

ABSTRACT

Utilizing Electrons in the Search for Associated Higgs Boson Production with the ATLAS Detector: Higgs decaying to a tau pair and vector boson decaying leptonically

Savannah J. Thais

2019

The Higgs boson was discovered by the ATLAS and CMS collaborations in 2012 using data from $\sqrt{s}=8\text{TeV}$ proton-proton collisions at the LHC. Since the initial discovery of the $H \rightarrow 4l$ and $H \rightarrow \gamma\gamma$ decays, multiple other Higgs analyses of production modes and decay channels have reached discovery significance. This thesis describes the ongoing search for the still unobserved vector boson ($V = W^\pm, Z$) associated Higgs production with the Higgs decaying to a tau lepton pair using 139 fb^{-1} of $\sqrt{s}=13\text{ TeV}$ proton-proton collision data collected by the ATLAS detector during Run-2. This analysis requires the vector boson to decay leptonically in order to reduce background contributions from hadronic activity in the detector and this thesis focuses primarily on the usage of final state electrons in the analysis.

The primary backgrounds in all analysis categories are misidentified (or ‘faked’) objects; these contributions are estimated using a data-driven technique which relies on Machine Learning (ML) for object identification and reconstruction. ML is used broadly in High Energy Physics analyses and this work is introduced, with a focus on techniques for improving electron identification through image processing.



Utilizing Electrons in the Search for Associated Higgs Boson Production with the ATLAS Detector: Higgs decaying to a tau pair and vector boson decaying leptonically

A DISSERTATION
PRESENTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF
YALE UNIVERSITY
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
SAVANNAH J. THAIS

DISSERTATION DIRECTOR: PROFESSOR SARAH DEMERS

JUNE 2019

©2019 – SAVANNAH J. THAIS
ALL RIGHTS RESERVED.

Contents

LISTING OF FIGURES	xiv
LISTING OF TABLES	xvi
ACKNOWLEDGEMENTS	xviii
o INTRODUCTION	i
1 THEORY OVERVIEW	3
1.1 The Standard Model	4
1.1.1 Particles	6
1.1.2 Forces and Interactions	7
1.2 Mathematical Formalism	8
1.2.1 Electroweak	9
1.2.1.1 QED	9
1.2.1.2 The Weak Interaction	10
1.2.1.3 Electroweak Unification	11
1.2.2 QCD	12
1.3 The Higgs Boson	13
1.3.1 Electroweak Symmetry Breaking	13
1.3.2 Yukawa Coupling	16
1.3.3 Higgs Production Methods	17
1.3.3.1 Gluon-Gluon Fusion	17
1.3.3.2 Vector Boson Fusion	18
1.3.3.3 Associated Production with a Vector Boson	19
1.3.3.4 Top Fusion	19
1.3.4 Higgs Boson Decays	20
1.3.4.1 Massive Fermions	20
1.3.4.2 Massive Bosons	22
1.3.4.3 Massless Particles	22
2 THE LHC AND THE ATLAS DETECTOR	23
2.1 Particle Colliders	24
2.1.1 The Large Hadron Collider	26
2.1.2 The LHC Accelerator Complex	28

2.2	Particle Detection	30
2.2.1	The ATLAS Detector	31
2.2.1.1	Inner Detector	33
2.2.1.2	Calorimeters	35
2.2.1.3	Muon Spectrometer	40
2.2.1.4	Magnet System	42
3	ATLAS COMPUTING AND SOFTWARE	44
3.1	Data Acquisition and Storage	45
3.1.1	Triggers and DAQ	46
3.1.1.1	Trigger Menus	48
3.1.2	Storage	50
3.2	Simulation	51
3.2.1	Event Generation	52
3.2.2	Detector Simulation	53
3.3	Data Processing	55
3.3.1	Reconstruction	56
3.3.1.1	Tracking	56
3.3.1.2	Vertexing	58
3.3.1.3	Calorimeter Clustering	60
3.3.2	Identification	62
3.3.2.1	Muons	62
3.3.2.2	Jets	63
3.3.2.3	Tau Leptons	67
3.3.2.4	Missing Transverse Energy	68
4	MACHINE LEARNING	71
4.1	Central Concepts	72
4.2	Algorithms	74
4.2.1	Boosted Decision Trees	74
4.2.2	Neural Networks	76
4.2.2.1	Convolutional NNs	79
4.2.2.2	Recurrent NNs	81
4.2.2.3	Adversarial Networks	83
4.3	Applications in LHC Physics	84
4.3.1	Reconstruction, Identification, and Calibration	85
4.3.1.1	BDTs and Standard NNs	85
4.3.1.2	Images and CNNs	87
4.3.1.3	RNNs	89
4.3.2	Analyses	90
4.3.2.1	$ZZ_d \rightarrow lll$ Search	91

4.3.3	Simulation	91
4.3.4	Systematics	93
5	ELECTRONS IN ATLAS	96
5.1	Electron Identification (ID) Software	97
5.1.1	The Likelihood Function	98
5.1.1.1	Electron Likelihood	100
5.1.1.2	Operating Points	103
5.1.2	Probability Density Functions (PDFs)	105
5.1.2.1	Tag & Probe Method	105
5.1.2.2	Smoothing PDFs	106
5.1.2.3	Data vs Monte Carlo PDFs	107
5.1.2.4	Correcting PDFs	109
5.2	Low p_T Electron ID	110
5.2.1	Motivation	110
5.2.2	Software Development	112
5.2.3	Variable Optimization	113
5.3	Machine Learning and Electron ID	116
5.3.1	Future Studies	120
6	$VH, H \rightarrow \tau\tau$ ANALYSIS	122
6.1	Motivation	123
6.2	Definitions	123
6.2.1	Object Criteria	125
6.2.1.1	Taus	125
6.2.1.2	Muons	125
6.2.1.3	Electrons	125
6.2.1.4	Jets	127
6.2.1.5	MET	127
6.2.2	Overlap Removal	128
6.2.3	Mass Reconstruction	128
6.2.3.1	Missing Mass Calculator	128
6.2.3.2	Late Transverse Projected Mass	130
6.3	Analysis Categories	131
6.3.1	$WH, \tau_{lep}\tau_{had}$	133
6.3.2	$WH, \tau_{had}\tau_{had}$	134
6.3.3	$ZH, \tau_{lep}\tau_{had}$	134
6.3.4	$ZH, \tau_{had}\tau_{had}$	135
6.4	Run-1 Results	135

7	BACKGROUND ESTIMATION IN $VH, H \rightarrow \tau\tau$ ANALYSIS	138
7.1	Background Types	139
7.1.1	Irreducible Backgrounds	140
7.1.2	Fake Backgrounds	140
7.2	The Fake Factor Method	141
7.2.1	Method Derivation	141
7.2.1.1	Two Object Case	142
7.2.1.2	Three and Four Object Cases	146
7.3	Fake Factor Measurements	147
7.3.1	Fake Region	147
7.3.2	Fake Candidate Selection	148
7.3.3	Fake Electrons	148
7.3.3.1	Type and Origin Studies	151
7.3.4	Fake Taus and Muons	155
7.3.4.1	Fake Taus	155
7.3.4.2	Fake Muons	157
7.3.5	MC Corrections	157
7.4	Fake Factor Software	158
7.4.1	Closure Tests	159
7.4.1.1	WH hadhad Channel	159
7.4.1.1.1	Pre-selection	159
7.4.1.1.2	$Z \rightarrow \tau\tau$ Control Region	161
8	CONCLUSIONS	164
8.1	$VH, H \rightarrow \tau\tau$ Analysis	165
8.1.1	Analysis Status	165
8.1.2	Expected Events	166
8.1.3	Machine Learning in the Analysis	167
8.2	Final Remarks	168
	APPENDIX A AUTHOR'S INDIVIDUAL CONTRIBUTIONS	169
A.1	Electron ID	169
A.1.1	Related Presentations	170
A.2	$VH, H \rightarrow \tau\tau$ Analysis	171
A.2.1	Related Presentations	171
A.3	Machine Learning	172
A.3.1	Related Presentations	172
A.4	Outreach	173
A.4.1	Related Presentations	173
	REFERENCES	186

Listing of figures

1.1	Summary of several SM production cross section measurements, corrected for leptonic branching fractions, and compared to the corresponding theoretical predictions [6].	4
1.2	Table formation of the Standard Model which includes the mass, charge, and spin of all known SM particles.	5
1.3	The SM interaction vertices. α describes the relative strength of the force as it is the dimensionless transformation of the coupling constant g [2].	8
1.4	An illustration of the Higgs potential in the case that $\mu^2 < 0$ [20].	14
1.5	Dominant SM Higgs production cross-sections as a function of Higgs mass [21].	17
1.6	Feynman diagram of gluon-gluon fusion Higgs boson production with a top or bottom quark loop	18
1.7	Feynman diagram of vector boson fusion Higgs boson production	18
1.8	Feynman diagram of associated Higgs boson production with a vector boson	19
1.9	Feynman diagram of top fusion Higgs boson production	20
1.10	Branching ratios for the main decays of a SM Higgs boson as a function of Higgs mass. The theoretical uncertainties are shown as bands [12].	21
2.1	Standard Model production cross-section predictions at hadron-hadron colliders[36]	25
2.2	Left: total integrated luminosity delivered by the LHC (green), recorded by ATLAS (yellow) and usable for physics (blue). Right: Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy p-p collisions [41].	27
2.3	LHC planned operating schedule until 2038. EYETS stands for Extended Year End Technical Stop, a lengthened version of the annual brief shutdown during LHC running to allow for repairs. LS stands for Long Shutdown.	28
2.4	HL-LHC luminosity projection after completing Run 2	28
2.5	The CERN Accelerator Complex	29
2.6	The luminosity-weighted distribution of the mean number of interactions per crossing for the Run 2 pp collision data at 13 TeV center-of-mass energy [41]	31
2.7	Diagram of the various subsystems of the ATLAS detector [49]	32
2.8	Illustration of impact parameters in the transverse plane (left) and R-Z plane (right).	33
2.9	The r - z cross-section of one quadrant of the ATLAS Inner Detector. The top figure is the whole inner detector and the lower figure is a magnified view of only the pixel detector [50].	34
2.10	Cut-away view of the ATLAS calorimeter system [49]	36
2.11	The accordion geometry used in the EM calorimeter [53].	36

2.12	Cummulative amount of material, in units of interaction length, for different components of the ATLAS calorimeter. Tile_i refers to layers of the TileCal, HEC_i refers to layers of the Hadronic End Caps, and FCAL_i refers to layers of the Forward Calorimeter [49].	39
2.13	Schematic diagram of a ϕ wedge of the Tile Calorimeter [54].	40
2.14	Cut-away view of the ATLAS muon system [49].	41
2.15	A schematic picture showing a quarter-section of the muon system in a plane containing the beam axis [56].	42
2.16	Schematic diagram of the ATLAS magnet system	43
3.1	A high-level summary of the ATLAS data flow [58].	45
3.2	Schematic diagram of the ATLAS TDAQ System in Run 2. The FTK is still in the commissioning phase [60].	46
3.3	Schematic view of the trigger towers used as input to the L_1 Calorimeter trigger algorithms	47
3.4	Physics trigger group rates for the L_1 (left) and HLT (right) as a function of the number of luminosity blocks from a fill taken in July 2016 [62].	50
3.5	The flow of the ATLAS simulation software, from event generators (top left) through reconstruction (top right). Algorithms are placed in square-cornered boxes and persistent data objects are placed in rounded boxes. The optional pile-up portion of the chain, used only when events are overlaid, is dashed [68].	51
3.6	Pictorial representation of a generated ttH event. The central hard interaction is shown as the central red circle, which is followed by the subsequent decays of the Higgs and top quarks (smaller red circles). Additional hard radiation due to QCD is produced (blue) as well as a secondary hard interaction (purple). Finally, hadronization (light green) and hadron decay (dark green) occurs. Photon radiation is represented in yellow and occurs at multiple stages [69].	53
3.7	Illustraion of a single particle space-point (left) and merged particle space-point (right) on a pixel sensor [85].	57
3.8	Single track reconstruction efficiency as a function of particle p_T for a ρ (orange), three prong τ (red), five prong τ (green), and B^0 (blue) [85].	58
3.9	The number of vertices reconstructed as a function of the average interactions per bunch crossing for two fills taken in 2018 [89].	59
3.10	Clusters for a simulated dijet event created in the FCAL using the topological clustering algorithm [91].	61
3.11	Depiction of the four types of reconstructed muons	63
3.12	Muon reconstruction efficiency as a function of η measured in $Z \rightarrow \mu\mu$ events for muons with $p_T > 10$ GeV [97].	64
3.13	The four main jet clustering algorithms performed on the same data with the same input radius [98].	66

3.14	Efficiency for hadronic tau identification (open) and combined reconstruction and identification (closed) for one-prong (left) and three-prong (right) taus [104].	68
3.15	Distribution of MET for an inclusive sample of $Z \rightarrow \mu\mu$ events extracted from data and compared to MC [105].	70
4.1	Depiction of an algorithmically constructed decision tree.	75
4.2	Schematic diagram of a boosted forest of decision trees [119].	76
4.3	Schematic diagram of a simple fully-connected neural network.	77
4.4	Diagram showing the application of a single convolutional filter to a single window of an input image [125].	79
4.5	Two examples of interpretable convolutional filters for image processing [125]. . .	80
4.6	A graphical representation of max pooling [125].	80
4.7	Schematic depiction of a full convolutional neural network built for image classification.	81
4.8	A single layer recurrent neural network unfolded to show a series of time steps in data training.	82
4.9	A single LSTM node across three time steps. In this example, sigmoid and tanh activation functions are used and c_i represents the previous state of information within the node.	83
4.10	An example adversarial neural network where the primary task is classification [127].	84
4.11	Performance of BDTs and deep NNs (DNNs) for W-jet tagging (left) and top-jet tagging (right). For both plots the black curve represents the standard cut-based method [131].	86
4.12	Average of several pre-processed W-jet images. Here the Q_1 and Q_2 axes represent the transformed $\eta - \phi$ space [137].	87
4.13	Performance comparison of CNNs (purple dashed and black curves), a Max-Out network (blue dashed curve) and cut-based discriminators using jet substructure variables (red, blue dotted, and green dashed curves) on the W- vs QCD-jet discrimination task [138].	88
4.14	Schematic of QCD-motivated RNN-based jet-clustering and classification [142]. .	89
4.15	ROC-curve comparison of $H \rightarrow ZZ^* \rightarrow llll$ event selection algorithms performance.	92
4.16	Comparison of LAGAN generated (solid lines) and Pythia generated (dotted lines) jet kinematics (jet mass (left), (jet p_T (center), and n-subjettiness (right)) for signal (W-jets, red lines) and background (QCD jets, blue lines) [146].	93
4.17	Approximate median significance as a function of decision threshold for networks trained with different values of λ [127].	94
5.1	Schematic depiction of an electron traversing the ATLAS detector.	97

5.2	The transformed LH-based identification discriminant d'_L for reconstructed electron candidates. The black histogram is for prompt electrons in a $Z \rightarrow ee$ simulation sample, and the red (dashed-line) histogram is for backgrounds in a generic two-to-two process simulation sample (both simulation samples are described in Section 5.1.2). The histograms are normalised to unit area.	99
5.3	Example of a distribution of an electron variable - f_1 described in Table 5.1 - that would be inefficient if used in a cut-based ID but improves the LH based ID [150].	99
5.4	Ratio of background efficiencies for a LH based algorithm to that of the closest-efficiency cut-based selections as a function of η (left) and E_T (right) [151].	100
5.5	Electron identification efficiencies for $Z \rightarrow ee$ events as a function of E_T (left) and pseudo-rapidity (right). The efficiencies are shown for data (full) and MC (open) for three LH based operating points [154].	104
5.6	The KDE Method. Left: a stylized depiction of the standard KDE method (top) and Adaptive KDE method (bottom). Right: an example of a raw variable distribution (grey) and the KDE-smoothed PDF (black).	108
5.7	Comparisons of data-driven (blue) and MC-based (red) PDFs in the $7 \text{ GeV} < p_T < 10 \text{ GeV}$ and $0.8 < \eta < 1.15$ bin for two variables. $\Delta\eta_1$ on the left and eProbabilityHT on the right.	108
5.8	The f_3 (left) and R_{had} (right) PDFs for data (black dashed line) and MC (solid red) in the $30 \text{ GeV} < p_T < 40 \text{ GeV}$ and $0.8 < \eta < 1.15$ bin. The MC is shown before (solid red) and after (blue line) corrections have been applied.	110
5.9	Left: An example plot of the output discriminant for probes (data and $Z \rightarrow ee$ MC) in the $10 \text{ GeV} < p_T < 15 \text{ GeV}$ and $0.80 < \eta < 1.15$ bin. The data peak at high discriminant value comes from electrons but the larger peak at low discriminant value is background. Right: the same plot for $40 \text{ GeV} < p_T < 45 \text{ GeV}$. The fraction of background in the data sample for this p_T range is negligible.	111
5.10	Production of a pair of squarks with an ISR jet[158].	111
5.11	Pseudo-proper time for all probe electron candidates. The prompt signal component is shown by the dashed blue line and the non-prompt signal component is shown by the light blue dashed line.	112
5.12	Dielectron invariant-mass fit for all probe electron candidates. The pseudo-proper time is required to be $-1 \text{ ps} < \tau < 0.2 \text{ ps}$. Dots with error bars represent the opposite-sign (OS) pairs for data, the fitted J/ψ signal is shown by the dashed blue and the $\psi(2S)$ by the dashed light blue lines. The sum of the background contributions is depicted as a purple dotted line.	113
5.13	Electron reconstruction and identification efficiencies as a function of transverse energy E_T , integrated over the full pseudo-rapidity range.	114
5.14	n-1 LH ROC curves. The nominal LH with all variables included is the dashed black line.	115

5.15	Comparisons of signal (red) and background (black) PDF shapes for new low p_T variables in the $4.5 \text{ GeV} < p_T < 7 \text{ GeV}$ and $1.81 < \eta < 2.01$ bin. Top left: charge $\cdot d_0$, top right: $\Delta(\Delta(\phi))$, bottom: $\Delta(\text{curvature})/\text{error}(\text{curvature})$	116
5.16	A sample electron image for each of the four EM calorimeter layers.	117
5.17	The same electron image from Figure 5.16 after being up-sampled to a uniform granularity of 128×128	118
5.18	Network architecture for image-based electron ID and calibration.	118
5.19	Classifier output of the image-based electron ID CNN for signal (red) and background (blue).	119
5.20	Normalized difference between predicted particle energy and truth particle energy for the image-based calibration CNN (orange) and the standard ATLAS electron calibration algorithm (blue).	119
5.21	A schematic depiction of a 3D convolution for use in a CNN.	120
6.1	Cross-sections times branching fractions for relevant Higgs production and decay modes normalized to the SM predicted values. The values are obtained by a simultaneous fit to all channels [162].	124
6.2	Progressive cut flow for a portion of the ZH lep-had analysis category (after requiring exactly 3 leptons and one hadronic tau). Cut 2 is the p_T cut of interest and other cuts are defined in Table 6.1.	126
6.3	Examples of the probability distribution functions of ΔR for 1-prong hadronic tau decays (left), 3-prong hadronic tau decays (middle) and leptonic tau decays (right) [170].	129
6.4	Run 1 MMC distribution after signal region cuts for the ZH lep-had channel. [171].	129
6.5	Depiction of particle partitioning in the M_{2T} method [172].	131
6.6	Run 1 M_{2T} distributions for three Higgs mass values after signal region cuts for the WH lep-had channel (left) and WH had-had channel (right) [171].	132
6.7	Feynman diagram for the leading order quark-initiated WH production.	133
6.8	Feynman diagram for the leading order quark-initiated ZH production. There are also contributions from gluon initiated triangle and box diagrams.	134
6.9	The Run-1 95% CIs on the ratio of measured cross-section to SM cross-section for each channel individually and all four combined [171].	136
6.10	The Run-1 signal strength μ for each analysis channel individually and all four combined [171].	137
7.1	Computing flow diagram for the MC-based irreducible background estimation (left) and data-driven fake background estimation (right).	140
7.2	Di-muon visible mass in the $Z \rightarrow \mu\mu$ Fake Region for events with an anti-selected electron (left) and a selected electron (right).	148
7.3	Measured electron fake rate for the ZH channels.	150
7.4	Measured electron fake rate for the WH channels.	151

7.5	Electron origins (as defined in Table 7.7) in the ZH Fake Region. Selected electrons on the left and anti-selected electrons on the right.	153
7.6	Electron origins (as defined in Table 7.7) in the ZH Signal Region. Selected electrons on the left and anti-selected electrons on the right	153
7.7	Electron origins (as defined in Table 7.7) in the WH Fake Region. Selected electrons on the left and anti-selected electrons on the right.	154
7.8	Electron origins (as defined in Table 7.7) in the WH Signal Region. Selected electrons on the left and anti-selected electrons on the right	154
7.9	Run-1 distribution of the hadronic tau E_T as closure test for the fake factor method in the WH lep-had channel [171].	160
7.10	WH had-had pre-selection closure test tau-1 η distribution.	161
7.11	WH had-had pre-selection closure test lepton p_T distribution.	161
7.12	WH had-had pre-selection closure test MET distribution distribution.	162
7.13	WH had-had $Z \rightarrow \tau\tau$ closure test lepton p_T distribution.	162
7.14	WH had-had $Z \rightarrow \tau\tau$ closure test MET distribution.	163

Listing of tables

2.1	Coverage, granularity, and number of readout channels of the ATLAS calorimeter system [49].	38
3.1	The main ATLAS triggers for the 2017 trigger menu with observed rates [61]. . . .	49
3.2	Numbers of materials and volumes used to construct the ATLAS detector simulated geometry [68].	54
3.3	ID efficiencies for prompt muons from W decays and the misidentification rates for hadron decays computed from a $t\bar{t}$ MC sample [97].	64
4.1	Performance for the various event selection algorithms trained to select exotic Higgs events. NN refers to the shallow NN, DN refers to the deep NN, and complete means the combination of high-level and low-level features [143].	90
5.1	Definitions of electron discriminating variables, the types of backgrounds the variables help to discriminate against (light-flavor (LF), converted photons (γ), or heavy-flavor (HF)), and if a variable is used as a likelihood PDF ($\mathcal{L}$) or used as a rectangular cut (C). The * refers to the fact that the E/p and w_{stot} variables are only used for electrons with $p_T > 150$ GeV for the Tight identification operating point, and are not used for the looser operating points.	101
5.2	Electron transverse energy binning used for the electron likelihood PDFs and discriminant cut values.	103
5.3	Electron pseudorapidity binning used for the electron likelihood PDFs and discriminant cut values.	103
5.4	Summary of Tag & Probe selection for $Z \rightarrow ee$ events	106
5.5	Summary of Tag & Probe selection for $J/\psi \rightarrow ee$ events.	107
6.1	Cut definitions for Figure 6.2. See Section 6.3.1 for additional details.	127
6.2	Run 2 triggers used in the four analysis channels along with corresponding trigger and offline p_T thresholds.	132
6.3	The Run-1 expected and observed significance for each of the four analysis channels [171].	136
7.1	Fake factor method definitions.	142
7.2	Truth-level event types and their probabilities of entering the Signal Region. . . .	143
7.3	Possible reconstructions for truth event types.	144

7.4	The number of events that are reconstructed a certain way.	144
7.5	Measured electron fake rate for the ZH channels.	149
7.6	Measured electron fake rate for the WH channels.	150
7.7	Electron origin numbering scheme.	152
7.8	Measured 1-prong tau fake rates for the ZH channels.	155
7.9	Measured 3-prong tau fake rates for the ZH channels.	156
7.10	Measured 1-prong tau fake rates for the WH channels.	156
7.11	Measured 3-prong tau fake rates for the WH channels.	156
7.12	Preliminary measured muon fake rates.	157
8.1	Expected event yield in the full Run-2 data-set for the ZH and WH channels of the $VH, H \rightarrow \tau\tau$ analysis.	166

TO MENTAL HEALTH PROFESSIONALS EVERYWHERE.

Acknowledgments

THOUGH MANY REGARD THE PhD AS A SOLITARY ENDEAVOR (and it has certainly felt that way during my last few weeks of writing), in my experience that couldn't be farther from the truth. I have gone through a graduate program at an institution I didn't even dream of attending while growing up; this document and this experience have only been possible because of the support, encouragement, and love of many people.

I have had the great privilege to work with not one, but two incredible advisors during my time at Yale. I thank Professor Keith Baker for always believing in me, helping me to learn and grown as a scientist, and encouraging me during the most difficult periods of this journey. I thank Professor Sarah Demers for being the kind of person someone dreams of having advise their thesis; thank you for helping me build a research path I love, for all the work you do in this department, ATLAS, and beyond, and for being an incredible mentor and role-model as a scientist and woman. I also thank Professor Thomas Appelquist and Professor Bonnie Flemming for serving on my thesis committee. Thank you to my UChicago physics family, Young-kee, David, Bjoern, Yangyang, and Jeff, for building my foundations and giving me a second home at CERN.

I'm grateful to the colleagues I have worked with during the past five years for inspiring me with their dedication and brilliant ideas. I have the deepest admiration for Joey Riechert and Lucas Flores and appreciate all the time we spent working on electrons together. Similarly, I thank Eleonora Benhar for being the best post-doc I could have ever asked for during those years. Thank you to my research group Mariel, Emma, and Lotte for owning this analysis and battling the xTau Framework with me.

Thank you to all my dear friends in ATLAS and CERN: its been a joy to learn with you, make change, and see the world. I am especially thankful for my angry lady crew, Julia, Lesya, Leo, Jennet, and Rachel; you are all bad-asses and the physics world is lucky to have you. Thank you to the two people who have been like my ATLAS older siblings, Micky and Dan; I can't begin to express how much you both have done for me. Dan, you single-handedly talked me out of dropping out and Micky, I honestly don't know who I would be as a physicist and ML practitioner without you.

Thank you to all the people who have made New Haven feel like home over the past years. Kat, Kelly, Yuri, Josh, Darryl, Danny, Deeha, Matt, Margot, Zak, and so many others. I'm thankful to have been a part of communities full of beautiful individuals: The Grad Women's Rugby Team, Danceworks, and Senate.

I extend my deepest gratitude to all the people in IOP, DBT, and MHC who have spent the time to help me heal and grow; thank you to Frances Addachi, Kelly Workman, Ashley Heely, Emily Cooley, and Ryan Wallace. Above all, thank you to Paul Hoffman, your words and work will remain with me for the rest of my life.

I am grateful to my parents for always providing opportunities for me to grow, explore, and chase my dreams. I thank my father for reminding me that education is the one thing no one can take away from you and my mother for showing me how to teach and how to learn. To my sister, thank you for always being proud of me and for showing me it's ok to forge your own path. I also thank the artists whose words I've included in these chapters for giving me a connection to home even when I'm far away.

To my best friends, Annina, Nataly, and Annie, it took all four of us to get me through this journey. You may not have helped me with QFT homework or edited my conference slides, but I can say with absolute certainty that I would not be here without each of you. I love you so much.

Finally, thank you to Sohan. You have held my hand through the lowest points of my life, loved me when I didn't deserve it, and never stopped believing in me. And always told me how smart I was (for an experimentalist).

We do not believe in ourselves until someone reveals that deep inside us something is valuable, worth listening to, worthy of our trust, sacred to our touch. Once we believe in ourselves we can risk curiosity, wonder, spontaneous delight or any experience that reveals the human spirit.

E. E. Cummings

Make voyages. Attempt them. There is nothing else.

Tennessee Williams



Introduction

THE LARGE HADRON COLLIDER AND THE ATLAS DETECTOR ARE SOME OF THE MOST IMPRESSIVE FEATS OF THE HUMAN SCIENTIFIC ENDEAVOR. They are marvels of innovation, design, engineering, and computing, yes, but also of community, collaboration, and curiosity. Many unique physical and software components and an extraordinary group of individuals from diverse backgrounds are necessary to build, maintain, and run these machines and to make productive use of the data they provide.

The ATLAS computing system alone contains a remarkable number of unique tools each designed to accommodate task-specific constraints and challenges. These range from hardware triggers that must process nearly 60 million GB of data per second, to reconstruction algorithms that ef-

ficiently find patterns in complex, chaotic environments, to physics analyses that seek to precisely separate a rare signal from abundant backgrounds while limiting uncertainty.

This is, then, an incredibly exciting research area for students interested in software and computing development. In many cases, it is possible to incorporate novel ideas and innovative methods in this development while still maintaining the advantages and robustness of previous designs. In particular, it is possible to use industry standard Machine Learning techniques in many ATLAS tools. In fact, these techniques are becoming increasingly prevalent within the ATLAS experiment, and in some cases, in particular the planned High Luminosity LHC, they are necessary to address computing challenges.

This work is one example of working to solve some of these challenges. It has involved extensive software development to extend current particle ID functionality (thereby enabling a range of new physics searches), to model complex high-dimensional backgrounds, and to develop and enhance Run-2 physics analyses. More broadly, it represents one individual's efforts to help develop improved computing practices and tools within ATLAS and understand the effects of new technologies on High Energy Particle Physics.

This thesis includes introductions to the mathematical formulation of particle physics and machine learning as well as an overview of the current ATLAS computing systems and challenges it faces. It then describes studies undertaken at the intersection of these three areas to improve physics measurements and search for new processes.

*It's not the beauty of a building you should look at; it's
the construction of the foundation that will stand the test
of time.*

David Allen Coe

1

Theory Overview

THE CHARACTERISTICS AND INTERACTIONS OF PARTICLES CAN BE DESCRIBED BY MATHEMATICAL LAWS. In the age of modern particle physics, there is substantial interplay between the theoretical development of these laws and their experimental validation. Experiments like those at the Large Hadron collider test initial models and experimental results are used to retune theories and construct new ones. This chapter describes those particles which have been experimentally verified and the forces and interactions which govern them. Further descriptions and a more in-depth mathematical formulation can be found in [1], [2], [3], [4], and [5].

1.1 THE STANDARD MODEL

The Standard Model (SM) of particle physics describes the fundamental particles and forces that make up all visible matter in the universe. It is considered one of the most comprehensive and powerful scientific theories ever developed due to the wide range of phenomena it describes and its predictive success. As shown in Figure 1.1, many experimental results have been consistent with SM predictions.

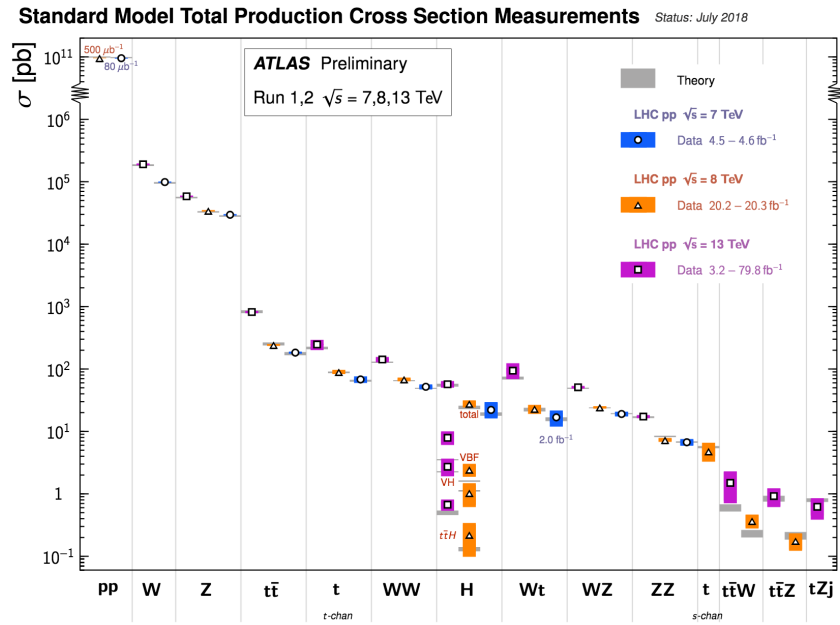


Figure 1.1: Summary of several SM production cross section measurements, corrected for leptonic branching fractions, and compared to the corresponding theoretical predictions [6].

The SM particles can be summarized in a table structure (Figure 1.2), much like the Periodic Table of Elements. The particles are organized by type. For the first three columns, the top two rows describe the quarks while the bottom two rows describe the leptons; generations of particles increase from left to right. These particles are described in additional detail in Section 1.1.1. The fourth col-

umn describes the gauge bosons which mediate three of the four fundamental forces: electromagnetism, the strong force, and the weak force. These are discussed further in Section 1.1.2. Finally, the Higgs Boson is included on the far right. The Higgs couples to the quarks, bosons, and charged leptons and this interaction gives rise to particle masses. The Higgs is described in Section 1.3.

mass →	$\approx 2.3 \text{ MeV}/c^2$	$\approx 1.275 \text{ GeV}/c^2$	$\approx 173.07 \text{ GeV}/c^2$	0	$\approx 126 \text{ GeV}/c^2$
charge →	$2/3$	$2/3$	$2/3$	0	0
spin →	$1/2$	$1/2$	$1/2$	1	0
	u up	c charm	t top	g gluon	H Higgs boson
QUARKS	$\approx 4.8 \text{ MeV}/c^2$ $-1/3$ $1/2$	$\approx 95 \text{ MeV}/c^2$ $-1/3$ $1/2$	$\approx 4.18 \text{ GeV}/c^2$ $-1/3$ $1/2$	0 0 1	
	d down	s strange	b bottom	γ photon	
	$0.511 \text{ MeV}/c^2$ -1 $1/2$	$105.7 \text{ MeV}/c^2$ -1 $1/2$	$1.777 \text{ GeV}/c^2$ -1 $1/2$	$91.2 \text{ GeV}/c^2$ 0 1	
	e electron	μ muon	τ tau	Z Z boson	
LEPTONS	$< 2.2 \text{ eV}/c^2$ 0 $1/2$	$< 0.17 \text{ MeV}/c^2$ 0 $1/2$	$< 15.5 \text{ MeV}/c^2$ 0 $1/2$	$80.4 \text{ GeV}/c^2$ ± 1 1	
	ν_e electron neutrino	ν_μ muon neutrino	ν_τ tau neutrino	W W boson	
				GAUGE BOSONS	

Figure 1.2: Table formation of the Standard Model which includes the mass, charge, and spin of all known SM particles.

Although the SM has been extensively verified and has impressive predictive power, it does not fully explain all observed phenomena. For instance, there is currently no quantum theory of gravity and no explanation for the observed matter-antimatter asymmetry of the universe. Additionally, physicists have observed strong astrophysical evidence for dark matter [7] and dark energy [8] yet the SM fails to provide any particle-based rationalization. Furthermore, within the SM itself there is no motivation for the relatively light Higgs mass (compared to the Planck mass) or the relatively

large top quark mass (compared to the mass of the other 5 quarks). Collectively, these and other phenomena have motivated a robust theoretical and experimental interest in Beyond the Standard Model (BSM) physics. Many proposed BSM particles are potentially accessible at the LHC (eg [9]), but those theories are generally beyond the scope of this thesis.

1.1.1 PARTICLES

The SM describes two types of particles: fermions and bosons. Standard quantum mechanical descriptors like mass, charge (or lack of charge), and spin are not sufficient to entirely characterize all the particles of SM, and so physicists have introduced additional quantum numbers, flavor and color, to further described internal degrees of freedom [10]. Additionally, as indicated by the Charge, Parity, and Time Symmetry Theorem [11], all SM particles have a corresponding anti-particle of the same mass but opposite charge.*

The fundamental SM fermions all have spin $1/2$ and can be categorized by the types of interactions they experience. All fermions experience the weak force. There are six SM leptons, some of which are charged and some of which are neutral. The charged leptons are the electron (e), muon (μ), and tau (τ), which all have charge e and consequently also experience the electromagnetic force. The uncharged leptons are the electron neutrino (ν_e), the muon neutrino (ν_μ), and the tau neutrino (ν_τ). Due to their lack of charge they do not experience the electromagnetic force. Additionally, their masses are extremely small and are not described by the SM. All leptons are colorless and therefore do not experience the strong force.

There are six types of quarks: up (u), down (d), charm (c), strange (s), top (t), and bottom (b). They all carry a fractional charge which allows them to experience the electromagnetic force in addition to the weak force. Additionally, each quark carries a color charge which allows them to experi-

*This is not necessarily true for neutrinos as the SM provides no mechanism for neutrino mass creation. This is an ongoing area of study, see [12] for additional information

ence the strong force. Due to color confinement, only color neutral combinations can exist as stable particles. Consequently, quarks combine in quark-antiquark pairs called mesons and quark triplets called baryons. Mesons and baryons are collectively referred to as hadrons, the most common of which are protons and neutrons.

The fermions are organized into three generations. The electron, electron neutrino, up quark, and down quark form the first generation and interact to form all visible, stable, matter in the universe. The second and third generations have progressively higher mass and are unstable.* Thus far, no searches for additional generations of fermions have been successful and there is compelling indirect evidence that no further generations exist [13].

There are four spin 1 gauge bosons in the SM: the $W^\pm$, the Z^0 , the photon (γ), and the gluon (g). The gauge bosons are referred to as “force carriers” because they are the particles exchanged in order to mediate the three fundamental forces in the SM. The $W^\pm$ and Z^0 are massive and mediate the weak force. Both the photon and gluon are massless; the photon mediates the electromagnetic force while the gluon mediates the strong force.

1.1.2 FORCES AND INTERACTIONS

There are four fundamental forces in nature (electromagnetic, strong, weak, and gravitational) but only three are included in the SM. There is presently no quantum formulation of gravity which is very weak at the quantum scale.

As described above, each of the three SM forces are mediated by a gauge boson. To first order these forces can be represented as interaction vertices between the mediating particle and the fermions which experience that force (Figure 1.3). The strength of a fundamental interaction between a fermion and gauge boson is determined by the coupling constant of the interaction vertex.

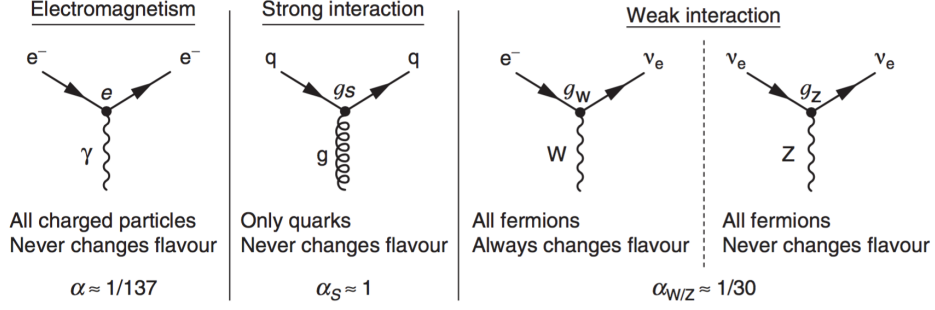


Figure 1.3: The SM interaction vertices. α describes the relative strength of the force as it is the dimensionless transformation of the coupling constant g [2].

1.2 MATHEMATICAL FORMALISM

Mathematically, each of the three SM forces is described by a quantum field theory (QFT), where the mediating particle is treated as a vector field. These QFTs can be represented in group theory as symmetry groups which describe the invariant transformations of each force. Collectively, the SM obeys a $SU(3) \times SU(2) \times U(1)$ group symmetry. The $SU(3)$ group arises from Quantum Chromo-dynamics (QCD), the QFT describing the strong force. The $SU(2) \times U(1)$ groups arise from Electroweak theory, the QFT which collectively describes the weak and electromagnetic forces.

Additionally, physicists utilize Lagrangian formalism to describe the kinetic and potential terms that regulate the interactions of these fields and their relative strengths. The Lagrangian formulations of each force begin with the Dirac equation for the free propagation of a fermion of mass m : $(i\gamma^\mu \partial_\mu - m)\psi$. Here $\gamma_\mu = (\gamma^0, \gamma^1, \gamma^2, \gamma^3)$, the Dirac gamma matrices which satisfy the anti-commutation relations $\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}$. ψ is a spinor field solution to the Dirac equation and the Lagrangian describing ψ is $\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi$. $\bar{\psi}$ is the Dirac adjoint $\bar{\psi} = \psi^\dagger \gamma^0$ and $\psi^\dagger$ is the Hermitian conjugate. This also defines the probability density $\bar{\psi}\psi$ and the probability current $j^\mu = \bar{\psi}\gamma^\mu\psi$.

The mathematical formulation of each force is outlined in the following sections.

1.2.1 ELECTROWEAK

Electroweak theory is the unification of Quantum Electrodynamics (QED), the QFT of the electromagnetic force, and the Weak theory. It represents the electromagnetic and weak forces as low energy manifestations of the same underlying force. Above the unification energy of 246 GeV, they would merge into a single electroweak force [14].

1.2.1.1 QED

In the development of QED the free particle Lagrangian discussed above is subject to the additional requirement of gauge invariance to spinor phase transformations: $\psi \rightarrow e^{iq\phi(x)}\psi$. This invariance is achieved by introducing the covariant derivative $D_\mu = \partial_\mu - iqA_\mu$. The covariant derivative includes an additional vector field A_μ to cancel the extra phase term in the transformed Lagrangian. In order for the Lagrangian to be fully invariant, A_μ must transform as $A_\mu \rightarrow \partial_\mu\phi$.

