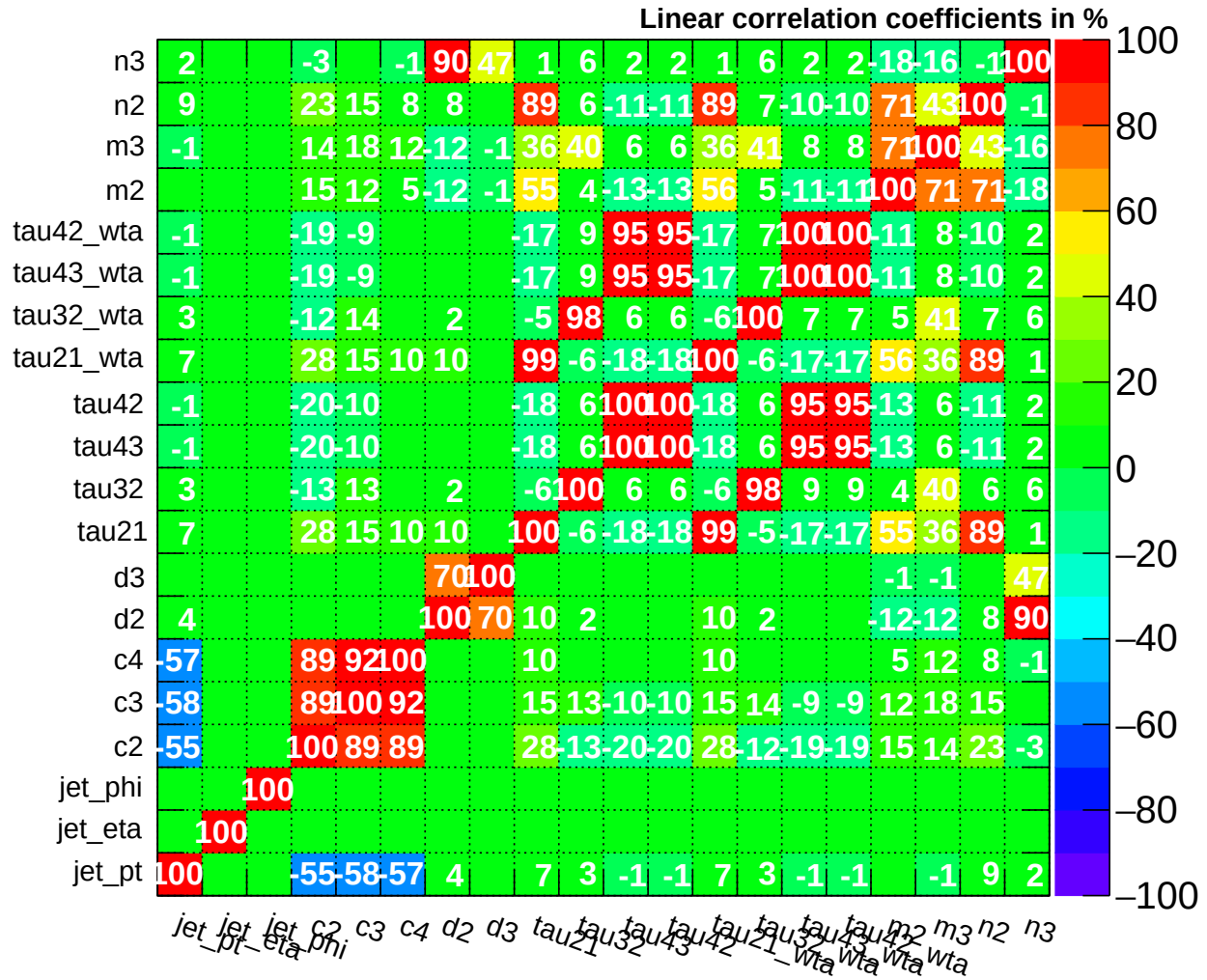


Figure B.1: Correlation matrix of the testing features. The features found in Table 4.1 are a subset of features in the matrix that have minimal correlations with other features.