Thus, the full QED Lagrangian becomes

$$\mathcal{L} = \bar{\psi}i\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - q\bar{\psi}\gamma^\mu A_\mu\psi - \frac{1}{4}\mathcal{F}^{\mu\nu}\mathcal{F}_{\mu\nu}.$$

Here $\mathcal{F}^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ is the anti-symmetric gauge invariant electromagnetic field strength tensor. The first term in the Lagrangian is the fermionic kinetic term, the second term is the fermion mass term, the third term is the interaction term, and the final term is the kinetic term for the A_μ field. The field kinetic term can be understood in terms of classical electromagnetic theory as giving rise to the electric and magnetic fields: $-\frac{1}{4}\mathcal{F}^{\mu\nu}\mathcal{F}_{\mu\nu} = \frac{E^2 - B^2}{2}$.

The gauge field A_μ is mediated by the photon. According to the Lagrangian, the photon must be massless as including a mass term for A_μ would violate gauge invariance. Additionally, QED is an abelian gauge theory which prohibits any photon self-interaction terms. QED is represented by the

$U(1)$ gauge group symmetry.

1.2.1.2 THE WEAK INTERACTION

Dirac spinors can be decomposed into two irreducible representations called left-handed and right-handed Weyl spinors: $\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$. The projections of a Dirac spinor onto its left and right states are obtained by the operators $P_L = \frac{1-\gamma^5}{2}$ and $P_R = \frac{1+\gamma^5}{2}$ respectively. Here γ^5 is the matrix:

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} -\mathbb{I}_2 & 0 \\ 0 & \mathbb{I}_2 \end{pmatrix}.$$

The Weyl spinors behave the same under rotation transformations but experience opposite transformations under Lorentz boosts. The operation of flipping the signs of spatial coordinates is referred to as Parity (P). Invariance under parity transformations, also called chiral symmetry, is an important property in QFT. The Weyl spinors can be connected through a parity transformation as $P(\psi_{R/L}(t, \vec{x})) = \psi_{L/R}(t, -\vec{x})$.

Historically, many physicists believed parity would be conserved in all particle interactions. However the 1957 observation of β -decay in polarised Cobalt-60 demonstrated that parity was in fact not conserved in the weak interaction. This indicates that the weak interaction vertex and corresponding probability current must have a different form than that of QED.

While any bilinear covariant derivative can be used to construct SM Lagrangian terms, experimental data indicated the need for both vector currents and parity violating axial currents in equal parts to fully explain β -decay. Thus the weak interaction vertex could take the form of either $\gamma^\mu(1 - \gamma^5)$ or $\gamma^\mu(1 + \gamma^5)$. Experimentally it was shown that the weak charged current is a vector minus axial (V-A) interaction with a coupling vertex factor $-ig'/2\sqrt{2} \times \gamma^\mu(1 - \gamma^5)$ where g' is the weak coupling constant. The left projection operator P_L appears in the vertex factor indicating that only the

left-handed chiral component of fermions interact with the charged weak force.

This formulation of the weak force is represented by the $SU(2)_L$ local gauge symmetry group. $SU(2)$ is generated by the Pauli matrices $\sigma^a/2$ for $a = \{1, 2, 3\}$. This introduces three massless vector fields W_μ^a which can be combined to form the fields of the physically observable $W^\pm$ bosons as $W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp iW_\mu^2)$. The remaining field W_μ^3 is an independent neutral weak field. $SU(2)_L$ is a non-abelian group which allows $W^\pm$ self-interactions.

1.2.1.3 ELECTROWEAK UNIFICATION

The Glashow model of electroweak unification replaces the general $U(1)$ symmetry of QED with a new $U(1)_Y$ local gauge group which is symmetric under the transformation $\psi(x) \rightarrow \exp[ig' \frac{Y}{2} \zeta(x)]\psi(x)$. This gives rise to a new gauge field B_μ which couples to the new “weak hypercharge”, Y , rather than the electromagnetic charge e .

Electroweak unification describes the mixing of this new B_μ field with the weak neutral field W^3 by introducing the new covariant derivative $D_\mu = \partial_\mu - ig\sigma^a W_{\mu a} + ig'\frac{1}{2}Y B_\mu$. In order to preserve full $SU(2)_L \times U(1)_Y$ symmetry, the W_μ^a and B_μ fields have strength tensors $W^{a\mu\nu} = \partial^\mu W^{a\nu} - \partial^\nu W^{a\mu} + gf^{abc}W^{b\mu}W^{c\nu}$ and $B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu$ respectively. In the equation for the W_μ^a field strength tensor, f is the totally anti-symmetric structure constant. Finally, this leads to the unified electroweak Lagrangian:

$$\begin{aligned} \mathcal{L} = & -\frac{1}{4}B^{\mu\nu}B_{\mu\nu} - \frac{1}{4}W^{a\mu\nu}W_{a\mu\nu} \\ & + \bar{\psi}_L \gamma^\mu (\partial_\mu - ig \frac{\sigma^a}{2} W_\mu^a + ig' \frac{1}{2} Y B_\mu) \psi_L + \bar{\psi}_R \gamma^\mu (\partial_\mu + ig' Y B_\mu) \psi_R. \end{aligned}$$

The fields of the physically observable neutral Z^0 boson and photon can be written as mixed states of the W_μ^3 and B_μ fields. The corresponding mixed fields are $Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W$ and $A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W$ respectively. Here, θ_W is the Weinberg weak mixing angle which

is related to the coupling strengths of the electromagnetic and weak force as $\cos\theta_W = \frac{g}{\sqrt{g^2+g'^2}}$.

θ_W is a free parameter in the SM which can be constrained experimentally.

Fermionic mass terms like $m\bar{\psi}\psi$ seen in the original formulation of QED are no longer allowed as the electroweak Lagrangian treats left and right Weyl spinors differently. Additionally, the four electroweak gauge bosons must be massless to preserve gauge symmetry of the Lagrangian. It is clear in nature that some SM fermions and bosons do in fact have mass, and this seeming contradiction is reconciled by the Higgs field described in Section 1.3.

1.2.2 QCD

QCD is the QFT of the strong force. The strong force is only experienced by quarks which carry an additional quantum number called color. Thus, QCD is subject to the additional constraint that color charge is conserved. There are three color quantum numbers referred to as red, blue, and green, and so QCD is based on the $SU(3)_C$ symmetry group.

The QCD Lagrangian must be invariant under gauge transformations of the form $U = \exp(-i \sum_{j=1}^8 \theta_j \lambda_j / 2)$ where θ_j are arbitrary parameters and λ_j are the Gell-Mann matrices. This requires the introduction of the covariant derivative $D_\mu = \partial_\mu + i\alpha_S t_i G_\mu^i$ where G_μ^i are the 8 gluon fields (which mediate the strong force) and α_S is the gluon field coupling strength. The strength tensor of the gluon field is $\mathcal{G}_{\mu\nu}^i = \partial_\mu G_\nu^i - \partial_\nu G_\mu^i + \alpha_S f_{ijk} G_\mu^j G_\nu^k$ where f_{ijk} are the $SU(3)$ structure constants and $t_i = \lambda_i / 2$.

The full QCD Lagrangian is then

$$\mathcal{L} = -\frac{1}{4} \mathcal{G}_{\mu\nu}^i \mathcal{G}^{i\mu\nu} + \bar{\psi}_{fa} (i\gamma^\mu \partial_\mu \delta_{ab} - \alpha_S \gamma^\mu t_{ab}^i G_\mu^i - m_f \delta_{ab}) \psi_{fa}.$$

Here, f labels quark flavor ($f \in [1, 6]$), a labels color, $i \in [1, 8]$ indexes the gluons, and t_{ab}^i are the generators of $SU(3)$ which satisfy the commutation relation $[t_a, t_b] = i f_{abc} t_c$. QCD is a non-

abelian theory which allows for gluon self-interactions.

1.3 THE HIGGS BOSON

This mathematical description of the SM, as presented above, is not self-consistent. The physically observed masses of the gauge bosons break the gauge symmetry of the $SU(2)_L \times U(1)_Y$ electroweak interaction. The Brout-Englert-Higgs mechanism, described further in Section 1.3.1, was proposed in 1964 to spontaneously break electroweak symmetry (EWS) [15] [16] [17]. This mechanism also predicted the existence of a spin 0 scalar particle, generally referred to as the Higgs boson or simply the Higgs.

A Higgs-like particle with a mass of 125 GeV was discovered by the ATLAS and CMS experiments in 2012 [18] [19]. Thus far, all measurements of this particle have been consistent with the SM and the Brout-Englert-Higgs mechanism of EWS breaking (EWSB). This result solidifies physicists' understanding of how bosons and fermions acquire mass and places constraints on some free parameters of the SM. The study of the Higgs continues to be a high priority research area at the LHC as verifying all of its production and decay modes and measuring any self-couplings is critical to further understand its role in the SM.

1.3.1 ELECTROWEAK SYMMETRY BREAKING

In order to keep the mathematical description of the SM renormalizable EWS must be broken spontaneously, meaning the SM Lagrangian is invariant under the electroweak transformation but the individual spinor states are not. The mathematical mechanism of EWSB is introduced below and described in further detail in [20].

To spontaneously break EWS, a new complex scalar field Φ is introduced. To preserve the invariance of the EW Lagrangian, Φ must be an $SU(2)_L \times U(1)_Y$ multiplet; the simplest multiplet, the

isospin dublet:

$$\Phi = \begin{pmatrix} \Phi^+ \\ \Phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \Phi_1 + i\Phi_2 \\ \Phi_3 + i\Phi_4 \end{pmatrix}$$

can be used to represent the new field. The Lagrangian of the new field is $\mathcal{L} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi)$ where D_μ is the covariant derivative of EWS and $V(\Phi)$ is the potential of the scalar field. The potential of the Higgs field is $V(\Phi) = \mu^2 \Phi^2 + \lambda \Phi^4$. Requiring $\lambda > 0$ and $\mu^2 < 0$ allows a physical solution with spontaneous symmetry breaking as shown in Figure 1.4 as the true minimum of the potential is given by $\Phi_{min}^2 = \frac{-\mu^2}{\lambda}$.

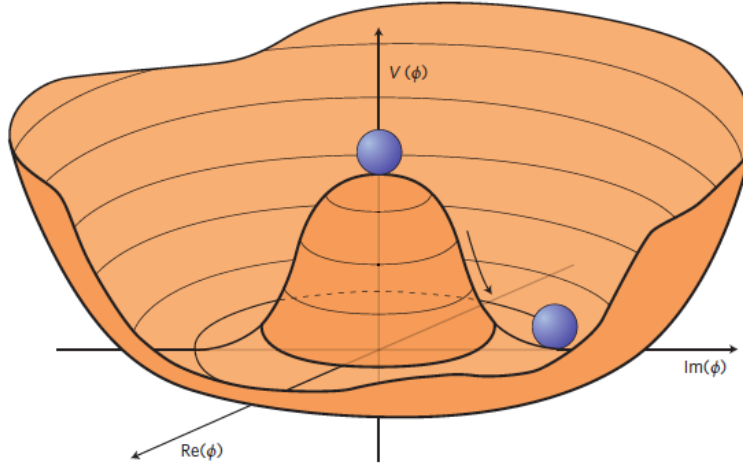


Figure 1.4: An illustration of the Higgs potential in the case that $\mu^2 < 0$ [20].

This minimization of Φ^2 in fact defines a circle of infinite degenerate solutions, and without loss of generality one can choose the intuitive representation $\Phi_1 = \Phi_2 = \Phi_4 = 0$ and $\Phi_3 = v = \sqrt{-\mu^2/\lambda}$. Here, v is the vacuum expectation value of the Higgs field. Thus, the field can be written as

$$\Phi_{vacuum} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}.$$

The resulting particle spectrum can be understood by examining the Lagrangian under small perturbations from the minimum of the form $\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v+h \end{pmatrix}$. Consider the kinetic term for the Higgs field $(D^\mu \Phi)^\dagger (D_\mu \Phi)$:

$$\begin{aligned} D_\mu \Phi &= \frac{1}{\sqrt{2}} [ig \frac{1}{2} \sigma^a W_{\mu a} + ig' \frac{1}{2} Y B_\mu] \begin{pmatrix} 0 \\ v+h \end{pmatrix} \\ &= \frac{i}{2\sqrt{2}} [g \left(\begin{pmatrix} 0 & W_1 \\ W_1 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -iW_2 \\ iW_2 & 0 \end{pmatrix} + \begin{pmatrix} W_3 & 0 \\ 0 & -W_3 \end{pmatrix} \right) + g' \begin{pmatrix} Y B_\mu & 0 \\ 0 & Y B_\mu \end{pmatrix}] \begin{pmatrix} 0 \\ v+h \end{pmatrix} \\ &= \frac{i(v+h)}{2\sqrt{2}} \begin{pmatrix} g(W_1 - iW_2) \\ -gW_3 + g'Y B_\mu \end{pmatrix} \end{aligned}$$

and so

$$(D^\mu \Phi)^\dagger (D_\mu \Phi) = \frac{(v+h)^2}{8} [g^2(W_1^2 + W_2^2) + (-gW_3 + g'Y B_\mu)^2].$$

A new orthonormal basis for the Lagrangian can be defined in terms of the fields of the physically observable $W_\mu^\pm$, Z_μ , and A_μ bosons defined in Sections 1.2.1.2 and 1.2.1.3. Finally, the Higgs Lagrangian becomes:

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \partial_\mu h \partial^\mu h - \lambda v^2 h^2 \\ &+ \frac{1}{2} \left(\frac{g'^2 v^2}{4} ((W_\mu^+)^2 + (W_\mu^-)^2) + \frac{v^2}{4} (g^2 + g'^2) Z_\mu^2 + 0 \cdot A_\mu^2 \right) \\ &+ \frac{g^2 v}{4} ((W_\mu^+)^2 + (W_\mu^-)^2) h + \frac{g^2}{8} ((W_\mu^+)^2 + (W_\mu^-)^2) h^2 \\ &\frac{v^2}{4} (g^2 + g'^2) Z_\mu^2 h = \frac{1}{8} (g^2 + g'^2) Z_\mu^2 h^2 \\ &- \lambda v h^3 - \frac{\lambda}{4} h^4. \end{aligned}$$

Here the first line of the Lagrangian gives the kinetic and mass terms of the Higgs field, the second line gives the mass terms of the $W^\pm$ and Z^0 bosons, the third line gives the coupling of the $W^\pm$ bosons to the Higgs field, the fourth line gives the coupling of the Z^0 boson to the Higgs field, and the final line gives the self-interaction terms of the Higgs. This formulation of the Lagrangian defines the masses of the gauge bosons as

$$m_{W^\pm} = \frac{1}{2}vg, \quad m_Z = \frac{1}{2}v\sqrt{g^2 + g'^2}.$$

Additionally, it defines the Higgs mass as $m_H = v\sqrt{2\lambda}$. It is clear that there are no mass or Higgs coupling terms for the A_μ photon fields and thus the photon remains massless.

1.3.2 YUKAWA COUPLING

The mass of SM fermions is described through their Yukawa couplings. The Yukawa interaction describes how fermionic Dirac spinor ψ interfaces with the scalar Higgs field Φ . This can be expressed mathematically as a cubic interaction of the form $g_Y \bar{\psi} \Phi \psi$ where g_Y is the Yukawa coupling constant. The Lagrangian for this interaction is then $\mathcal{L} = \frac{1}{2}D^\mu \Phi D_\mu \Phi - V(\Phi) + \bar{\psi}(iD_\mu \gamma^\mu - m)\psi - g_Y \bar{\psi} \Phi \psi$. Expanding the interaction term using the specific Higgs potential described above yields fermionic mass terms of the form

$$m_i \bar{\psi}_i \psi_i = \frac{g_{Y,i} v}{\sqrt{2}}.$$

The individual Yukawa couplings $g_{Y,i}$ are free parameters in the SM that can be constrained experimentally.

1.3.3 HIGGS PRODUCTION METHODS

There are four main Higgs production mechanisms that can be studied using proton-proton collisions at the LHC. All four have now been observed and the measured production cross-sections are consistent with SM predictions. Each mode is introduced briefly below and Figure 1.5 shows the SM production cross-sections for a $\sqrt{s} = 13\text{TeV}$ proton-proton collider as a function of the Higgs mass.

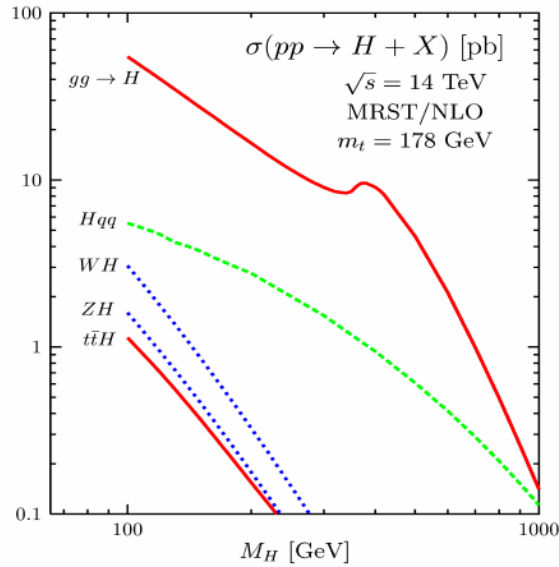


Figure 1.5: Dominant SM Higgs production cross-sections as a function of Higgs mass [21].

1.3.3.1 GLUON-GLUON FUSION

As shown in Figure 1.5, gluon-gluon fusion (ggf) is the dominant production mode for the 125 GeV Higgs. The ggF process must occur through a quark loop, as demonstrated in Figure 1.6, because the Higgs only couples to massive particles. As shown in Section 1.3.2, the Higgs-fermion coupling strength increases with fermion mass and thus the quark loop is most often a top or bottom loop.

This production mode has been observed for multiple Higgs decay modes and contributed substantially to the initial Higgs discovery [18].

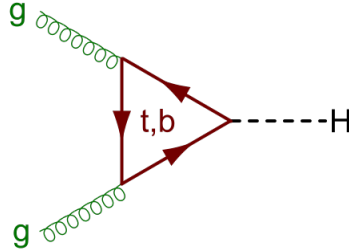


Figure 1.6: Feynman diagram of gluon-gluon fusion Higgs boson production with a top or bottom quark loop

1.3.3.2 VECTOR BOSON FUSION

The second most common Higgs production mode at the LHC is Vector Boson Fusion (VBF). This process is shown to first order in Figure 1.7. VBF is quark initiated; the quarks within the colliding protons exchange virtual $W^\pm$ or Z^0 and this exchange emits a Higgs. The interacting quarks do not need to be the same type which increases the cross-section of this process. As with ggF, VBF production has been observed for multiple Higgs decay modes and was central to the Higgs discovery [18].

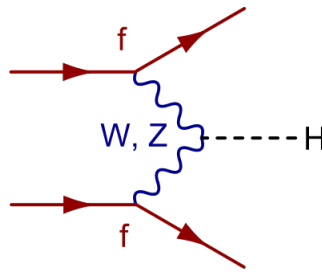


Figure 1.7: Feynman diagram of vector boson fusion Higgs boson production

1.3.3.3 ASSOCIATED PRODUCTION WITH A VECTOR BOSON

The next most common Higgs production mode, and the focus of the analysis described in Chapters 6 and 7, is associated production of the Higgs with a $W^\pm$ or Z^0 boson. This process, shown in Figure 1.8, is commonly referred to as VH where V represents vector boson. The leading production process is quark initiated; a quark and anti-quark merge to form a virtual $W^\pm$ or Z^0 which, if sufficiently energetic, then radiates a Higgs. As shown in Figure 1.5, the WH cross-section is slightly larger than the ZH cross-section due to the lower mass of the W.

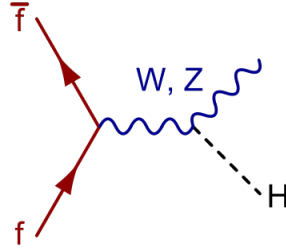


Figure 1.8: Feynman diagram of associated Higgs boson production with a vector boson

1.3.3.4 TOP FUSION

The final relevant production mode is top fusion, ttH. As shown in Figure 1.9, ttH is gluon initiated and each initial gluon decays into a top anti-top pair (two of which then fuse to form the Higgs). Due to the high mass of the top quark, this production mechanism requires a large amount of energy and consequently has the lowest cross-section. The ttH mode was observed independently in 2018 [22].

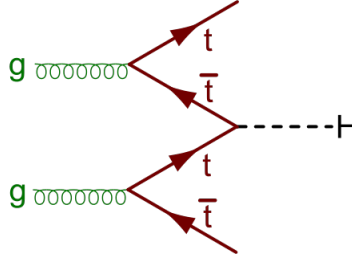


Figure 1.9: Feynman diagram of top fusion Higgs boson production

1.3.4 HIGGS BOSON DECAYS

As shown in Sections 1.3.1 and 1.3.2, the Higgs couples to all massive SM particles, except neutrinos, and thus there are many potential decay modes for the Higgs. Although incorporating SM conservation laws of charge, color, and flavor reduces that number somewhat, there are still numerous allowed decays. The main Higgs decays accessible at a $\sqrt{s} = 13\text{TeV}$ collider are summarized in Figure 1.10. The likelihood of each decay, referred to as the branching ratio (BR), depends on the particle's coupling strength to the Higgs, the differences in mass, and the mass of the Higgs itself. Each of these decay modes is introduced briefly below and additional information, including second-order-decay-product-influenced measurement strategies, can be found in [12].

1.3.4.1 MASSIVE FERMIONS

One way for the Higgs to decay is into a fermion-antifermion pair. As derived in Section 1.3.2, the fermionic Higgs coupling strengths are proportional to the fermion masses and thus decays to heavy fermions have a higher BR. A decay to a top-antitop pair is disallowed for a 125GeV Higgs due to mass constraints, and so the most common Higgs decay is to a bottom-antibottom pair. This decay has a BR of 0.584. Despite being the most likely Higgs decay mode, it was not observed independently until Run 2 due to difficulties triggering on, reconstructing, and identifying b-quarks

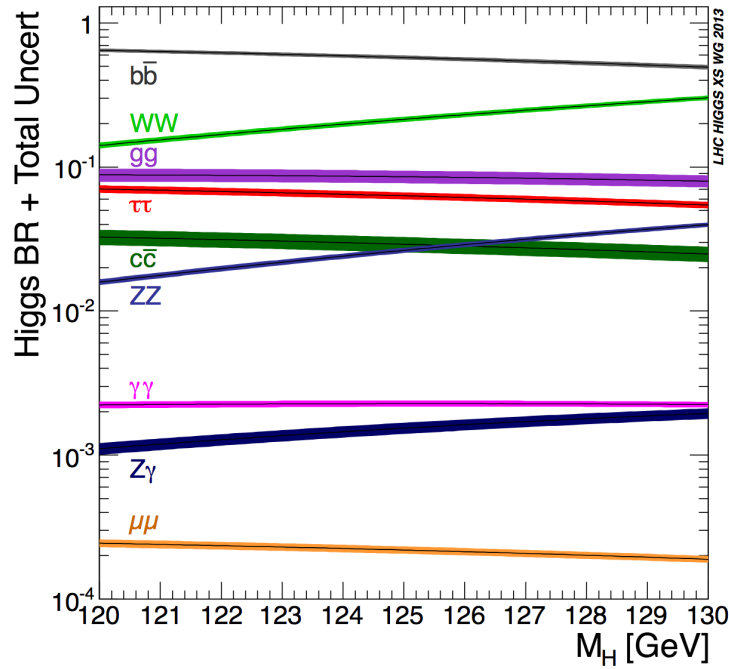


Figure 1.10: Branching ratios for the main decays of a SM Higgs boson as a function of Higgs mass. The theoretical uncertainties are shown as bands [12].

[23]. The next most common quark-antiquark decay is to a charm-anticharm pair which has a BR of 0.029. This decay mode has not been independently observed [24]. Higgs decays to lighter quarks have a BR too low for observation at the LHC.

The most likely leptonic decay of the Higgs, and the subject of the analysis described in Chapters 6 and 7, is to a $\tau^+\tau^-$ pair due to the relatively high tau mass. The BR for this decay is 0.0627 and it has been observed for three of the four production modes described in Section 1.3.3 [25]. Higgs decay to a $\mu^+\mu^-$ pair has a BR of only 0.000218 which positions it as the least likely decay mode for feasible observation at the LHC. Measuring this decay channel is an important goal for the HL-LHC [26]. The BR for Higgs to an e^+e^- is generally believed to be too low for potential observation at the LHC.

1.3.4.2 MASSIVE BOSONS

Additionally, the Higgs may split into a pair of massive gauge bosons. The BR for Higgs to W^+W^- is 0.214, making it the second most common Higgs decay. It was initially observed for the ggF and VBF production modes in Run 1 [27] and for ttH [22] and VH [28] production in Run 2. The BR for Higgs to a ZZ pair is substantially lower at only 0.0262. However, the $H \rightarrow ZZ \rightarrow llll$ channel, often referred to as the ‘golden channel’, has a very clean signature due to the presence of four leptons and thus was one of the initial Higgs discovery channels [18].

1.3.4.3 MASSLESS PARTICLES

The Higgs may also decay into massless gauge bosons (photons and gluons), but these processes require an intermediate loop of virtual heavy quarks or massive bosons. The $H \rightarrow gg$ is the most likely of these decays with a BR of 0.086. Gluons carry color charge and thus this decay must be mediated by a quark loop, most often a top loop. This decay is in fact the inverse of the ggF production mode described in Section 1.3.3.1. This decay is difficult to measure at colliders.

The Higgs decay to two photons ($\gamma\gamma$) may proceed through fermion or boson loops and has a much smaller BR, only 0.0027. However, like the $H \rightarrow ZZ \rightarrow llll$, it is a feasible signature to measure at colliders due to excellent photon momentum reconstruction and identification. This decay was also one of the initial Higgs discovery channels [18]. A decay to $Z\gamma$ is also allowed in the SM, with a BR of 0.00154 and requires a virtual massive fermion or boson loop similar to the $\gamma\gamma$ decay. This decay mode has not yet been observed [29].

The SM does not provide a mechanism for generating neutrino masses, but there are multiple BSM extensions that predict Higgs decays to neutrinos. Many of these theories create signatures that are potentially observable at LHC detectors; see for example [30] and [31].

It's important to surround yourself with people you can learn from.

Reba McEntire

2

The LHC and the ATLAS Detector

THE FORCES AND INTERACTIONS DESCRIBED IN THE PREVIOUS CHAPTER CAN BE TESTED EXPERIMENTALLY BY COMPARING THEORETICAL PREDICTIONS TO OBSERVATIONS IN DATA. A range of particle physics experiments have been built over the past century to characterize the fundamental constituents of matter, test the Standard Model (SM), and probe Beyond the Standard Model (BSM) theories [32]. Currently, the majority of these experiments utilize methods at the Intensity Frontier [33], focusing on high-intensity beams to study rare-processes, or the Energy Frontier [34], focusing on high-energy interactions. The primary experimental technology probing the Energy Frontier is particle colliders [35]. All work in this thesis relies on data and software from the ATLAS Experiment at the Large Hadron Collider (LHC), a proton-proton collider located

at CERN.

2.1 PARTICLE COLLIDERS

A particle collider uses electromagnetic fields to accelerate and eventually collide highly collimated beams of charged particles. By colliding leptons, nucleons, hadrons, or other particles at ultra-relativistic speeds, colliders are able to produce extremely heavy or rare particles which allows physicists to test new theories and better characterize the interactions of known particles.

There are several relevant metrics for describing the performance of a collider. Perhaps the most important is the center of mass energy of the collisions, E_{CM} . According to the De Broglie relation, $\lambda = h/p$, higher particle momenta decreases the distance scales that can be probed by collisions, giving researchers access to the small-scale structure of matter. Furthermore, to produce heavy particles, the energy of the collision must be greater than rest mass of the particle of interest. The production cross-section, σ , of fundamental particles increases with increasing E_{CM} , as shown in Figure 2.1. Thus, maximizing E_{CM} is generally a goal of collider development.

Another important characterization is luminosity, $\mathcal{L}$, which relates the production cross-section of a process to the observed production rate, R :

$$R = \frac{dN}{dt} = \sigma \times \mathcal{L}$$

All modern colliders use bunched beams (described in additional detail in 2.1.2). The luminosity of collider with two beams of bunch size n_1 and n_2 is expressed as

$$\mathcal{L} = f_{coll} \frac{n_1 n_2}{4\pi\sigma_x^* \sigma_y^*}.$$

Here, σ_1^* and σ_2^* describe the transverse beam size in the bend and vertical directions and f_{coll} is

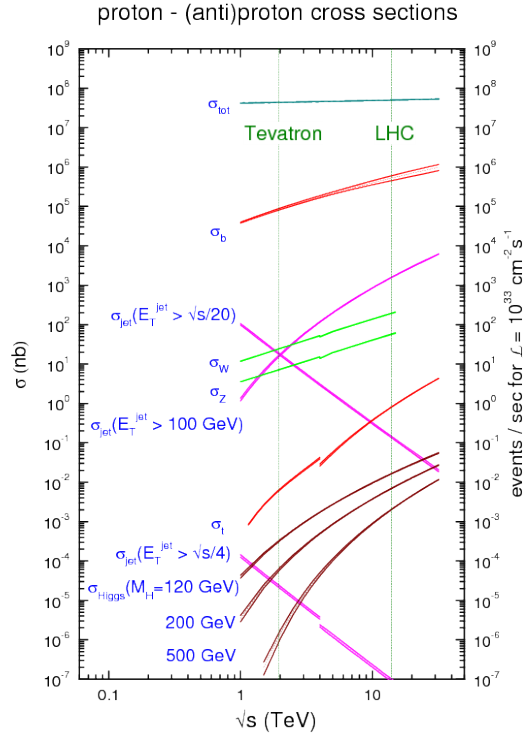


Figure 2.1: Standard Model production cross-section predictions at hadron-hadron colliders[36]

the bunch crossing frequency of the collider. This luminosity equation relies on the assumptions that the bunches are identical in transverse profile, that the profiles are Gaussian and independent of position along the bunch, and the particle distributions are not altered during bunch crossing. Luminosity can be affected by non-zero beam crossing angles, Θ_C , and bunch length σ_z . The luminosity will be reduced by a factor $1/(1 + \phi^2)^{1/2}$ where $\phi = \Theta_C \sigma_z / (2\sigma_x^*)$ [12]. As with E_{CM} , maximizing luminosity is generally a goal of collider design.

2.1.1 THE LARGE HADRON COLLIDER

The LHC is currently the world's largest and most powerful particle accelerator. It is located at the European Organization for Nuclear Physics (CERN). Initial designs began in the 1980's, while CERN's previous collider, the Large Electron-Positron collider (LEP) was still under construction and taking data. The design was officially approved in 1994 and the first beams were circulated in 2008 [37]. CERN currently has 23 Member States, 7 associated Member States, and 6 Observer States, including the United States [38]. These states, and many others with scientific contacts or cooperation agreements, finance and operate the LHC and receive access to the collected data. Thus, CERN continues to support its missions of strong international collaboration, scientific discovery, and technological innovation.

The LHC was constructed in the existing 27 km circular LEP tunnel under the French/Swiss border and consists of two parallel beampipes surrounded by magnets. It is used mainly to accelerate and collide protons, though other particles such as lead ions can be used for specific studies. The two beams run in opposite directions and intersect at four points along the ring where the detectors (ATLAS, CMS, ALICE, and LHCb) are located. The LHC magnet system consists of over 10,000 superconducting magnets including dipole magnets to guide the beams in a circular path, quadrupole magnets to focus the beams, and higher pole order magnets for small field corrections [39].

The initial plan for Run 1 of the LHC was to deliver data with $E_{CM} = \sqrt{s} = 14$ TeV beginning in 2008. Unfortunately a week after initial beams were circulated in September 2008, a magnet quench occurred damaging 53 superconducting magnets and releasing approximately 6 tons of liquid helium. The projected timeline was adjusted, and physics data-taking began in 2009 at a reduced energy $\sqrt{s} = 7$ TeV [40]. The energy was increased to $\sqrt{s} = 8$ TeV in 2012 and data taking continued until early 2013.

The LHC was shutdown from 2013-2015 to allow repairs and upgrades to the accelerator and detectors. Run 2 of the LHC began in 2015 at $\sqrt{s} = 13$ TeV and continued until December 2019. During Run 2, the LHC delivered an integrated luminosity of 156 fb^{-1} . Of this, 147 fb^{-1} was recorded by ATLAS (the detector used in this thesis) and 139 fb^{-1} passed quality requirements for use in physics analyses. A breakdown of the Run 2 luminosity, including luminosity delivered per year, is shown in Figure 2.2.

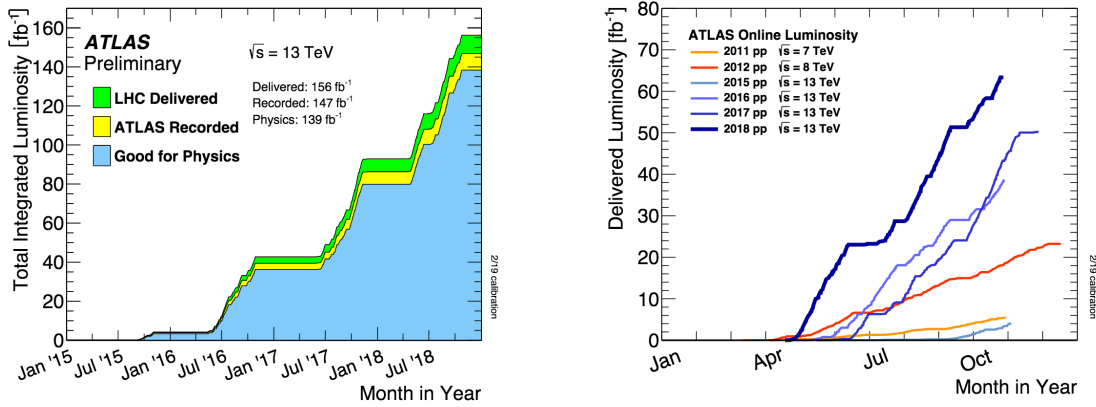


Figure 2.2: Left: total integrated luminosity delivered by the LHC (green), recorded by ATLAS (yellow) and usable for physics (blue). Right: Cumulative luminosity versus day delivered to ATLAS during stable beams and for high energy p-p collisions [41].

The LHC is currently in a second shutdown phase and Run 3 at $\sqrt{s} = 14$ TeV is scheduled to begin in 2021. The tentative full-term schedule for the LHC is shown in Figure 2.3, and the predicted luminosity growth in Figure 2.4. Run 3 is expected to continue until 2024, at which time a third shut-down will begin to allow commissioning of the High-Luminosity LHC (HL-LHC) which will proceed in three runs from 2026-2038 [42].



Figure 2.3: LHC planned operating schedule until 2038. EYETS stands for Extended Year End Technical Stop, a lengthened version of the annual brief shutdown during LHC running to allow for repairs. LS stands for Long Shutdown.

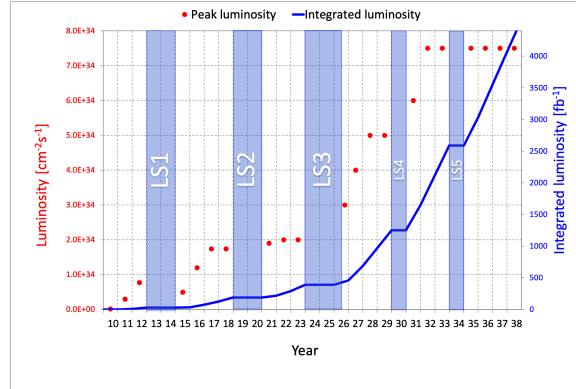


Figure 2.4: HL-LHC luminosity projection after completing Run 2

2.1.2 THE LHC ACCELERATOR COMPLEX

The LHC is supplied with high-energy protons through a multi-stage injector chain illustrated in Figure 2.5. For proton-proton runs (as are used for this dissertation) hydrogen atoms from a concentrated source are passed through Linac2, a linear accelerator containing a strong electric field which removes the electrons from the atoms. This results in a pure proton beam with an energy

of 50 MeV. This beam is then sent into the Proton Synchrotron Booster (PSB) which consists of 4 superimposed synchrotronic rings which further accelerate the beam to 1.4 GeV. The beam then enters the Proton Synchrotron (PS), a circular accelerator that increases the beam energy to 25 GeV. The PS feeds into the Super Proton Synchrotron (SPS), a 7km circular accelerator, where it reaches an energy of 450 GeV. Finally, the beam enters the main LHC ring where it is split into two beams that travel in opposite directions. Within the main ring the protons are grouped into bunches of approximately 115 billion protons each. This is a consequence of the oscillating frequency of the radio frequency cavities used to accelerate the protons and also allows collisions to occur at discrete intervals. Within the main ring, protons are further accelerated to their maximum energy of 6.5 TeV (in Run 2) and then circulate within the ring for several hours as collisions occur at the interaction points [37].

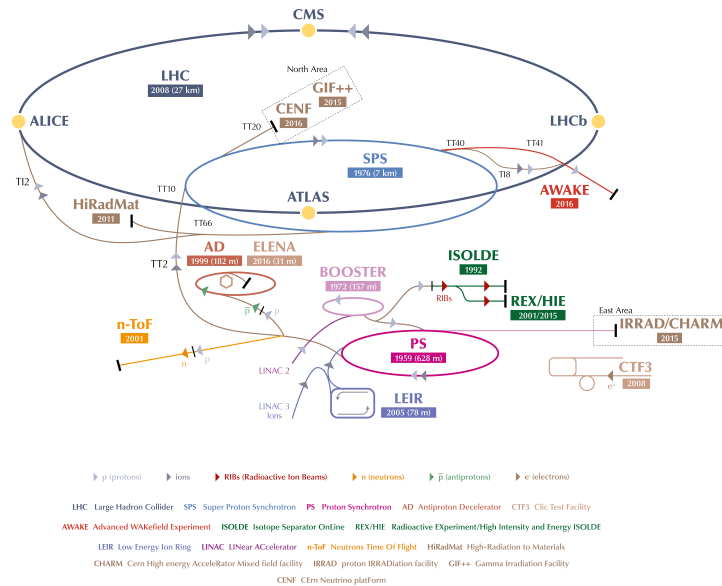


Figure 2.5: The CERN Accelerator Complex

The protons in the LHC circulate in a defined and discrete bunch structure. During normal operation, bunches of roughly 10^{11} protons are separated by 25 nanosecond gaps, and a full LHC ring holds 2808 bunches. The bunches begin on the order of a few centimeters in size and are compressed to around $20\text{ }\mu\text{m}$ near the interaction points to increase the potential for collisions. This bunch structure allows for an interaction frequency of 40 MHz; however, to provide ramp up time for beam injection or magnet quenching, the actual interaction frequency is closer to 30 MHz [43].

During bunch crossings, the interaction of interest is the hard-scattering collision where the quarks within the colliding protons transfer large amounts of momentum resulting in a system of large mass [44]. Although it is rare for more than one hard-scattering event to occur in a single bunch crossing, multiple soft interactions are expected. The particles resulting from these soft interactions overlap with particles from the hard-scattering event, a phenomena referred to as pileup. LHC analyses distinguish between two types of pileup: instantaneous, which involves particles from the same bunch crossing, and out-of-time, which involves particles from previous or following bunch crossings. Both types of pileup contribute to crowding in the detectors and complicate particle reconstruction [45]. The average pileup for Run 2 is shown in Figure 2.6.

2.2 PARTICLE DETECTION

After the initial hard-scattering event a variety of particles are produced either as direct products of the collision or as decays of the initial particles. Detectors at the interaction points aim to measure these particles. Only particles which have a long enough lifetime and interact with SM particles are detectable. However after reconstructing these particles (as described in Chapter 3), imposing additional constraints such as momentum and energy conservation, and accounting for inefficiencies in the detector, physicists can find evidence for other particles such as neutrinos or BSM particles that likely occurred before or alongside the detected particles.

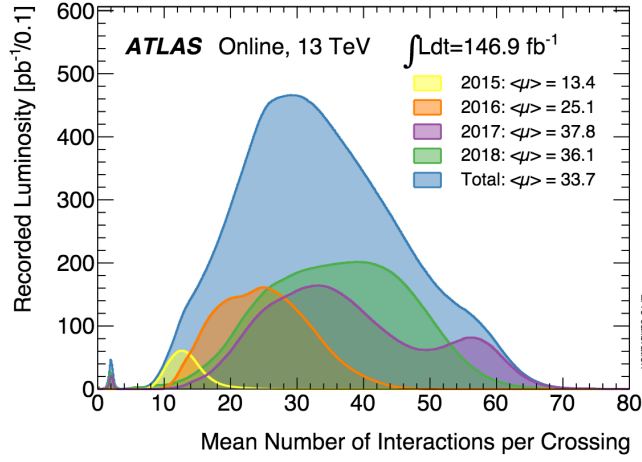


Figure 2.6: The luminosity-weighted distribution of the mean number of interactions per crossing for the Run 2 pp collision data at 13 TeV center-of-mass energy [41]

Modern detectors at collider facilities rely on two primary technologies. Semiconductor detectors use materials such as silicon to create diodes and charged particles passing through the material create currents that can be tracked and measured. Calorimeters are typically designed to absorb particles and use scintillating materials to measure the resulting energy showers. Additional common methods of particle detection include Transition Radiation counters or Cherenkov Light measurements [46].

2.2.1 THE ATLAS DETECTOR

ATLAS (A Torroidal LHC ApparatuS) is one of the two general physics detectors at the LHC (along with CMS) and was designed to effectively detect various SM particles for refined SM measurements and new physics searches. It is one of the largest particle detector ever built at 46m long, 25m in diameter, and it weighs over 7,000 tons [47][48].

A schematic of the detector is shown in Figure 2.7. ATLAS consists of 4 subsystems (described below) which wrap concentrically around the beamline. The detector can be divided into two geo-

metric components: the barrel, which wraps around the beamline, and the endcaps which are perpendicular to the beamline at either end of the detector.

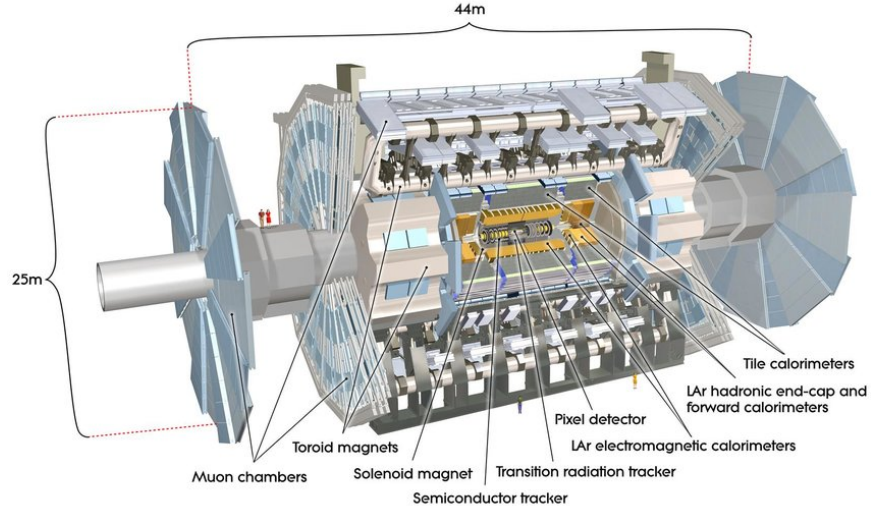


Figure 2.7: Diagram of the various subsystems of the ATLAS detector [49]

A right-handed coordinate system is used to describe position within ATLAS. The origin of the coordinate system is set to be the event interaction point; from this origin the beam defines the $\hat{z}$ direction and the $\hat{x}$ - $\hat{y}$ plane is perpendicular to the beamline. The positive $\hat{x}$ direction is towards the center of the LHC ring and the positive $\hat{y}$ direction is straight up. Vector components in the $\hat{x}$ - $\hat{y}$ plane are called transverse. Transverse components are particularly important in physics analyses because they are invariant to boosts in the $\hat{z}$ direction that come from particles' initial velocities.

The cylindrical symmetry of the detector allows for the definition of additional angular coordinates. The azimuthal angle ϕ is measured around the beam axis and is zero towards the center of the LHC ring. The polar angle θ is the angle from the beam axis and is zero along the beamline. A common reparameterization of the polar angle is the pseudo-rapidity $\eta = -\ln(\tan(\theta/2))$. Particles along the beamline thus have a pseudo-rapidity $\eta = \infty$ and particles perpendicular to it have

pseudo-rapidity $\eta = 0$. For massive particles, the standard rapidity $y = 1/2 \ln[(E + p_Z)/(E - p_Z)]$ can also be used. Angular distances in the η - ϕ plane are described by $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ [49].

Finally, the centrality of objects within a detector can be described in terms of impact parameters. The impact parameter d_0 is the signed distance from a charged track to the z axis while z_0 is the z -coordinate of the track at the point of closest approach to the global z axis (Figure 2.8). These parameters are calculated during track fitting as described in Chapter 3.

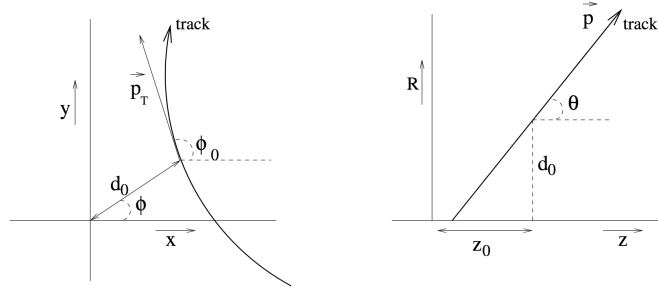


Figure 2.8: Illustration of impact parameters in the transverse plane (left) and R-Z plane (right).

2.2.1.1 INNER DETECTOR

The Inner Detector is closest to the interaction point and consequently is highly sensitive and compact. It is designed to allow for robust track-finding algorithms, precise momentum resolution, and accurate vertex reconstruction. The Inner Detector, shown in Figure 2.9, consists of 3 specialized, high granularity detectors (all surrounded by a 2T magnetic field) that measure the momentum, direction, and charge of charged particles. It provides full coverage in ϕ for $|\eta| < 2.5$ [50].

The innermost portion of the Inner Detector is the Pixel Detector. It is the highest granularity subsystem of the Inner Detector and includes 4 concentric cylindrical layers around the beamline and 3 parallel layers in each endcap. Each layer consists of 10 cm^2 modules of 46,080 silicon pixels, each with its own readout channel. This corresponds to a position resolution of $14 \times 115 \mu\text{m}^2$ and

allows for high definition track reconstruction with up to 4 hits per track (described further in Chapter 3) near the interaction point, yielding precise impact parameter reconstruction. The innermost cylindrical layer, called the Insertable B-Layer (IBL), was added for Run 2 to mitigate radiation damage suffered in Run 1 [51].

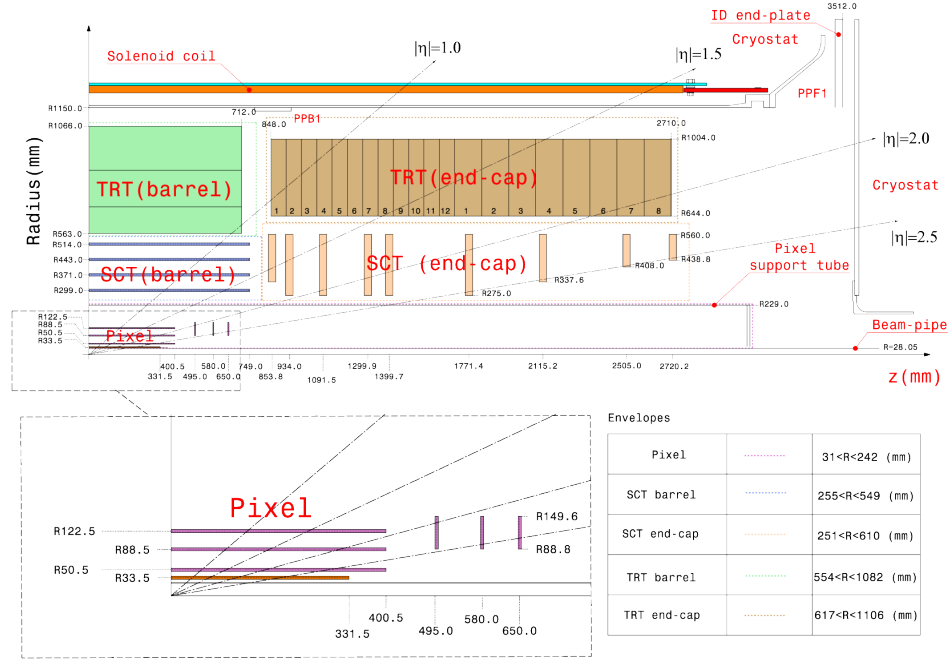


Figure 2.9: The r - z cross-section of one quadrant of the ATLAS Inner Detector. The top figure is the whole inner detector and the lower figure is a magnified view of only the pixel detector [50].

The semiconductor tracker (SCT) is silicon microstrip detector with 4088 double sided modules distributed between 4 barrel layers and 9 endcap layers. There is a readout component every $80 \mu\text{m}$ providing a resolution of $17 \mu\text{m}$. This allows up to 8 precision hits per track which contributes to improved momentum measurement and vertex reconstruction. In both the Pixel Detector and SCT, hits are defined as electron-hole drifts in the applied electric field resulting from electrons freed from orbit by incident particles.

Finally, the Transition Radiation Detector (TRT) is made of 4mm diameter gas-filled drift-tube straws interleaved with radiators (fibers in the barrel and foils in the endcaps). There are 50,000 straws in the barrel and 250,000 straws in the endcaps, each end of each straw is read out separately, allowing a position resolution of 0.17 mm. In addition to providing continuous tracking, the TRT improves the identification of charged particles like electrons and pions. The fibers and foils have different dielectric constants than the straws, and when charged particles pass through the boundaries, transition radiation photons are produced. The amount of transition radiation produced corresponds inversely to the incident particle's mass. Each straw readout provides two independent thresholds to distinguish tracking hits (lower threshold) from transition radiation hits (higher threshold) [52].

2.2.1.2 CALORIMETERS

Calorimeters measure the energy of particles by absorbing them. ATLAS has two sampling calorimeter subsystems, shown in Figure 2.10, which absorb particles and electronically sample the resulting energy shower distributions. Together these calorimeters cover the range $|\eta| < 4.9$; the $|\eta|$ region matched to the inner detector has a finer granularity for precision measurements of charged particles, while the extended $|\eta|$ range has a coarser granularity suitable for hadronic jet reconstruction [49].

The electro-magnetic (EM) calorimeter is designed to absorb and measure charged particles, mainly electrons and photons. It consists of a barrel component covering $|\eta| < 1.5$ and two endcaps covering $1.4 < |\eta| < 3.2$. The barrel and endcaps contain 3 layers of accordion shaped lead plates which are separated by layers of liquid argon (LAr) and readout electrodes (Figure 2.11). The accordion structure allows full azimuthal coverage with no cracks [53].

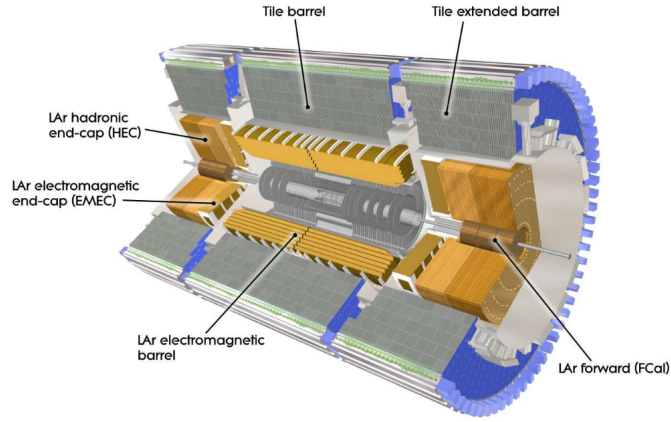


Figure 2.10: Cut-away view of the ATLAS calorimeter system [49]

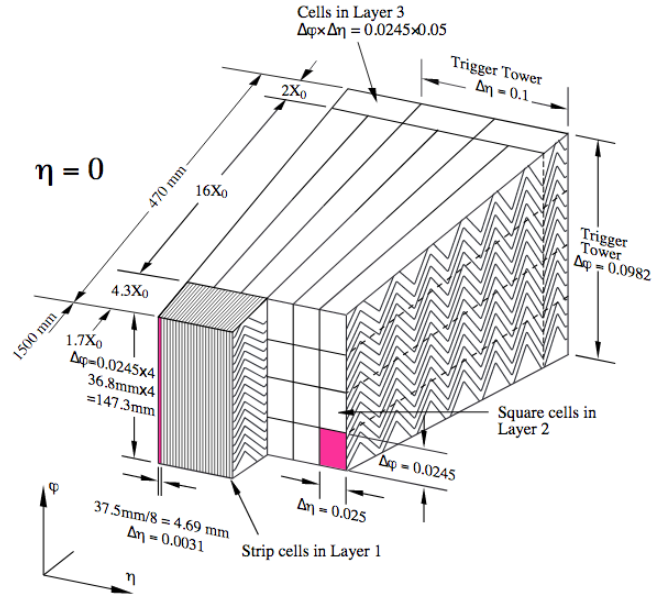


Figure 2.11: The accordion geometry used in the EM calorimeter [53].

As charged particles move through the EM calorimeter, their interactions with the lead absorbers create showers of charged particles through bremsstrahlung radiation and electron-positron pair

production. These particle showers then ionize the liquid argon and an applied high-voltage in the LAr filled gap guides the resulting ions and electrons to the readout electrodes. This process can repeat in multiple layers of the calorimeter with each successive shower having decreased energy [53]. The types and amounts of materials used in the EM calorimeter determine its radiation length, X_0 , which is defined as the average distance an electron can travel before its energy is reduced by a factor $1/e$. The radiation length is given in terms of atomic weight A and atomic number Z as $X_0 = \frac{716.4A}{Z(Z+1)\ln(287/\sqrt{Z})}$ [12]. Lead is a high- Z material and hence allows for a small radiation length.

The first layer of the EM calorimeter extends to a depth of $4.3X_0$ and consists of strip towers segmented in η . The strips are finely segmented (ie 8 strips in front of a central cell) in the central η region ($|\eta| < 1.4$) and become coarser as η increases. The full calorimeter granularity is described in Table 2.1. The second layer is made of square towers of size $\eta \times \phi = 0.025 \times 0.025$ for $|\eta| < 2.5$ and size $\eta \times \phi = 0.1 \times 0.1$ for $|\eta| > 2.5$. The second layer extends to a depth of $16X_0$. The third layer covers only $|\eta| < 2.5$ with towers of size $\eta \times \phi = 0.05 \times 0.025$ and extends to a depth of $2X_0$. Additionally, for $|\eta| < 1.8$, there is a presampler consisting of an active LAr layer to correct for energy lost upstream in the calorimeter [49]. Collectively, these layers provide sufficient material to absorb nearly all electrons and photons. A summary of the absorption material in all calorimeters is shown in Figure 2.12.

The hadronic calorimeters are positioned outside the EM calorimeters and consist of a central barrel in the region $|\eta| < 1.0$, two extended barrels in the regions $0.8 < |\eta| < 1.7$, and two endcaps in the region $1.5 < |\eta| < 3.2$ [49]. The granularity of these components are described in Table 2.1. The hadronic calorimeters are used to absorb and record showers of hadronic particles which interact via the strong force.

The central and extended barrels are collectively referred to as the Tile Calorimeter (TileCal), and consist of scintillating tiles separated by steel plates. Hadronic showers begin in the EM calorime-

		Barrel	End-cap	
EM calorimeter				
Number of layers and $ \eta $ coverage				
Presampler	1	$ \eta < 1.52$	1	$1.5 < \eta < 1.8$
Calorimeter	3	$ \eta < 1.35$	2	$1.375 < \eta < 1.5$
	2	$1.35 < \eta < 1.475$	3	$1.5 < \eta < 2.5$
			2	$2.5 < \eta < 3.2$
Granularity $\Delta\eta \times \Delta\phi$ versus $ \eta $				
Presampler	0.025×0.1	$ \eta < 1.52$	0.025×0.1	$1.5 < \eta < 1.8$
Calorimeter 1st layer	$0.025/8 \times 0.1$	$ \eta < 1.40$	0.050×0.1	$1.375 < \eta < 1.425$
	0.025×0.025	$1.40 < \eta < 1.475$	0.025×0.1	$1.425 < \eta < 1.5$
			$0.025/8 \times 0.1$	$1.5 < \eta < 1.8$
			$0.025/6 \times 0.1$	$1.8 < \eta < 2.0$
			$0.025/4 \times 0.1$	$2.0 < \eta < 2.4$
			0.025×0.1	$2.4 < \eta < 2.5$
			0.1×0.1	$2.5 < \eta < 3.2$
Calorimeter 2nd layer	0.025×0.025	$ \eta < 1.40$	0.050×0.025	$1.375 < \eta < 1.425$
	0.075×0.025	$1.40 < \eta < 1.475$	0.025×0.025	$1.425 < \eta < 2.5$
			0.1×0.1	$2.5 < \eta < 3.2$
Calorimeter 3rd layer	0.050×0.025	$ \eta < 1.35$	0.050×0.025	$1.5 < \eta < 2.5$
Number of readout channels				
Presampler	7808		1536 (both sides)	
Calorimeter	101760		62208 (both sides)	
LAr hadronic end-cap				
$ \eta $ coverage			$1.5 < \eta < 3.2$	
Number of layers			4	
Granularity $\Delta\eta \times \Delta\phi$			0.1×0.1	$1.5 < \eta < 2.5$
			0.2×0.2	$2.5 < \eta < 3.2$
Readout channels			5632 (both sides)	
LAr forward calorimeter				
$ \eta $ coverage			$3.1 < \eta < 4.9$	
Number of layers			3	
Granularity $\Delta x \times \Delta y$ (cm)			FCal1: 3.0×2.6	$3.15 < \eta < 4.30$
			FCal1: $\sim$ four times finer	$3.10 < \eta < 3.15,$ $4.30 < \eta < 4.83$
			FCal2: 3.3×4.2	$3.24 < \eta < 4.50$
			FCal2: $\sim$ four times finer	$3.20 < \eta < 3.24,$ $4.50 < \eta < 4.81$
			FCal3: 5.4×4.7	$3.32 < \eta < 4.60$
			FCal3: $\sim$ four times finer	$3.29 < \eta < 3.32,$ $4.60 < \eta < 4.75$
Readout channels			3524 (both sides)	
Scintillator tile calorimeter				
	Barrel		Extended barrel	
$ \eta $ coverage	$ \eta < 1.0$		$0.8 < \eta < 1.7$	
Number of layers	3		3	
Granularity $\Delta\eta \times \Delta\phi$	0.1×0.1		0.1×0.1	
	0.2×0.1		0.2×0.1	
Readout channels	5760		4092 (both sides)	

Table 2.1: Coverage, granularity, and number of readout channels of the ATLAS calorimeter system [49].

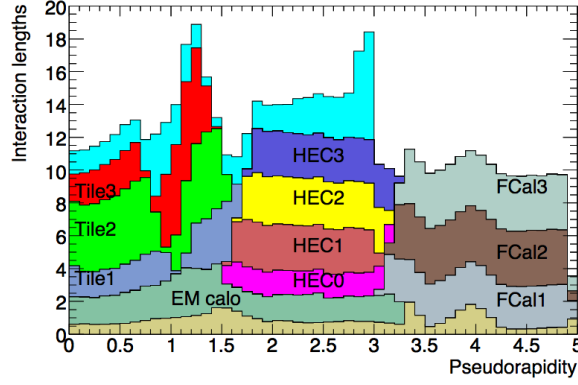


Figure 2.12: Cumulative amount of material, in units of interaction length, for different components of the ATLAS calorimeter. Tile_i refers to layers of the TileCal, HEC_i refers to layers of the Hadronic End Caps, and FCal_i refers to layers of the Forward Calorimeter [49].

ter where strong interactions with the argon and lead nuclei produce additional hadrons, and this process continues into the TileCal; some particles in hadronic showers may also interact electromagnetically.

Each cylinder in the TileCal contains 64 modules, and the tiles in each module are placed in the plane perpendicular to the beamline and staggered in depth (Figure 2.13). Similar to the EM calorimeter, charged particle interactions in the scintillating tiles create an electronic signal. This signal is read out by wave length shifting fibers on either end of the module and sent to photomultiplier tubes (PMTs). Each tile is readout by two PMTs [54].

The depth of hadronic calorimeters is characterized by the nuclear interaction length, λ , which describes the mean distance travelled by a hadronic particle before undergoing an inelastic nuclear interaction. The 4.7 to 1 ratio of steel to scintillator in the TileCal creates a nuclear interaction length of $\lambda=20.7$ cm. The EM calorimeter provides nearly 2λ of material and the hadronic calorimeter provides over 8λ altogether.

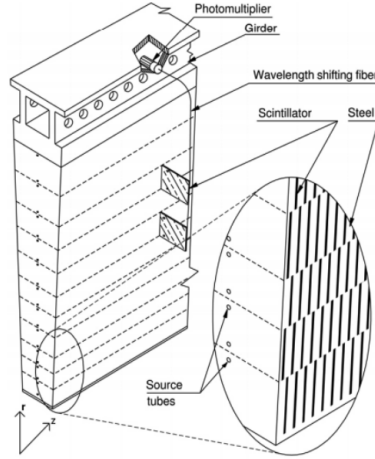


Figure 2.13: Schematic diagram of a ϕ wedge of the Tile Calorimeter [54].

The hadronic endcaps function similarly to the EM calorimeter. They consist of flat copper plates separated by LAr and are segmented into four longitudinal layers [53]. The granularity is described in Table 2.1.

Additionally, ATLAS contains a forward hadronic calorimeter (FCAL) covering $3.1 < |\eta| < 4.9$. The FCAL consists of 3 longitudinal layers in which cylindrical LAr gaps are arranged in matrices of copper (first layer) or tungsten (second and third layers) with electrode readouts. The FCAL is used for reconstruction of forward jets and operation in high pile-up environments [53].

2.2.1.3 MUON SPECTROMETER

Muons are minimally ionizing particles which allows them to often pass through the calorimeters with limited interaction. The Muon Spectrometer is the outermost subsystem of ATLAS and was designed to capture muons that have punched through the calorimeters and to provide standalone muon reconstruction, momentum measurement, and triggers. The muon spectrometer is shown in Figure 2.14 and consists of three concentric cylinders in the barrel region ($|\eta| < 1$) and four end cap

disks on either side ($1 < |\eta| < 2.7$)[55].

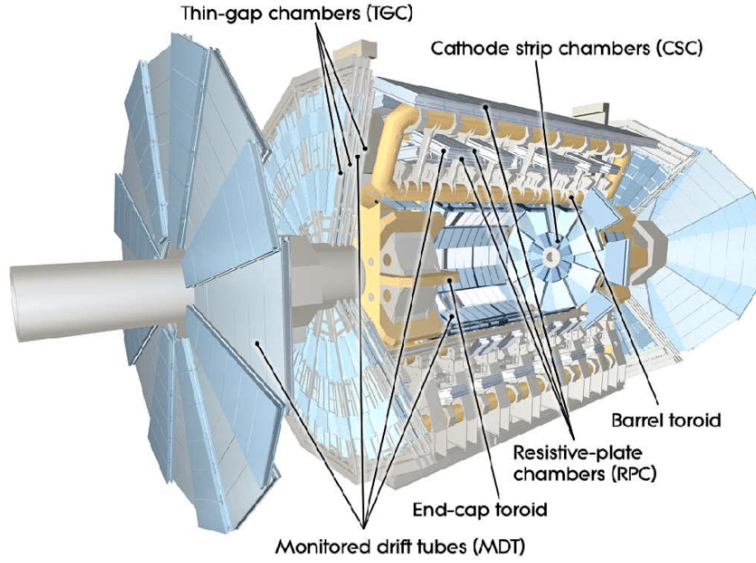


Figure 2.14: Cut-away view of the ATLAS muon system [49].

Over most of the $|\eta|$ range, muon track coordinates are measured by Monitored Drift Tubes (MDTs). Charged particles passing through the MDTs will ionize gas atoms within the tubes and these ions will then drift and be readout by a tungsten anode wire. The MDT system consists of 1,171 chambers and 354,240 individual tubes which provide a resolution of $80 \mu\text{m}$. At larger pseudo-rapidities ($2 < |\eta| < 2.7$) fine-granularity Cathode Strip Chambers (CSCs) are used in the first layer of the muon spectrometer to provide additional discriminating power in high occupancy environments. The CSCs are segmented in ϕ in 8 chambers and each chamber consists of an anode wire oriented in the radial direction with multiple cathode strips laid perpendicularly across. Crossing muons will deposit charges on multiple strips and interpolating between these charges provides a position measurement. The CSCs provide a resolution of $60 \mu\text{m}$ [55].

The muon trigger system covers the range $|\eta| < 2.4$ and supports precise momentum measurements and track information orthogonal to that provided by the MDTs and CSCs. In the barrel,

two layers of Resistive Plate Chambers (RPCs) are assembled with the middle layer of the MDTs and a third layer of RPCs is placed outside the final MDT layer (Figure 2.15). RPCs consist of a small gas-gap between two parallel resistive plates; charged particles passing the gas-gap will create a shower of electrons that drift to the readout anode. Each RPC unit contains two gas-gaps, one oriented in the ϕ direction and the other in the η direction. In the endcaps, three layers of Thin Gap Chambers (TGCs) are positioned perpendicular to the beam axis. TGCs are multi-wire proportional chambers; the wires are held at high-voltage and connected to grounded cathode planes which collect ionization charges. The high voltage and close proximity of the wires in the TGCs allows excellent time resolution [55].

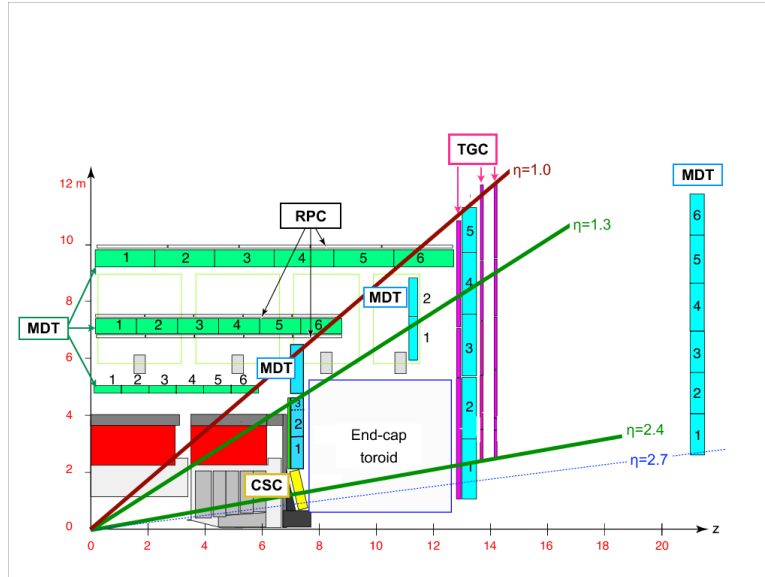


Figure 2.15: A schematic picture showing a quarter-section of the muon system in a plane containing the beam axis [56].

2.2.1.4 MAGNET SYSTEM

The ATLAS Magnet System bends the tracks of charged particles to allow momentum measurements. The track bending is caused by the Lorentz force: $F = qE + q(v \times B)$ and is hence proportional to the particle's velocity. The magnet system is shown in Figure 2.16. It has three components: a 2T solenoid magnet encompassing the inner detector, a 0.5T toroid in the barrel region of the muon spectrometer, and 2.1T torroids in each of the muon spectrometer endcaps [57].

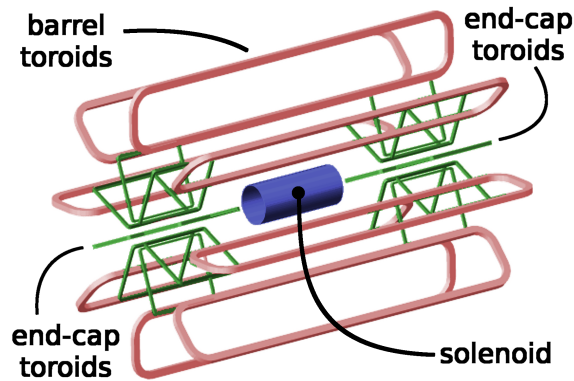


Figure 2.16: Schematic diagram of the ATLAS magnet system

The central solenoid is a conduction cooled superconducting solenoid with minimum radial thickness. The barrel and endcap toroids each contain 8 air-core superconducting coils. These unique toroid magnets are what gives the ATLAS detector its name.

Framing reality is one of the only ways we can be sure it exists.

Sturgill Simpson

3

ATLAS Computing and Software

THE ATLAS DETECTOR PRODUCES PETABYTES OF DATA EVERY YEAR WHILE ATLAS RESEARCHERS PRODUCE ADDITIONAL PETABYTES OF SIMULATED DATA FOR VALIDATION STUDIES AND ANALYSIS DESIGN. The production, processing, and analysis of this data requires massive computing resources and the development specialized software. The ATLAS data ecosystem is described in the following chapter and is summarized in Figure 3.1.

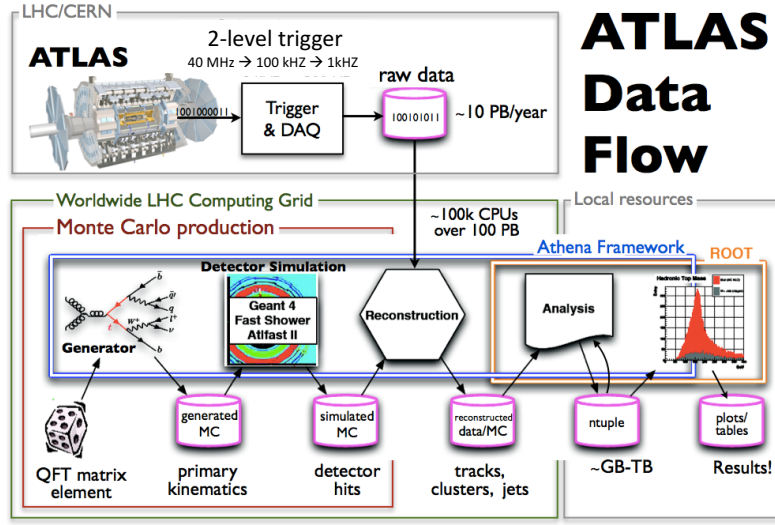


Figure 3.1: A high-level summary of the ATLAS data flow [58].

3.1 DATA ACQUISITION AND STORAGE

The first stage of processing for collision data is the Trigger and Data Acquisition System, collectively referred to as TDAQ. When the LHC is running at optimal performance, it produces over 30 million bunch crossings per second within the ATLAS detector. Each physics event requires approximately 25 megabytes to store the information from all readouts; this corresponds to a data production of nearly 60 million GB per second which, given current technology is impossible to store without some sort of size reduction [59]. Furthermore, as described in Chapter 2, a large portion of the data does not correspond to the hard-scattering events of interest. Thus, ATLAS relies heavily on the trigger system, a 2 level computing system which selects 0.1% of events for storage and further analysis [60].

3.1.1 TRIGGERS AND DAQ

The ATLAS TDAQ is divided into 2 distinct levels as shown in Figure 3.2. The Level 1 Trigger (L1), a hardware trigger which uses a subset of information from the calorimeters and muon spectrometer, and the software-based High Level Trigger (HLT), which refines L1 decisions by applying reconstruction algorithms that closely mirror the offline algorithms while accounting for trigger specific time and CPU constraints [60].

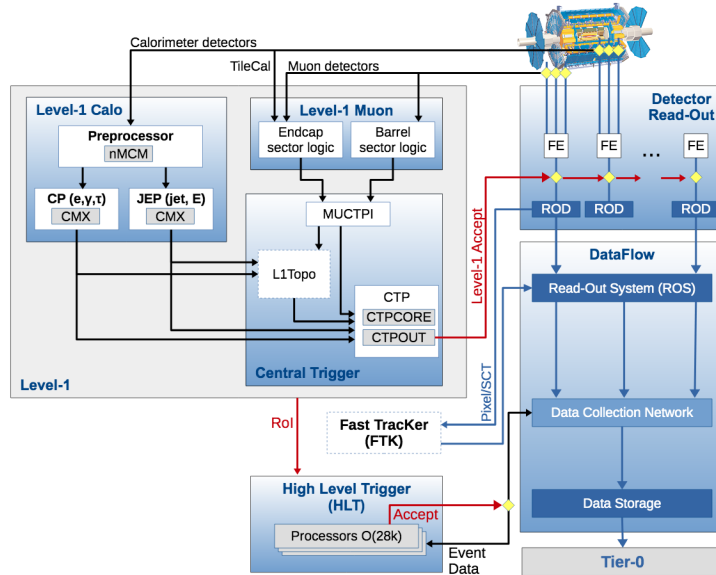


Figure 3.2: Schematic diagram of the ATLAS TDAQ System in Run 2. The FTK is still in the commissioning phase [60].

The L1 trigger runs on custom FPGAs (Field Programmable Gate Arrays); decisions are completed within $2 \mu\text{s}$ of the event occurring and it has a total output frequency of less than 100kHz.

The final L_1 decision is determined by the central trigger processor (CTP) which receives information from the calorimeters and muon spectrometer, as well as information from Minimum Bias Trigger Scintillators, the LUCID Cherenkov Counter, and the Zero-Degree Calorimeter [49]. The L_1 calorimeter algorithms utilize Region of Interests (ROIs) formed from trigger tower clusters (Figure 3.3) whose energy exceeds a predefined threshold. ROIs for EM objects and taus are formed and processed separately from ROIs for jets, and this information is merged in the L_1 Topo module and sent to the CTP. The L_1 muon algorithms relies on signals from the muon trigger chambers. Information from the barrel and endcap chambers are processed separately to find hit patterns consistent with high- p_T muons originating from the interaction point and these signals are then combined and passed to the CTP. The CTP implements a trigger ‘menu’ of different combinations of object types and energies, and surviving events are then buffered in the Read Out System (ROS) and eventually sent to the HLT [60].

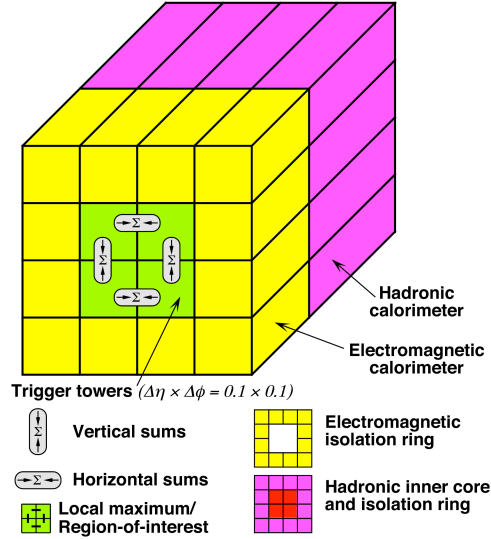


Figure 3.3: Schematic view of the trigger towers used as input to the L_1 Calorimeter trigger algorithms

The HLT is a large farm of CPUs which further analyze the ROIs defined by the L_1 trigger; it

has an output frequency of around 1kHz. The HLT runs multiple feature extraction algorithms on these ROIs to request event-data fragments and then performs simplified reconstruction algorithms (also referred to as online reconstruction) on these data subsets. The reconstructed features are passed to a boolean decision tree which determines if the trigger conditions are satisfied and, if so, the event data is processed through the Data Acquisition System where it is stored at the experiment site and sent through the Tier 0 computing facility (described further in Section 3.1.2) [61].

3.1.1.1 TRIGGER MENUS

The full collection of all L1 and HLT triggers are referred to as the trigger menu, which allows physics analyzers to select events with desired signatures. The frequency at which a trigger flags events for storage is called the trigger rate; prescales are used in some triggers to reduce the rate of data output by randomly rejecting a defined fraction of passed events. The trigger menu is divided into several categories:

- primary triggers: used for physics analyses and typically un-prescaled. These triggers cover all physics signatures relevant to ATLAS research including electrons, photons, muons, taus, jets, and missing energy.
- support triggers: used for efficiency and performance measurements or monitoring; typically highly prescaled.
- alternative triggers: used to implement new or experimental reconstruction algorithms that differ from primary or support trigger design
- backup triggers: have tighter selections and lower rates than primary and secondary triggers
- calibration triggers: used for detector calibration. They often operate at a very high rate but store small events with only the information relevant to calibration [60].

The main triggers for 2017 are shown in Table 3.1, along with their corresponding rates. The physics trigger group rates for L1 and HLT as a function of luminosity blocks is shown in Figure 3.4.

Trigger	Typical offline selection	Trigger Selection		Level-1 Peak Rate (kHz)	HLT Peak Rate (Hz)
		Level-1 (GeV)	HLT (GeV)	$L = 1.7 \times 10^{34} \text{ cm}^{-2} \text{ s}^{-1}$	
Single leptons	Single isolated μ , $p_T > 27 \text{ GeV}$	20	26 (i)	15	180
	Single isolated tight e , $p_T > 27 \text{ GeV}$	22 (i)	26 (i)	28	180
	Single μ , $p_T > 52 \text{ GeV}$	20	50	15	61
	Single e , $p_T > 61 \text{ GeV}$	22 (i)	60	28	18
	Single τ , $p_T > 170 \text{ GeV}$	100	160	1.2	47
Two leptons	Two μ , each $p_T > 15 \text{ GeV}$	2×10	2×14	1.8	26
	Two μ , $p_T > 23, 9 \text{ GeV}$	20	22, 8	15	42
	Two very loose e , each $p_T > 18 \text{ GeV}$	2×15 (i)	2×17	1.7	12
	One e & one μ , $p_T > 8, 25 \text{ GeV}$	$20 (\mu)$	7, 24	15	5
	One e & one μ , $p_T > 18, 15 \text{ GeV}$	15, 10	17, 14	2.0	4
	One e & one μ , $p_T > 27, 9 \text{ GeV}$	22 (e, i)	26, 8	28	3
	Two τ , $p_T > 40, 30 \text{ GeV}$	20 (i), 12 (i) (+jets, topo)	35, 25	5	61
	One τ & one isolated μ , $p_T > 30, 15 \text{ GeV}$	12 (i), 10 (+jets)	25, 14 (i)	2.1	10
	One τ & one isolated e , $p_T > 30, 18 \text{ GeV}$	12 (i), 15 (i) (+jets)	25, 17 (i)	4	15
Three leptons	Three loose e , $p_T > 25, 13, 13 \text{ GeV}$	$20, 2 \times 10$	$24, 2 \times 12$	1.3	< 0.1
	Three μ , each $p_T > 7 \text{ GeV}$	3×6	3×6	0.2	6
	Three μ , $p_T > 21, 2 \times 5 \text{ GeV}$	20	$20, 2 \times 4$	15	8
	Two μ & one loose e , $p_T > 2 \times 11, 13 \text{ GeV}$	$2 \times 10 (\mu)$	$2 \times 10, 12$	1.8	0.3
	Two loose e & one μ , $p_T > 2 \times 13, 11 \text{ GeV}$	$2 \times 8, 10$	$2 \times 12, 10$	1.7	0.1
One photon	One loose γ , $p_T > 145 \text{ GeV}$	22 (i)	140	28	43
Two photons	Two loose γ , $p_T > 55, 55 \text{ GeV}$	2×20	50, 50	2.6	6
	Two medium γ , $p_T > 40, 30 \text{ GeV}$	2×20	35, 25	2.6	17
	Two tight γ , $p_T > 25, 25 \text{ GeV}$	2×15 (i)	2×20 (i)	1.7	14
Single jet	Jet ($R = 0.4$), $p_T > 435 \text{ GeV}$	100	420	3.3	33
	Jet ($R = 1.0$), $p_T > 480 \text{ GeV}$	100	460	3.3	24
	Jet ($R = 1.0$), $p_T > 450 \text{ GeV}$, $m_{\text{jet}} > 50 \text{ GeV}$	100	$420, m_{\text{jet}} > 40$	3.3	29
E_T^{miss}	$E_T^{\text{miss}} > 200 \text{ GeV}$	50	110	5	110
Multi-jets	Four jets, each $p_T > 125 \text{ GeV}$	3×50	4×115	0.5	16
	Five jets, each $p_T > 95 \text{ GeV}$	4×15	5×85	5	10
	Six jets, each $p_T > 80 \text{ GeV}$	4×15	6×70	5	4
	Six jets, each $p_T > 60 \text{ GeV}$, $ \eta < 2.0$	4×15	$6 \times 55, \eta < 2.4$	5	15
b -jets	One b ($\epsilon = 40\%$), $p_T > 235 \text{ GeV}$	100	225	3.3	15
	Two b ($\epsilon = 60\%$), $p_T > 185, 70 \text{ GeV}$	100	175, 60	3.3	12
	One b ($\epsilon = 40\%$) & three jets, each $p_T > 85 \text{ GeV}$	4×15	4×75	5	15
	Two b ($\epsilon = 70\%$) & one jet, $p_T > 65, 65, 160 \text{ GeV}$	$2 \times 30, 85$	$2 \times 55, 150$	1.2	15
	Two b ($\epsilon = 60\%$) & two jets, each $p_T > 65 \text{ GeV}$	$4 \times 15, \eta < 2.5$	4×55	3.2	13
B -Physics	Two μ , $p_T > 11, 6 \text{ GeV}$	11, 6	11, 6 (di- μ)	2.5	47
	Two μ , $p_T > 6, 6 \text{ GeV}$, $2.5 < m(\mu, \mu) < 4.0 \text{ GeV}$	$2 \times 6 (J/\psi, \text{topo})$	$2 \times 6 (J/\psi)$	1.6	48
	Two μ , $p_T > 6, 6 \text{ GeV}$, $4.7 < m(\mu, \mu) < 5.9 \text{ GeV}$	$2 \times 6 (B, \text{topo})$	$2 \times 6 (B)$	1.6	5
	Two μ , $p_T > 6, 6 \text{ GeV}$, $7 < m(\mu, \mu) < 12 \text{ GeV}$	$2 \times 6 (Y, \text{topo})$	$2 \times 6 (Y)$	1.4	10
Total Rate				85	1550

Table 3.1: The main ATLAS triggers for the 2017 trigger menu with observed rates [61].