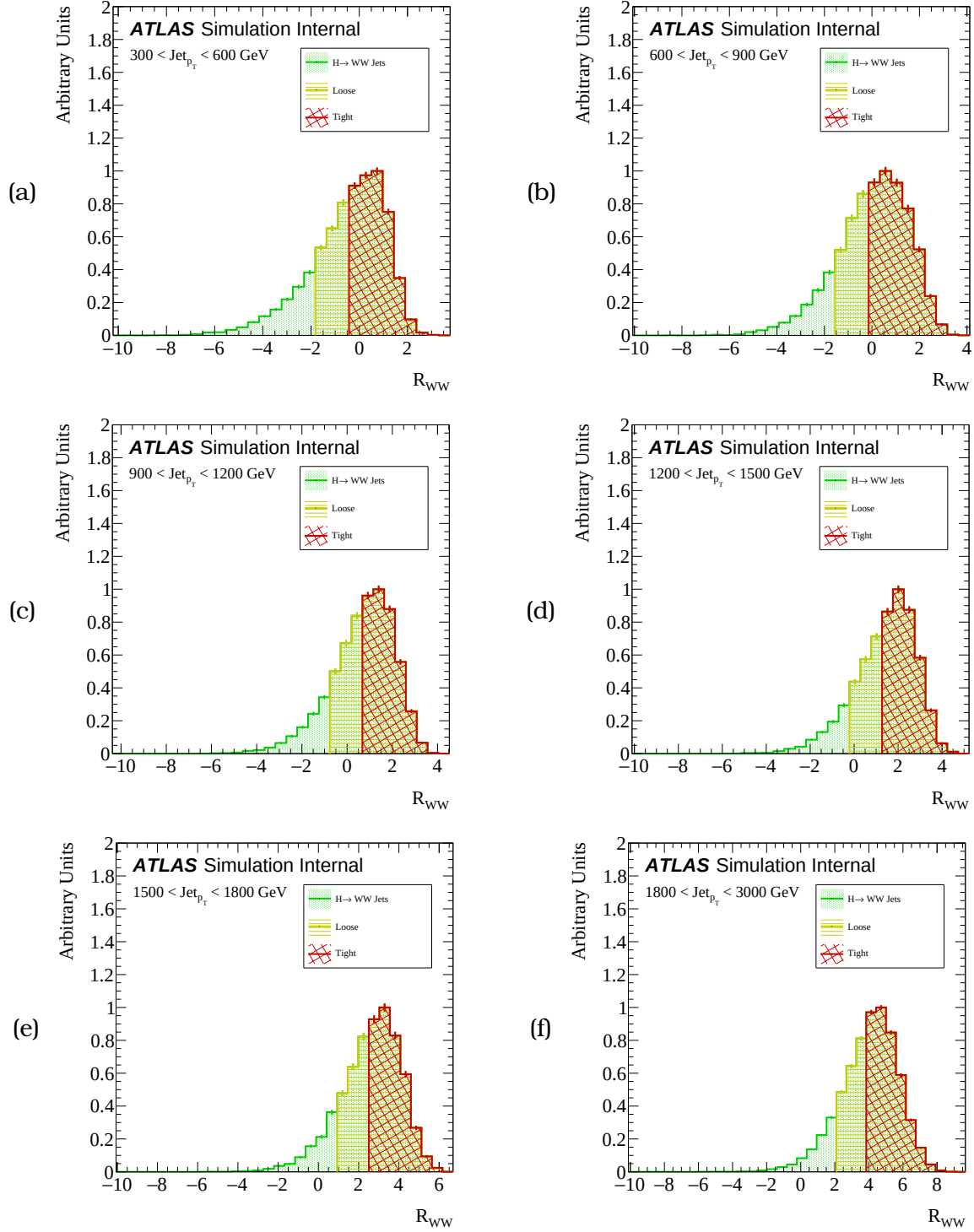


Figure B.2: The performance of the tagger rejecting QCD dijet MC versus signal efficiency. Compared here are two of the more powerful input parameters of the network and conventionally considered for signal variable discrimination.

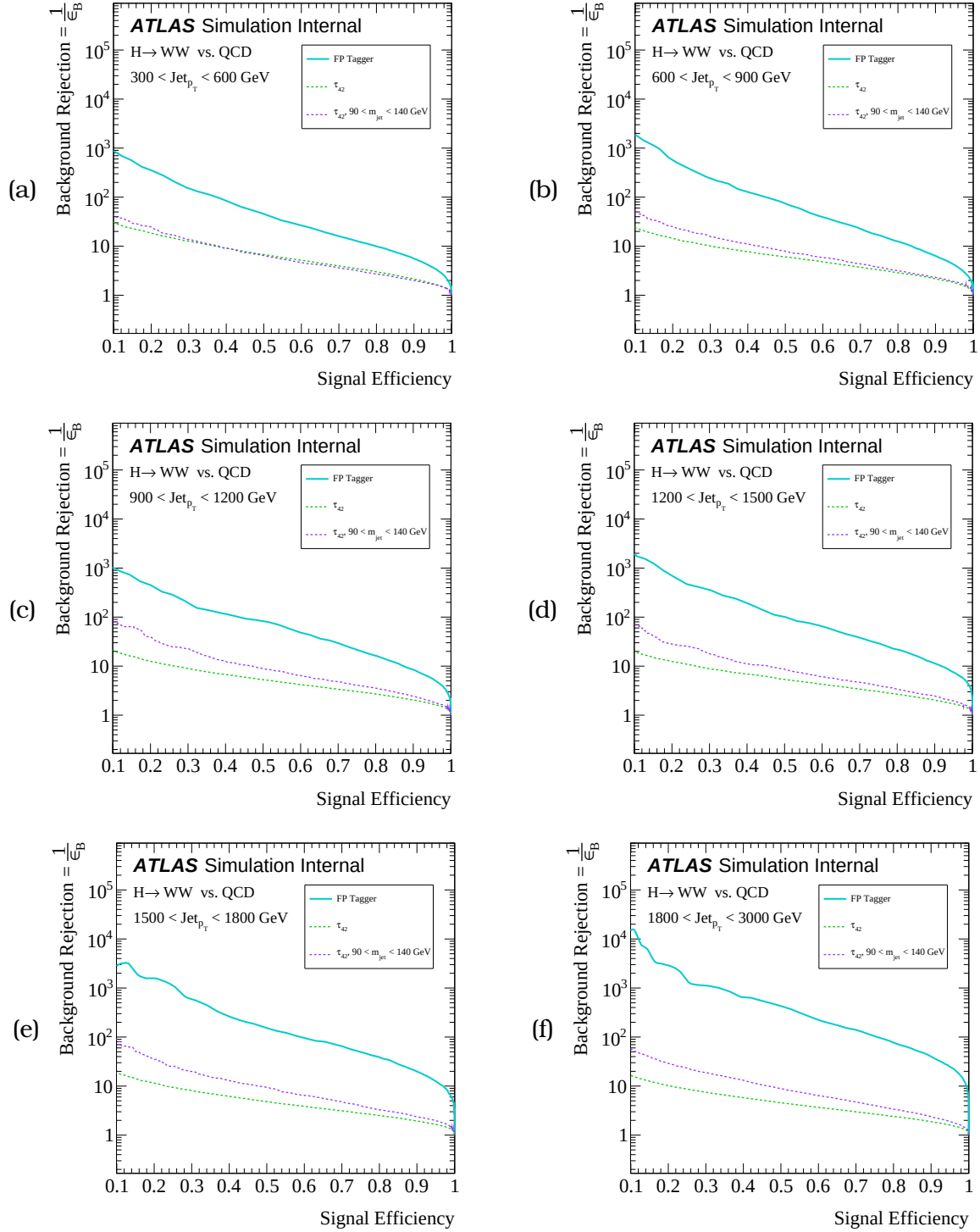


Figure B.3: The performance of the tagger rejecting QCD dijet MC versus signal efficiency. Compared here are two of the more powerful input parameters of the network and conventionally considered for signal variable discrimination.

Appendix C

SIGNAL SAMPLE VARIATIONS

AMI Tag	Model	Description
s3295	FTFP_BERT_ATL_noDiffraction	Physics modeling variations with differences in the assumptions of the scattering of the MC simulation
s3296	QGSP_FTFP_BERT_ATL	
s3297	FTFP_BERT_ATL_rescattering	
s3298	FTFP_BERT_ATL_chipsXS	
s3299	FTFP_BERT_ATL_HP	
s3373	ATLAS-R2-2016-01-00-02	Variations defining different detector models with changes to ID material quantities , geometries, and materials in the transition regions between the calorimeters.
s3374	ATLAS-R2-2016-01-00-14	
s3375	ATLAS-R2-2016-01-00-15	
s3376	ATLAS-R2-2016-01-00-16	

Table C.1: A brief description of each of the physics list variations, model inputs, as well as the detector differences summarized from [154].

Appendix D

TIGGERS

Large-R Jet Triggers
HLT_j480_a10t_lcw_jes_L1J100
HLT_j480_a10r_L1J100
HLT_j480_a10_lcw_subjes_L1J100
HLT_j460_a10r_L1J100
HLT_j460_a10_lcw_subjes_L1J100
HLT_j460_a10_lcw_subjes_L1SC111
HLT_j460_a10t_lcw_jes_L1SC111
HLT_j460_a10t_lcw_jes_L1J100
HLT_j440_a10t_lcw_jes_40smcINF_L1J100
HLT_j440_a10t_lcw_jes_L1J100
HLT_j440_a10_lcw_subjes_L1J100
HLT_j440_a10r_L1J100
HLT_j420_a10t_lcw_jes_35smcINF_L1SC111
HLT_j420_a10t_lcw_jes_35smcINF_L1J100
HLT_j420_a10t_lcw_jes_40smcINF_L1J100
HLT_j420_a10t_lcw_jes_L1J100
HLT_j420_a10r_L1J100
HLT_j420_a10_lcw_L1J100
HLT_j400_a10r_L1J100
HLT_j400_a10_lcw_L1J100
HLT_j390_a10t_lcw_jes_30smcINF_L1J100
HLT_j360_a10_sub_L1J100
HLT_j360_a10_lcw_L1J100
HLT_j360_a10r_L1J100

Table D.1: Complete list of the unrescaled single large-R jet triggers.

Appendix E

EXPERIMENTAL UNCERTAINTIES

Systematic	Description
Energy	
EG_RESOLUTION_ALL	
EG_SCALE_AF2	
EG_SCALE_ALL	
Electron	
EL_EFF_ID_TOTAL_1NPCOR_PLUS_UNCOR	
EL_EFF_Iso_TOTAL_1NPCOR_PLUS_UNCOR	
EL_EFF_Reco_TOTAL_1NPCOR_PLUS_UNCOR	
Jet	
JET_BJES_Response	
JET_EffectiveNP_Detector1	
JET_EffectiveNP_Detector2	
JET_EffectiveNP_Mixed1	
JET_EffectiveNP_Mixed2	
JET_EffectiveNP_Mixed3	
JET_EffectiveNP_Modelling1	
JET_EffectiveNP_Modelling2	
JET_EffectiveNP_Modelling3	
JET_EffectiveNP_Modelling4	
JET_EffectiveNP_Statistical1	
JET_EffectiveNP_Statistical2	
JET_EffectiveNP_Statistical3	
JET_EffectiveNP_Statistical4	
JET_EffectiveNP_Statistical5	
JET_EffectiveNP_Statistical6	

JET_EtaIntercalibration_Modelling
 JET_EtaIntercalibration_NonClosure_highE
 JET_EtaIntercalibration_NonClosure_negEta
 JET_EtaIntercalibration_NonClosure_posEta
 JET_EtaIntercalibration_TotalStat
 JET_Flavor_Composition
 JET_Flavor_Response
 JET_JER_DataVsMC_AFII
 JET_JER_EffectiveNP_1
 JET_JER_EffectiveNP_2
 JET_JER_EffectiveNP_3
 JET_JER_EffectiveNP_4
 JET_JER_EffectiveNP_5
 JET_JER_EffectiveNP_6
 JET_JvtEfficiency
 JET_Pileup_OffsetMu
 JET_Pileup_OffsetNPV
 JET_Pileup_PtTerm
 JET_Pileup_RhoTopology
 JET_PunchThrough_AFII
 JET_RelativeNonClosure_AFII
 JET_SingleParticle_HighPt

Muon

MUON_EFF_ISO_STAT
 MUON_EFF_ISO_SYS
 MUON_EFF_RECO_STAT
 MUON_EFF_RECO_STAT_LOWPT
 MUON_EFF_RECO_SYS
 MUON_EFF_RECO_SYS_LOWPT
 MUON_EFF_TTVA_STAT
 MUON_EFF_TTVA_SYS
 MUON_ID
 MUON_MS
 MUON_SAGITTA_RESBIAS
 MUON_SAGITTA_RHO
 MUON_SCALE

Tracks
PU_PRW_DATASF
TRK_BIAS_D0_WM
TRK_BIAS_QOVERP_SAGITTA_WM
TRK_BIAS_Z0_WM
TRK_EFF_LOOSE_GLOBAL
TRK_EFF_LOOSE_TIDE
TRK_FAKE_RATE_LOOSE
TRK_FAKE_RATE_LOOSE_TIDE

Table E.1: Complete list of systematic uncertainties included in the analysis.

ACRONYMS

- 2HDM** Two-Higgs-doublet model. 22, 100
- ALICE** A Large Ion Collider Experiment. 26, 27
- ATLAS** A Toroidal LHC ApparatuS. vii, viii, xvi, 1, 17–19, 21–24, 26, 29–33, 36–44, 47–51, 55–58, 60–62, 85, 94, 114, 115, 131
- AUC** Area Under Curve. xxiii, 76, 77
- BERT** Bertini Nucleon-nucleon Scattering model. 42
- BR** Branching Ratio. xv, xxiii, 16, 17, 21, 23, 61
- BSM** Beyond the Standard Model. vii, xx, 1, 17, 19, 41, 52, 59, 117, 131
- C** Charge Conjugation. 6
- C/A** Cambridge/Aachen. 49, 50, 78
- CERN** Conseil Européen pour la Recherche Nucléaire. xv, 14, 24, 26–28
- CKM** Cabibo-Kobayashi-Maskawa. 12, 14
- CMS** Compact Muon Solenoid. 1, 26
- CNN** Convolutional Neural Network. 71, 72
- COM** Center-of-Mass. 4, 5, 25, 27, 28, 30
- CP** Charge/Parity Symmetry. 6, 14
- CPT** Charge, parity, and time reversal. 6
- CSC** Cathode Strip Chambers. 36
- EBH** Englert–Brout–Higgs. 2
- EM** Electromagnetic. 5, 7, 8, 13, 33, 34, 61
- EW** Electroweak. 2, 6–8, 11, 13
- FASER** Forward Search Experiment. 26
- FCal** Liquid Argon Forward Calorimeter. xvi, 34, 35, 48
- FTF** Fritiof string model. 42
- Geant4** GEometry ANd Tracking. 42, 43, 94
- ggF** gluon-gluon Fusion. xv, xx, 15, 18, 19, 23, 131
- GSC** Global Sequential Calibration. 53
- HEC** Liquid Argon Hadronic End Cap Calorimeter. 34, 35
- HEP** High Energy Physics. 65, 66, 73, 139, 142
- HL-LHC** High Luminosity LHC. xvi, 28, 41
- HLT** high level trigger. 39, 102
- ID** Inner Detector. xvi, 29, 31–33, 36, 44, 45, 60, 61, 158
- JER** Jet Energy Resolution. 44, 52, 54, 55, 114–116
- JES** Jet Energy Scale. xvii, 44, 52–54, 102, 114–116
- KK** Kaluza-Klein. 15, 22
- LAr** Liquid Argon. 33–35, 61
- LHC** Large Hadron Collider. vii, xv, 1, 15–17, 23–25, 27–30, 38–41, 59, 115, 140
- LHCb** Large Hadron Collider beauty. 26
- LHCf** LHC Forward. 26
- Linac** Linear accelerator. 28
- LO** Leading Order. 15, 75
- MC** Monte Carlo. xvii–xxi, xxiii, xxiv, 22, 42, 43, 52–55, 74, 77, 87, 92–96, 99, 100, 103, 104, 106, 112, 116, 121, 132, 134, 158
- MDT** Monitored Drift Tubes. 36
- MET** Missing Transverse Energy. 33, 35, 44, 62
- MLP** Multilayer Perception. xvii, 67
- MoEDAL** Monopole and Exotics Detector At the LHC. 26
- MS** Muon Spectrometer. 36, 39, 60, 61
- MVA** Multivariate analysis. 62, 66, 87