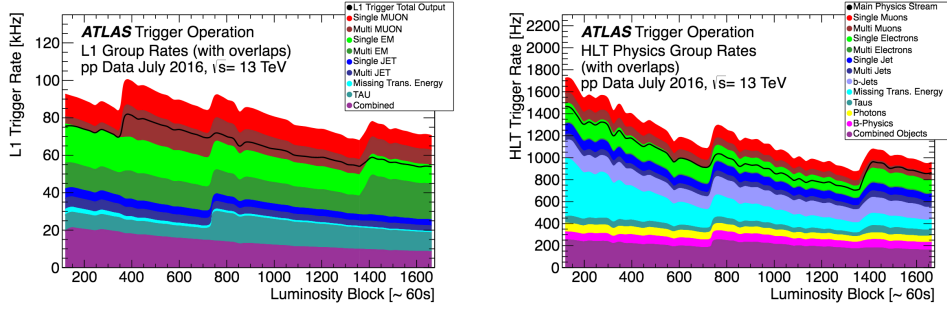


Figure 3.4: Physics trigger group rates for the L1 (left) and HLT (right) as a function of the number of luminosity blocks from a fill taken in July 2016 [62].

3.1.2 STORAGE

The Worldwide LHC Computing Grid (WLCG) provides global resources to store, distribute, process, and analyze the over 50 PB of data produced by the LHC every year [63]. WLCG is divided into 4 levels called Tiers. All data recorded by ATLAS and other LHC experiments first enters the CERN Data Center, referred to as Tier 0, where raw detector data is reconstructed into physics objects. The ATLAS algorithms run at Tier 0 are described further in Section 3.3. Tier 0 is connected by 10 GB/s optical fibers to 13 Tier 1 centers around the world. Tier 1 centers are responsible for storing raw and reconstructed data, performing large-scale data reprocessing, and storing simulated data produced at Tier 2 centers. The work described in this dissertation was primarily conducted using the BNL ATLAS Tier 1 center [64].

The 155 Tier 2 systems are typically located at universities or other scientific facilities with sufficient computing power. Tier 2s are used primarily for generation and reconstruction of simulated data. Finally, Tier 3 systems are used by individual physicists or research groups to access the WLCG and conduct individual analysis tasks [65].

All ATLAS WLCG data is managed by the Rucio Distributed Data Management System [66]. Rucio ensures that all ATLAS researchers can manage and process data in a heterogeneous dis-

tributed environment. The system manages more than 300 PB of data contained in more than 830 million files using over 130 global data centers [67].

3.2 SIMULATION

Simulated data is used to understand and inform detector hardware, design reconstruction and analysis software, and interpret predictions and results of various theories. The ATLAS simulation infrastructure is comprised of a range of software which generates particle collision simulations and carries them through the process of hadronization, detector interactions, and digitization. The output of the simulation chain is identical in structure to the output of the TDAQ system. The entire simulation flow is shown in Figure 3.5. The ATLAS simulation chain relies on randomized methods of Monte Carlo sampling, and simulated data is often referred to as simply ‘Monte Carlo’ (MC).

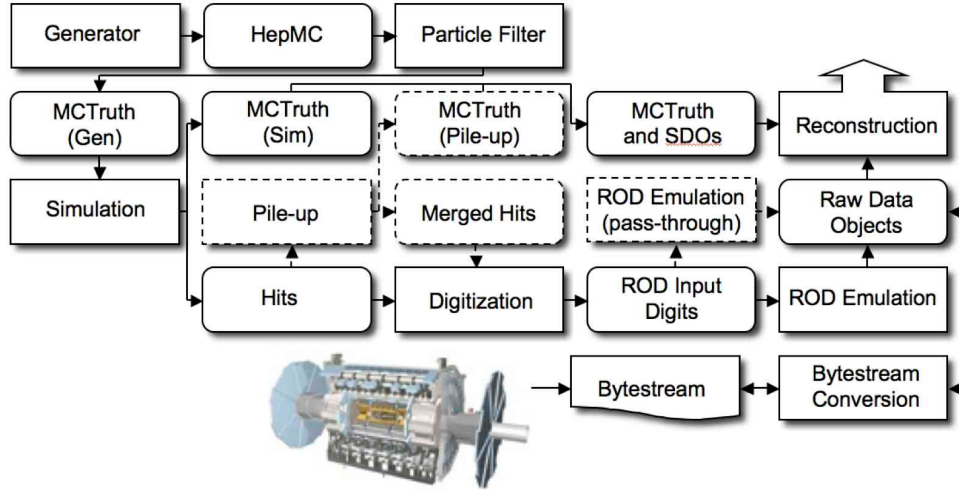


Figure 3.5: The flow of the ATLAS simulation software, from event generators (top left) through reconstruction (top right). Algorithms are placed in square-cornered boxes and persistent data objects are placed in rounded boxes. The optional pile-up portion of the chain, used only when events are overlaid, is dashed [68].

3.2.1 EVENT GENERATION

Physics generators that simulate the desired hard-scattering processes and subsequent decays are the first stage of MC simulation. These generators produce a list of particles created by proton-proton collisions that are expected to exist long enough to propagate through the detector. The generators also record “truth” information listing every incoming and outgoing particle regardless of whether it is passed to the later simulation stages [68].

ATLAS simulation utilizes several externally supported general purpose MC generators which employ similar computation strategies. The first step is generating the central hard-scattering process. This relies on matrix-element based computations and these matrix-elements can be hard coded by the developer or calculated to a certain order in perturbation theory by the generator. The evolution of the produced particles due to QCD (Chapter 1) is then described by parton showering which connects the hard scale of colored parton creation to the hadronization scale. At the hadronization scale, on the order of a few Λ_{QCD} , QCD partons are transformed into primary hadrons by applying phenomenological fragmentation models based on effective field theories. Generators also account for QED bremsstrahlung radiation and potential secondary hard or semi-hard interactions of remaining hadron remnants. The combination of these event generation stages is represented pictorially for a single ttH event in Figure 3.6 [69].

General purpose generators that function as described above include Pythia [70], Herwig [71], MadGraph [72], and Sherpa [69]. These generators are typically FORTRAN or C++ based and their outputs are all translated into the standard HepMC event record format [73]. The analyses presented in this dissertation primarily utilize Pythia and Sherpa generated MC samples. These general purpose generators described above are often interfaced with specialized generators to improve the description of certain final states. These include Tauola for tau lepton decays [74], EvtGen for B meson and hadron decays [75], AlpGen for hadronization in final states with multiple well-separated

jets [76], MC@NLO for top events [77], and AcerMC for W and Z decays with several jets [78].

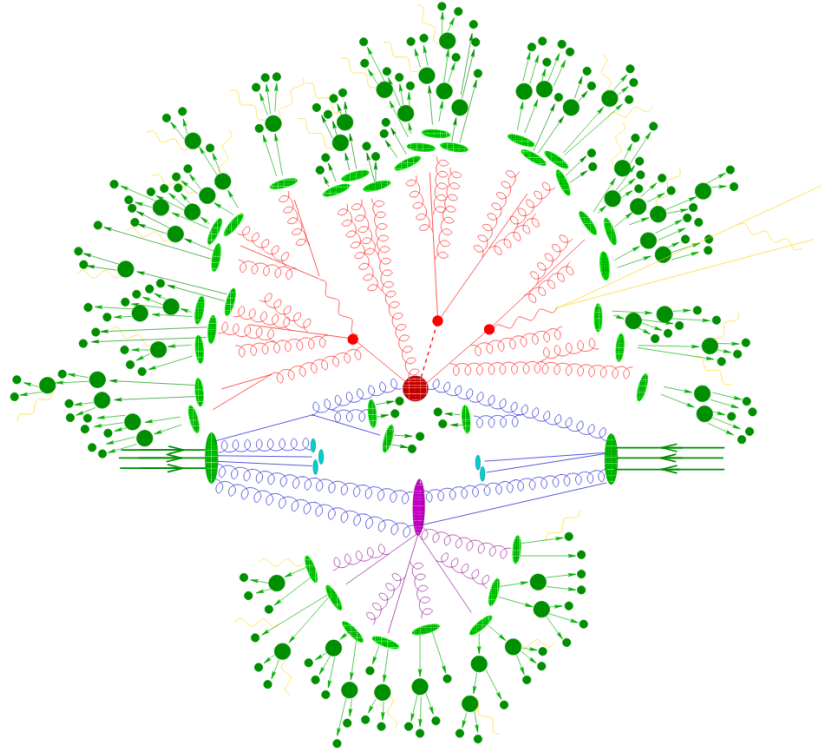


Figure 3.6: Pictorial representation of a generated $t\bar{t}H$ event. The central hard interaction is shown as the central red circle, which is followed by the subsequent decays of the Higgs and top quarks (smaller red circles). Additional hard radiation due to QCD is produced (blue) as well as a secondary hard interaction (purple). Finally, hadronization (light green) and hadron decay (dark green) occurs. Photon radiation is represented in yellow and occurs at multiple stages [69].

3.2.2 DETECTOR SIMULATION

The particle lists created by event generators are further processed through a detector simulation which describes how the particles interact with various physical components of the ATLAS detector and how those interactions are read out by detector electronics.

Subsystem	Materials	Solids	Logical vol.	Physical vol.	Total vol.
Beampipe	43	195	152	514	514
BCM	40	131	91	453	453
Pixel	121	7,290	8,133	8,825	16,158
SCT	130	1,297	9,403	44,156	52,414
TRT	68	300	357	4,034	1,756,219
LAr calorimetry	68	674	639	106,519	506,484
Tile calorimetry	8	51,694	35,227	75,745	1,050,977
Inner detector	243	12,501	18,440	56,838	1,824,614
Calorimetry	73	52,366	35,864	182,262	1,557,459
Muon system	22	33,594	9,467	76,945	1,424,768
ATLAS total	327	98,459	63,769	316,043	4,806,839

Table 3.2: Numbers of materials and volumes used to construct the ATLAS detector simulated geometry [68].

The ATLAS Simulation group supports a centralized simulated detector geometry written in GEANT4 [79], a widely used scientific simulation toolkit. The ATLAS detector description contains nearly 5 million volumes comprised of over 300 different materials. A summary of these volumes is shown in Table 3.2. The detector simulation is revised as updates are made to the detector or when additional simulation functionality becomes available [80].

GEANT4 also provides physics models to simulate different processes of particle interaction with materials including the photoelectric effect, Compton scattering, bremsstrahlung, ionization, multiple scattering, decays, nuclear interactions and more. These physics models rely on iterative Monte Carlo sampling. The models are applied as appropriate to generated particles traveling through the simulated detector geometry. Truth information, taken from the event generation record, including real tracks and decays of particles is also stored.

The final simulation step is digitization, which transforms the hits and energy deposits created by the detector simulation into Raw Data Objects identical in structure to the electronic readouts of the physical detector. Pileup overlay, detector noise, and other backgrounds such as cosmic ray

muons are also added to the simulation during digitization. Truth information from the digitization stage is stored as Simulated Data Objects (SDOs) which act as maps from generated truth particles to simulated detector readouts. SDOs can be used later to validate reconstruction algorithms and quantify their performance [68].

The full simulation chain described above is incredibly time and CPU intensive and as such is not possible to provide complete simulations with high statistics for all relevant physics studies. A variety of programs, collectively referred to as Fast Simulations, exist to complement the full simulation in cases when full accuracy isn't necessary. Approximately 75% of full simulation time is spent simulating electromagnetic particles; the Fast G4 Simulation expedites this process by replacing low energy electromagnetic particles with pre-simulated showers [81]. The ATLFAST-II package provides large simulation statistics by directly simulating the input to reconstruction algorithms rather than separate detector geometry and digitization steps. It contains two modules: FATRAS [82], which speeds up Inner Detector and Muon Spectrometer simulations by generating only the track input required by reconstruction algorithms, and FastCaloSim [83], which reduces particle-calorimeter interaction simulation time by directly calculating the energy of single particle showers.

3.3 DATA PROCESSING

Both recorded and simulated data are processed through the ATLAS reconstruction and object identification chain before being used in physics analyses. Reconstruction algorithms are typically detector-component specific and, when combined, create particle candidates. Identification algorithms specific to individual SM particles are supported by dedicated Working Groups within the collaboration and further classify particle candidates. Datasets with labeled particles, trigger history, and in the case of MC, truth information, are then used by individual researchers for physics analyses.

3.3.1 RECONSTRUCTION

The particle identification algorithms described in Section 3.3.2 typically take partially reconstructed objects as inputs, rather than directly using detector read-out information. These reconstructed objects take the form of tracks left in the Inner Detector, interaction vertices, energy deposit clusters in the calorimeters, or matched combinations of these. The common algorithms used to reconstruct these objects are described below.

3.3.1.1 TRACKING

The path a particle takes through the inner detector is called a track, and track finding algorithms seek to reconstruct these paths by connecting hits in various layers of the inner detector. These algorithms must provide parameters to fully describe the track (impact parameters d_0 and z_0 , angles ϕ and θ , and particle charge and momentum) and account for various particle-material interactions such as scattering, ionization loss, bremsstrahlung, and hadronization. The full track finding process proceeds in several steps [84].

First, a connected component analysis is used to cluster pixels and strips in a given detector component together. These clusters are formed when the combined deposited energy in cells sharing an edge or corner exceeds a certain tunable threshold. These clusters are then transformed into 3-D objects called space-points; in the pixel detector one cluster becomes one spacepoint while in the SCT clusters from both sides of a strip are required [85]. Spacepoints can be classified as single-particle (all energy comes from a single particle) or merged (multiple particles deposit energy in the same cells) as shown in Figure 3.7.

Space-points are then combined in sets of three, called track seeds, which serve as the basis for track candidates. Track seeding allows full track finding algorithms to consider a large number of possible tracks while still providing an initial momentum estimate. The quality of final tracks de-

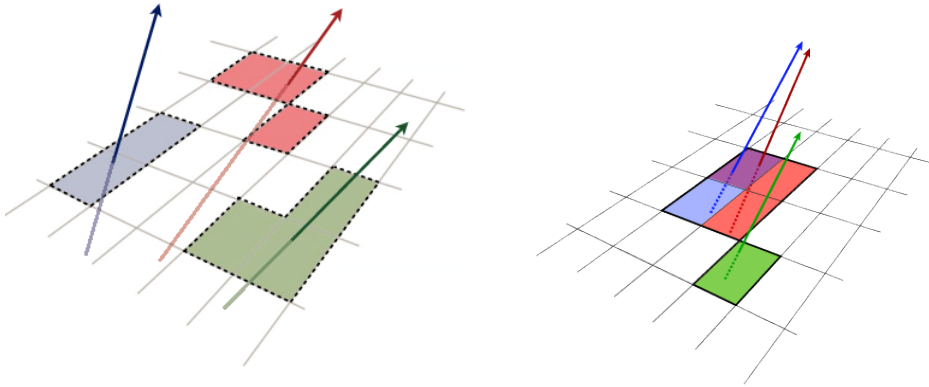


Figure 3.7: Illustration of a single particle space-point (left) and merged particle space-point (right) on a pixel sensor [85].

depends on which sub-detector the track seed originates in, so seeds are ordered to prioritize SCT-only, then pixel-only, then mixed-detector seeds. The track seeds are then passed to a combinatorial Kalman filter [86] which builds full track candidates by adding additional compatible space-points from remaining layers of the inner detector.

An ambiguity solver algorithm is used to process cases where multiple track candidates share the same space point. Tracks are scored by considering track momentum, track holes, track fit accuracy, and assigned space-points. Tracks are then processed in descending order of score and shared space-points are either assigned as uniquely corresponding to that track or merged with another track, the space-point is removed from the track, or the track is rejected entirely. Track candidates which pass the ambiguity solver are assigned a final neural network calculated parameterized fit using all available track information [85]. Neural networks are described further in Chapter 4.

Additional TRT-based track seeding is used in certain ROIs corresponding to electromagnetic calorimeter clusters created by converted photons [87]. Muon specific reconstruction and identification algorithms also utilize track fitting in the Muon Spectrometer, and this is described further in Section 3.3.2.1.

ATLAS track reconstruction is highly efficient as shown in Figure 3.8. However, the detector resolution limits the performance potential of current techniques in high pile-up environments. Revising the ATLAS tracking software, potentially by incorporating additional ML techniques, is an on-going research area for the HL-LHC.

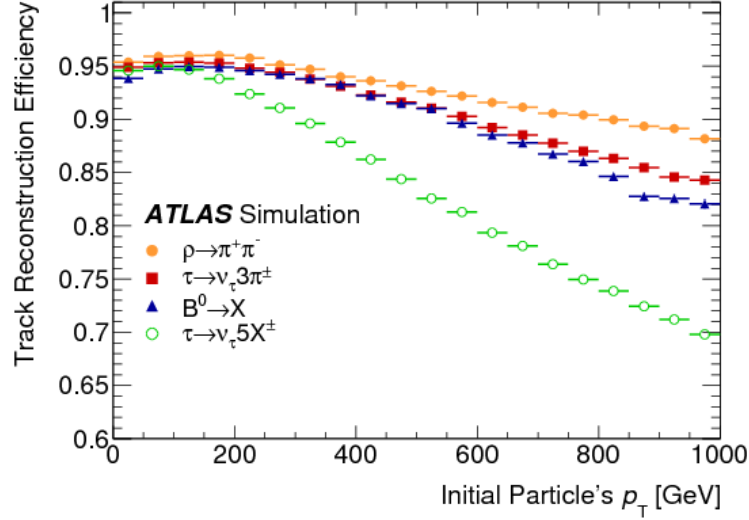


Figure 3.8: Single track reconstruction efficiency as a function of particle p_T for a ρ (orange), three prong τ (red), five prong τ (green), and B^0 (blue) [85].

3.3.1.2 VERTEXING

The points where multiple tracks intersect are called vertices. Vertices are found using an iterative algorithm which seeds vertices using the z-position at the beamline of the reconstructed tracks by applying a χ^2 fit to the seed and nearby tracks. Tracks originating more than 7σ away from the vertex are used to seed a new vertex and this process is repeated until all possible vertices are defined. All vertices are required to have at least two associated tracks. Vertices are associated with interactions by calculating the sum of the squares of p_T of associated tracks, and at least 50% of the energy of an interaction must be accounted for by the assigned tracks. This ensures that final vertex position

is influenced most heavily by tracks of particles coming from that interaction. The vertex with the highest $(p_T)^2$ sum is called the primary vertex.[88]

A new vertexing algorithm is being developed based on image processing ML techniques. This algorithm simultaneously identifies all potential vertices in a given event using plots of the identified tracks as input. This algorithm is more robust to pile-up and reduces the CPU time required for vertexing, both of which are important considerations for HL-LHC algorithm design [87].

The vertex reconstruction efficiency for 2018 is shown in Figure 3.9 by comparing the average number of reconstructed vertices to the average number of interactions per bunch crossing (μ). The efficiency decreases with increasing μ due to shadowing, where two or more interactions are too close in proximity to be resolved individually.

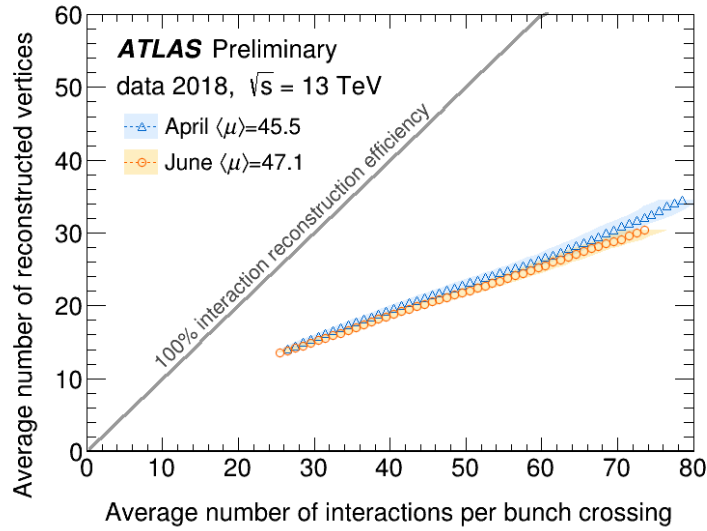


Figure 3.9: The number of vertices reconstructed as a function of the average interactions per bunch crossing for two fills taken in 2018 [89].

3.3.1.3 CALORIMETER CLUSTERING

Particles deposit energy in a spread of calorimeter cells both laterally and longitudinally. Dedicated algorithms group the cells corresponding to an individual particle together and assign a cumulative energy value to the constructed cluster. ATLAS uses two types of clustering algorithms which are described below.

The “sliding window” clustering algorithm sums cells within a fixed-size section of the calorimeter. This algorithm proceeds in three steps. First, the calorimeters are divided into a grid of sections of size $\Delta\eta \times \Delta\phi$, called towers, and the energy of all cells in all layers within the tower is summed. Then a window of fixed-size $N_{\eta}^{window} \times N_{\phi}^{window}$ is moved across the grid of towers and when the energy contained within the window is a local maximum and exceeds a preset threshold, the group of towers within the window is categorized as a precluster. The location of the precluster is set as the energy-weighted center of all cells within a separate fixed-size window positioned on the tower at the center of the sliding window. The final cluster is formed by adding all additional cells within an $N_{\eta}^{cluster} \times N_{\phi}^{cluster}$ window of the seed position.

Sliding window clustering is used to form the clusters used in the electron, photon, tau, and jet identification algorithms. The sliding window, position window, cluster, and tower sizes, as well the energy threshold can be adjusted to form separate input clusters for these different objects [90].

Topological clustering is an alternative algorithm which iteratively adds neighboring cells to an initial seed cluster when the energy in the new cell is above an expected noise threshold. This method results in final clusters of varying size, in contrast to the sliding window algorithm. All cells with a signal-to-noise above an initial threshold (related to the expected electronics noise from gain and detector conditions) are identified as cluster seeds. The seeds are sorted in descending order of signal strength, and all neighboring cells are considered for addition to the cluster. A neighboring cell is added if it has not already been used as a seed and its signal-to-noise ratio is above the set neigh-

bor threshold. If a cell is adjacent to more than one seed, the two seeds are merged. An example of a topologically clustered jet event is shown in Figure 3.10.

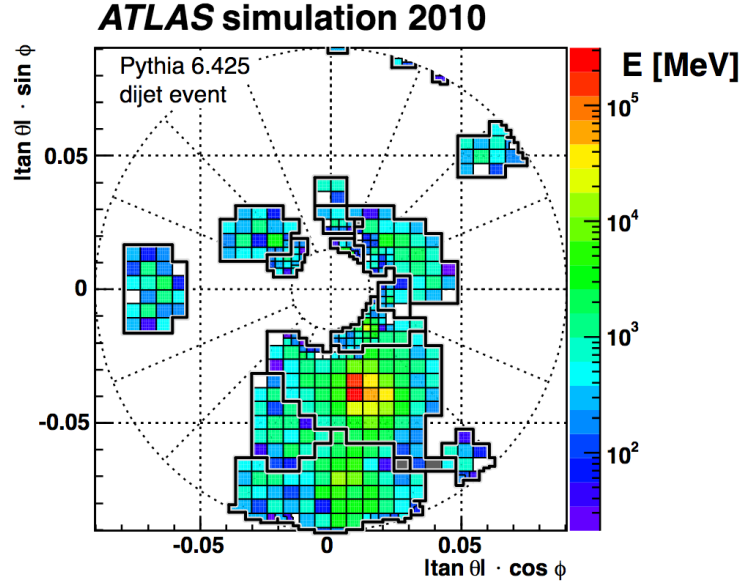


Figure 3.10: Clusters for a simulated dijet event created in the FCAL using the topological clustering algorithm [91].

Topological clustering is very effective at suppressing noise in large clusters, and thus is used for jet and missing energy algorithms. The signal-to-noise thresholds for seeding and adding neighbors can be adjusted for different applications, as can the number of neighboring cells considered. Additional details can be found in [91].

The final energy assigned to the constructed cluster depends on the clustering algorithm used and the type of object the cluster is eventually assigned to. The calibration methods are described in detail in [92] for electrons and photons and in [93] for taus and jets.

3.3.2 IDENTIFICATION

The individual algorithms used to fully reconstruct and identify the objects relevant to the $VH, H \rightarrow \tau\tau$ analysis described in Chapters 6 and 7 are outlined below. The software used for identifying electrons is described in detail in Chapter 5.

3.3.2.1 MUONS

Accurately reconstructing and identifying muons is essential to a variety of ATLAS analyses including the Higgs boson discovery [94] and subsequent measurements, other SM measurements (eg [95]), and BSM searches ([96] and many others). The dedicated muon algorithms must work for a variety of muons ranging from soft, non-isolated muons produced in jets to high p_T muons from W/Z decays or new physics. The algorithms described below make use of all components of the ATLAS detector: tracks in the Inner Detector, track stubs in the Muon Spectrometer, and energy deposits in the calorimeters and Muon Spectrometer.

In addition to Inner Detector tracks, which are constructed as described in Section 3.3.1.1, muon algorithms also make use of tracks in the Muon Spectrometer (MS). Tracks are first reconstructed in individual components of the MS using a Hough transform in the MDT segments and associated trigger chambers and a combinatorial search in the CSC. Full track candidates are then formed by matching segments in different layers of the MS using a segment-seeded combinatorial search that begins with seeds in the middle MS layers. Finally, track candidates are fit using a global χ^2 fit [97].

The MS-track candidates are then combined with Inner Detector tracks and calorimeter clusters to form four types of muon candidates (demonstrated graphically in Figure 3.11):

- Combined muons: tracks in the MS and Inner Detector are combined using a global refit. The $VH, H \rightarrow \tau\tau$ analysis described in this dissertation uses only combined muons.
- Segment-tagged muons: tracks in the Inner Detector are classified as muon candidates if they can be extrapolated and matched to at least one track segment in the MS.

- Calorimeter-tagged muons: tracks in the Inner Detector are classified as muon candidates if they can be matched to a calorimeter cluster consistent with a minimum-ionizing particle.
- Extrapolated or stand-alone muons: a track in the MS is extrapolated to the Inner Detector.

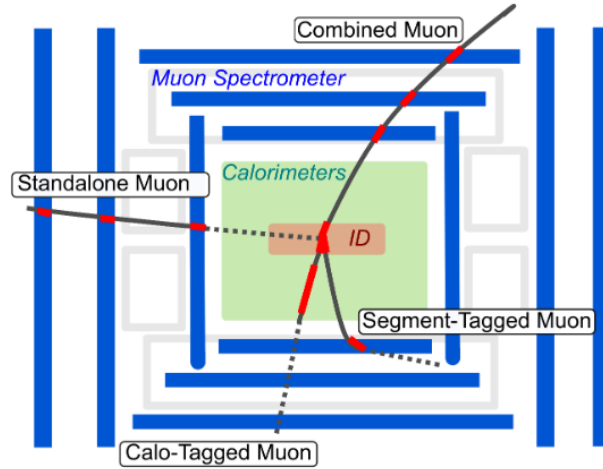


Figure 3.11: Depiction of the four types of reconstructed muons

Overlap between the muon types is resolved in the order they are listed above [97]. An example of muon reconstruction efficiency is shown in Figure 3.12.

Muon candidates are then officially identified as muons by placing requirements on fit quality and momentum reconstruction (further described in [97]). The muon Working Group in ATLAS supports four muon identification (ID) working points (Loose, Medium, Tight, and High p_T) to suit the efficiency and purity needs of different analyses. The efficiencies of the various operating points for muons originating in W decays is shown in Table 3.3.

3.3.2.2 JETS

Jets are hadronic showers of particles that originate from some initial parton splitting. Jets are therefore not well-defined particles like electrons or muons. The goal of jet reconstruction algorithms

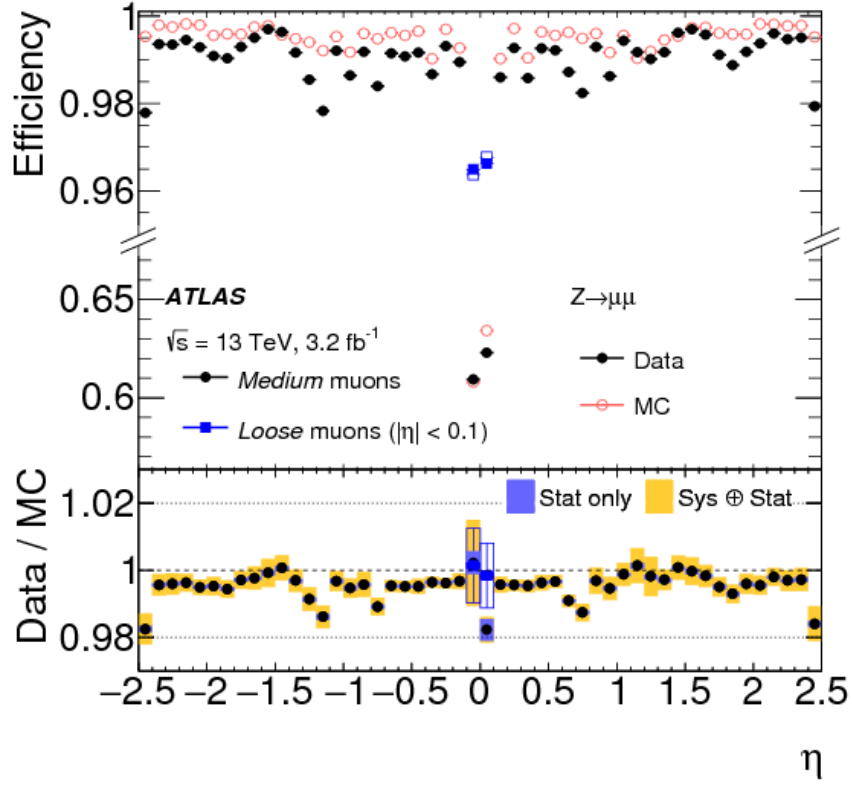


Figure 3.12: Muon reconstruction efficiency as a function of η measured in $Z \rightarrow \mu\mu$ events for muons with $p_T > 10$ GeV [97].

Selection	$4 < p_T < 20$ GeV		$20 < p_T < 100$ GeV	
	$\epsilon_{\mu}^{\text{MC}} [\%]$	$\epsilon_{\text{Hadrons}}^{\text{MC}} [\%]$	$\epsilon_{\mu}^{\text{MC}} [\%]$	$\epsilon_{\text{Hadrons}}^{\text{MC}} [\%]$
Loose	96.7	0.53	98.1	0.76
Medium	95.5	0.38	96.1	0.17
Tight	89.9	0.19	91.8	0.11
High- p_T	78.1	0.26	80.4	0.13

Table 3.3: ID efficiencies for prompt muons from W decays and the misidentification rates for hadron decays computed from a $t\bar{t}$ MC sample [97].

is to group related energy deposits into a single collection and assign an accurate energy to the constructed collection.

Many algorithms exist to group related phase-space objects, however many of these techniques are not well suited to the specific problem of jet clustering. This is because jet clustering algorithms in ATLAS must satisfy two physics derived properties in order for the reconstructed jets to be usable in final physics analyses. These properties are Infrared (IR) Safety, which means the calibrated energy of the constructed jet must be invariant to the addition of infinitely soft radiation, and Collinear Safety, which means jet properties must remain identical if a single parton is substituted for two or more collinear partons with identical momentum. Collectively, these requirements are referred to as IRC safety.

There are two primary classes of jet algorithms utilized by ATLAS, both of which take topoclusters (Section 3.3.1.3) as inputs. Sequential clustering algorithms group topoclusters together using two distance metrics: $d_{ij} = \min(p_{ti}^a p_{tj}^a) \times \frac{R_{ij}^2}{R}$ where a is an algorithm-dependent constant, R_{ij} is the η - ϕ distance between the two clusters, and R is a set maximum radius for the reconstructed jets, and $d_{iB} = p_{ti}^a$, the momentum-space distance between the beam axis and the new cluster. The algorithms create an ordered list of d_{ij} and d_{iB} and add clusters j to jet i until d_{iB} is the minimum in the list. There are three main sequential clustering algorithms which differ primarily on the choice of exponent a , and thus the distance metric. The k_T algorithm has $a = 2$, which allows low p_T clusters to dominate, the *anti* - k_T algorithm has $a = -2$, which allows high p_T clusters to dominate, and the Cambridge/Aachen algorithm has $a = 0$, removing momentum dependence from the distance metric entirely [98].

The second class of clustering algorithms is cone algorithms, which, as the name suggests, are based on the assumption that the particles in a jet will manifest in conical regions and thus group clusters in strict circular regions in η - ϕ space. SIScone is the only IRC safe cone algorithm, and is described in detail in [98]. Figure 3.13 compares the four main jet clustering algorithms when applied

to the same event with the same maximum radius R ; the shapes, sizes, and numbers of jets change with each algorithm.

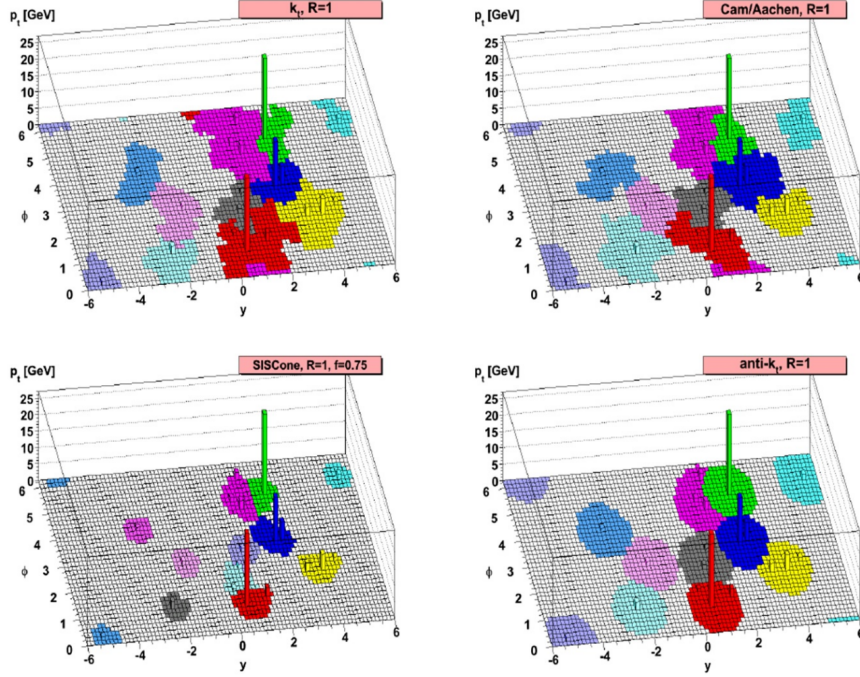


Figure 3.13: The four main jet clustering algorithms performed on the same data with the same input radius [98].

After clustering, jets may undergo additional corrections. To mitigate pile-up effects on momentum calculations, all jets are corrected by a standard p_T offset, determined in MC as a function of primary vertices, and an additional jet-area based correction. Additionally, all jets undergo energy and pseudorapidity calibration based on the relation of reconstructed to truth jets in MC samples. Jets may also undergo pruning, which removes sub-jets which contribute only a small fraction of energy, origin correction, and pile-up track subtraction [99].

After clustering and corrections, jets can be processed by a variety of tagging algorithms that exploit jet-substructure information and other kinematic variables to determine the likely initial par-

tile causing the hadronization. These algorithms are described, for example, in [100] for b-jets and [101] for top quarks and W bosons. There is substantial recent work regarding novel, ML based jet representation and tagging methods, which is discussed further in Chapter 4.

3.3.2.3 TAU LEPTONS

Tau leptons are produced in a range of physics processes, and in particular are of central importance to the $VH, H \rightarrow \tau\tau$ analysis presented in Chapters 6 and 7. In contrast to electrons and muons, taus have a relatively large mass (1.77 GeV) and a proper decay length of only 87 μm . Taus can decay leptonically ($\tau \rightarrow l\nu_l\nu_\tau$) or hadronically ($\tau \rightarrow \text{hadrons } \nu_\tau$), and in both cases the decay occurs before any interaction with the detector. Leptonic tau decays are therefore reconstructed using the electron and muon algorithms and dedicated tau algorithms focus only on hadronic tau decays. The hadronic decay products of taus are typically charged and neutral pions and occasionally kaons. In order to conserve the initial charge of the tau, there is always an integer number of charged tracks in a tau decay and hadronic taus are classified by the number of such track (referred to as substructure ‘prongs’) they contain. One prong decays represent 72% of all hadronic tau decays, and three prong decays represent an additional 22% [12].

The tau reconstruction algorithm begins with anti- k_T seeded jets with a p_T of at least 10 GeV. A vertex is assigned by choosing the track vertex with the largest fraction of momentum from tracks within $\Delta R < 0.2$ of the tau jet. A dedicated energy calibration, which considers the number of primary vertices in the event and the calibrated topocluster energies, is applied as described in [102]. This calibration functions effectively at high p_T , but is less accurate at low p_T . A newly developed energy calibration based on Boosted Regression Trees is applied to taus with $p_T < 100$ GeV, as described in [103].

The primary backgrounds for tau identification are jets of energetic hadrons produced by fragmented quarks and gluons. A Boosted Decision Tree (BDT, described further in Chapter 4) based

identification is used to distinguish hadronic taus from these backgrounds. The BDT utilizes cluster and track kinematics and secondary vertex information. Separate BDTs are trained for 1-prong and 3-prong taus. Additionally, reconstructed one-prong taus within $\Delta R < 0.4$ of a reconstructed and identified electron are rejected to further reduce backgrounds. The tau Working Group supports 3 BDT-based ID operating points to address a variety of analysis needs [104]. The tau reconstruction and ID efficiencies are shown in Figure 3.14. A Recurrent Neural Network (RNN, described further in Chapter 4) based tau ID is also being developed, although it was not fully utilized in Run 2.

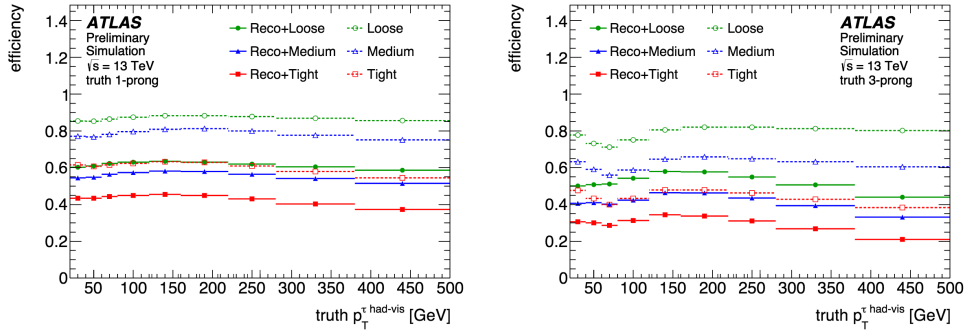


Figure 3.14: Efficiency for hadronic tau identification (open) and combined reconstruction and identification (closed) for one-prong (left) and three-prong (right) taus [104].

3.3.2.4 MISSING TRANSVERSE ENERGY

The law of conservation of momentum dictates that the net momentum in the transverse plane of a direct proton-proton collision should be zero. After reconstructing all tracks and objects in an event, any remaining momentum imbalance in the transverse plane is classified as missing transverse energy (MET). MET may be understood as indicating the presence of SM neutrinos, which are undetectable by ATLAS, or new BSM physics (also undetectable by ATLAS). However, it is important to also account for reconstruction and identification inefficiencies and other systematic effects.

The MET calculation is one of the more challenging reconstruction problems in ATLAS as it re-

quires input from all detector subsystems and all other reconstruction and identification algorithms. The calculation combines contributions from hard objects (fully reconstructed and identified particles and jets) and soft signals (reconstructed charged particle tracks associated with the hard scatter primary vertex). The basic equation for calculating the components of MET is

$$E_{x,(y)}^{miss} = - \sum_{i \in \{\text{hard objects}\}} p_{x(y),i} - \sum_{j \in \{\text{soft signals}\}} p_{x(y),j}.$$

This calculation requires that all contributing objects are reconstructed from mutually exclusive detector signals. Thus, overlapping reconstructed objects are rejected in a defined order; the most commonly used order is electrons, photons, taus, muons, jets, unused tracks, although this can be adjusted for different analysis requirements [105]. A demonstration of MET reconstruction performance is shown in Figure 3.15.

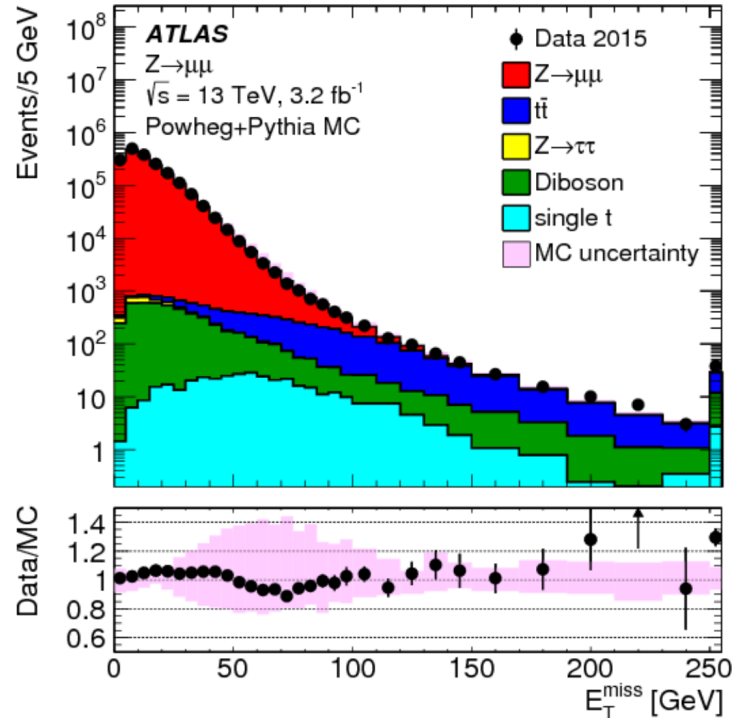


Figure 3.15: Distribution of MET for an inclusive sample of $Z \rightarrow \mu\mu$ events extracted from data and compared to MC [105].