NIST National Institute of Standards and Technology. 143
NP Nuisance Parameter. xxi, 119, 124, 133, 136
NPs Nuisance Parameters. 118, 124
P Parity Transformation. 6
PD Pixel Detector. 31, 32, 45
PDF Parton Distribution Function. xx, 26, 42, 118, 120
PDF Probability Density Function. 100, 117, 118, 120, 121
PFlow Particle Flow. 56, 57
PS Proton Synchrotron. 28, 116
PSB Proton Synchrotron Booster. 28

QCD Quantum Chromodynamics. xvii, xviii, xxi, xxii, 11, 62, 64, 73, 74, 81, 84, 88, 91–95, 100, 104, 106, 145, 156, 157
QED Quantum Electrodynamics. 6
QM Quantum Mechanics. 7

ReLU Rectifier activation function. xvii, 68, 83, 85
RF Radiofrequency. 28
RPC Resistive Plate Chambers. 36

SCT semi-conductor tracker. 31, 32, 45
SF Scale Factor. 96

SGD Stochastic Gradient Descent. 70
SM Standard Model. vii, viii, xv, xx, 1, 2, 4–9, 12–19, 22–24, 41, 60, 117, 120, 121, 131
SPS Super Proton Synchrotron. 28
SSB Spontaneous Symmetry Breaking. 2, 8, 13, 18

T Time Reversal. 6
TAR Track Assisted Reclustered algorithm. xvii, 44, 56–59, 65, 98, 100–102, 115
TAS Track-assisted subjet algorithm. 58
TCC Track-Calo-Cluster. 56, 57
TGC Thin Gap Chambers. 36
TOTEM TOTAl cross section, Elastic scattering and diffraction dissociation Measurement at the LHC. 26
TRT Transition Radiation Tracker. 31–33

UFO Unified Flow Objects. 56

VBF Vector Boson Fusion. 15
VEV Vacuum Expectation Value. xv, 9, 10
VH Vector Boson Associated. 15
VR Variable Radius. 56, 101

WP Working Point. xviii, xix, xxiii, xxiv, 60, 89, 90, 93, 105, 116, 154

REFERENCES

- [1] The ATLAS Collaboration. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. In: *arXiv* (2012). eprint: 1207.7214 (page 1).
- [2] The CMS Collaboration. “Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC”. In: *arXiv* (2012). eprint: 1207.7235 (page 1).
- [3] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *JINST* 3 (2008), S08001 (pages 1, 24, 25, 29).
- [4] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509 (page 1).
- [5] Claudia Marcelloni. “John Ellis, CERN theorist, working on the blackboard. John Ellis, théoricien, travaillant au tableau noir”. 2007 (page 2).
- [6] E. Noether. “Invariante Variationsprobleme”. In: *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 1918 (1918), pp. 235–257 (page 4).
- [7] D. Goldberg. “The standard model in a nutshell”. In: *Contemporary Physics* 59 (Nov. 2017), pp. 1–1 (page 4).
- [8] J. D. Jackson and L. B. Okun. “Historical roots of gauge invariance”. In: *arXiv* (2000). eprint: hep-ph/0012061 (page 5).
- [9] Heilbron J. L. “Electricity in the 17th and 18th Centuries: A Study of Early Modern Physics”. In: *The British Journal for the History of Science* 13.3 (1980), 270–272 (page 5).
- [10] Enrico Fermi. “Versuch einer Theorie der β -Strahlen. I”. In: *Zeitschrift für Physik* 88 (1934), pp. 161–177 (page 6).
- [11] C. S. Wu et al. “Experimental Test of Parity Conservation in Beta Decay”. In: *Phys. Rev.* 105 (4 1957), pp. 1413–1415 (page 6).
- [12] J. H. Christenson et al. “Evidence for the 2π Decay of the K_2^0 Meson”. In: *Phys. Rev. Lett.* 13 (4 1964), pp. 138–140 (page 6).
- [13] S. M. Bilenky and J Hosek. “Glashow-Weinberg-Salam theory of electroweak interactions and the neutral currents”. In: *Phys. Rep.; (Netherlands)* 90:2 (1982) (page 7).
- [14] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Reading, USA: Addison-Wesley, 1995. ISBN: 978-0-201-50397-5 (page 8).