Figure out who you are and do it on purpose.

Dolly Parton

4

Machine Learning

Machine Learning (ML) encompasses a variety of statistical techniques and models that allow algorithms to learn patterns and behaviors without being explicitly programmed. The study and development of these techniques has been on-going since the 1960s, however the subset of ML called Deep Learning (DL) which allowed effective training of very large, complex algorithms over high-dimensionality data-sets only emerged in 2012 [106]. Today, DL, and more broadly ML, is used to solve complex computing problems in a variety of fields including image processing ([107] [108]), language processing ([109] [110]), resource distribution [111], security [112], and content personalization [113].

In many ways, ML and physics share similar goals: building mathematical models to fit a set of observations and making predictions in related environments using those models. Given these

shared goals and the complex and intensive computing needs of LHC experiments (Chapter 3), it is no surprise that ML for HEP has evolved into its own field of research. This chapter describes the current status of these efforts: important ML concepts are defined in Section 4.1 commonly used algorithms are introduced in Section 4.2 and novel applications in HEP are presented in Section 4.3.

4.1 CENTRAL CONCEPTS

ML tasks can be separated into two classes. Supervised learning methods are used to map input data to a discrete set of labels or some continuous output distribution; they are developed using data-sets where the ground truth (i.e. the correct classification of a data point) is known. Unsupervised learning methods aim to infer natural structure within the input data and are developed without any ground truth information. Much of the current research on ML for HEP utilizes supervised learning and so the remainder of this chapter will focus only on supervised algorithms.

Supervised learning tasks typically take the form of a search for a function $f : X \rightarrow Y$ which maps the input data X (also called training data) to a lower-dimensional space of target labels Y with a primary goal of minimizing the loss function $L(y, f(x))$. The specific form of the loss function depends on the individual task, the selected algorithm, and the training procedure. Common, appropriate loss functions are included in the algorithm descriptions in the following sections.

In an ideal case, the ML algorithm would find the function f that minimizes L over all possible values of (x, y) . In practice, however, this is impossible with current computing architectures due to the high dimensionality of typical training data-sets and the enormous allowed function space. Thus, the algorithm training procedure seeks to minimize L over the training data-set values within a reduced function space f_ϕ defined by algorithm-specific parameters ϕ .

The performance of a trained ML algorithm can be evaluated using a range of criteria and selecting a task appropriate evaluation metric is a critical part of ML design. One of the most common

evaluation methods for classification algorithms is the Confusion Matrix [114], an $N \times N$ matrix where N is the number of class labels. The matrix values (n, m) represent the number of data points whose ground truth is n and predicted label is m . This leads to several easily commutable evaluation metrics like accuracy (percentage of correct predictions), class precision (the percentage of a predicted class whose ground truth is the same label), and class sensitivity (the percentage of a ground truth class whose predicted labels are correct). In binary classification tasks, the Confusion Matrix can be used to construct the Receiver Operating Characteristic curve (ROC curve) which plots the positive class sensitivity vs 1-negative class sensitivity. The area under the ROC curve (AUC-ROC) serves as an additional evaluation metric; the closer the AUC-ROC is to one, the better the performance.

The ultimate goal of a trained (supervised) ML algorithm is to perform well on new data not encountered during the training process; this is referred to as generalization and a failure to do this is called over-training to indicate that the algorithm has learned non-generalizable information present only in the training data. Over-training is often only identified by running the trained algorithm over a portion of the input data that was not used for training (called a test set); thus, test set validation is a critical component of ML model design. Test set validation can be incorporated directly into training through the method of cross-validation [115], where the algorithm is trained multiple times with a different randomized subset of training data (validation set) omitted each time; the validation set is used to quantify the error for that particular trained model and the final algorithm is an average of the various models.

Additionally, a variety of techniques can be used to reduce the risk of over-training, collectively referred to as regularization methods [116]. These are generally incorporated into the loss function formulation and penalize the algorithm for having large function parameters. They can also take the form of drop-out methods that randomly set some function parameters to zero. Increasing the size of the training data-set is another way to reduce the risk of over-training, but this increases the

required computing time to train the algorithm.

Algorithm optimization often involves tuning specific meta-variables called hyperparameters. These include the learning rate (what amount the function parameters are adjusted in each training iteration (epoch)), drop-out rate, regularization weight (how important the regularization goal is compared to other components of the loss function), initialization of the function parameters, and other task specific parameters. The most common method for tuning hyperparameters is a grid search where the model is trained multiple times where the hyperparameters are selected from an $N \times M$ matrix that gives n possible values for each of m hyperparameters.

4.2 ALGORITHMS

HEP utilizations of ML generally implement one of two algorithm types: Boosted Decisions Trees (BDTs) and Artificial Neural Networks (NNs). The basics construction and training procedures for each are outlined in the following sections. For all algorithms, the variables represented in the training data-set are referred to as features.

4.2.1 BOOSTED DECISION TREES

The fundamental component of a BDT is a decision tree as shown in Figure 4.1. Decision trees can be constructed algorithmically for class labeling tasks according to the following steps:

1. All data (containing examples of all classes) is equally weighted and input at the root node.
2. At each branching, all possible linear cuts on all features are considered and the cut that best separates one class from the rest of the data is selected.
3. Step 2 is continued until a stopping criteria is reached (typically when a node has fewer than x data points or the tree reaches some specified depth).
4. For each final node, the label of the the class with the most (ground truth labeled) data points in that node is assigned to all data points in that node.

Although simple to understand, an individual decision tree does not typically achieve high levels of accuracy.

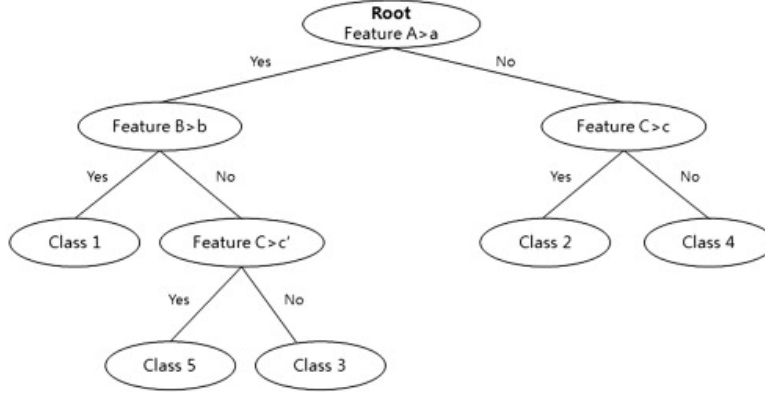


Figure 4.1: Depiction of an algorithmically constructed decision tree.

In ML, an ensemble method is an algorithm that iteratively combines weak learners to form a single strong learner. A BDT is then an ensemble method that applies ‘boosting’ to a ‘forest’ of decision trees. This method was first developed in 1995 in an effort to conceptualize decision trees as an ML optimization problem with a loss function, generally the mean squared error $L = \frac{1}{n} \sum_i (y_i - \hat{y}_i)^2$ for a data-set of size n with i truth labels y_i and learned labels $\hat{y}_i$ [117]. Boosting describes the training procedure of not retraining the individual tree but instead creating a new tree where the previously mis-classified data-points are given a higher weight in the root node. All the trees are then combined to create the final model and the loss function is evaluated over this ensemble; this process is demonstrated in Figure 4.2. There are a variety of different boosting methods, but gradient boosting, as described in [118], is the most common.

BDTs are preferred in some HEP ML tasks because they are relatively easy to construct, tend to perform well with little optimization, and are faster to train than most other algorithms. Additionally, the classification decisions are easy to understand, and least for each individual tree. On the other hand, BDTs are highly prone to over-training, particularly if the training data is high

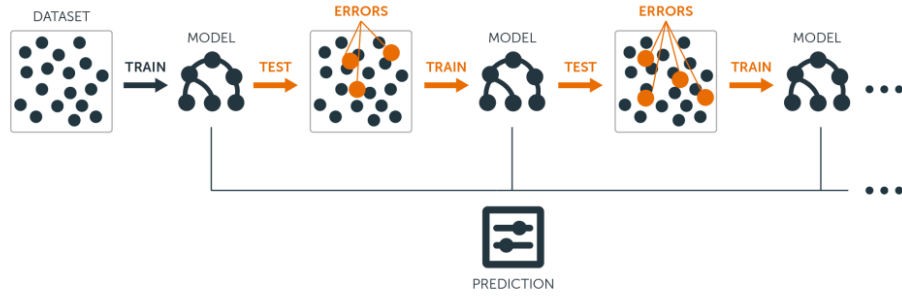


Figure 4.2: Schematic diagram of a boosted forest of decision trees [119].

dimensional. Consequently, developing BDTs for HEP tasks often necessitates intensive feature-engineering where physicists construct training variables to preserve maximal information with limited dimensionality. Additionally, there are several regularization methods which are applicable to BDTs including shrinkage (where the up-weighting of mis-classified data points is reduced as the number of trees grows).

4.2.2 NEURAL NETWORKS

NNs are ML algorithms whose data processing methods were inspired by the brain’s biological neural network. NNs consist of ‘artificial neurons’ called nodes arranged in a layered structure. Information is transmitted between layers through series of weighted connections (based on neural synapses). These connections can be represented as a matrix whose values are the connection weights. Learning appropriate weights to model the input data is the primary goal of NN training. A simple fully-connected NN architecture is shown in Figure 4.3; the first layer is composed of input nodes with one for each feature of the training data, the middle layers are called hidden layers, and the final layer is composed of output nodes which give the final results (e.g. predicted label in classification tasks) [120].

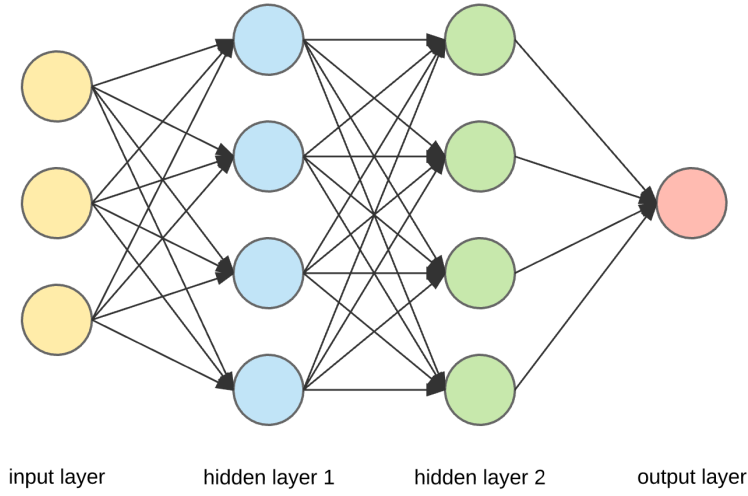


Figure 4.3: Schematic diagram of a simple fully-connected neural network.

In contrast to BDTs, NNs create a non-linear transformation of the input data because each node has an activation function. Like neural synapses, these activation functions allow information to be transmitted between nodes only if a certain threshold is reached. Mathematically the transformations between layers (h) is formulated as

$$h_{i+1} = g_i(W_i h_i + b)$$

where g_i is the activation function for that layer, W_i is the matrix of connection weights, and b_i is an optional vector of biases which would also be learned during training. Common activation functions include step functions, sigmoids, tanh, and Rectified Linear Units (ReLU) of the form $R(x) = \max(0, x)$.

Common loss functions for NNs include Mean Squared Error, Mean Squared Logarithmic Error, Mean Absolute Error, Cross-Entropy ($L = \frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$), and Negative Logarithmic Likelihood ($L = -\frac{1}{n} \sum_{i=1}^n \log(\hat{y}_i)$). The choice of loss function ultimately depends on the training task and the chosen activation function.

NNs are trained using a method called backpropagation [121]:

1. The NN weights are randomly initialized.
2. The training data is processed through the initial NN.
3. The predicted output values are compared to the ground truth values and the resulting error is calculated according to the the loss function.
4. The NN weights are adjusted to reduce the value of L .
5. Steps 1-4 are repeated until some stopping criteria is reached (typically number of epochs or minimal variation in L over several epochs).

The weight updates in the backpropagation training process is typically done using stochastic gradient descent:

$$w_{ij}(e + 1) = w_{ij}(e) - r \frac{\partial L}{\partial w_{ij}} - \zeta(e)$$

where e indicates the epoch, r is the learning rate and ζ is a stochastic term. This requires that the loss function is differentiable.

In practice, deep NNs (containing many hidden layers) often encounter the so-called ‘vanishing gradient problem’ [122]: as the error is backpropagated through the layers, the gradient rapidly approaches zero making it difficult to improve model performance. In recent years several computing innovations have been developed that allow for accurate training of very deep NNs. These include additional computing power from GPUs, dropout, and pre-training to initialize the network weights with e.g. autoencoders [123].

In addition to the basic structure outlined above, deep NNs can contain other modular differentiable components. The most common such components are convolutional filters and recurrent units, both of which are described below.

4.2.2.1 CONVOLUTIONAL NNs

The development of convolutional NNs (CNNs) was transformative for ML based image processing [124]. Training images are represented as 2D matrices of pixel values. The primary component of a CNN is a convolutional layer that applies a set of learnable filters as convolutions to the training data. Each filter is applied to a small portion of the input and then passed as a sliding window over the width and height (and depth in the case of 3D convolutions) of the input. This tiling allows the filter to learn information about the image regardless of rotations or distortions. An example of an individual convolutional filter is shown in Figure 4.4.

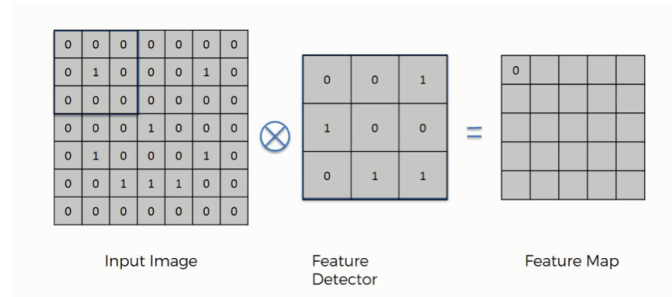


Figure 4.4: Diagram showing the application of a single convolutional filter to a single window of an input image [125].

In the case of image processing, certain filters (referred to as feature detectors) can apply human interpretable image transformations as shown in Figure 4.5. However, most learned convolutions are not meaningful to humans.

Mathematically, the transfer of information from a convolutional layer is expressed as

$$h_{ij} = g(k_j \cdot x_i + b_j)$$

where k_j is an individual filter and i indexes the patches k_j is applied to. Applying the same filter in patches over the entire input image has a similar effect to weight-sharing in a standard NN. This

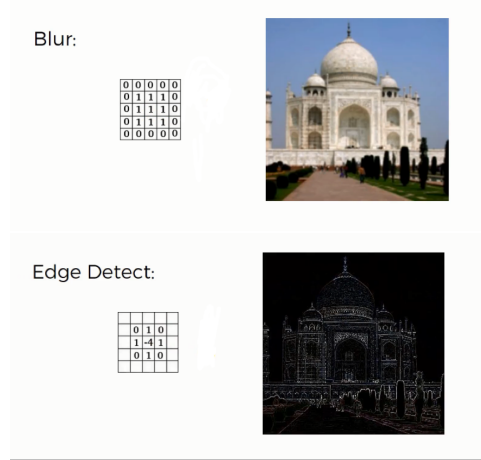


Figure 4.5: Two examples of interpretable convolutional filters for image processing [125].

reduces the number of learned parameters in the model and thus decreases training time. This also allows components of an input image to be recognized regardless of their position in the image.

Convolutional layers are typically followed by non-linear down-sampling layers called pooling layers, which reduce the dimensionality of the data by combining the outputs of nodes from the previous layer. Pooling is particularly important for image classification problems where the detection of individual image components often matters less than their spatial relation to other components. Pooling also reduces the memory usage and training time of CNNs. A common pooling method is max pooling where the maximum value in a certain window of the convolved input is passed to the next layer as shown in Figure 4.6.

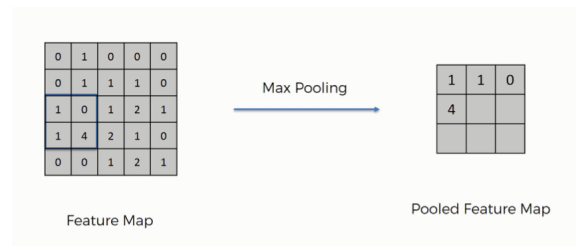


Figure 4.6: A graphical representation of max pooling [125].

CNNs usually contain many iterations of convolutional layers followed by pooling layers. This allows the model to learn a hierarchical representation of the training data moving from low- to mid- to high-level components. The series of convolutional and pooling layers can be followed by a flattening layer which compresses the convolved inputs into a 1D vector, and one or more fully connected layers that use the transformed input to complete the training task. An example of an entire image classification CNN incorporating all the components described above is shown in Figure 4.7. CNNs can be trained with backpropagation just as standard NNs.

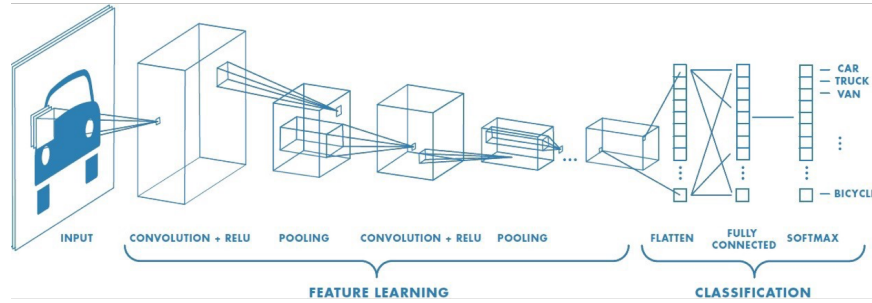


Figure 4.7: Schematic depiction of a full convolutional neural network built for image classification.

4.2.2.2 RECURRENT NNs

Neither standard NNs nor CNNs can handle variable length training data without altering certain data points by dropping information or adding dummy values. This functionality is critical for tasks like language processing or, in the case of HEP, allowing different numbers of clusters or tracks to contribute to object reconstruction. Additionally, neither type of network preserves order information of the inputs which is crucial for tasks like speech recognition or track-finding.

Recurrent nodes for NNs were developed in the 1980s to allow information preservation NN training and accommodate variable length inputs [126]. A basic recurrent node is described mathe-

matically as

$$h_i + 1 = g_i(W h_i + V h_{i-1} + b)$$

where W is the standard weight matrix connecting one layer to another and V is the recurrent weight matrix that allows processed data to be fed back into the same layer. h_{i-1} can represent either a previous feature or a previous time step. NNs containing one or more layers of recurrent nodes are referred to as Recurrent Neural Networks (RNNs); a simple, single layer RNN is shown in Figure 4.8.

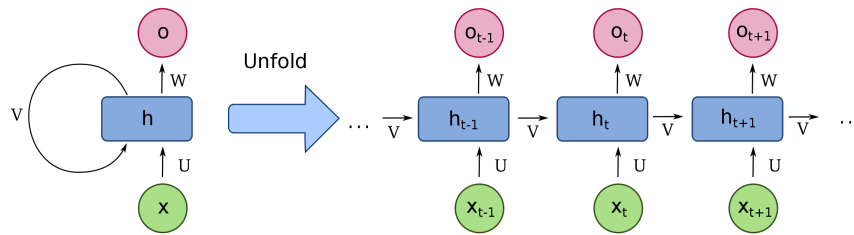


Figure 4.8: A single layer recurrent neural network unfolded to show a series of time steps in data training.

In practice, deep RNNs tend to experience vanishing or exploding gradients which make it computationally difficult to preserve information from distant input features or time steps. In the past 20 years, different techniques have been developed to allow the training of deep RNNs. The most common technique is introducing gating to the recurrent nodes which allow the transformation matrices to be applied selectively. Long Term Short Term Memory Units (LSTMs) and Gated Recurrent Units (GRUs) are popular gated nodes. LSTMs, shown in Figure 4.9, include three gates that control the flow of information withing the node. These are the input gate I , output gate O , and forget gate F . Each gate is parameterized by its activation function and its input and recurrent weight matrices W and V . GRUs are simplified LSTMs with no output gates.

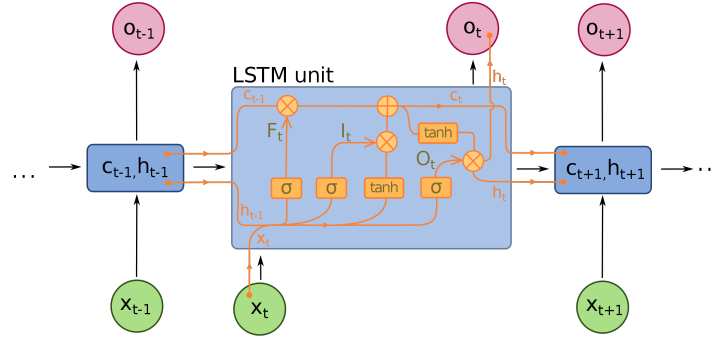


Figure 4.9: A single LSTM node across three time steps. In this example, sigmoid and tanh activation functions are used and c_i represents the previous state of information within the node.

4.2.2.3 ADVERSARIAL NETWORKS

Unlike CNNs and RNNs, adversarial NNs do not introduce new node types to the standard NN but rather combine two NNs into a single classifier. This is done by pitting two networks against each other in a non-cooperative ‘game’. The primary network receives the training data as input and attempts to complete the primary algorithm task (e.g. classification). The adversary network receives the output of the primary network as input and attempts to predict something else (Figure 4.10). The two networks are trained together by combining their loss functions as $L = L_f(X) - L_r(f(X), X)$ where f refers to the primary network and r refers to the adversary. Typically optimizing fully for both objects is impossible and a hyperparameter λ is introduced to scale the importance of the adversary task in the loss function.

Adversarial neural networks are commonly used to generate new data. In this case, the primary task is generating a new data point and the adversary task is predicting if the input data is real or generated. These types of networks are called Generative Adversarial Networks (GANs). Another common use of adversarial neural networks is to force a classifier to be robust to a certain feature of the training data which may not be explicitly included, but can nonetheless be constructed from the

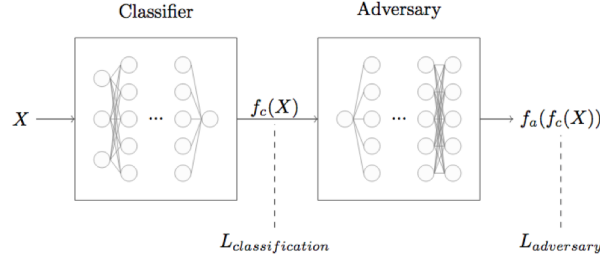


Figure 4.10: An example adversarial neural network where the primary task is classification [127].

input features. In this case, the primary task is classification while the adversary task is predicting the feature of concern.

4.3 APPLICATIONS IN LHC PHYSICS

Many computing tasks at various stages of the LHC computing flow (Chapter 3) rely on large volumes of high dimensionality data, which makes them appropriate use cases for a variety of ML algorithms. Implementing these algorithms come with the additional benefit of being able to utilize information that traditional cut-based methods cannot (i.e. variables whose distributions overlap and low-level information like detector outputs). ML can also reduce algorithm dependency on systematics like pile-up or simulation mis-modeling.

Furthermore, many ML algorithms, once trained, are faster to run than cut-based methods. Consequently, LHC physicists have already found many uses ML in detector trigger systems. This area of study is of particular importance for the development of the HL-LHC. Excellent surveys of the current status of ML for HEP research can be found in [128], [129], and [130] and primary LHC-based research areas are described below.

4.3.1 RECONSTRUCTION, IDENTIFICATION, AND CALIBRATION

Particle and event reconstruction, identification, and calibration is the focus of a great deal of LHC-based ML research. Many of these tasks can be formulated as classification problems making them excellent candidates for supervised learning models.

4.3.1.1 BDTs AND STANDARD NNs

Many of the cut-based methods originally developed for these tasks make use of physics-motivated constructed variables, making them a natural use case for BDTs and NNs. Additionally, these methods are relatively easy to understand and develop; thus they were the focus of initial LHC ML research.

An excellent example of BDTs and NNs for particle ID the study of W- and top-jet tagging in ATLAS [131]. Here, W-jets and top-jets refer to jets originating from Ws or tops that are ‘fat’ (large radius) or boosted so as to be reconstructed as a jet rather than individual decay products. The training data-set was constructed by reconstructing jets in MC samples with the standard anti- k_t clustering, reconstructing ‘truth-jets’ using truth-level information of long-lived particles, and then matching the clustered jets and truth-jets to form a labeled data-set. The BDTs and NNs were trained using standard physics-motivated jet substructure variables where the final set of variables was chosen through an iterative process. Both algorithms were optimized with a hyperparameter grid search and tested for over-training with cross-validation. As shown in Figure 4.11, both the BDTs and NNs outperform the previous cut based method. Similar studies for other ID tasks are summarized in [132].

Track-reconstruction for LHC experiments is highly CPU and time intensive. ML has already found some use in current tracking methods; for example, in ATLAS NNs are used to assign track energies in the case where multiple tracks pass through the same pixel cluster [133]. Nonetheless, the

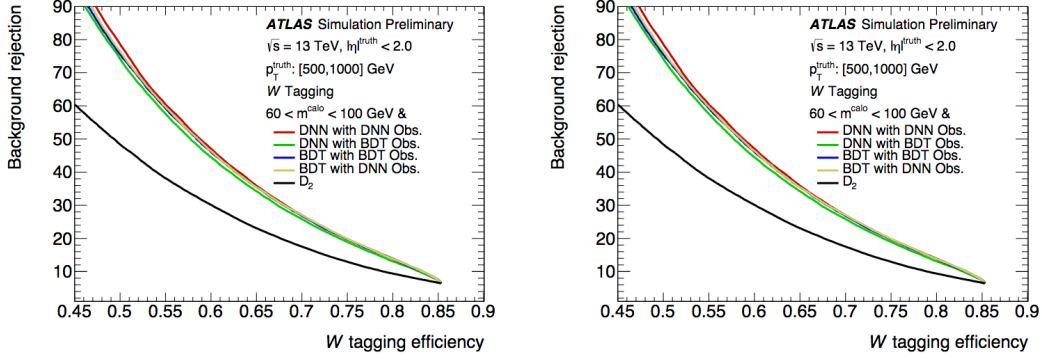


Figure 4.11: Performance of BDTs and deep NNs (DNNs) for W-jet tagging (left) and top-jet tagging (right). For both plots the black curve represents the standard cut-based method [131].

most CPU intensive portion of track finding is the initial track candidate seeding. As described in Chapter 3, in ATLAS this is currently done with a Kalman filter; unfortunately this is impractical in terms of computing needs and decreasing accuracy in the high pileup environment of the HL-LHC. Many on-going research projects seek to solve this problem, including crowd-sourced computing challenges like [134].

Adjusting the weights of MC events to better match data distributions and correcting the assigned energy of reconstructed objects are important tasks in LHC computing flows. Typically this is done by calculating a re-weighting factor by comparing data and MC in the first case or reconstructed and truth jets in the latter case. Studies like [135] have shown that using BDTs for MC re-weighting is both faster and more accurate; in fact, discriminators have a more difficult time distinguishing between ML re-weighted MC and data than between data and previous MC re-weighting techniques. BDT-based calibration is implemented for many physics objects in ATLAS (e.g. electrons and photons [136]) with excellent performance.

4.3.1.2 IMAGES AND CNNs

In 2014, the process of representing jets as images was developed [137]. Cells or towers in a detector calorimeter are represented as individual pixels and the value of the pixel is the amount of energy deposited in that cell/tower. In practice, this representation requires some amount of pre-processing to account for detector geometry and image sparsity. This typically involves noise reduction, ROI finding to select a certain window of cells to represent, cell geometry normalization, and alignment to exploit detector $\eta - \phi$ symmetry. Additionally, most jet images studies have focused on jets only in the barrel region to avoid the granularity differences between the barrel and endcap calorimeters. A composite of several pre-processed W-jet images is shown in Figure 4.12.

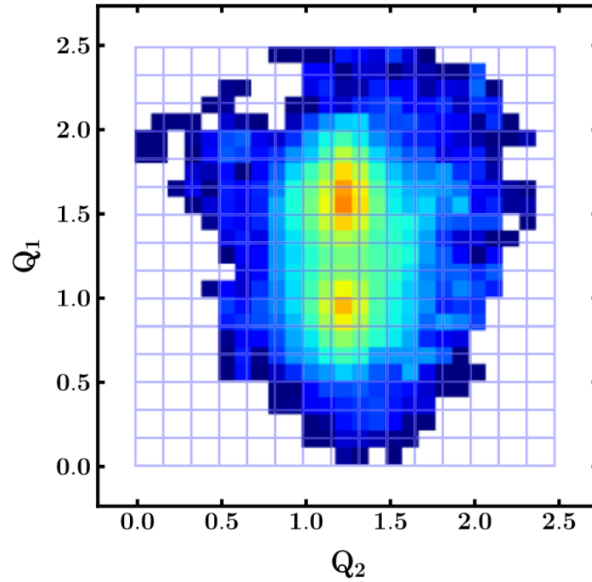


Figure 4.12: Average of several pre-processed W-jet images. Here the Q_1 and Q_2 axes represent the transformed $\eta - \phi$ space [137].

There are numerous studies of image-based jet classification using CNNs [132]. One of the most successful studies sought to distinguish W-jets from background QCD-jets [138] using a CNN with

ReLU activations and a ‘Max-Out’ network that consisted of two max pooling layers and two fully connected layers. Both networks were trained with 8 million samples and an additional 2 million test samples generated using Pythia and pre-processed as described above. Additionally, both networks were optimized using a limited hyperparameter grid search. As shown in Figure 4.13, both network types outperformed cut-based combinations of standard jet substructure variables.

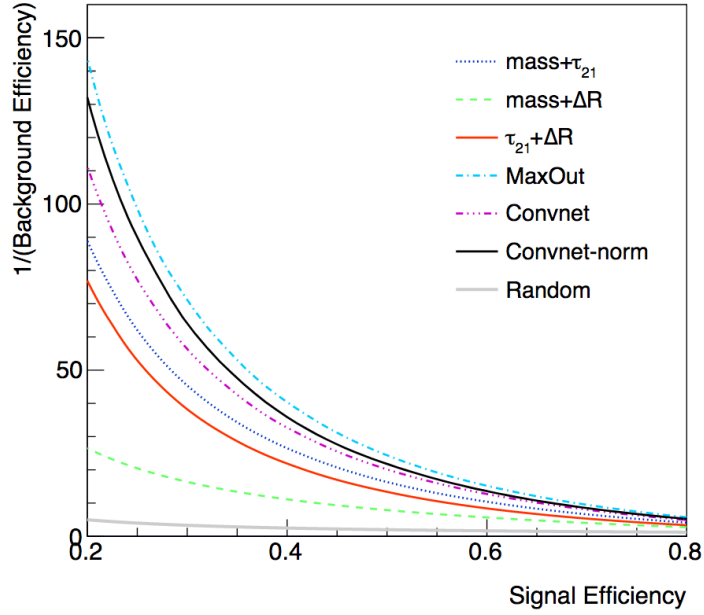


Figure 4.13: Performance comparison of CNNs (purple dashed and black curves), a Max-Out network (blue dashed curve) and cut-based discriminators using jet substructure variables (red, blue dotted, and green dashed curves) on the W- vs QCD-jet discrimination task [138].

This method is also being studied for electron ID in the EM calorimeter as described in Chapter

4.3.1.3 RNNs

RNNs have been used in a variety of particle ID and tagging applications at the LHC. These methods are advantageous because they allow the input of variable length track and cluster lists, thus allowing maximal information preservation. Successful examples include boosted top-tagging [139], ATLAS b-tagging [140], and the new ATLAS RNN tau ID and trigger [141].

Language processing inspired RNNs have found uses in jet clustering [142]. This method relies on the physically motivated assumption that the particles produced in a jet should be formed in some QCD-dictated order. In this application, the RNN takes the 4-momenta of particles in an event as input and recursively learns an embedded representation of the jets in the event that is then classified by a fully-connected NN (Figure 4.14). In the same W- vs QCD-jet tagging task described in Section 4.3.1.2 jets constructed and classified with this method achieved the same accuracy as the Max-Out network but required substantially less training time.

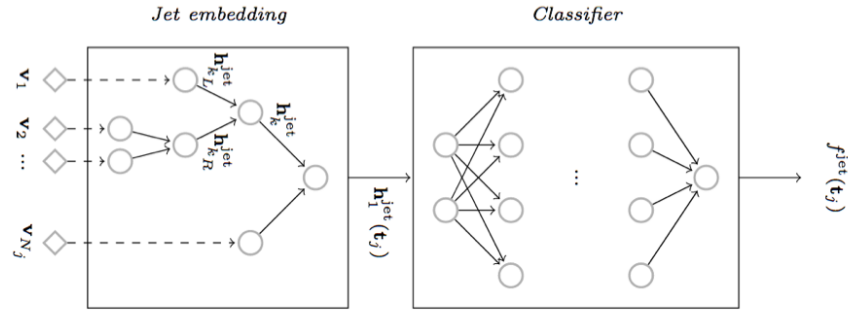


Figure 4.14: Schematic of QCD-motivated RNN-based jet-clustering and classification [142].

4.3.2 ANALYSES

ML, particularly NNs, has been utilized to improve the event selection efficiency and background rejection for a range of LHC analyses. One of the earliest studies used NNs to select events containing different exotic particles. This study compared a shallow NN trained on ‘high-level’ features (i.e. reconstructed boson masses, MET, etc) to a deep NN trained on the same high-level features and a deep NN trained on ‘low-level’ features (i.e. individual particle p_T , number of jets, etc) [143]. As shown in Table 4.1, the deep NN trained on low-level features outperformed the networks trained on physics-motivated variables, indicating that some information is lost in the construction of these variables.

Technique	AUC		
	Low-level	High-level	Complete
BDT	0.73 (0.01)	0.78 (0.01)	0.81 (0.01)
NN	0.733 (0.007)	0.777 (0.001)	0.816 (0.004)
DN	0.880 (0.001)	0.800 (< 0.001)	0.885 (0.002)
Technique	Discovery significance		
	Low-level	High-level	Complete
NN	2.5σ	3.1σ	3.7σ
DN	4.9σ	3.6σ	5.0σ

Table 4.1: Performance for the various event selection algorithms trained to select exotic Higgs events. NN refers to the shallow NN, DN refers to the deep NN, and complete means the combination of high-level and low-level features [143].

Deep NNs have been used in various other LHC searches and analyses. However, these methods do not always yield higher signal selection efficiency or improved background rejection. The usefulness of these methods depends on the specific analysis and is an on-going area of research.

4.3.2.1 $ZZ_d \rightarrow lll$ SEARCH

ML event classification has also been studied to improve the selection efficiency in the ATLAS Run-2 $H \rightarrow ZZ_d \rightarrow lll$ search. This search targets a massive mediator of the hypothetical $U(1)_d$ dark gauge symmetry extension of the SM [144]. In contrast to related dark photon searches that rely on measuring displaced vertices or MET [145], the massive nature of the Z_d allows the use of the Higgs as a portal to this dark sector. This analysis selects $H \rightarrow ZZ^* \rightarrow lll$ events and fits the Z^* reconstructed mass spectrum. If no Z_d boson exists, this mass spectrum would be flat, but if the Z_d does exist, there will be an excess at the mass of the Z_d .

Efficiently selecting $H \rightarrow ZZ^* \rightarrow lll$ events is thus critical for increasing the sensitivity of this search. An initial Run-2 study explored using BDTs and NNs for event selection in an effort to improve efficiency. An extensive variable optimization study was conducted to select the training features; the final BDT was trained with p_T^{4l} , η^{4l} , and $p_T^{Z_2}$. The NN was trained with the same variables as the Run-1 cut-based analysis and a grid search over all hyperparameters was used to optimize the network architecture. The potential performance of a Run-2 cut-based selection was modeled using the rectangular cut functionality of TMVA built with the Run-1 variables. As shown in Figure 4.15, both ML methods out-performed the cut-based method.

4.3.3 SIMULATION

As described in Chapter 3, detector simulation is the most computationally intensive component of the ATLAS simulation chain. ML, particularly GANs, have been shown to help speed up this process for some detector-object interactions by generating images like those introduced in Section 4.3.1.2. In these methods, the primary task is generating a new detector image and the adversary task is discriminating between images generated with the full detector simulation and images generated by the primary network. A common issue when training GANs is so-called ‘mode-collapse’ where

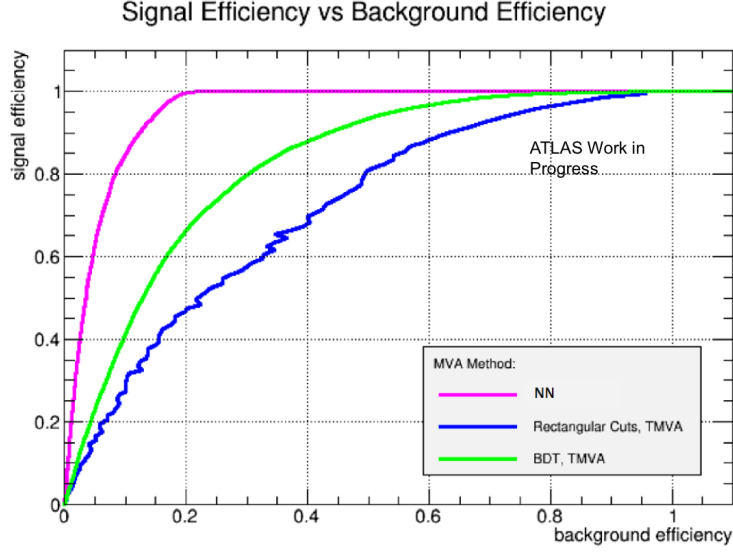


Figure 4.15: ROC-curve comparison of $H \rightarrow ZZ^* \rightarrow llll$ event selection algorithms performance.

the generator learns a small feature that is maximally confusing to the adversary and thus does not produce a variety of simulated examples. Mode-collapse can be alleviated by introducing a third network trained to complete some auxiliary task and incorporating its error into the overall loss function.

Location-aware GANs (LAGANs), which introduce locally connected layers to preserve location information, have been successfully used to generate jet images [146]. This implementation included an auxiliary task of distinguishing W-jets from QCD-jets. As shown in Figure 4.16, the LAGAN produced images with kinematic information that closely matched those produced by Pythia. Furthermore, the LAGAN simulation was an order of magnitude faster than the Pythia simulation. Reducing the computing load for MC simulation is critical to ensure successful physical analyses in the HL-LHC. A similar method was successful in generating electromagnetic calorimeter shower images [147].

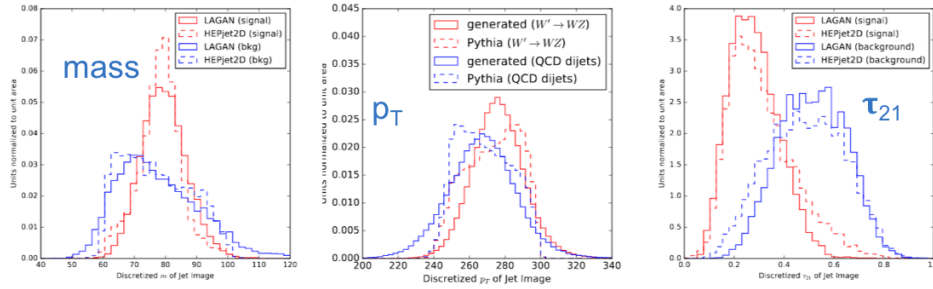


Figure 4.16: Comparison of LAGAN generated (solid lines) and Pythia generated (dotted lines) jet kinematics (jet mass (left), (jet p_T (center), and n-subjettiness (right)) for signal (W-jets, red lines) and background (QCD jets, blue lines) [146].

4.3.4 SYSTEMATICS

A central challenge in many LHC analyses is robustness of results to systematic uncertainties and changing conditions. Adversarial networks can also be used to impose physics motivated constraints (such as reducing pileup dependence or decorrelating classification results from some variable) on NN-based classifiers. In these cases, the primary task is the specific classification and the adversary task is to reproduce the feature of interest. This can be formulated mathematically by rewriting the underlying model of features X and labels Y as $p(X, Y, Z)$ where Z is a nuisance parameter and Y now depends on both X and Z . The goal of the network is then to find a function $f_\phi(X) = Y$ that is robust to the unknown value of Z . This typically results in a less accurate classifier, but has the advantage of reducing uncertainty on the final measurement.

The first use of this method for LHC physics used a simplified quantization of pileup as the nuisance parameter; Z could take on two discrete values: $Z = 0$ for no pileup and $Z = 1$ for a pileup of 50 [127]. Figure 4.17 shows the approximate median significance (AMS) of the final classification as a function of the selected decision threshold for a variety of adversary importance weights (λ). The network with $\lambda = 10$ achieves the highest significance, illustrating the advantages of some

sacrifices in classifier accuracy for reducing pileup dependence.

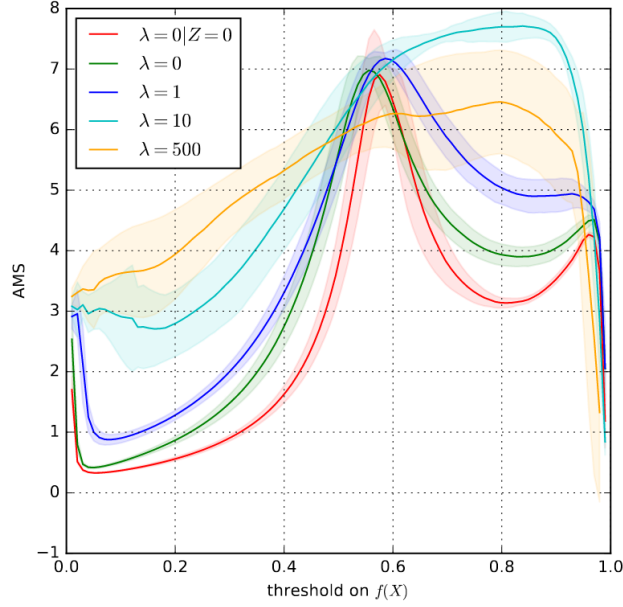


Figure 4.17: Approximate median significance as a function of decision threshold for networks trained with different values of λ [127].

Another study sought to decorrelate the output of jet classifiers from jet mass [148]. This is useful because many current jet taggers distort the jet mass distribution which increases the uncertainty of background models in many analyses thereby decreasing the final result significance. In this study, the primary task was distinguishing W-jets from QCD-jets and the adversary task was reproducing the jet mass. This study also found improved final significance when sacrificing some classification power in favor of reducing jet mass dependence, even when compared to non-ML methods for preventing distortions of the jet mass spectrum.

It is clear that there are substantial opportunities to improve physics results at the LHC using ML. The following chapters detail additional studies applying ML to electron ID and the $VH, H \rightarrow \tau\tau$ analysis.

There's always one more way to try things and that's your way, and you have a right to try it.

Waylon Jennings

5

Electrons in ATLAS

RECONSTRUCTING AND IDENTIFYING ELECTRONS IN THE ATLAS DETECTOR IS ESSENTIAL TO MEASURING MANY IMPORTANT PHYSICS PROCESS WITH LEPTONS IN THE FINAL STATE. The reconstruction and identification algorithms make use of characteristic detector signatures including tracks in the inner detector and energy deposits in the electromagnetic calorimeter.

The electron identification algorithm is employed in two phases in the ATLAS computing flow. A simplified version is first used in the online HLT system to identify electron candidates as trigger objects. The full version is used in offline data processing to create final analysis objects. Both implementations of the algorithm serve to distinguish prompt, isolated electrons from background objects; these include hadronic jets and non-prompt electrons from photon conversions and heavy

flavor decays. The inputs to both algorithms are particle candidates consisting of clusters in the electromagnetic calorimeter matched to tracks in the Inner Detector.

5.1 ELECTRON IDENTIFICATION (ID) SOFTWARE

The current electron ID algorithm (described in detail in [149]) uses physics-motivated variables constructed from a combination of calorimeter shower shape information, inner detector track information, and track cluster matching information. Figure 5.1 shows how an electron typically travels through the ATLAS detector to create these variables. On average an electron will hit the IBL pixel layer, 3 other pixel layers, 4 silicon-strips, and 30 TRT straws; it will then pass through the solenoid and deposit energy in some or all of the 4 layers of the electromagnetic calorimeter. A small amount of energy may also be deposited in the hadronic calorimeter.

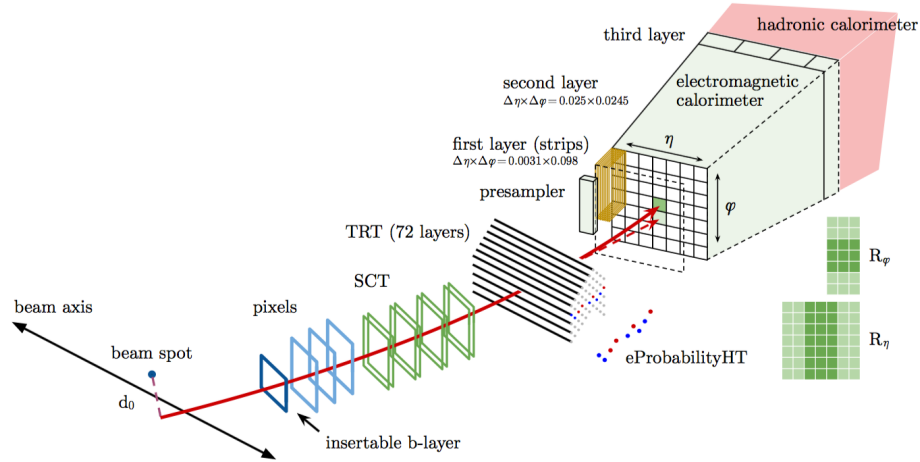


Figure 5.1: Schematic depiction of an electron traversing the ATLAS detector.

5.1.1 THE LIKELIHOOD FUNCTION

The electron ID algorithm is, at the time of this dissertation, a Likelihood (LH) method; LHs are Multivariate Analysis Techniques used to separate signal from background. This method proceeds in three steps. First, signal and background Probability Density Functions (PDFs) of a set of n variables are constructed from purified samples. The PDFs can be constructed from data or MC simulation depending on the use case; these methods are discussed further in Section 5.1.2. These PDFs can be written as $P_{s,i}(x_i)$ ($P_{b,i}(x_i)$), where s indicates signal and b indicates background. P is the value of the signal (background) PDF of the i_{th} variable evaluated using object x . After PDF construction, information from all n variables is combined and all objects are given a discriminant score d_L according to the equation:

$$d_L = \frac{L_S}{L_S + L_B} \quad L_S(x) = \prod_{i=1}^n P_{s,i}(x_i) \quad L_B(x) = \prod_{i=1}^n P_{b,i}(x_i).$$

Finally, a cut value on this discriminant is selected and objects which pass this cut are classified as electrons. For the electron LH, the signal and background discriminants peak sharply at 1 and 0 respectively, and so an additional transformation function $d'_L = \tau^{-1} \ln(d_L^{-1} - 1)$ is applied to further spread the distributions. The value of τ is variable and determines the width of the applied spread; τ is set to 15 for all electron LH calculations. An example of the transformed discriminants can be seen in Figure 5.2.

A main benefit of the LH method is that it is possible to include variables whose signal and background PDFs overlap in such a way that traditional cut-and-count methods could not be applied, but nonetheless posses discriminating power due to shape or other differences. One such variable, f_1 , which describes how electrons typically deposit energy in the calorimeter, is shown in Figure 5.3. Additionally, because the LH combines many variables it is able to recover objects from the tails of certain distributions that look reasonable in other PDFs. Together, these effects lead to background

rejection up to 2 times more efficient than the Run 1 cut-based algorithm (see Figure 5.4).

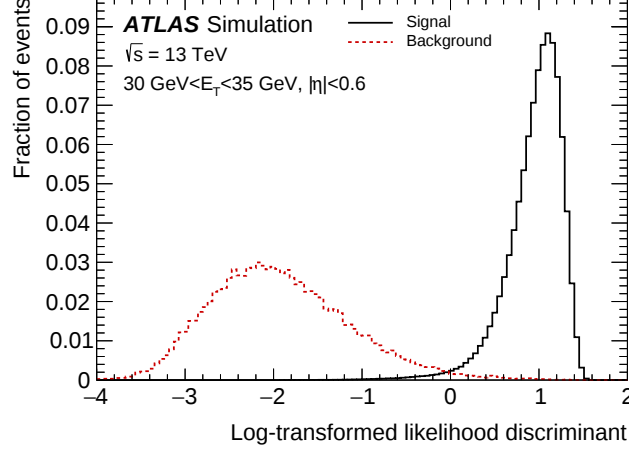


Figure 5.2: The transformed LH-based identification discriminant d'_L for reconstructed electron candidates. The black histogram is for prompt electrons in a $Z \rightarrow ee$ simulation sample, and the red (dashed-line) histogram is for backgrounds in a generic two-to-two process simulation sample (both simulation samples are described in Section 5.1.2). The histograms are normalised to unit area.

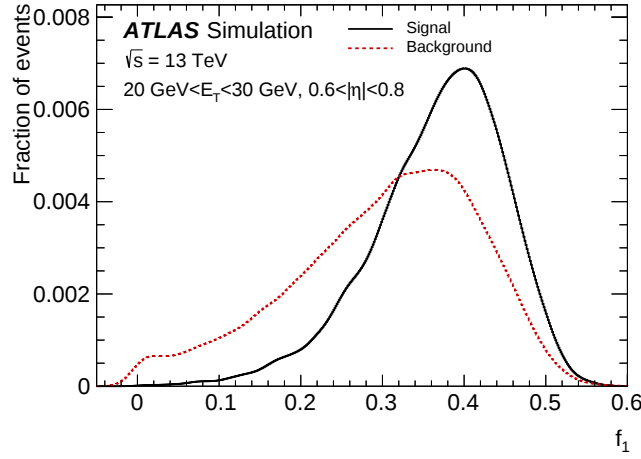


Figure 5.3: Example of a distribution of an electron variable - f_1 described in Table 5.1 - that would be inefficient if used in a cut-based ID but improves the LH based ID [150].

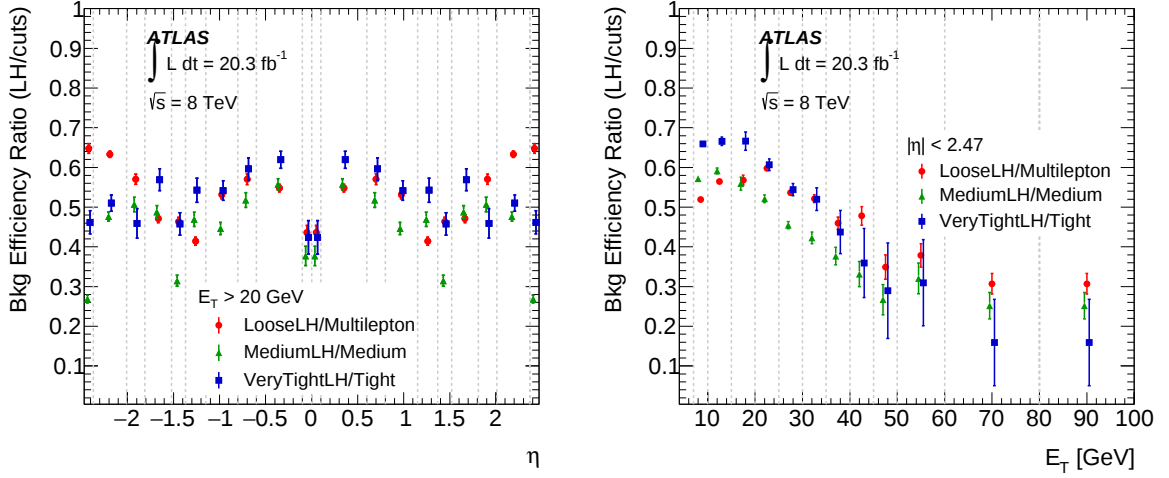


Figure 5.4: Ratio of background efficiencies for a LH based algorithm to that of the closest-efficiency cut-based selections as a function of η (left) and E_T (right) [151].

During the original Run 1 development of the LH-based ID, other ML algorithms including BDTs and NNs were considered. Ultimately, the LH technique was selected due to its simplicity, interpretability, and processing speed [152]. In order to improve the electron ID efficiency for full Run 2 analyses and future Run 3 and HL-LHC analyses, more complex ML techniques are being studied. This is described further in Section 5.3.

5.1.1.1 ELECTRON LIKELIHOOD

The electron LH ID uses a combination of up to 19 detector related variables. The motivation for these variables is described below, and a summary can be found in Table 5.1.

R_{had} , the ratio of energy deposited in the hadronic calorimeter to energy in the electromagnetic calorimeter distinguish electrons from hadrons based on how far they traverse through the detector. Calorimeter shower width variables w_{tot} , $w_{\eta 2}$, R_{η} , and R_{ϕ} further separate electrons and hadrons as hadrons typically have much wider showers than electrons. Additionally, the other calorimeter

Table 5.1: Definitions of electron discriminating variables, the types of backgrounds the variables help to discriminate against (light-flavor (LF), converted photons (γ), or heavy-flavor (HF)), and if a variable is used as a likelihood PDF ($\mathcal{L}$) or used as a rectangular cut (C). The * refers to the fact that the E/p and w_{stot} variables are only used for electrons with $p_T > 150$ GeV for the Tight identification operating point, and are not used for the looser operating points.

Type	Description	Name	Rejects			Usage
			LF	γ	HF	
Hadronic leakage	Ratio of E_T in the first layer of the hadronic calorimeter to E_T of the EM cluster. (used over the range $ \eta < 0.8$ or $ \eta > 1.37$)	R_{had1}	x		x	$\mathcal{L}$
	Ratio of E_T in the hadronic calorimeter to E_T of the EM cluster. (used over the range $0.8 < \eta < 1.37$)	R_{had}	x		x	$\mathcal{L}$
Back layer of EM calorimeter	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter. This variable is only used for $p_T < 80$ GeV due to known inefficiencies at high p_T , and is also removed from the LH for $ \eta > 2.37$, where it is poorly modeled by the MC.	f_3	x			$\mathcal{L}$
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\sum E_i \eta_i^2)/(\sum E_i) - ((\sum E_i \eta_i)/(\sum E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	$w_{\eta 2}$	x	x		$\mathcal{L}$
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centered at the electron cluster position	R_ϕ	x	x	x	$\mathcal{L}$
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position	R_η	x	x	x	$\mathcal{L}$
Strip layer of EM calorimeter	Shower width, $\sqrt{(\sum E_i (i - i_{\max})^2)/(\sum E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η , and $i_{\max}$ is the index of the highest-energy strip	w_{stot}	x	x	x	C*
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	E_{ratio}	x	x		$\mathcal{L}$
	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter	f_1	x			$\mathcal{L}$
Track conditions	Number of hits in the innermost pixel layer; discriminates against photon conversions	n_{Blayer}		x		C
	Number of hits in the pixel detector	n_{Pixel}		x		C
	Number of total hits in the pixel and SCT detectors	n_{Si}		x		C
	Transverse impact parameter with respect to the beam-line	d_0		x	x	$\mathcal{L}$
	Significance of transverse impact parameter defined as the ratio of d_0 and its uncertainty	$ d_0/\sigma_{d_0} $		x	x	$\mathcal{L}$
	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum	$\delta p/p$	x			$\mathcal{L}$
TRT	Likelihood probability based on transition radiation in the TRT	$eProbability_{HT}$	x			$\mathcal{L}$
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	$\Delta\eta_1$	x	x		$\mathcal{L}$
	$\Delta\phi$ between the cluster position in the middle layer of the calorimeter and the track extrapolated from the middle layer to the perigee, where the track momentum is rescaled to the cluster energy before extrapolating	$\Delta\phi_{res}$	x	x		$\mathcal{L}$
	Ratio of the cluster energy to the track momentum	E/p	x	x		C*

depth variables f_1 and f_3 describe shower shapes as the object proceeds through the electromagnetic calorimeter. E_{ratio} checks for multiple incident particles by comparing the difference in the two largest energy clusters in the silicon strip layer of the electromagnetic calorimeter to their sum.

Variables d_0 and $|d_0/\sigma d_0|$ describe the object's transverse and longitudinal impact parameters which helps distinguish possible electrons from b- and c-jets. $\Delta p/p$ characterizes an object's bremsstrahlung radiation loss and separates electrons from charged hadrons which lose much less energy in the Inner Detector.

Finally, the variable $eProbabilityHT$ exploits transition radiation information from the TRT. Light objects like electrons radiate more photons in the TRT than heavier objects like pions and muons, and the radiated photons produce additional high-threshold (HT) hits in the TRT. In Run 1, the variable F_{HT} , or the ratio of HT hits to total TRT hits was used in the electron LH. Leaks in the TRT gas system necessitated replacement of several xenon modules with argon modules and resulted in a decreased probability of HT hits in the TRT. Consequently, a tool was developed to regain the information from the original F_{HT} variable. The HT probability of each TRT hit is calculated using the electron hypothesis and pion hypothesis separately. $eProbabilityHT$ describes the ratio of these two probabilities. The development and implementation of this tool is described in further detail in [153].

The distributions of the variables described above vary with both particle E_T and η . Consequently, PDFs and LHs are constructed separately for a 10x9 set of p_T and η bins. The η bin limits are taken from the detector geometry and the p_T bin limits were set after a study of the rate of shower-shape change. The bin definitions are shown in Tables 5.2 and 5.3.

The LH ID is used both in offline data processing and online for HLT electron triggers. The trigger LH differs from the offline algorithm described above in a few small ways. First, the $\Delta p/p$ variable is not used in the HLT because the track-fitting algorithm is too computationally intensive. Furthermore, impact parameter variables d_0 and $|d_0/\sigma d_0|$ are not used to allow analyses which rely

Bin boundaries in p_T [GeV]								
PDFs	4.5	7	10	15	20	30	40	∞

Table 5.2: Electron transverse energy binning used for the electron likelihood PDFs and discriminant cut values.

Bin boundaries in $ \eta $									
0.0	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47

Table 5.3: Electron pseudorapidity binning used for the electron likelihood PDFs and discriminant cut values.

on leptonically decaying τ leptons or other exotic particles (such as the $VH, H \rightarrow \tau\tau$ described in Chapters 6 and 7) to also utilize the HLT electron triggers. Finally, PDFs for the HLT LH are constructed separately due to different variable resolutions compared to the offline ID. An additional HLT LH ID exists which uses only calorimeter variable information. This ID is used as a highly efficient preselection before track reconstruction algorithms are run; events which fail this preselection do not undergo track finding, reducing the amount of computational resources needed.

5.1.1.2 OPERATING POINTS

The electron ID LH is used to support five operating points which provide ATLAS analyses flexibility to balance signal selection efficiency and background rejection. These operating points are referred to as Very Loose, Loose, LooseAndBLayer, Medium, and Tight. Background rejection efficiency increases from the VeryLoose to Tight operating points and each operating point is a subset of all looser operating points.

The LH discriminant cut defining each operating point was initially selected in Run 1 to reproduce the efficiency of the original cut-based electron ID. Subsequent retunings and developments of the LH algorithm have sought to maintain roughly these same efficiencies while further improv-

ing background rejection. The VeryLoose operating point was added in Run 2 primarily for use in background estimation techniques that benefit from a reduced electron ID (such as the Fake Factor Method described in Chapter 7).

In addition to passing the LH discriminant cut for each operating point, electron candidates are also required to pass good track-quality requirements of at least one pixel hit and at least seven silicon hits to be classified as an electron. The Loose, Medium, and Tight operating points require an additional pixel hit to reduce the contribution of converted photons. To further remove conversions, the Medium and Tight operating points require a hit in the inner-most pixel layer, the IBL. The LooseAndBLayer operating point was designed to include this additional requirement while maintaining the Loose operating point efficiency. The efficiency of the Loose, Medium, and Tight operating points for 2017 data and MC can be seen in Figure 5.5. The efficiency is slightly higher for MC due to the tuning procedure which incorporates pile-up profiles derived from MC as described in [149]. This discrepancy is accounted for the the data to MC correction scale factors.

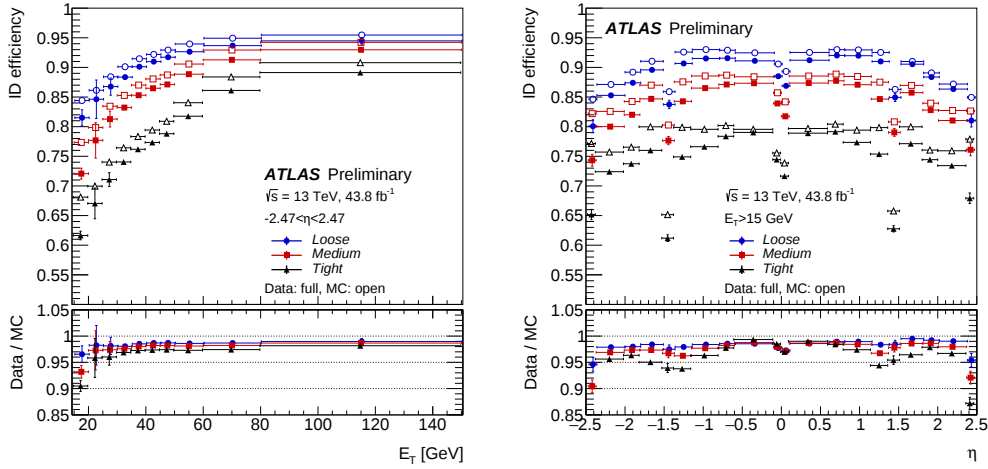


Figure 5.5: Electron identification efficiencies for $Z \rightarrow ee$ events as a function of E_T (left) and pseudo-rapidity (right). The efficiencies are shown for data (full) and MC (open) for three LH based operating points [154].

5.1.2 PROBABILITY DENSITY FUNCTIONS (PDFs)

The PDFs used to build the LH ID are constructed from either MC of relevant processes or purified data samples. An individual PDF is created by building a histogram of a single variable for the entire MC or data sample; all variables used are described in Section 5.1.1.1. Signal PDFs are constructed using the Tag-and-Probe method described below, while background PDFs are constructed from objects which pass one of the electron/photon supporting triggers. Additional cuts ($M_T < 40$ GeV, $MET < 25$ GeV) are included to remove any objects resulting from W decays, and an invariant-mass < 40 GeV or > 140 GeV cut is included to remove contributions from Z decays.

5.1.2.1 TAG & PROBE METHOD

In order to increase purity in the signal electron ID PDFs while maintaining unbiased electron candidates, a Tag & Probe method is used to construct the final PDFs from the original sample. With this method, developing a new set of ID operating points relies on the previous iteration.

All ID PDFs are constructed using electrons from processes where an object (a Z boson or a J/ψ meson) decays into two electrons. After event level selections for the relevant process are applied, one of the electrons (the “tag”) is required to pass the Tight operating point. A second electron in the event which, when combined with the tag electron, forms an invariant mass within 10 GeV (± 3 GeV) of the Z mass (J/ψ mass) is then labeled the “probe” electron. The probes are thus unbiased and used to build the LH PDFs. A full list of event selection and Tag & Probe level cuts can be found in Table 5.4 for $Z \rightarrow ee$ and Table 5.5 for $J/\psi \rightarrow ee$. Additional details regarding the Tag & Probe method for $J/\psi \rightarrow ee$ events can be found in Section 5.2 and the overall method is further described in [151] and [149].

Tag & Probe Selection for $Z \rightarrow ee$ events
Single electron trigger fired
Tag electron with $p_T > 25 \text{ GeV}$
Tag electron $ \eta < 1.37$ OR ($ \eta > 1.52$ AND $ \eta < 2.47$)
Tag electron passes Tight identification
$\Delta R(\text{tag electron, trigger electron}) < 0.10$
Tag electron passes LAr object quality requirement
Probe electron with $p_T > 15 \text{ GeV}$
Probe electron $ \eta < 2.47$
Probe electron passes VeryLoose identification
Probe electron passes LAr object quality requirement
$80 \text{ GeV} < m_{ee} < 100 \text{ GeV}$
Tag electron and probe electron have opposite electric charge

Table 5.4: Summary of Tag & Probe selection for $Z \rightarrow ee$ events

5.1.2.2 SMOOTHING PDFs

The PDFs used to construct the LH should contain only physically meaningful features and be free of fluctuations due to binning and statistics. Low statistics in a particular PDF could cause similar electron candidates to be assigned vastly different final discriminant values, and in particular a bin with 0 value would lead to undefined discriminates for some electron candidates. To mitigate these issues, Adaptive Kernel Density Estimation (KDE) is applied to all signal and background PDFs before the final LH discriminant is calculated.

KDE serves to smooth a variable distribution by treating each bin in the PDF as a δ -function, replacing each δ -function with a “kernel” function, then summing all the kernel functions to form the final PDF. For the purposes of the LH ID, the kernel function is a Gaussian with a tunable width parameter. In the adaptive KDE method, the Gaussian width parameter is increased in bins with low statistics. A visualization of both methods is presented in Figure 5.6, along with an example of a smoothed and unsmoothed PDF. The electron LH PDFs are smoothed using the TMVA adaptive

Tag & Probe Selection for $J/\psi \rightarrow ee$ events.
Di-electron trigger fired
Tag electron with $p_T > 4.5 \text{ GeV}$
Tag electron $ \eta < 1.37$ OR ($ \eta > 1.52$ AND $ \eta < 2.47$)
Tag electron passes Tight identification
$\Delta R(\text{tag electron, trigger electron}) < 0.10$
Tag electron passes LAr object quality requirement
Probe electron with $p_T > 4.5 \text{ GeV}$
Probe electron $ \eta < 2.47$
Probe electron passes VeryLoose identification
Probe electron passes LAr object quality requirement
$2.8 \text{ GeV} < m_{ee} < 3.3 \text{ GeV}$
Tag electron and probe electron have opposite electric charge
$\Delta R(\text{tag electron, probe electron}) > 0.10$
$-1 < \tau < 0.2$

Table 5.5: Summary of Tag & Probe selection for $J/\psi \rightarrow ee$ events.

KDE tool [155].

5.1.2.3 DATA VS MONTE CARLO PDFs

The LH PDFs can be constructed from data or MC events. Using MC allows for the use of truth-matching, where the truth labels described in Chapter 2 are checked to guarantee the objects are or are not electrons. This ensures highly pure signal and background PDFs. However, estimations and imperfect detector modeling in the MC simulation chain results in shape differences between data-driven and MC-driven PDFs, rendering a MC-driven LH less effective when applied to actual data (see Figure 5.7). These effects can be reduced by adjusting the MC PDFs with data-driven corrections as described in Section 5.1.2.4, but are not entirely eliminated. On the other hand, data-driven PDFs require additional work to eliminate all contributions from non-desired physics processes and measuring the resulting purity is difficult. In particular, for $J/\psi \rightarrow ee$ PDFs, significant contribu-

tions from non-prompt J/ψ s (from heavy flavor decays) remain after Tag & Probe selections and mass cuts. The current method for purifying data-driven $J/\psi \rightarrow ee$ PDFs is described in Section

5.2.2

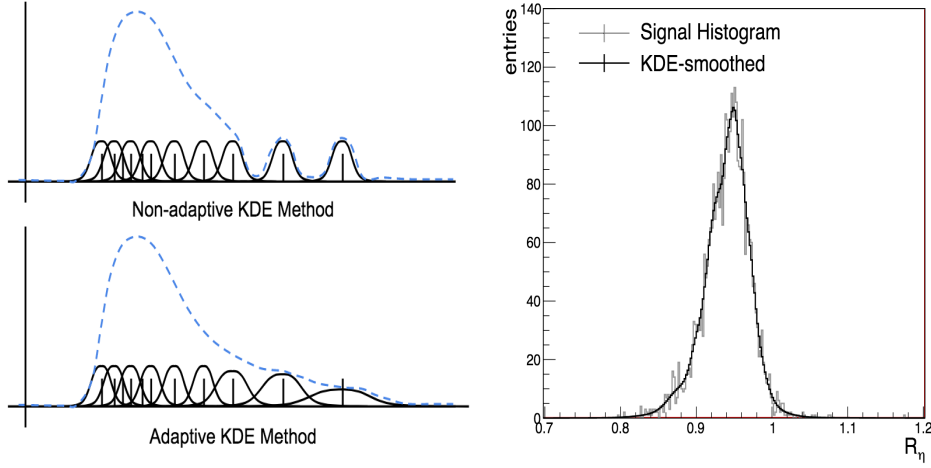


Figure 5.6: The KDE Method. Left: a stylized depiction of the standard KDE method (top) and Adaptive KDE method (bottom). Right: an example of a raw variable distribution (grey) and the KDE-smoothed PDF (black).

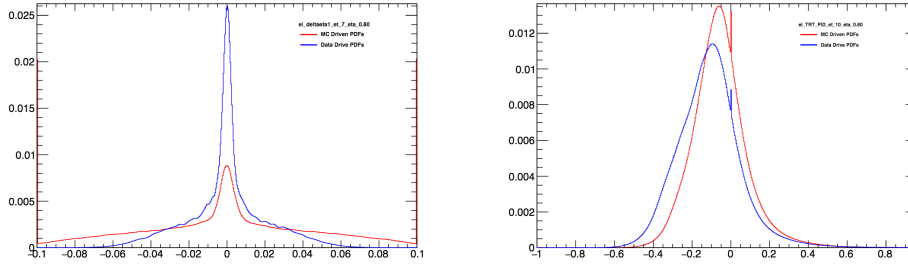


Figure 5.7: Comparisons of data-driven (blue) and MC-based (red) PDFs in the 7 GeV $< p_T < 10$ GeV and $0.8 < |\eta| < 1.15$ bin for two variables. $\Delta\eta_1$ on the left and eProbabilityHT on the right.

In Run 1, data-driven PDFs were used to build the electron ID LH. However, there was no software implemented to construct $J/\psi \rightarrow ee$ PDFs pure enough to build an efficient LH discriminant. Consequently, the $Z \rightarrow ee$ PDFs from the lowest p_T bin with high enough statistics (15-20 GeV) were used for all lower p_T bins. This resulted in reduced electron ID efficiency below 15 GeV. For the first portion of Run 2, MC PDFs were used for the full p_T range of the electron LH. After sufficient Run 2 data was collected and methods for constructing purified $J/\psi \rightarrow ee$ PDFs were developed, a data-driven electron LH ID was introduced and will be used for all full Run 2 analyses.

5.1.2.4 CORRECTING PDFs

When MC PDFs are used in the electron ID LH they must be corrected to more closely match the distributions seen in data. Most of the first order differences between MC and data PDFs take the form of a constant offset or a width difference (i.e. the Full Width at Half Maximum, FWHM).

These differences are corrected by adjusting each electron's PDF entry, ν , to $\nu_{MC}^* = \nu_{MC} - a$ for a constant offset a , or $\nu_{MC}^* = (\nu_{MC} - \bar{\nu}_{data,MC}) * w$ for width parameter w and mean value $\bar{\nu}_{data,MC}$. Optimal values for a are found by minimizing the χ^2 test statistic, $\chi^2 = \sum_{bins} (n_{data} - n_{MC})^2 / (\sigma_{data} + \sigma_{MC})$ while the optimal values for w are taken to be the ratio of FWHM between data and MC.

Corrections are calculated and applied separately for each η bin, but do not depend on p_T . The corrections derived from signal samples are also applied to background samples. An example of MC PDFs before and after these corrections can be seen in Figure 5.8. These corrections can be applied before or after KDE smoothing as the differences between the results of the two orderings was found to be negligible [149].

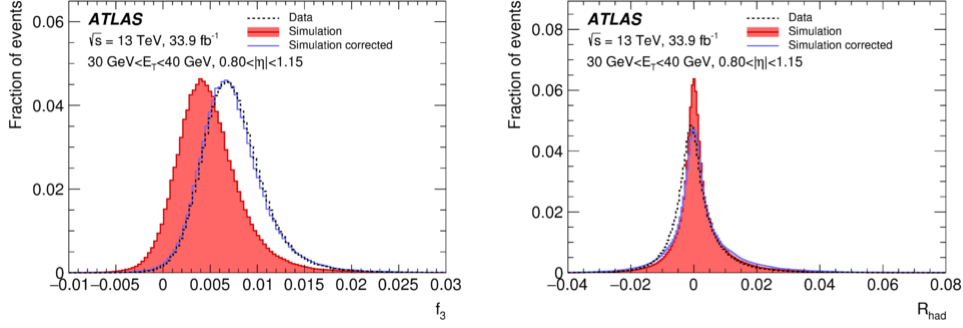


Figure 5.8: The f_3 (left) and R_{had} (right) PDFs for data (black dashed line) and MC (solid red) in the $30 \text{ GeV} < p_T < 40 \text{ GeV}$ and $0.8 < |\eta| < 1.15$ bin. The MC is shown before (solid red) and after (blue line) corrections have been applied.

5.2 LOW p_T ELECTRON ID

Electrons with a p_T between 4.5 and 15 GeV are referred to as “low p_T ” electrons. These electrons are treated separately from the “nominal p_T ” range electrons ($p_T \geq 15 \text{ GeV}$) because probe electrons from the physics process, $Z \rightarrow ee$, used to build and tune nominal p_T PDFs and LHs has very low statistics below 15 GeV. This leads to increased relative background contamination in the PDFs (as seen in Figure 5.9) and results in decreased ID performance [152].

5.2.1 MOTIVATION

Effective identification of low p_T electrons is critical to a variety of ATLAS analyses. One such class of analyses is searches for compressed Supersymmetry which frequently look for event topologies where the supersymmetric particles are produced in association with a colored Initial State Radiation (ISR) jet (see Fig 5.10). If the ISR jet has a high momentum, the recoil of the supersymmetric particles creates a signature of a back-to-back ISR jet and large MET coupled with very soft leptons (as low as 4 GeV) [156]. Similar signatures are seen in direct Dark Matter production searches where the recoiling object can be a photon [157] or a hadronic jet [158]. Finally, low p_T electrons are impor-

tant to precision measurements of the $H \rightarrow ZZ^* \rightarrow 4l$ decay channel where the off-shell Z boson can be low mass and subsequently decay to soft electrons [159].

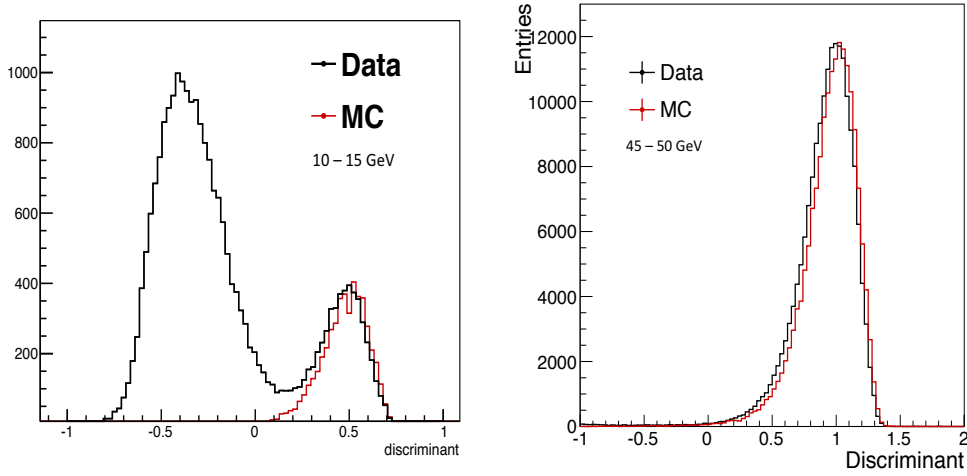


Figure 5.9: Left: An example plot of the output discriminant for probes (data and $Z \rightarrow ee$ MC) in the $10 \text{ GeV} < p_T < 15 \text{ GeV}$ and $0.80 < |\eta| < 1.15$ bin. The data peak at high discriminant value comes from electrons but the larger peak at low discriminant value is background. Right: the same plot for $40 \text{ GeV} < p_T < 45 \text{ GeV}$. The fraction of background in the data sample for this p_T range is negligible.

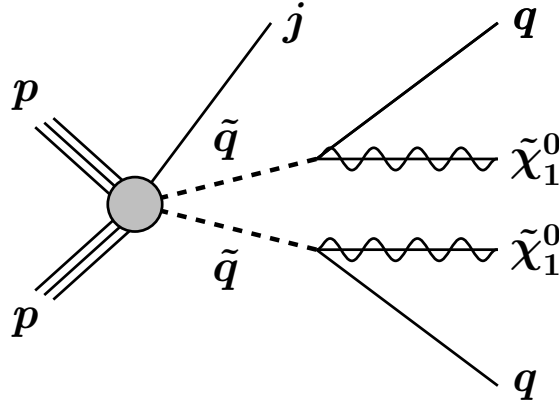


Figure 5.10: Production of a pair of squarks with an ISR jet[158].

5.2.2 SOFTWARE DEVELOPMENT

The technical capability to construct data-driven low p_T PDFs from $J/\psi \rightarrow ee$ events was developed in Run 2. The PDFs are constructed using a Tag & Probe selection and events are collected using secondary di-electron trigger. Tag electrons are required to pass the Tight Operating Point and have a p_T of at least 4.5 GeV (reduced from 7 GeV in Run 1 in order to access softer electrons). The probe electron must form an invariant mass with the tag within a 0.5 GeV window of the J/ψ mass (3.8 - 4.3 GeV) and the tag and probe must be separated by $\Delta R > 0.1$ to avoid overlap between the two objects.

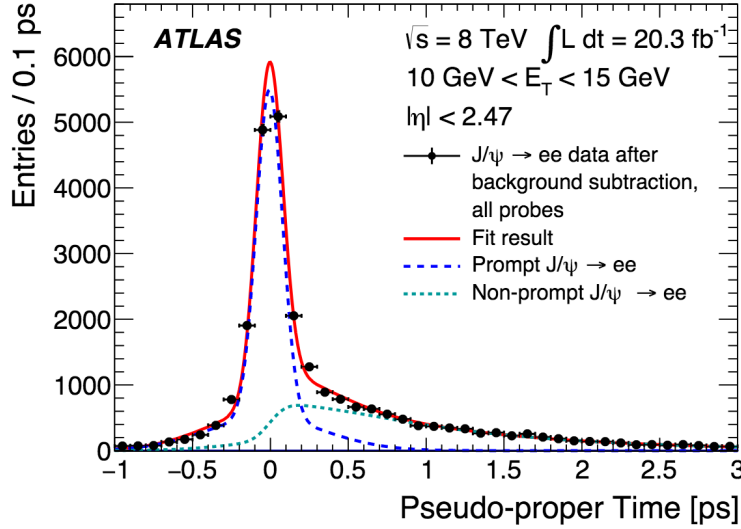


Figure 5.11: Pseudo-proper time for all probe electron candidates. The prompt signal component is shown by the dashed blue line and the non-prompt signal component is shown by the light blue dashed line.

The main challenge in constructing data-driven $J/\psi \rightarrow ee$ PDFs is the presence of non-prompt J/ψ candidates (mainly from b-quark decays) after the selections described above are applied. To remove these contributions, a cut is placed on the pseudo-proper time of the J/ψ candidates. Pseudo-proper time, τ , is defined as $\tau = (L_{xy} * m^{J/\psi}) / (p_T^{J/\psi})$ where L_{xy} is the distance from the primary

event vertex to the J/ψ vertex. An example τ distribution for probe electrons can be seen in Figure 5.11. For Run 2, probe electrons are required to have a pseudo-proper time $-1 \text{ ps} < \tau < 0.2 \text{ ps}$ and the dielectron invariant mass distribution for signal and background using this cut is shown in Figure 5.12.

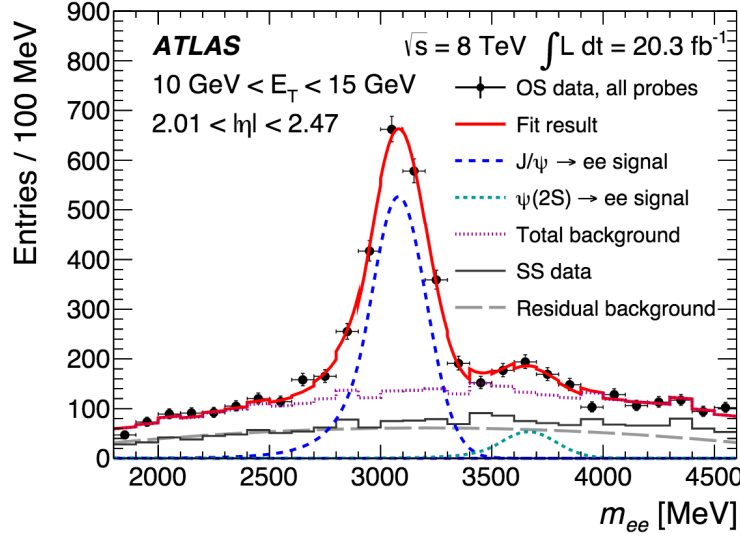


Figure 5.12: Dielectron invariant-mass fit for all probe electron candidates. The pseudo-proper time is required to be $-1 \text{ ps} < \tau < 0.2 \text{ ps}$. Dots with error bars represent the opposite-sign (OS) pairs for data, the fitted J/ψ signal is shown by the dashed blue and the $\psi(2S)$ by the dashed light blue lines. The sum of the background contributions is depicted as a purple dotted line.