- [15] Wojciech Bizoń, Ulrich Haisch, and Luca Rottoli. “Constraints on the quartic Higgs self-coupling from double-Higgs production at future hadron colliders”. In: *arXiv* (2018). eprint: 1810.04665 (page 9).
- [16] Francesco Armando Di Bello. “Search for boosted Higgs boson and other resonances decaying into b -quark pairs using the ATLAS detector and studies of CMOS pixel sensors for the HL-LHC”. Presented 02 Sep 2019. 2019 (page 11).
- [17] J. Charles et al. “CP Violation and the CKM Matrix: Assessing the Impact of the Asymmetric B Factories”. In: *arXiv* (2004). CV. eprint: hep-ph/0406184 (page 12).
- [18] L. Boyle, Wikipedia, the free encyclopedia. *Standard Model Of Particle Physics Most Complete Diagram*. [CC BY-SA 4.0]. 2014 (page 13).
- [19] V. Gribov and B. Pontecorvo. “Neutrino astronomy and lepton charge”. In: *Physics Letters B* 28.7 (1969), pp. 493–496 (page 14).
- [20] Kamiokande Collaboration and Y. Fukuda et Al. “Evidence for oscillation of atmospheric neutrinos”. In: *arXiv* (1998). eprint: hep-ex/9807003 (page 14).
- [21] Planck Collaboration et al. “Planck 2018 results. I. Overview and the cosmological legacy of Planck”. In: *arXiv* (2018). eprint: 1807.06205 (page 14).
- [22] Steven Weinberg. “The cosmological constant problem”. In: *Rev. Mod. Phys.* 61 (1 1989), pp. 1–23 (page 15).
- [23] Lisa Randall and Raman Sundrum. “A Large Mass Hierarchy from a Small Extra Dimension”. In: *arXiv* (1999). eprint: hep-ph/9905221 (page 15).
- [24] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. Geneva: CERN, 2016 (page 16).
- [25] S Heinemeyer et al. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. Ed. by S Heinemeyer. CERN Yellow Reports: Monographs. Geneva: CERN, 2013 (page 16).
- [26] Atlas and Cms Collaborations. “Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments”. In: *arXiv* (2015). eprint: 1503.07589 (page 16).
- [27] D. de Florian et al. “Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector”. In: *arXiv* (2016). eprint: 1610.07922 (page 16).
- [28] *HHGDD < AtlasProtected < TWiki*. [Online; accessed 12. Oct. 2021]. 2021 (page 17).
- [29] ATLAS Collaboration. “Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *arXiv* (2019). eprint: 1906.02025 (page 18).
- [30] *Summary of non-resonant and resonant Higgs boson pair searches from the ATLAS experiment*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2021-031>. Geneva: CERN, 2021 (page 18).

- [31] ATLAS Collaboration. “Search for Higgs boson pair production in the $\gamma\gamma b\bar{b}$ final state with 13 TeV pp collision data collected by the ATLAS experiment”. In: *arXiv* (2018). eprint: 1807.04873 (page 18).
- [32] *Combination of searches for non-resonant and resonant Higgs boson pair production in the $b\bar{b}\gamma\gamma$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}b\bar{b}$ decay channels using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, 2021 (page 19).
- [33] *S to HH(S) to bbWW diagrams*. 2021 (pages 20, 21).
- [34] Daniel A. Camargo et al. “A Two Higgs Doublet Model for Dark Matter and Neutrino Masses”. In: *arXiv* (2019). eprint: 1901.05476 (page 22).
- [35] Th. Kaluza. “On the Unification Problem in Physics”. In: *arXiv* (2018). eprint: 1803.08616 (page 22).
- [36] Oskar Klein. “Quantentheorie und fünfdimensionale Relativitätstheorie”. In: *Z. Phys.* 37.12 (1926), pp. 895–906 (page 22).
- [37] ATLAS Collaboration. *Search for Higgs boson pair production in the two bottom quarks plus two photons final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. Geneva: CERN, 2021 (page 23).
- [38] B. Di Micco et al. “Higgs boson potential at colliders: status and perspectives”. In: *arXiv* (2019). eprint: 1910.00012 (page 23).
- [39] Stefan Höche. “Introduction to parton-shower event generators”. In: *arXiv* (2014). eprint: 1411.4085 (page 26).
- [40] Esma Mobs. “The CERN accelerator complex - 2019. Complexe des accélérateurs du CERN - 2019”. In: (2019). General Photo (page 27).
- [41] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *JINST* 3 (2008), S08003 (pages 26, 29, 33).
- [42] CMS Collaboration. “The CMS experiment at the CERN LHC”. In: *JINST* 3 (2008), S08004 (page 26).
- [43] The ALICE Collaboration. “The ALICE experiment at the CERN LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08002–S08002 (page 26).
- [44] The LHCb Collaboration. “The LHCb Detector at the LHC”. In: *Journal of Instrumentation* 3.08 (2008), S08005–S08005 (page 26).
- [45] The TOTEM Collaboration. “The TOTEM Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (2008), S08007–S08007 (page 26).
- [46] The LHCf Collaboration et al. “The LHCf detector at the CERN Large Hadron Collider”. In: 3.08 (2008), S08006–S08006 (page 26).
- [47] James Pinfold. “The MoEDAL Experiment at the LHC - Searching for Physics Beyond the Standard Model”. In: *EPJ Web Conf.* 182 (2018), 02096. 7 p (page 26).
- [48] FASER Collaboration et al. “FASER: ForwArD Search ExpeRiment at the LHC”. In: *arXiv* (2019). eprint: 1901.04468 (page 26).

- [49] O. Aberle et al. *High-Luminosity Large Hadron Collider (HL-LHC): Technical design report*. CERN Yellow Reports: Monographs. Geneva: CERN, 2020 (page 28).
- [50] L Arnaudon et al. *Linac4 Technical Design Report*. Tech. rep. revised version submitted on 2006-12-14 09:00:40. Geneva: CERN, 2006 (page 28).
- [51] The ATLAS TDAQ Collaboration. “The ATLAS Data Acquisition and High Level Trigger system”. In: *J. Instrum.* 11.06 (2016), P06008–P06008 (page 29).
- [52] Izaak Neutelings. *B tagging jets*. [Online; accessed 23. Oct. 2021]. 2021 (pages 30, 67).
- [53] Norbert Wermes and G Hallewel. *ATLAS pixel detector: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1998 (page 31).
- [54] ATLAS Collaboration. “Performance of the ATLAS Transition Radiation Tracker in Run 1 of the LHC: tracker properties”. In: *arXiv* (2017). eprint: 1702.06473 (page 32).
- [55] *ATLAS inner detector: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997 (page 32).
- [56] Bartosz Mindur. *ATLAS Transition Radiation Tracker (TRT): Straw tubes for tracking and particle identification at the Large Hadron Collider*. Tech. rep. Geneva: CERN, 2016 (page 32).
- [57] *ATLAS tile calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996 (pages 33, 35).
- [58] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1996 (page 34).
- [59] ATLAS Collaboration. “Jet energy scale and resolution measured in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *arXiv* (2020). eprint: 2007.02645 (pages 34, 52).
- [60] The ATLAS Collaboration. “Commissioning of the ATLAS Muon Spectrometer with Cosmic Rays”. In: *arXiv* (2010). eprint: 1006.4384 (page 35).
- [61] *ATLAS muon spectrometer: Technical Design Report*. Technical design report. ATLAS. Geneva: CERN, 1997 (page 36).
- [62] Y. Arai et al. “ATLAS muon drift tube electronics”. In: *JINST* 3 (2008), P09001 (page 36).
- [63] *ATLAS magnet system: Technical Design Report, 1*. Technical design report. ATLAS. Geneva: CERN, 1997 (page 36).
- [64] ATLAS Collaboration. *ATLAS Magnet System: Technical Design Report, 1*. ATLAS-TDR-6. 1997 (page 36).
- [65] Joao Pequenaio and Paul Schaffner. “How ATLAS detects particles: diagram of particle paths in the detector”. 2013 (page 37).
- [66] F. Lasagni Manghi. *LUCID: The ATLAS Luminosity Detector*. Tech. rep. Geneva: CERN, 2018 (page 38).