5.2.3 VARIABLE OPTIMIZATION

Despite transitioning to data-driven $J/\psi \rightarrow ee$ PDFs for the low p_T electron ID, the ID efficiency remains lower in this region (Figure 5.13). This is likely because this region was not specifically optimized in Run 1, and the same variables chosen for nominal p_T are used at low p_T with no further study.

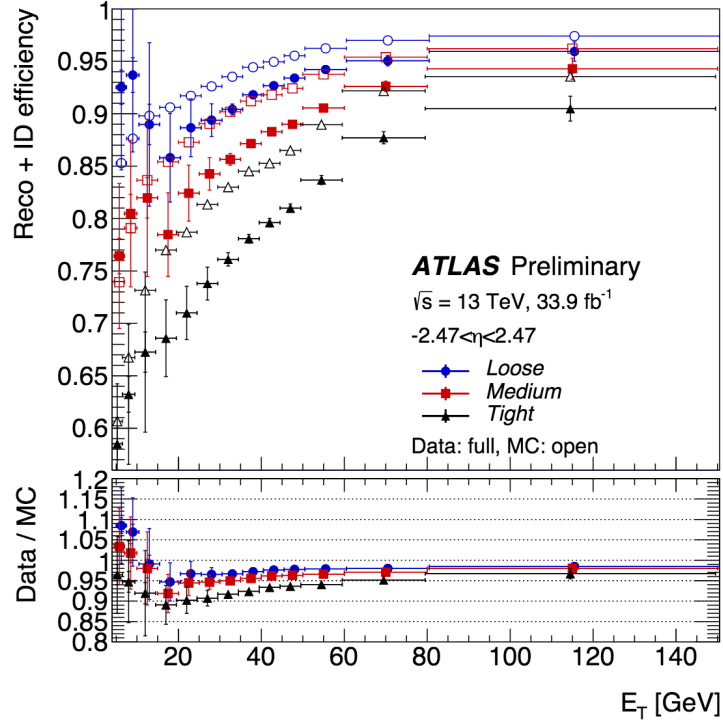


Figure 5.13: Electron reconstruction and identification efficiencies as a function of transverse energy E_T , integrated over the full pseudo-rapidity range.

An “n-1” LH study was conducted to determine if all LH variables remain efficient in the low p_T range. In this study a set of nearly identical LH operating points were constructed, with the only difference being that each LH removes just one variable from the nominal collection of discriminating variables. The performance of these alternative LH operating points can then be compared to the nominal LH performance by examining the receiver operating characteristic (ROC) curve (Figure 5.14). If a given ROC curve is higher than the nominal ROC curve (dotted black curve), then the identification performance for this alternative operating point is better than the nominal operating point, and thus it may be beneficial to remove that particular variable. Several variables including f_1 (hot pink line), f_3 (olive green line), $w_{\eta 2}$ (black solid line), and $\Delta\phi_{res}$ (aqua blue line) consis-

tently improved performance when removed across most low p_T and η bins studied, and could be removed from future iterations of the low p_T LH.

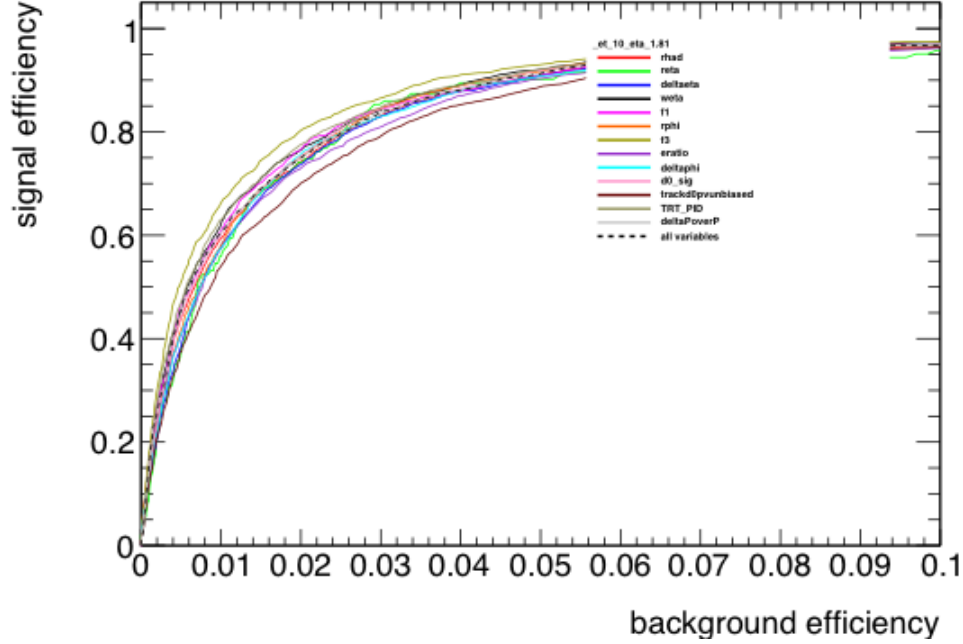


Figure 5.14: n-1 LH ROC curves. The nominal LH with all variables included is the dashed black line.

In addition to optimizing the variables already used in the low p_T region, several new variables are also being studied to further improve low p_T electron ID efficiency. These variables are:

- $\text{charge} \cdot d_0$: this would be used in place of the current transverse impact parameter variable d_0 .
- $\Delta(\text{curvature})/\text{error}(\text{curvature})$: curvature is defined as $1/p_T$ and is well measured at low p_T . This variable is sensitive to the amount of energy lost due to bremsstrahlung radiation, which is particularly important in the low p_T region. This variable is a possible replacement for $\Delta p/p$ which is better measured at higher p_T .
- $\Delta(\Delta\phi) = \Delta\phi_{LM} - \Delta\phi_{FM}$: this variable measures the change in the track-calorimeter match variable $\Delta\phi$ between its value in the first and last layers of the calorimeter. This variable may be useful for identifying converted photons.

Comparisons of the PDF shapes of signal and background for these variables can be seen in Figure 5.15. There are clear shape differences in both $\Delta(\text{curvature})/\text{error}(\text{curvature})$ and $\Delta(\Delta\phi)$, indicating these variables would likely improve the signal selection efficiency of the low p_T LH.

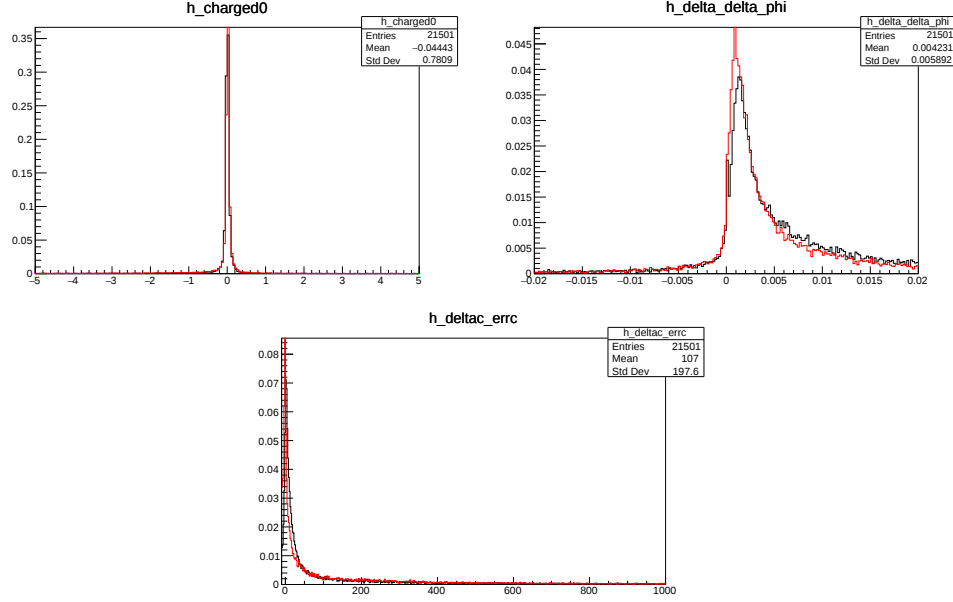


Figure 5.15: Comparisons of signal (red) and background (black) PDF shapes for new low p_T variables in the $4.5 \text{ GeV} < p_T < 7 \text{ GeV}$ and $1.81 < |\eta| < 2.01$ bin. Top left: charge $\cdot d_0$, top right: $\Delta(\Delta\phi)$, bottom: $\Delta(\text{curvature})/\text{error}(\text{curvature})$

5.3 MACHINE LEARNING AND ELECTRON ID

CNN-driven particle ID using images formed from the hadronic calorimeter has been quite successful (Chapter 4). This technique is currently being studied for application to electron and photon ID in the LAr EM calorimeter.

The creation of electron images in the EM calorimeter is more difficult than forming images in the hadronic calorimeter because the detector granularity varies substantially between layers and between the barrel and endcap regions (Chapter 2). A data set of over 2 million events and corre-

sponding images is supported by the ATLAS E/Gamma Working Group. In order to preserve the most information possible, the data-set for these studies contains different granularity images for each layer of the EM calorimeter. The images are formed by first taking all cells in a 7×11 window of the second calorimeter layer (ECAL2) centered on the highest energy cell in the electron ROI. All cells in the other three layers that overlap with this window are included in the image for that layer. A sample image is shown in Figure 5.16. As is done in many of the jet image applications, initial studies are restricted to the barrel region of the detector. In cases where the ECAL2 7×11 window extends beyond the barrel region the non-barrel pixels are set to zero.

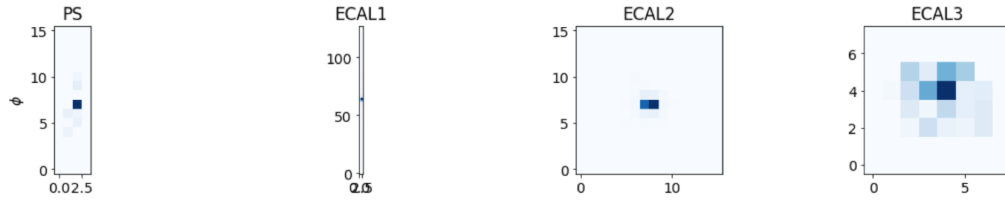


Figure 5.16: A sample electron image for each of the four EM calorimeter layers.

With this method of image construction the CNN architecture design must account for the different layer granularities. For initial studies, the images were all up-sampled to a common granularity of 128×128 as shown in Figure 5.17 using a Keras 2D upsampling convolutional layer [160]. The four images are then concatenated into a 128×128 stack and processed through a convolutional layer, ReLu activation layer, and a maxpooling layer. The processed images are then flattened and concatenated with additional scalar features of the particle including charge, η , ϕ , and TRTPID. This vector is then input into two fully connected NNs, one of which predicts a classification value (truth is 1 for signal and 0 for backgrounds) and one of which predicts a scale factor to calibrate the particle energy. The network was entirely constructed in Keras with a TensorFlow [161] back end and the full architecture is shown in Figure 5.18.

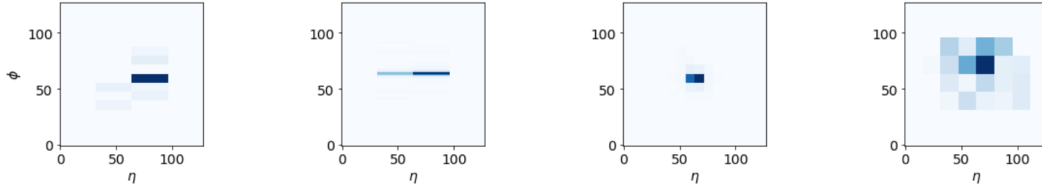


Figure 5.17: The same electron image from Figure 5.16 after being up-sampled to a uniform granularity of 128×128 .

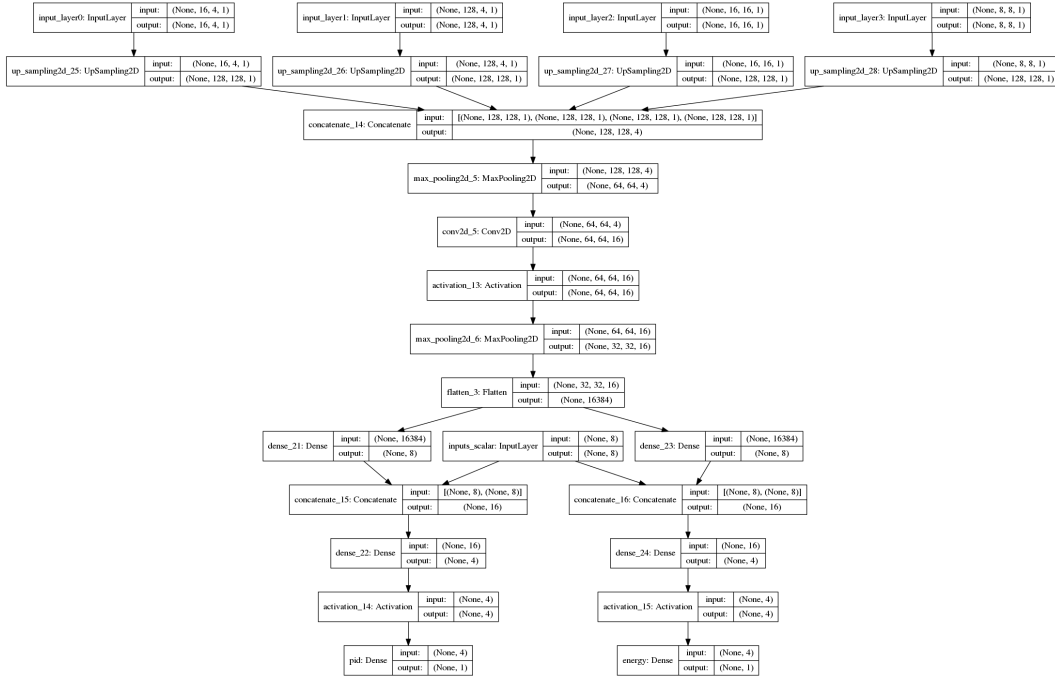


Figure 5.18: Network architecture for image-based electron ID and calibration.

The network was trained using 1 million signal (prompt electron) samples and 100,000 background ($2 \rightarrow 2$ QCD processes) samples where 20% was held-out for the test-set and 20% for the validation set. The training process used standard backpropagation with a binary cross-entropy loss function. Results of the two tasks are shown in Figures 5.19 and 5.20; the classification task achieves excellent signal-background separation and the calibration task actually outperforms the current

ATLAS E/Gamma algorithm [92].

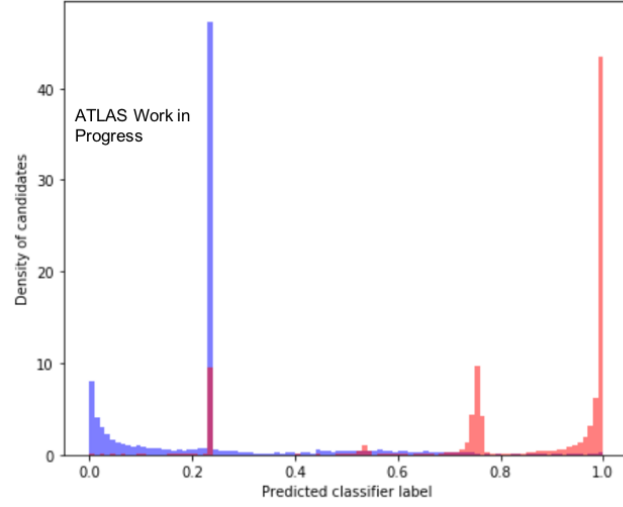


Figure 5.19: Classifier output of the image-based electron ID CNN for signal (red) and background (blue).

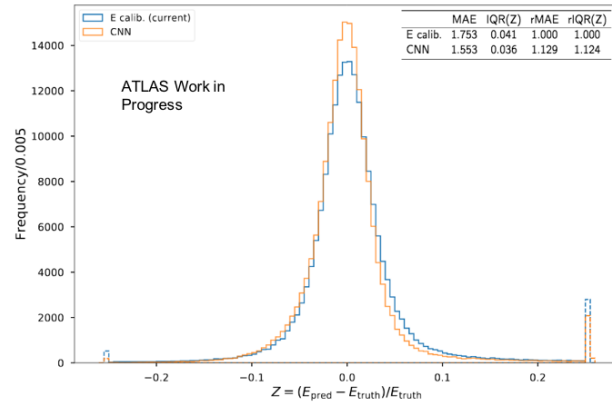


Figure 5.20: Normalized difference between predicted particle energy and truth particle energy for the imaged-based calibration CNN (orange) and the standard ATLAS electron calibration algorithm (blue).

5.3.1 FUTURE STUDIES

This area of research is on-going and there are several techniques being studied to improve upon the performance described above. For instance, there are other possible ways of addressing the granularity mis-match such as applying a separate convolution to each layer image and concatenating them just before the task networks or up-sampling and applying a 3D convolution (Figure 5.21) that would allow the network to learn relationships between the EM calorimeter layers. Additionally, track information can be included in either by including it in the scalar feature concatenation or by creating track images using the Inner Detector outputs.

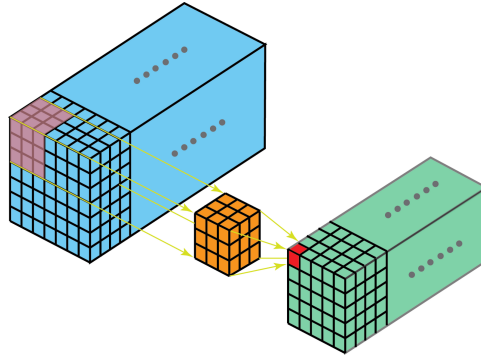


Figure 5.21: A schematic depiction of a 3D convolution for use in a CNN.

Including other types of information like timing and cell quality could further improve network performance. There are important questions that must be addressed before an image-based electron ID could be fully incorporated into the ATLAS software flow. This includes how to handle the barrel to endcap transition when constructing images and understanding whether including η values allows the network to account for the fact that the layer images do not actually lie directly on top of each other in the physical detector. Furthermore, as discussed in Section 5.1.2.3 the ultimate goal is to develop an algorithm that is fully data-driven. This presents additional challenges when training ML algorithms as they typically require very high training sample purity in order to achieve reason-

able performance. Methods for purifying the data-sets currently used for the LH-based ID without biasing any kinematics are being explored.

By the time you get close to the answers it's nearly all over.

Merle Haggard

6

$VH, H \rightarrow \tau\tau$ Analysis

UNDERSTANDING THE MECHANISM OF ELECTROWEAK SYMMETRY BREAKING (CHAPTER 1), TESTING SM COUPLING STRENGTH PREDICTIONS, AND FURTHER CHARACTERIZING THE HIGGS BOSON ARE PRIMARY GOALS OF THE PHYSICS PROGRAM AT THE LHC. ATLAS supports a breadth of searches aimed at accomplishing these objectives including a variety of SM Higgs production and decay mode pairs and BSM models. One such search, focused on Higgs production in association with a vector boson (VH) and Higgs to tau tau decays, is presented in the following chapters.

6.1 MOTIVATION

The specific Higgs mode of interest in this analysis can be referred to as $VH, H \rightarrow \tau\tau$. As shown in Figure 6.1, VH production and $H \rightarrow \tau\tau$ decays have both been observed with the ATLAS detector, but the combination of the two has not. In fact, $VH, H \rightarrow \tau\tau$ is the only channel amongst the intersection of the four main production modes and five most common decay modes that has yet to be measured using Run-2 data from the ATLAS detector [162]*. Thus, this analysis is a crucial missing piece to developing a complete understanding of the SM Higgs.

Although ggF and VBF are the dominant Higgs production modes, VH is a particularly interesting channel for studying $H \rightarrow \tau\tau$ decays. Both Z and $W^\pm$ bosons can decay to light leptons (e and μ) at branching ratios suitable for observation at the LHC: approximately 3% for $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-$ and 10% for $W^\pm \rightarrow e^\pm\nu$ and $W^\pm \rightarrow \mu^\pm\nu$ [12]. This allows the $VH, H \rightarrow \tau\tau$ analysis to exploit the highly efficient electron and muon triggers and IDs. This results in relatively low background contributions and an unbiased $H \rightarrow \tau\tau$ decay. Thus this analysis provides an excellent opportunity to increase the precision of the $H \rightarrow \tau\tau$ branching ratio measurement. The Run-1 version of this search did not find statistically significant evidence for these processes (see Section 6.4) but increased increased luminosity and analysis design improvements make this a promising search for the full Run-2 dataset.

6.2 DEFINITIONS

This analysis encompasses four final states, or subchannels, defined by the decay products of the vector boson and taus; they can hence be referred to as VH and WH (for the boson type) and lep-had or had-had for the tau decay products. The selection requirements for the four channels are detailed

The $VH, H \rightarrow WW^$ [163] and $ggF, H \rightarrow \tau\tau$ [164] measurements are available but not included in the published combination shown in Figure 6.1.

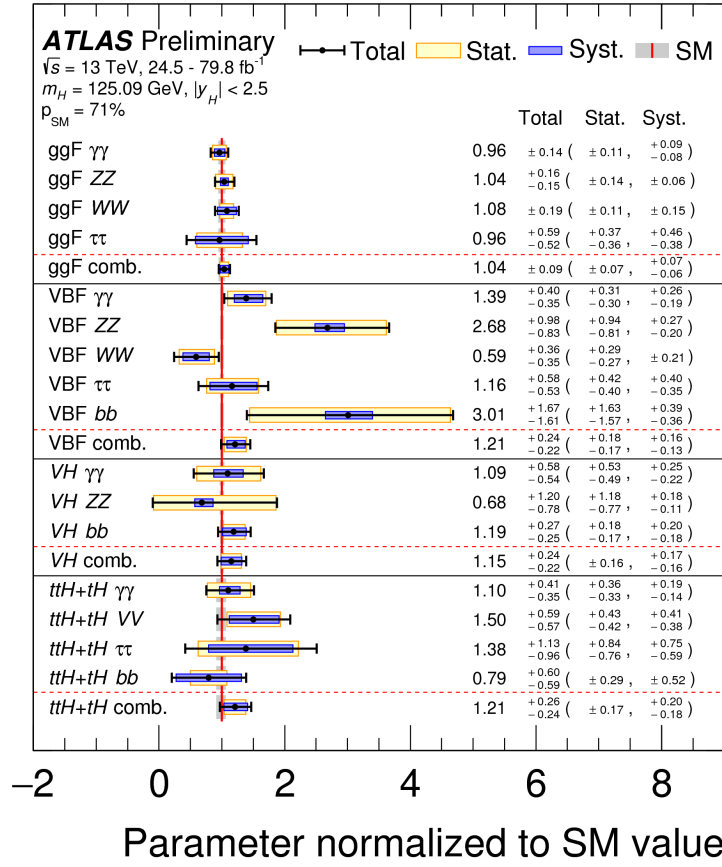


Figure 6.1: Cross-sections times branching fractions for relevant Higgs production and decay modes normalized to the SM predicted values. The values are obtained by a simultaneous fit to all channels [162].

in Section 6.3. The basic object requirements is shared between all the channels and each channel uses a variable related to the reconstructed Higgs mass as the final distribution to fit (Section 6.2.3.1). The criteria for all relevant objects and the methods of mass estimation are described below.

6.2.1 OBJECT CRITERIA

In most cases, the object requirements are chosen to be as close to the Run-1 criteria as possible [165]. All requirements are selected from the recommendations supported by the relevant ATLAS Working Groups. Additionally, when possible, they are chosen to be compatible with the ggF/VBF $H \rightarrow \tau\tau$ Run-2 measurement.

6.2.1.1 TAUS

This analysis uses taus that are seeded by anti- k_t $R=0.4$ jets and further reconstructed and identified as described in Chapter 3. All taus must have a minimum p_T of 25 GeV and be exactly 1- or 3-pronged. 1-pronged taus must have clusters and a leading track with $|\eta| < 2.47$ and 3-pronged taus must be withing $|\eta| < 2.5$. Additionally, taus must pass the medium BDT-based operating point as defined by the Tau Working Group [166]; this corresponds to an efficiency of approximately 55-60%. The final analysis will likely incorporate the new RNN-based tau ID (Chapter 3).

6.2.1.2 MUONS

Muons used in this analysis must be combined muons which are reconstructed and identified as described in Chapter 3. They must have a minimum p_T of 3.5 GeV and be in the region $|\eta| < 2.5$. They are required to pass the Loose operating point and track quality criteria as defined by the Muon Working Group [167]. This corresponds to an efficiency of over 95%.

6.2.1.3 ELECTRONS

Electrons are reconstructed and identified as described in Chapter 5. They are required to be in the region $|\eta| < 2.47$ and pass the Loose LH-based operating point and object quality criteria as set by the EGamma Working Group [168]. They are additionally required to pass calorimeter cluster and

track isolation requirements that the sum of transverse energy in the calorimeter (tracks) within a cone of size $R=0.4$ ($R=0.2$) around the electron cluster to be less than 8% of the electron E_T . Collectively this provides an efficiency between 80 and 90% depending on the p_T and $|\eta|$ of the electron. Electrons in the ‘crack’ region of $1.37 < |\eta| < 1.52$ where the EM Calorimeter transitions from the barrel to endcap regions are required to pass the Medium LH-based operating point; these electrons are only used for overlap removal as described in Section 6.2.2 and not as analysis category objects due to decreased electron ID efficiency in the crack (see Chapter 5).

In order to exploit the lowered electron ID p_T threshold for Run-2, the minimum electron p_T for this analysis was lowered from 10 GeV (the Run-1 requirement) to 5 GeV. Preliminary studies, as shown in Figure 6.2, indicate that this change allows additional events to pass the category requirements for the lep-had analysis channels. If further studies find that including these lower p_T electrons does not substantially increase systematic uncertainties this would increase statistics in these channels.

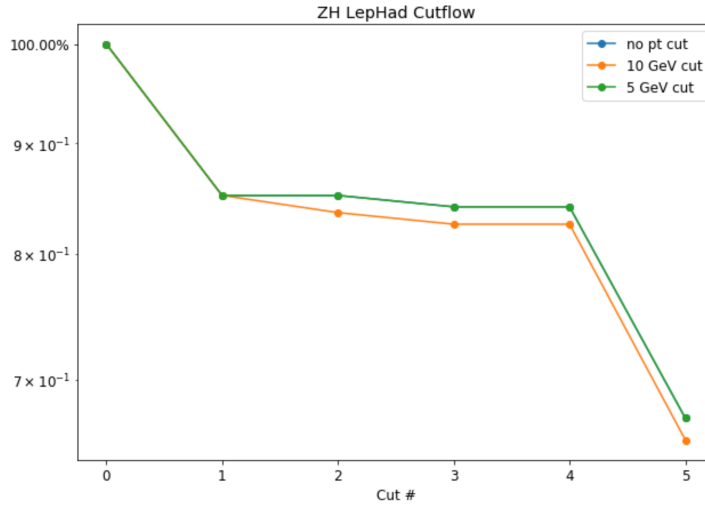


Figure 6.2: Progressive cut flow for a portion of the ZH lep-had analysis category (after requiring exactly 3 leptons and one hadronic tau). Cut 2 is the p_T cut of interest and other cuts are defined in Table 6.1.

Cut #	Definition
0	All 3 leptons are tight
1	Two same flavor oppositely signed leptons with $80\text{GeV} < m_{ll} < 100\text{GeV}$
2	Leptons pass min p_T (10 GeV for the green curve and 5 GeV for the orange curve)
3	Tau passes BDT medium and is oppositely charged from 3rd lepton)
4	Tau $p_T < 25\text{ GeV}$
5	$60\text{GeV} < p_T(\text{tau} + 3\text{rd lepton})$

Table 6.1: Cut definitions for Figure 6.2. See Section 6.3.1 for additional details.

6.2.1.4 JETS

Jets are not considered as analysis category objects but multiple control regions and the WH analysis channels include b-jet vetos; thus it is necessary to define analysis-wide jet requirements. This analysis uses anti- k_t reconstructed jets (Chapter 3) with a maximum radius of $R=0.4$. Only jets with $p_T > 30\text{ GeV}$ and $|\eta| < 2.5$ are used. Jet energies are calibrated according to the recommendations of the JetETMiss Working Group [169]. A fixed-cut b-tagging requirement corresponding to an 85% efficiency is used.

6.2.1.5 MET

Requirements on MET are not included in the four analysis category definitions but MET is used to separate the ZH and WH fake factor measurement regions (see Chapter 7) and as input to the mass estimation algorithms described in Section 6.2.3.1. MET is calculated as described in Chapter 3 using calorimeter energy deposits associated to and calibrated according to the reconstructed objects outlined above. Additional energy deposits that are not associated to a physics object are scaled by a dedicated algorithm to improve resolution in high-pileup events.

6.2.2 OVERLAP REMOVAL

In some cases, certain tracks and calorimeter clusters may be reconstructed as multiple objects and thus it is necessary to define a prioritization order. In the case where objects which pass the requirements outlined above overlap within a ΔR of 0.2 only one of them is preserved for input to the analysis category definitions. The overlap is resolved in the following order: muons > electrons > taus > jets. This procedure uses muons with a reduced minimum p_T of 2 GeV and precedes the electron isolation requirement application.

6.2.3 MASS RECONSTRUCTION

After estimating background contributions (Chapter 7) and accounting for systematic and statistical uncertainties, the final result in all four analysis channels is obtained from a fit of the estimated di-tau invariant mass spectrum; the signal distribution should correspond to the Higgs mass of 125 GeV. However, as all event categories contain one or more neutrinos in the final state, the mass distribution must be calculated in a way that accounts for the missing energy of the neutrinos. This is done separately for the ZH and WH categories using the methods described below.

6.2.3.1 MISSING MASS CALCULATOR

In the ZH analysis categories, neutrinos only come from the tau decays and there can be two (had-had case) or three (lep-had case) final state neutrinos. For these channels, the di-tau mass is reconstructed using the Missing Mass Calculator (MMC) [170]. This method takes the x - and y -components of the event's missing p_T and the visible mass of the tau pair as inputs. This system is inherently under-constrained because the individual orientation components of the multiple final state neutrinos are all unknown. To manage this, the MMC method implements a scan over all possible neutrino momentum values and the most likely di-tau mass is selected based on the fact that

the probability of a given distance (ΔR) between the visible and invisible decay products of a tau decay is known. A sample of the ΔR probability distributions used to develop the MMC likelihood are shown in Figure 6.3 and an example of a Run 1 MMC distribution is shown in Figure 6.4 where a clear separation between the $H \rightarrow \tau\tau$ mass and $Z \rightarrow \tau\tau$ mass is visible.

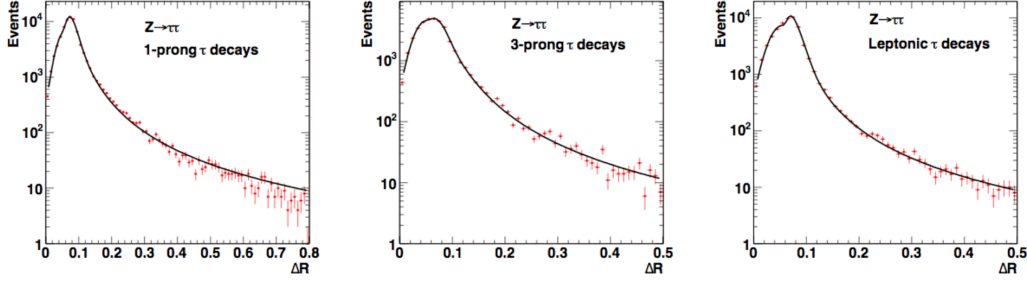


Figure 6.3: Examples of the probability distribution functions of ΔR for 1-prong hadronic tau decays (left), 3-prong hadronic tau decays (middle) and leptonic tau decays (right) [170].

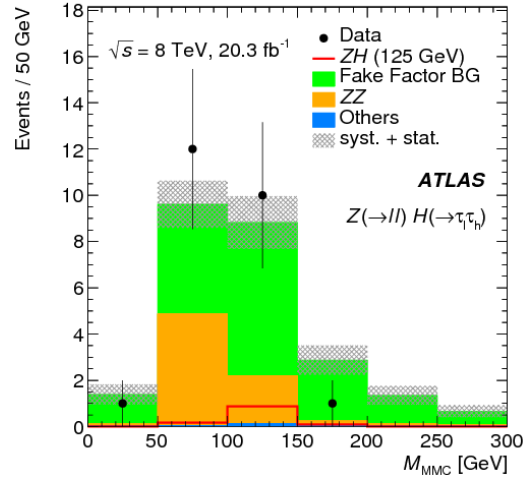


Figure 6.4: Run 1 MMC distribution after signal region cuts for the ZH lep-had channel. [171].

6.2.3.2 LATE TRANSVERSE PROJECTED MASS

In both the WH categories there is an additional neutrino from the W decay which makes the MMC method sub-optimal. Instead, the Late Projected Transverse Mass (M_{2T}) method is used as described in [172]. This method calculates a lower bound for the di-tau combined mass in a given event by minimizing over the allowed phase-space of possible neutrino momentum. The resulting distribution is bounded above by the Higgs mass.

The final state objects in the event are partitioned first into groups based on their parent particle (W or H in this case) and then into visible and invisible categories (Figure 6.5). The transverse momentum projection of each partition is determined: $\vec{p}_{aT} = \sum_{i \in V_a} \vec{p}_{iT}$ for the visible daughters of parent particle a and $\vec{q}_{aT} = \sum_{i \in I_a} \vec{q}_{iT}$ for the invisible daughters of parent particle a . The transverse energy projection of each partition can then be calculated: $e_{aTv(i)} = \sqrt{(\sum_{i \in V_a(I_a)} E_i)^2 - (\sum_{i \in V_a(I_a)} p_{iz})^2}$ and the transverse energy-momentum vectors are determined: $p_{aT}^\alpha = (e_{aTv}, \vec{p}_{aT})$ for the visible components of particle a and $q_{aT}^\alpha = (e_{aTi}, \vec{q}_{aT})$ for the invisible components of particle a . The late-projected transverse mass of particle a is then $M_{aT} = \sqrt{g_{\alpha\beta}(p_{aT}^\alpha + q_{aT}^\alpha)(p_{aT}^\beta + q_{aT}^\beta)}$ where $g_{\alpha\beta}$ is the Minkowski metric. The M_{2T} method then considers the largest parent mass ($\max_a[M_{aT}]$) and minimizes this value over all possible values of the invisible particles' (neutrinos') momenta. This defines the value

$$M_{2T} = \min_{\sum \vec{q}_{iT} = \vec{p}_T^{miss}} (\max_a [M_{aT}]).$$

The possible values over which M_{2T} is maximized are constrained by requiring that the sum of the neutrino p_T is equal to the MET of the event and invariant mass of the lepton and the neutrino attributed to the W be as close as possible to the W mass of 80.4 GeV. As defined, this phase-space is 9 dimensional in the had-had case (3 neutrinos with 3 momentum components) and 12 dimensional in the lep-had case. In order to reduce the phase-space, the neutrino momentum is approximated

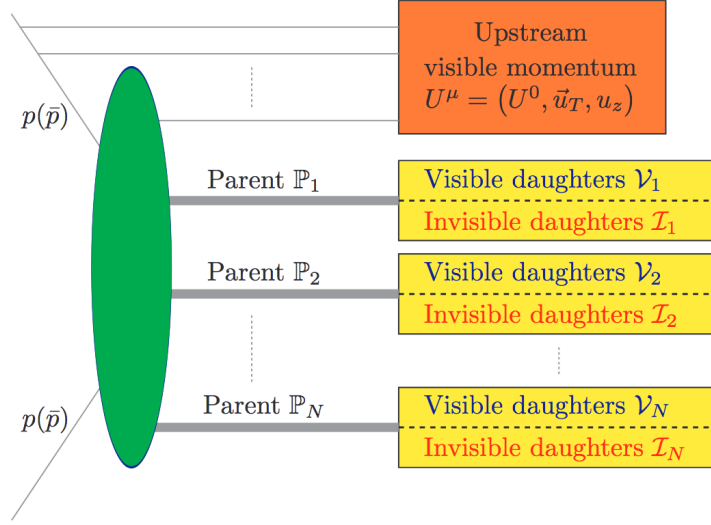


Figure 6.5: Depiction of particle partitioning in the M_{2T} method [172].

to be collinear with the visible tau decay products as $\vec{p}_v = (\frac{1}{x} - 1)\vec{p}_{vis}$ where $x = \frac{p_{vis}}{p_{vis} + p_v}$ is the fraction of the total tau momentum attributed to the visible decay products. The phase-space then becomes one-dimensional (x) for both the lep-had and had-had cases. An example of Run I M_{2T} distributions is shown in Figure 6.6.

6.3 ANALYSIS CATEGORIES

This analysis is divided into four separate categories each with its own event selection criteria; the categories are defined by the type of vector boson produced in association with the Higgs (Z or W) and the decay modes of the taus (one leptonic and one hadronic or both hadronic). Each set of category requirements was developed to provide highly efficient background rejection while maintaining as many signal events as possible. In Run I, this corresponded to an acceptance of 1.9% for the WH channels and 5.3% for the ZH channels. Any contributions from other Higgs processes after category selection was applied was found to be negligible [171].

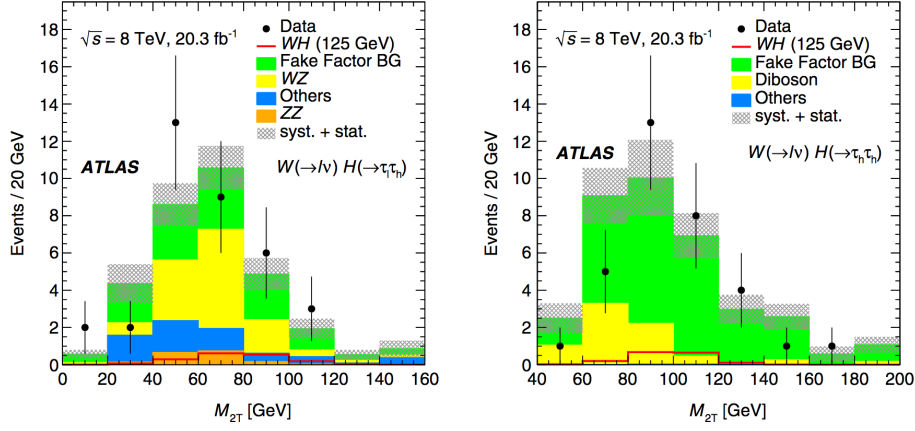


Figure 6.6: Run 1 M_{2T} distributions for three Higgs mass values after signal region cuts for the WH lep-had channel (left) and WH had-had channel (right) [171].

The common single-lepton and di-lepton triggers used to select events are shown in Table 6.2. The p_T of trigger objects used in the analysis is required to be 2 GeV above the trigger threshold in order to ensure maximal trigger efficiency.

Trigger	Trigger Threshold	Offline Threshold
HLT_2mu14	14 GeV	16 GeV
HLT_mu26_ivarmedium	26 GeV	28 GeV
HLT_2e17_lhvloose_nod0	17 GeV	19 GeV
HLT_e26_lhtight_nod0_ivarloose	26 GeV	28 GeV

Table 6.2: Run 2 triggers used in the four analysis channels along with corresponding trigger and offline p_T thresholds.

The event selection requirements for each of the four categories are provided in the following sections. There are a number of control regions associated with each channel which are primarily used to validate the fake factor method described in Chapter 7. These control regions are described thoroughly in [171].

6.3.1 $WH, \tau_{lep}\tau_{had}$

The WH channels are designed to select events where a Higgs boson is produced in association with a $W^\pm$ and the W then decays to a lepton and a light neutrino (Figure 6.7). In the WH lep-had channel, the Higgs decays to two taus, one of which decays hadronically while the other decays leptonically.

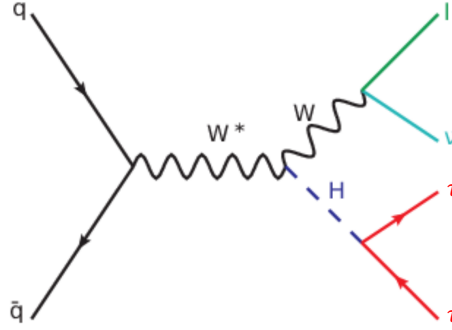


Figure 6.7: Feynman diagram for the leading order quark-initiated WH production.

This analysis category requires exactly two opposite-flavor, tightly identified light leptons and one BDT-Medium tau. The light lepton with lowest p_T is associated to the Higgs and the two light leptons are required to have the same charge to reduce backgrounds from $Z/\gamma^* \rightarrow \tau\tau + jets$ events and WW events where both W s decay leptonically. The leptons must also both be isolated. A b-jet veto is applied to reduce background from $t\bar{t}$ events. The tau is required to have opposite charge from the two light leptons and the total p_T of the three leptons must be greater than 80 GeV. This p_T requirement significantly reduces the Z +jets and multijets backgrounds which come mainly from low p_T jets mis-identified as taus. An additional requirement that $\Delta R(\tau_{lep}\tau_{had}) < 3.2$ further lowers these backgrounds.