- [67] Marco A. L. Leite. “Performance of the ATLAS Zero Degree Calorimeter”. In: *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. 2013, pp. 1–5 (page 38).
- [68] S. Abdel Khalek et al. “The ALFA Roman Pot detectors of ATLAS”. In: 11.11 (2016), P11013–P11013 (page 38).
- [69] L Adamczyk et al. *Technical Design Report for the ATLAS Forward Proton Detector*. Tech. rep. 2015 (page 38).
- [70] Takanori Kono. “Performance and evolution of the ATLAS TDAQ system with pp collisions at 7 TeV”. In: *2012 18th IEEE-NPSS Real Time Conference*. 2012, pp. 1–7 (page 39).
- [71] ATLAS Data Preparation Group. *intlumivsyar*. [Online; accessed 12. Oct. 2021]. 2019 (page 39).
- [72] ATLAS Collaboration. “ATLAS data quality operations and performance for 2015–2018 data-taking”. In: *arXiv* (2019). eprint: 1911.04632 (pages 39, 40).
- [73] Z Marshall. *Simulation of Pile-up in the ATLAS Experiment*. Tech. rep. Geneva: CERN, 2013 (page 39).
- [74] *Project Schedule | HL-LHC Industry*. [Online; accessed 12. Oct. 2021]. 2021 (page 41).
- [75] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *arXiv* (2014). eprint: 1405.0301 (pages 42, 99).
- [76] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *arXiv* (2010). eprint: 1002.2581 (pages 42, 100).
- [77] Torbjörn Sjöstrand et al. “An Introduction to PYTHIA 8.2”. In: *arXiv* (2014). eprint: 1410.3012 (pages 42, 100).
- [78] Enrico Bothmann et al. “Event Generation with Sherpa 2.2”. In: *arXiv* (2019). eprint: 1905.09127 (pages 42, 100).
- [79] Johannes Bellm et al. “Herwig 7.0/Herwig++ 3.0 release note”. In: *Eur. Phys. J. C* 76.4 (2016), p. 196. arXiv: 1512.01178 [hep-ph] (pages 42, 95, 100).
- [80] S. Agostinelli et al. “GEANT4 – a simulation toolkit”. In: *Nucl. Instrum. Meth. A* 506 (2003), p. 250 (page 42).
- [81] B. Andersson, G. Gustafson, and B. Nilsson-Almqvist. “A model for low-pT hadronic reactions with generalizations to hadron-nucleus and nucleus-nucleus collisions”. In: *Nuclear Physics B* 281.1 (1987), pp. 289–309 (page 42).
- [82] A. B. Kaidalov and K. A. Ter-Martirosian. “Multiple Production of Hadrons at High-Energies in the Model of Quark-Gluon Strings”. In: *Sov. J. Nucl. Phys.* 39 (1984), p. 979 (page 42).
- [83] E Abat and et.al Abdallah. *Response and Shower Topology of 2 to 180 GeV Pions Measured with the ATLAS Barrel Calorimeter at the CERN Test-beam and Comparison to Monte Carlo Simulations*. Tech. rep. Geneva: CERN, 2010 (page 42).

- [84] ATLAS Collaboration. “Performance of the ATLAS track reconstruction algorithms in dense environments in LHC Run 2”. In: *Eur. Phys. J. C* 77 (2017), p. 673. arXiv: 1704.07983 [hep-ex] (page 45).
- [85] ATLAS Collaboration. *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in High Pile-Up LHC Environment*. ATLAS-CONF-2012-042. 2012 (page 45).
- [86] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 262.2 (1987), pp. 444–450 (page 45).
- [87] R. Frühwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Methods Phys. Res., Sect. A* 262.2 (1987), pp. 444–450 (page 45).
- [88] ATLAS collaboration. “A neural network clustering algorithm for the ATLAS silicon pixel detector”. In: *arXiv* (2014). eprint: 1406.7690 (page 46).
- [89] T. G. Cornelissen et al. “Updates of the ATLAS Tracking Event Data Model (Release 13)”. In: (June 2007) (page 47).
- [90] ATLAS Collaboration. “Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton–proton collisions at the LHC”. In: *Eur. Phys. J. C* 77 (2017), p. 332. arXiv: 1611.10235 [hep-ex] (page 46).
- [91] ATLAS Collaboration. “Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1”. In: *Eur. Phys. J. C* 77 (2017), p. 490. arXiv: 1603.02934 [hep-ex] (pages 47, 48).
- [92] G. P. Salam. “Towards Jetography”. In: *arXiv* (2009). eprint: 0906.1833 (page 49).
- [93] J E Huth et al. “Toward a standardization of jet definitions”. In: (1990), 7 p (page 49).
- [94] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In: *arXiv* (2008). eprint: 0802.1189 (page 50).
- [95] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The Catchment Area of Jets”. In: *arXiv* (2008). eprint: 0802.1188 (pages 50, 58).
- [96] Yu. L. Dokshitzer et al. “Better Jet Clustering Algorithms”. In: *arXiv* (1997). eprint: hep-ph/9707323 (page 50).
- [97] ATLAS Collaboration. “Performance of jet substructure techniques for large- R jets in proton–proton collisions at $\sqrt{s} = 7$ TeV using the ATLAS detector”. In: *arXiv* (2013). eprint: 1306.4945 (page 51).
- [98] Ahmed Ali and Gustav Kramer. “Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD”. In: *arXiv* (2010). eprint: 1012.2288 (page 52).
- [99] ATLAS Collaboration. “Jet energy scale measurements and their systematic uncertainties in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”.

- In: *Phys. Rev. D* 96 (2017), p. 072002. arXiv: 1703.09665 [hep-ex] (pages 53, 54).
- [100] ATLAS Collaboration. “Jet energy resolution in proton-proton collisions at $\sqrt{s} = 7$ TeV recorded in 2010 with the ATLAS detector”. In: *arXiv* (2012). eprint: 1210.6210 (page 54).
 - [101] ATLAS Collaboration. “Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector”. In: *arXiv* (2019). eprint: 1910.04482 (page 55).
 - [102] ATLAS Collaboration. *Jet Calibration and Systematic Uncertainties for Jets Reconstructed in the ATLAS Detector at $\sqrt{s} = 13$ TeV*. ATL-PHYS-PUB-2015-015. 2015 (page 55).
 - [103] ATLAS Collaboration. *Variable Radius, Exclusive- k_T , and Center-of-Mass Subjet Reconstruction for Higgs($\rightarrow b\bar{b}$) Tagging in ATLAS*. ATL-PHYS-PUB-2017-010. 2017 (page 56).
 - [104] ATLAS Collaboration. *Boosted Object Tagging with Variable- R Jets in the ATLAS Detector*. ATL-PHYS-PUB-2016-013. 2016 (page 56).
 - [105] ATLAS Collaboration. “Jet reconstruction and performance using particle flow with the ATLAS Detector”. In: *Eur. Phys. J. C* 77 (2017), p. 466. arXiv: 1703.10485 [hep-ex] (page 56).
 - [106] ATLAS Collaboration. *Jet reclustering and close-by effects in ATLAS Run 2*. ATLAS-CONF-2017-062. 2017 (page 58).
 - [107] Matteo Cacciari and Gavin P. Salam. “Pileup subtraction using jet areas”. In: *arXiv* (2007). eprint: 0707.1378 (page 58).
 - [108] ATLAS Collaboration. *Track assisted techniques for jet substructure*. ATL-PHYS-PUB-2018-012. 2018 (pages 58, 59).
 - [109] ATLAS Collaboration. *Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run*. ATL-PHYS-PUB-2017-013. 2017 (page 60).
 - [110] ATLAS Collaboration. “Muon reconstruction performance of the ATLAS detector in proton-proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 76 (2016), p. 292. arXiv: 1603.05598 [hep-ex] (page 61).
 - [111] ATLAS Collaboration. “Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data”. In: *Eur. Phys. J. C* 77 (2017), p. 195. arXiv: 1612.01456 [hep-ex] (page 61).
 - [112] ATLAS Collaboration. “Electron and photon energy calibration with the ATLAS detector using LHC Run 1 data”. In: *Eur. Phys. J. C* 74 (2014), p. 3071. arXiv: 1407.5063 [hep-ex] (page 61).
 - [113] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13$ TeV”. In: *Eur. Phys. J. C* 79 (2019), p. 639. arXiv: 1902.04655 [hep-ex] (page 61).

- [114] ATLAS Collaboration. *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC*. ATL-PHYS-PUB-2015-045. 2015 (page 62).
- [115] ATLAS Collaboration. *Performance of Missing Transverse Momentum Reconstruction in ATLAS studied in Proton-Proton Collisions recorded in 2012 at $\sqrt{s} = 8$ TeV*. ATLAS-CONF-2013-082. 2013 (page 62).
- [116] J. Veatch. *Dedicated object reconstruction techniques for $bbVV$* . [Online; accessed 13. Oct. 2021]. 2018 (pages 64, 65).
- [117] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. <http://www.deeplearningbook.org>. MIT Press, 2016 (pages 66, 70, 72).
- [118] Xavier Glorot and Yoshua Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Ed. by Yee Whye Teh and Mike Titterton. Vol. 9. Proceedings of Machine Learning Research. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010, pp. 249–256 (page 68).
- [119] John Bridle. “Training Stochastic Model Recognition Algorithms as Networks can Lead to Maximum Mutual Information Estimation of Parameters”. In: *Advances in Neural Information Processing Systems*. Ed. by D. Touretzky. Vol. 2. Morgan-Kaufmann, 1990 (page 69).
- [120] Jack H. Collins, Kiel Howe, and Benjamin Nachman. “Anomaly Detection for Resonant New Physics with Machine Learning”. In: *arXiv* (2018). eprint: 1805.02664 (page 69).
- [121] *Identification of Boosted Higgs Bosons Decaying Into $b\bar{b}$ With Neural Networks and Variable Radius Subjets in ATLAS*. Tech. rep. Geneva: CERN, 2020 (page 69).
- [122] Kaiming He et al. “Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification”. In: *arXiv* (2015). eprint: 1502.01852 (page 71).
- [123] *Ian Goodfellow: Generative Adversarial Networks (GANs) | MIT | Artificial Intelligence Podcast*. [Online; accessed 8. Nov. 2021]. 2019 (page 72).
- [124] J. Thaler and K. Van Tilburg. “Identifying Boosted Objects with N-subjettiness”. In: *arXiv* (2010). eprint: 1011.2268 (page 75).
- [125] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. “Energy Correlation Functions for Jet Substructure”. In: *arXiv* (2013). eprint: 1305.0007 (page 75).
- [126] Ian Mout, Lina Necib, and Jesse Thaler. “New Angles on Energy Correlation Functions”. In: *arXiv* (2016). eprint: 1609.07483 (pages 75, 76).
- [127] Luke de Oliveira et al. “Jet-Images – Deep Learning Edition”. In: *arXiv* (2015). eprint: 1511.05190 (pages 77, 78).
- [128] Frederic A. Dreyer, Gavin P. Salam, and Gregory Soyez. “The Lund Jet Plane”. In: *arXiv* (2018). eprint: 1807.04758 (pages 78, 79).
- [129] Bo Andersson et al. “Coherence Effects in Deep Inelastic Scattering”. In: *Z. Phys. C* 43 (1989), p. 625 (page 78).

- [130] Frédéric A. Dreyer and Huilin Qu. “Jet tagging in the Lund plane with graph networks”. In: *arXiv* (2020). eprint: 2012.08526 (page 78).
- [131] François Chollet et al. *Keras*. <https://keras.io>. 2015 (page 83).
- [132] Martín Abadi et al. “TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems”. In: (2015). Software available from [tensorflow.org](https://www.tensorflow.org) (page 83).
- [133] Corinna Cortes, Mehryar Mohri, and Afshin Rostamizadeh. “L2 Regularization for Learning Kernels”. In: *arXiv* (2012). eprint: 1205.2653 (page 83).
- [134] Geoffrey E. Hinton et al. “Improving neural networks by preventing co-adaptation of feature detectors”. In: *arXiv* (2012). eprint: 1207.0580 (page 85).
- [135] Fabian Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830 (page 85).
- [136] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *arXiv* (2014). eprint: 1412.6980 (page 87).
- [137] Gilles Louppe, Michael Kagan, and Kyle Cranmer. “Learning to Pivot with Adversarial Networks”. In: *arXiv* (2016). eprint: 1611.01046 (page 91).
- [138] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. arXiv: 1405.0301 [hep-ph] (page 95).
- [139] Richard D. Ball et al. “Parton distributions with LHC data”. In: *Nucl. Phys. B* 867 (2013), p. 244. arXiv: 1207.1303 [hep-ph] (pages 95, 100).
- [140] ATLAS Collaboration. “Combination of searches for Higgs boson pairs in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *arXiv* (2019). eprint: 1906.02025 (page 98).
- [141] ATLAS Collaboration. *Updated Luminosity Determination in pp Collisions at $\sqrt{s} = 7$ TeV using the ATLAS Detector*. ATLAS-CONF-2011-011. 2011 (page 99).
- [142] D. J. Lange. “The EvtGen particle decay simulation package”. In: *Nucl. Instrum. Meth. A* 462 (2001), p. 152 (page 100).
- [143] Richard D. Ball et al. “Parton distributions for the LHC Run II”. In: *JHEP* 04 (2015), p. 040. arXiv: 1410.8849 [hep-ph] (page 100).
- [144] S. Abachi et al. “Search for the top quark in $p\bar{p}$ collisions at $\sqrt{s} = 1.8$ TeV”. In: *Phys. Rev. Lett.* 72 (14 1994), pp. 2138–2142 (page 112).
- [145] ATLAS Collaboration. *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at the LHC*. ATLAS-CONF-2019-021. 2019 (page 115).
- [146] ATLAS Collaboration. *Search for resonant pair production of Higgs bosons in the $b\bar{b}b\bar{b}$ final state using 139 fb^{-1} of $\sqrt{s} = 13$ TeV pp collision data with the ATLAS detector*. Tech. rep. Geneva: CERN, 2021 (page 116).
- [147] S. S. Wilks. “The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses”. In: *The Annals of Mathematical Statistics* 9.1 (1938), pp. 60–62 (page 117).

- [148] Abraham Wald. “Tests of statistical hypotheses concerning several parameters when the number of observations is large”. In: *Transactions of the American Mathematical Society* 54 (1943), pp. 426–482 (page 117).
- [149] Li-Gang Xia. “Improved Asymptotic Formulae for Statistical Interpretation Based on Likelihood Ratio Tests”. In: *arXiv* (2021). eprint: 2101.06944 (page 119).
- [150] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *arXiv* (2010). eprint: 1007.1727 (page 120).
- [151] Abraham Wald. “Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large”. In: *Transactions of the American Mathematical Society* 54.3 (1943), pp. 426–482 (page 120).
- [152] *Vol. 4 (2017): Proceedings of the 2015 European School of High-Energy Physics | CERN Yellow Reports: School Proceedings*. [Online; accessed 11. Oct. 2021]. 2015 (page 120).
- [153] Thomas Philippe Calvet. “Search for the production of a Higgs boson in association with top quarks and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC”. Presented 08 Nov 2017. 2017 (page 122).
- [154] Alexander Ekman. “Derivation of Improved K_L^0 Response Uncertainty for use in the Jet Energy Scale Calibration at ATLAS”. Master’s Thesis. MA thesis. 2019 (page 158).

Curriculum Vitae

Blake C. Forland

lyon.bcf@gmail.com

Education

Ph.D.	Indiana University, Physics	Dec. 2021
M.S.	Indiana University, Physics	Sep. 2014
B.S.	Metro State University of Denver, Physics	Dec. 2013

Research Experience

Graduate Research Assistant	Jan 2017 - Dec 2021
Indiana University, CERN, ATLAS Experiment	
Undergraduate Research Assistant	Jan 2011 - May 2013
Metro State University, High Altitude Observatory	
Research Experience for Undergraduate	2012
Harvard Center for Astrophysics Physics	

Publications

Gibson S., et al. *FORWARD: A toolset for multiwavelength coronal magnetometry*, Front. Astron. Space Sci., 3, 8, 2016

Bak-Steslicka, U., et al. *The spatial relation between EUV cavities and linear polarization signatures*, Proceedings IAU Symposium No. 300, 2013

Forland, B., et al. *Coronal cavity survey: morphological clues to eruptive magnetic topologies*, Topical Issue in Solar Physics on Coronal Magnetism, 288, 603, 2013.

Bak-Steslicka, U., et al. *The magnetic structure of solar prominence cavities: new observational signature revealed by coronal magnetometry*, ApJL, 770, 28, 2013.

Conferences & Talks

<i>HH Readiness for Run 3</i> , Chair Di-Higgs Workshop	2020
--	------

<i>Tagging with TAR jets</i> , Talk Higgs and Diboson Searches Plenary	2019,2020
<i>Analysis Improvement: Machine Learning Techniques</i> , Talk Di-Higgs Workshop	2019
<i>Tagging large-R jets with ML Combination</i> , Talk Interdisciplinary Machine Learning Workshop	2018
<i>A ML Based 4-prong Jet Tagger</i> , Talk Interdisciplinary Machine Learning Workshop	2017
<i>FORWARD Code: Now with widgets</i> , Poster International Astronomical Union 8	2013

Schools

Italian National Institute for Nuclear Physics School of Statistic	2019
Princeton Computational and Data Science School for High Energy Physics	2018
European School of High Energy Physics Invitation	2018

Teaching Experience

Graduate Assistant Instructor Indiana University	2014-2016
--	-----------

Leadership Experience

HEP Congressional Advocacy Liaison	2019
Physics Graduate Student Council Founder	2017
Graduate and Professional Student Government Vice President	2016-2017