6.3.2 $WH, \tau_{had}\tau_{had}$

The WH had-had category requires both Higgs associated taus to decay hadronically. In this case, the highest p_T lepton in the event is associated to the W. Events are first selected by requiring exactly two oppositely charged BDT-Medium taus and one tight, isolated light lepton. A b-jet veto is applied to reduce backgrounds from $t\bar{t}$ events. Events are required to have a combined transverse mass between the lepton and MET of at least 20 GeV to reject $Z \rightarrow \tau\tau$ background. The two taus are required to be separated by $0.8 < \Delta R < 2.8$ in order to reduce contributions from di-jet backgrounds. Finally, the scalar sum of the p_T of the three leptons must be at least 100 GeV to reduce the background from multi-jet events.

6.3.3 $ZH, \tau_{lep}\tau_{had}$

The ZH channels are designed to select events where a Higgs boson is produced in association with a Z boson and the Z then decays to a same-flavor opposite charge pair of light leptons (Figure 6.8). In the ZH lep-had channel, the Higgs decays to two taus, one of which decays hadronically while the other decays leptonically.

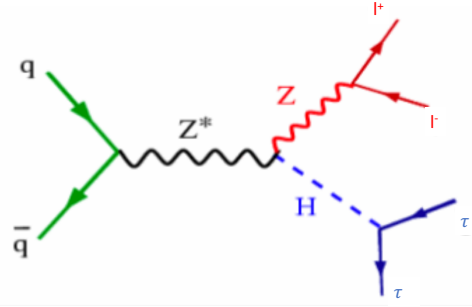


Figure 6.8: Feynman diagram for the leading order quark-initiated ZH production. There are also contributions from gluon initiated triangle and box diagrams.

Events in this category are required to have exactly three light leptons and one BDT-Medium tau.

The two same-flavor opposite-charge leptons with invariant mass closest to 91 GeV are associated to the Z and this mass must be between 80 and 100 GeV. The remaining lepton and the tau are associated to the Higgs and are required to have opposite charge. They must have a combined scalar p_T sum of at least 60 GeV to suppress Z+jets backgrounds.

6.3.4 $ZH, \tau_{had}\tau_{had}$

The ZH had-had category requires both Higgs-associated taus to decay hadronically. These events are required to have exactly two same-flavor opposite-charge light leptons and two BDT-Medium taus. The invariant mass of the two light leptons must be between 60 and 120 GeV. The two taus must be opposite charge and have a scalar p_T sum of at least 88 GeV to reduce Z+jets backgrounds.

6.4 RUN-1 RESULTS

In Run-1, the final observed signal strength μ of this analysis was determined using a binned global maximum-likelihood fit as described in [173] to the Higgs candidate mass spectra. As described in Section 6.2.3.1, the MMC distribution was fit for both ZH channels and the M_{2T} distribution was fit for both WH channels. A set of nuisance parameters $\vec{\theta}$ was used in the fit to account for systematic uncertainties and the expected number of signal and background events in each mass distribution bin was calculated as a function of these parameters. The test statistic q_μ was constructed as a function of the profile likelihood ratio and this statistic was used to measure the compatibility of the binned data with the background-only hypothesis and to set Confidence Level (CL) intervals on the final result. Finally, a significance was calculated to quantify the probability of obtaining the observed values of q_μ if the true signal strength was $\mu = 1$. Generally a significance of at least 3σ is required to qualify a process for observation and 5σ is required to qualify for discovery.

The results of the Run 1 analysis can be seen in Figures 6.9 and 6.10. The measured signal strength

for a 125 GeV Higgs, normalized to the SM expectation, was $\mu = 2.3 \pm 1.6$. The combined 95% confidence interval on the ratio of the observed cross-section to the SM predicted cross-section was 5.6. This was above the expected value for both the signal included and background-only hypotheses, but was nonetheless consistent with the expected values within uncertainty. A major challenge of the Run-1 analysis was low statistics in all analysis channels; this allowed slight data excesses in both had-had channels to substantially weaken the cross-section measurement limit. Finally, the expected and observed significance for each analysis channel is shown in Table 6.3.

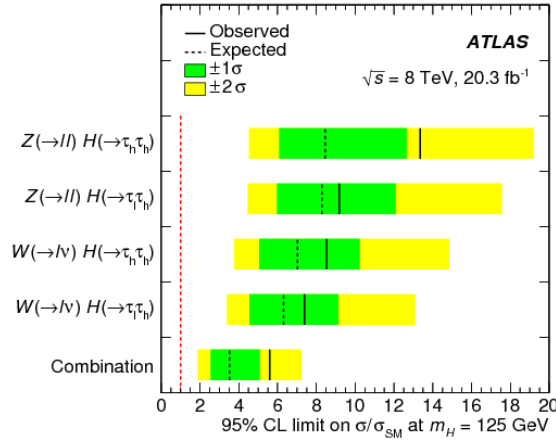


Figure 6.9: The Run-1 95% CIs on the ratio of measured cross-section to SM cross-section for each channel individually and all four combined [171].

Channel	Expected significance	Observed significance
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$	0.36σ	0.44σ
$W \rightarrow \mu\nu/e\nu, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$	0.32σ	0.60σ
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$	0.28σ	0.29σ
$Z \rightarrow \mu\mu/ee, H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$	0.32σ	1.38σ

Table 6.3: The Run-1 expected and observed significance for each of the four analysis channels [171].

The current status of the Run 2 analysis and planned improvements are described in Chapters 7 and 8.

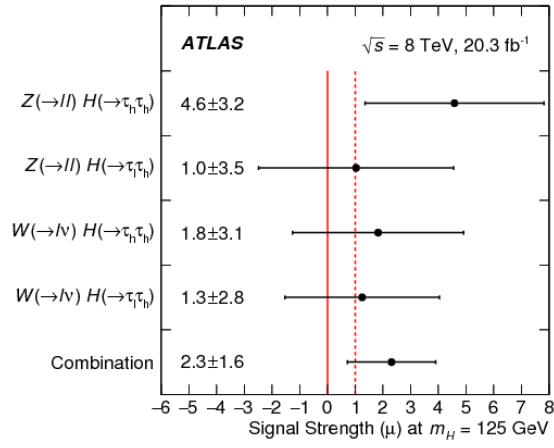


Figure 6.10: The Run-1 signal strength μ for each analysis channel individually and all four combined [171].

*It's going to take a lot of love to make things work out
right.*

Neil Young

7

Background Estimation in $VH, H \rightarrow \tau\tau$ Analysis

DESPITE THE TARGETED EVENT SELECTION REQUIREMENTS DEFINED IN CHAPTER 6 IT IS NOT POSSIBLE TO ELIMINATE ALL CONTRIBUTIONS FROM NON- $VH, H \rightarrow \tau\tau$ EVENTS IN THE FINAL DI-TAU MASS SPECTRUM. Events originating from a different process that results in identical final state particles with sufficiently similar kinematics may also pass the analysis channel selections. Additionally, the reconstruction and identification algorithms defined in Chapter 3 are not 100% accurate or efficient; it is possible that objects are mis-identified and events that do not truthfully

contain all required final state particles mistakenly pass the analysis channel selections.

Correctly modeling these effects, collectively referred to as analysis backgrounds, is crucial to ensure accurate analysis results and to limit uncertainty of the final measurement. Several techniques are utilized in ATLAS analyses to model background contributions including fake factors (Section 7.2), shape studies, and embedding methods [174].

Generally, data-driven methods for background estimation are preferred over MC-based methods. This is true for a variety of reasons including known mis-modeling effects in the current ATLAS simulation chain and the computing resources necessary to generate sufficient statistics to model specific background processes (see Chapter 3). Although incorporating production filters in the event generation stage can reduce the computing load in some cases they cannot be used for jet-based backgrounds like those that make up the majority of backgrounds in the $VH, H \rightarrow \tau\tau$ analysis. This is because the pile-up profile (which adds many jets to the event) is applied at a later simulation stage.

7.1 BACKGROUND TYPES

There are two types of backgrounds for this analysis: events in which all 3 or 4 final state leptons are actually produced by some non $VH, H \rightarrow \tau\tau$ process and events in which some of the leptons are actually misidentified jets. These are referred to as irreducible backgrounds and fake backgrounds respectively. An MC-based method is used to estimate the irreducible background while a data-driven method is used model the fake backgrounds. A depiction of each method is shown in Figure 7.1 where ϵ is the rate at which true, prompt objects pass the object criteria defined in Chapter 6 and f is the object specific fake-factor as defined in Section 7.2. The two methods are described in further detail in the following sections.

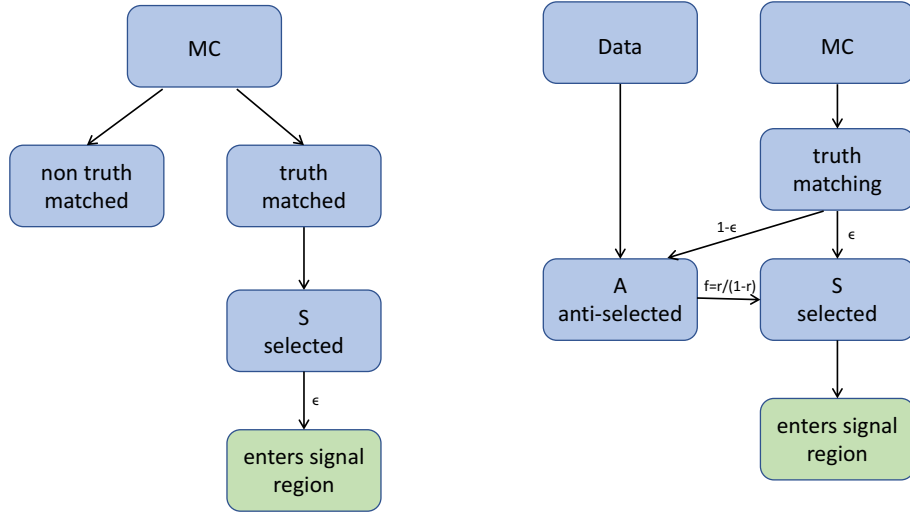


Figure 7.1: Computing flow diagram for the MC-based irreducible background estimation (left) and data-driven fake background estimation (right).

7.1.1 IRREDUCIBLE BACKGROUNDS

Irreducible backgrounds containing true, prompt electrons, muons, and taus come mainly from non-resonant diboson events. These backgrounds are determined entirely from MC. The analysis category selections described in Chapter 6 are applied to MC samples of these processes. Objects in events which pass the category selection are then truth matched to ensure they originate from vector boson and Higgs decays. Events that pass truth matching are used to populate the irreducible background distributions.

7.1.2 FAKE BACKGROUNDS

Events with fake objects are the primary source of background in all four analysis channels. These backgrounds are more difficult to estimate and are determined using the data-driven fake factor (FF) method which is the focus of the remainder of this chapter. The most common faked object

in this analysis is the hadronic tau, and in Run-1 the tau FF method was sufficient for background modeling in three of the four analysis channels. The electron FF was used only in the WH lep-had channel and the muon FF method was not used in the final analysis [171]. For Run-2 the FF method is being studied for all three object types and the closure tests introduced in Section 7.4.1 will aid in determining which methods to include in the final analysis. If it is found that one or two of the FF methods is sufficient to account for fake background contributions, the remaining methods will not be included to limit uncertainty on the final measurement.

7.2 THE FAKE FACTOR METHOD

Due to the relatively high number of final state objects (3 or 4 depending on the analysis category) and the potential for an event to include three different types of fake objects, modeling the fake background is one of the most challenging aspects of this analysis. The FF method was developed in Run-1 for this purpose. It is a data-driven extrapolation method in which events with fake objects taken from a kinematically similar region (the fake region) are used to calculate a scale factor (the fake factor) which is applied to the final signal distribution to estimate the number of fake events. The mathematical derivation of this method is detailed in Section 7.2.1, the FF measurements for each object type are described in Section 7.3, and method validation studies are introduced in Section 7.4.1.

7.2.1 METHOD DERIVATION

The full mathematical description of the FF method is built on a set of object descriptors related to the different phases of the calculation. These terms are summarized in Table 7.1. There are up to three identifiable final state objects in the WH channels and four in the ZH channels, thus it is necessary for the FF method to be able to account for up to four faked objects. To illustrate the method

construction the full derivation for the case in which only two final state objects can be faked (two object case) is presented in Section 7.2.1.1 and the final equations for the three and four object cases are presented in Section 7.2.1.2.

Term	Label	Definition
Event Types		
True Objects	T	‘real’ objects: this is truth-level information that is unknown in data but is useful for method derivation
Fake Objects	F	inverse of true objects
Reconstructions		
Selected	S	Object passes selection criteria (Chapter 6)
Anit-selected	A	Object passes fake criteria (Section 7.3)
Rates		
Fake rate	r	Rate at which Fake Objects (F) are reconstructed as selected (S)
Real efficiency	ϵ	Rate at which True Objects (T) are reconstructed as selected (S)
Fake Factor	f	$\frac{r}{1-r}$

Table 7.1: Fake factor method definitions.

7.2.1.1 TWO OBJECT CASE

The ultimate goal of the FF method is to calculate the number of events of a certain type (ie N_{FF} for number of events where both of the final state objects are actually fakes) that enter the signal region. However, this is truth-level information which is inherently unknowable in data. Thus, the FF method is used to compute these numbers in terms of quantities that are measurable in data.

First, consider event types that can enter the signal region through the FF method. These events must have at most one true object (if they had two true objects they would enter the signal region through normal selection). They can also include any number of fake objects but only a total of two objects can ultimately pass the object identification criteria (be selected). In this derivation it is assumed that true objects are always selected; as discussed in Chapter 3 this is not strictly true and a correction to account for this assumption is described in Section 7.3.5. The probability that a fake

object is selected is r and the probability it is not selected (anti-selected) is $1 - r$. The process for measuring r is described in Section 7.3. A summary of event types and their corresponding probabilities for entering the signal region is shown in Table 7.2. The fake rate for an object depends on its kinematics and so object ordering is necessary.

Event Type	Probability of Entering Signal Region
TF ₁	r
F ₁ F ₂	$r_1 r_2$
TF ₁ F ₂	$(r_1(1 - r_2)) + (r_2(1 - r_1))$
F ₁ F ₂ F ₃	$(r_1 r_2(1 - r_3)) + (r_1 r_3(1 - r_2)) + (r_2 r_3(1 - r_1))$
$\vdots$	$\vdots$

Table 7.2: Truth-level event types and their probabilities of entering the Signal Region.

It is clear from Table 7.2 that the complexity of the selection probabilities rapidly increases with the number of fake objects considered in the event due to the multiple allowed selection combinations. As shown in Section 7.3, in this analysis the fake rates for all object types are less than one. Thus, as the number of fake rates considered increases, the probability of a specific individual reconstruction decreases. Therefore, the remainder of this derivation will consider only first-order events (events with two objects). In the Run-1 analysis this approximation was sufficient for background modeling in all four signal channels and initial studies (Section 7.4.1) indicate this is also sufficient for the Run-2 analysis.

Now the event types can be written in terms of data-set qualities: whether the objects are reconstructed as selected or anti-selected. The three possible first-order two object cases and their reconstructions are shown in Table 7.3. The number of events with each of these reconstructions are shown in Table 7.4.

Event Type	Reconstructions
$F_1 T$	SS, AS
TF_2	SS, SA
$F_1 F_2$	SS, SA, AS, AA

Table 7.3: Possible reconstructions for truth event types.

Event Type	Reconstruction	Number of Events
$F_1 T$	SS	$r_1 \times N_{F_1 T}$
	AS	$(1 - r_1) \times N_{F_1 T}$
TF_2	SS	$r_2 \times N_{TF_2}$
	SA	$(1 - r_2) \times N_{TF_2}$
$F_1 F_2$	SS	$r_1 r_2 \times N_{F_1 F_2}$
	SA	$r_1 (1 - r_2) \times N_{F_1 F_2}$
	AS	$(1 - r_1) r_2 \times N_{F_1 F_2}$
	AA	$(1 - r_1)(1 - r_2) \times N_{F_1 F_2}$

Table 7.4: The number of events that are reconstructed a certain way.

Now the total number of each reconstruction can be found by summing across the underlying event types.

$$N_{SS} = (r_1 N_{F_1 T}) + (r_2 N_{TF_2}) + (r_1 r_2 N_{F_1 F_2})$$

$$N_{SA} = ((1 - r_2) N_{TF_2}) + (r_1 (1 - r_2) N_{F_1 F_2})$$

$$N_{AS} = ((1 - r_1) \times N_{F_1 T}) + ((1 - r_1) r_2 N_{F_1 F_2})$$

$$N_{AA} = (1 - r_1)(1 - r_2) N_{F_1 F_2}$$

Ultimately, only reconstructions with two selected objects (SS reconstructions) will pass the signal region selections. It is thus necessary to express N_{SS} in terms of the other measurable quantities N_{SA} , N_{AS} , and N_{AA} .

Begin by inverting the equation for N_{AA} :

$$N_{F_1 F_2} = \frac{N_{AA}}{(1 - r_1)(1 - r_2)} \quad (7.1)$$

Now consider the equation for N_{AS} :

$$\begin{aligned} N_{AS} &= ((1 - r_1)N_{F_1 T}) + ((1 - r_1)r_2N_{F_1 F_2}) \rightarrow \\ (1 - r_1)N_{F_1 T} &= N_{AS} - (1 - r_1)r_2N_{F_1 F_2} \rightarrow \\ N_{F_1 T} &= \frac{N_{AS}}{(1 - r_1)} - r_2N_{F_1 F_2} \end{aligned} \quad (7.2)$$

Substitute equation 7.1 into equation 7.2:

$$N_{F_1 T} = \frac{N_{AS}}{1 - r_1} - \frac{r_2 N_{AA}}{(1 - r_1)(1 - r_2)}$$

Define the fake factor f of an object as $f = r/(1 - r)$ and finally

$$N_{F_1 T} = \frac{N_{AS}}{1 - r_1} - \frac{f_2 N_{AA}}{1 - r_1}. \quad (7.3)$$

Repeating the same procedure for N_{SA} gives:

$$N_{TF_2} = \frac{N_{SA}}{1 - r_2} - \frac{f_1 N_{AA}}{1 - r_2}. \quad (7.4)$$

Finally, N_{SS} can be written in terms of equations 7.2 - 7.4.

$$\begin{aligned}
N_{SS} &= r_1 N_{F_1 T} + r_2 N_{T F_2} + r_1 r_2 N_{F_1 F_2} \rightarrow \\
N_{SS} &= r_1 \left[\frac{N_{AS}}{1 - r_1} - \frac{f_2 N_{AA}}{1 - r_1} \right] + r_2 \left[\frac{N_{SA}}{1 - r_2} - \frac{f_1 N_{AA}}{1 - r_2} \right] \\
N_{SS} &= f_1 [N_{AS} - f_2 N_{AA}] + f_2 [N_{SA} - f_1 N_{AA}] + f_1 f_2 N_{AA} \\
N_{SS} &= f_1 N_{AS} + f_2 N_{SA} - f_1 f_2 N_{AA} \tag{7.5}
\end{aligned}$$

Using the object specific fake factors it is possible to calculate the number of events with fake objects that enter the signal region in terms of individual event reconstructions.

7.2.1.2 THREE AND FOUR OBJECT CASES

The derivations for the three and four object cases proceed similarly and result in:

$$\begin{aligned}
N_{SSS} &= f_3 N_{SSA} - f_2 f_3 N_{SAA} + f_1 f_2 f_3 N_{AAA} - f_1 f_3 N_{ASA} \\
&\quad + f_2 N_{SAS} - f_1 f_2 N_{AAS} + f_2 N_{ASS}
\end{aligned}$$

for the three object case and

$$\begin{aligned}
N_{SSSS} &= f_1 N_{ASSS} + f_2 N_{SASS} + f_3 N_{SSAS} + f_4 N_{SSSA} - f_1 f_4 N_{ASSA} - f_2 f_4 N_{SASA} \\
&\quad - f_3 f_4 N_{SSAA} - f_2 f_3 N_{SAAS} - f_1 f_2 N_{AASS} - f_1 f_3 N_{ASAS} + f_1 f_2 f_4 N_{AASA} \\
&\quad + f_1 f_3 f_4 N_{ASAA} + f_2 f_3 f_4 N_{SAAA} + f_1 f_2 f_3 N_{AAAS} - f_1 f_2 f_3 f_4 N_{AAAA}
\end{aligned}$$

for the four object case.

7.3 FAKE FACTOR MEASUREMENTS

The fake rate is measured in a phase-space region that is enriched in fake object and should contain minimal true objects (referred to as the Fake Region). When measured in such a region the fake rate, r is defined as $r = \frac{N_S}{N_S + N_A}$. The fake factor is then $f = \frac{r}{1-r} = \frac{N_S}{N_A}$. The fake rates are sensitive to the underlying physics of the event and so in order to model the fake backgrounds accurately it is necessary to select a Fake Region that is as kinematically similar to the Signal Region as possible. For this analysis, the Fake Region is the $Z/\gamma^* \rightarrow \mu\mu$ region which is further defined in Section 7.3.1. After constructing the Fake Region using data, the number of selected and anti-selected are measured and the fake rates and factors are calculated.

The fake rates and factors of each object are measured separately for the ZH and WH channels due to a number of differences in the kinematics of the categories. This includes the presence of additional MET in the WH channels and differing object origins between the two categories.

7.3.1 FAKE REGION

The fake rates and factors for all three object types are measured in the $Z/\gamma^* \rightarrow \mu\mu + \text{jets}$ region (referred to as $Z \rightarrow \mu\mu$ for short). The muons that originate from the Z decay should be the only true leptons in the event and thus any additional lepton is likely a mis-identified jet (fake).

The $Z \rightarrow \mu\mu$ region is constructed by first requiring at least one of the two muon triggers used in the signal region (HLT_2mu14 and HLT_mu26_ivarmedium) is fired. Events are additionally required to have two isolated oppositely charged muons with a combined mass between 71 and 111 GeV and no b-jets. An example of the $Z \rightarrow \mu\mu$ purity in this region is shown in Figure 7.2: the vast majority of MC events that pass the Fake Region selection are $Z \rightarrow \mu\mu$ processes.

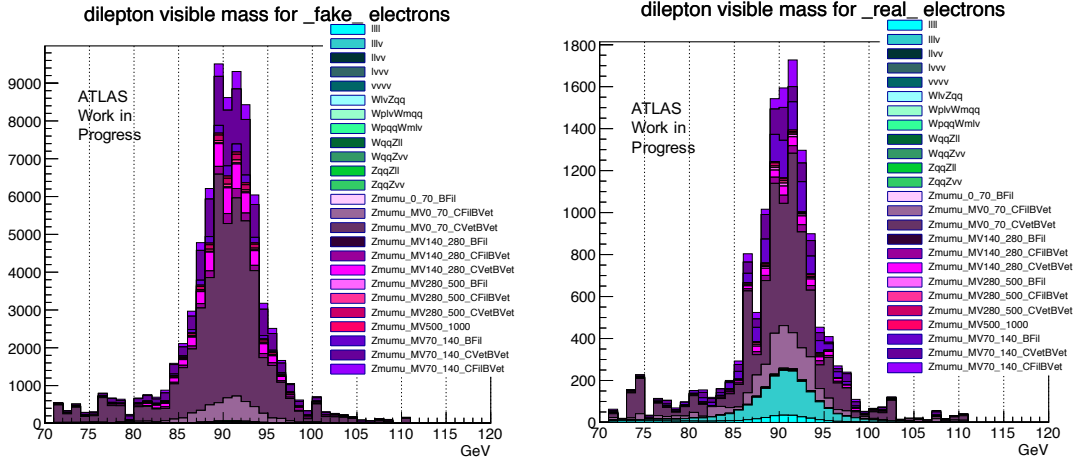


Figure 7.2: Di-muon visible mass in the $Z \rightarrow \mu\mu$ Fake Region for events with an anti-selected electron (left) and a selected electron (right).

7.3.2 FAKE CANDIDATE SELECTION

Anti-selected objects are required to pass the same object criteria described in Chapter 6 with the following differences:

- Anti-selected electrons must fail the Loose ID operating point rather than pass the Medium ID operating point.
- Anti-selected taus must fail rather than pass the BDT-Medium ID operating point but pass a loosened ID of $0.7 \times \text{BDT-Loose}$.
- Anti-selected muons must fail rather than pass the Gradient Isolation operating point.

The overlap removal resolution order for anti-selected objects is taus > electrons > muons. If two anti-selected objects of the same type overlap, the one with the greater p_T is chosen.

7.3.3 FAKE ELECTRONS

Fake electrons can originate from a variety of sources including mis-identified jets, light meson decays, heavy flavor hadron decays, and photon conversions. The electron ID algorithm depends heav-

ily on p_T and η and so the electron FRs and FFs are measured for a 2D phase-space of seven η bins and four p_T bins. The bin divisions were initially set in Run-1 to achieve roughly consistent statistical error across bins. This same binning is currently used for the Run-2 analysis with an additional low p_T bin. The FRs and FFs are measured separately for the ZH and WH channels due to different fake electron compositions as described further in the following section. The ZH Fake Region applies a tighter di-muon mass cut of 81-101 GeV while the WH Fake Region loosens the mass cut to 61-121 GeV.

The electron FRs measured in the full 2017 data-set are shown in Figures 7.3 and 7.4 and summarized in Tables 7.5 and 7.6. The highest p_T bin has low statistics and has been removed from the plots for ease of presentation. The final analysis will use the full Run-2 data-set which will improve the statistics in this bin and yield a more reliable FR measurement.

η	$5 < p_T < 10$	$10 < p_T < 15$	$15 < p_T < 20$	$20 < p_T$
-2.5 - -2.01	0.090	0.039	0.072	0.455
-2.01 - -1.52	0.083	0.139	0.243	0.337
-1.37 - -0.46	0.104	0.152	0.263	0.321
-0.46 - 0.46	0.179	0.279	0.275	0.297
0.46 - 1.37	0.106	0.154	0.280	0.313
1.52 - 2.01	0.093	0.129	0.257	0.295
2.01 - 2.5	0.085	0.039	0.077	0.381

Table 7.5: Measured electron fake rate for the ZH channels.

The FRs increase with increasing p_T and are highly symmetric in η (thus the analysis could use only four bins of $|\eta|$ as is done for taus).

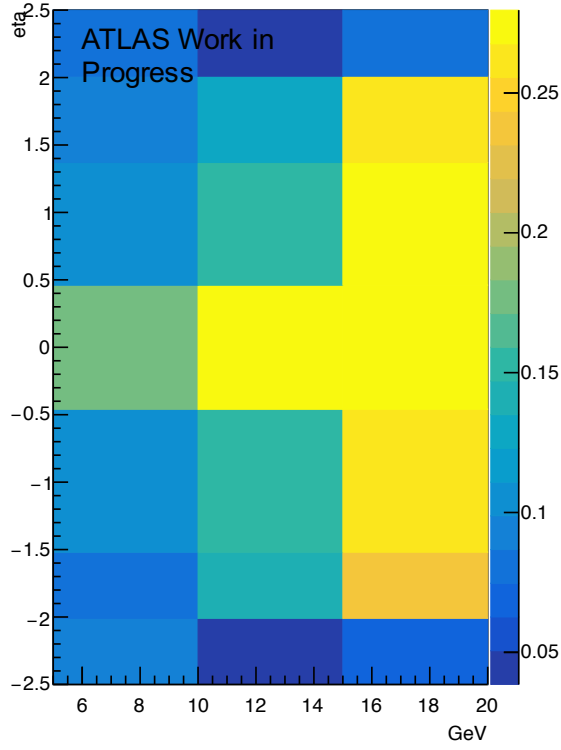


Figure 7.3: Measured electron fake rate for the ZH channels.

η	$5 < p_T < 10$	$10 < p_T < 15$	$15 < p_T < 20$	$20 < p_T$
-2.5 - -2.01	0.096	0.045	0.095	0.437
-2.01 - -1.52	0.086	0.155	0.245	0.330
-1.37 - -0.46	0.109	0.164	0.274	0.330
-0.46 - 0.46	0.181	0.292	0.295	0.308
0.46 - 1.37	0.109	0.171	0.307	0.310
1.52 - 2.01	0.097	0.141	0.286	0.308
2.01 - 2.5	0.092	0.045	0.099	0.374

Table 7.6: Measured electron fake rate for the WH channels.

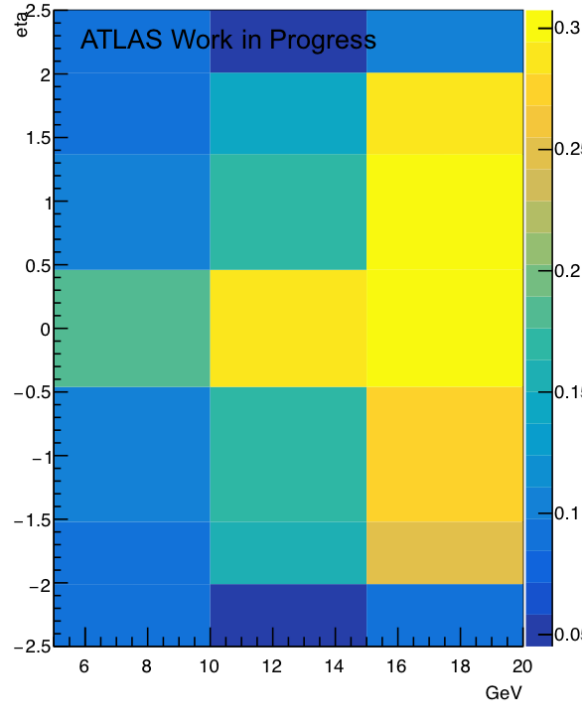


Figure 7.4: Measured electron fake rate for the WH channels.

7.3.3.1 TYPE AND ORIGIN STUDIES

In Run-1 it was found that the electron compositions varied significantly between the ZH and WH channels. In particular, the fake electrons in ZH channel were predominately mis-identified jets. The tighter di-muon mass requirement for the ZH FR measurement reduced the amount of final state radiation and thus the number of photon conversion fake electrons. On the other hand, fake electrons in the WH channel were equally split between mis-identified jets and converted photons. The low di-muon mass range (61-81 GeV) is enriched in $Z \rightarrow \mu\mu$ decays with additional final state radiation and thus loosened mass requirement for the WH FR measurement increased the number of photon conversion fake electrons.

Studies are on-going to understand if these same Fake Region definitions are necessary and suf-

ficient for the Run-2 electron FR measurement. The electron composition of the ZH Fake Region is shown in Figure 7.5 and the composition of the WH Fake Region is shown in Figure 7.7. The electron composition for the ZH and WH signal regions are shown in Figures 7.6 and 7.8. These plots all use the E/Gamma Working Group supported electron origin classification scheme which is summarized in Table 7.7.

Number	Electron Origin
0	Not Defined
Prompt Objects	
1	Prompt Electron
2	Single Muon
3	Single Photon
4	Single Tau
Conversions	
5	Photon Conversion
Particle Decays	
6	Dalitz Decay
8	Muon Decay
9	Leptonic Tau Decay
12	W Decay
13	Z Decay
14	Higgs Decay
28	J/Ψ Decay
Jets	
10	Top Jet
11	QCD Showering
23	Light Jet
24	Strange Jet
25	Charm Jet
26	Bottom Jet
27	$c\bar{c}$ Production

Table 7.7: Electron origin numbering scheme.

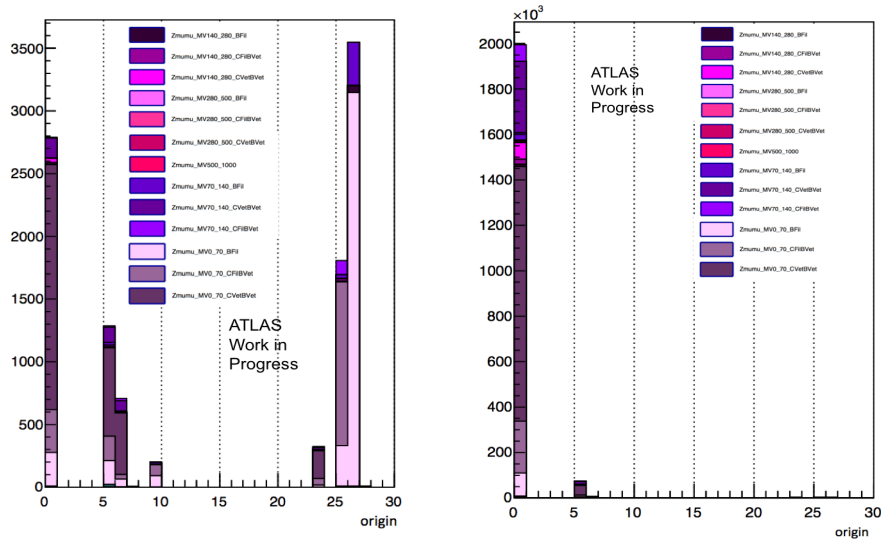


Figure 7.5: Electron origins (as defined in Table 7.7) in the ZH Fake Region. Selected electrons on the left and anti-selected electrons on the right.

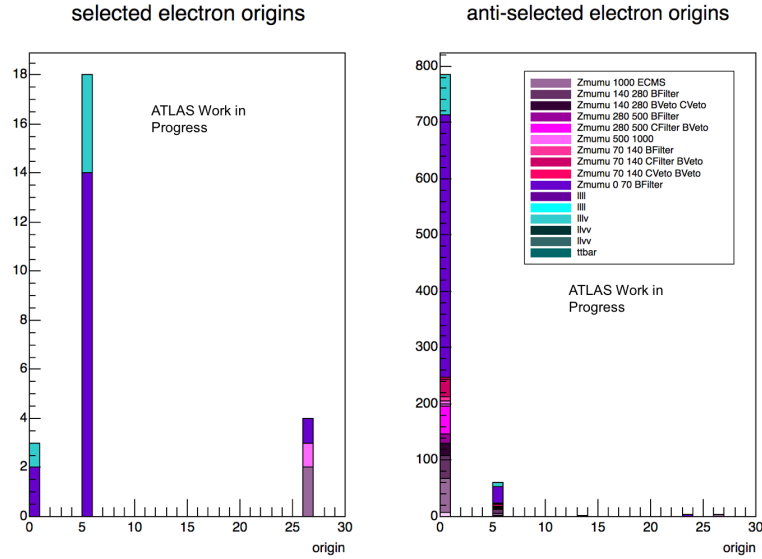


Figure 7.6: Electron origins (as defined in Table 7.7) in the ZH Signal Region. Selected electrons on the left and anti-selected electrons on the right

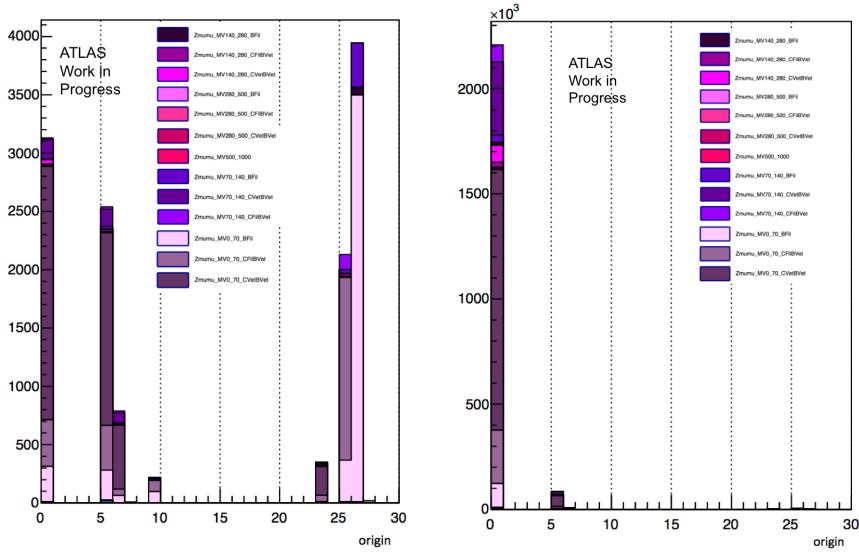


Figure 7.7: Electron origins (as defined in Table 7.7) in the WH Fake Region. Selected electrons on the left and anti-selected electrons on the right.

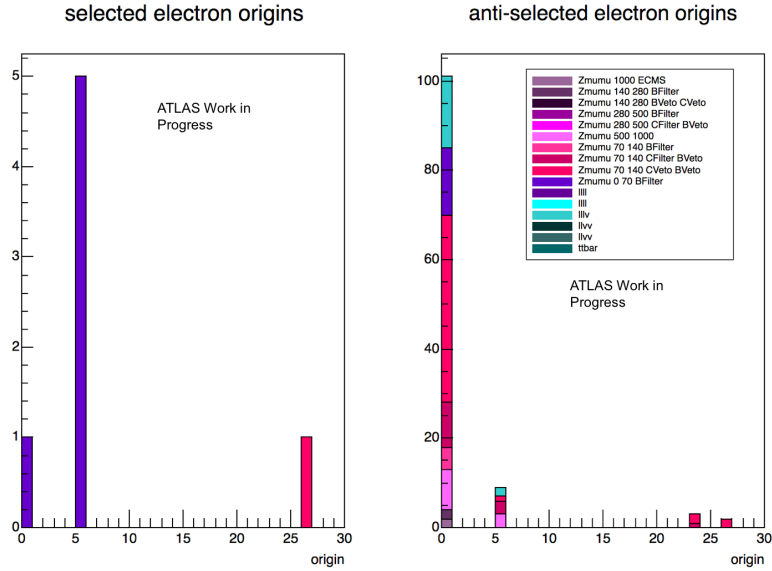


Figure 7.8: Electron origins (as defined in Table 7.7) in the WH Signal Region. Selected electrons on the left and anti-selected electrons on the right

Initial Run-2 studies indicate that both the ZH and WH signal region selected electron origins distributions are dominated by photon-conversions. In fact, both the selected and anti-selected origin distributions are similar across the two signal regions. This indicates it may not be necessary to measure the electron FRs and FFs separately for each category. On the other hand, the origin distributions in the fake region are dominated by mis-identified jets. Additionally work is necessary to ensure the electron composition in the Fake Region mimics that of the Signal Region.

7.3.4 FAKE TAUS AND MUONS

7.3.4.1 FAKE TAUS

The tau FRs and FFs are measured in a 2D phase-space of 5 p_T bins and 3 $|\eta|$ bins. The FFs and FRs are measured separately for the ZH and WH categories. In the fake tau measurement these categories are separated by requiring $MET > 20$ GeV for the WH measurement; there is no MET requirement for the ZH measurement. The FRs and FFs for 1-prong taus are typically higher than those of 3-prong taus and so they are measured separately for these categories as well. The tau FRs measured in the full 2017 data-set are summarized in Tables 7.8 - 7.11.

p_T (GeV)	$ \eta < 0.8$	$0.8 < \eta < 1.37$	$1.37 < \eta < 2.5$
$25 < p_T < 30$	0.168	0.146	0.158
$30 < p_T < 35$	0.161	0.151	0.149
$35 < p_T < 40$	0.157	0.149	0.140
$40 < p_T < 60$	0.131	0.114	0.118
$60 < p_T$	0.093	0.083	0.079

Table 7.8: Measured 1-prong tau fake rates for the ZH channels.

p_T (GeV)	$ \eta < 0.8$	$0.8 < \eta < 1.37$	$1.37 < \eta < 2.5$
$25 < p_T < 30$	0.032	0.034	0.033
$30 < p_T < 35$	0.031	0.031	0.033
$35 < p_T < 40$	0.030	0.031	0.030
$40 < p_T < 60$	0.025	0.023	0.021
$60 < p_T$	0.016	0.011	0.001

Table 7.9: Measured 3-prong tau fake rates for the ZH channels.

p_T (GeV)	$ \eta < 0.8$	$0.8 < \eta < 1.37$	$1.37 < \eta < 2.5$
$25 < p_T < 30$	0.133	0.102	0.132
$30 < p_T < 35$	0.125	0.109	0.121
$35 < p_T < 40$	0.117	0.113	0.112
$40 < p_T < 60$	0.106	0.091	0.098
$60 < p_T$	0.083	0.069	0.068

Table 7.10: Measured 1-prong tau fake rates for the WH channels.

p_T (GeV)	$ \eta < 0.8$	$0.8 < \eta < 1.37$	$1.37 < \eta < 2.5$
$25 < p_T < 30$	0.029	0.026	0.027
$30 < p_T < 35$	0.027	0.025	0.025
$35 < p_T < 40$	0.024	0.022	0.024
$40 < p_T < 60$	0.019	0.022	0.021
$60 < p_T$	0.014	0.012	0.013

Table 7.11: Measured 3-prong tau fake rates for the WH channels.

Jets can be formed through quark-initiated and gluon-initiated processes and the probability for a hadronic jet to be mis-identified as a tau depends on its origin. Thus the central challenge of modeling the fake tau background in this analysis is ensuring that the jet origin composition in the Fake Region matches that of the Signal Region as closely as possible. Studies to ensure this for the Run-2 analysis are on-going.

7.3.4.2 FAKE MUONS

The muon FRs and FFs are measured in a 2D phase-space of 3 p_T bins and 3 η bins. Initial muon fake rates measured in a subset of 2017 data are shown in Table 7.12.

p_T (GeV)	$-2.5 < \eta < -1.2$	$-1.2 < \eta < 1.2$	$1.2 < \eta < 2.5$
$p_T < 15$	0.666	0.422	0.700
$15 < p_T < 25$	0.631	0.440	0.691
$25 < p_T$	0.673	0.436	0.783

Table 7.12: Preliminary measured muon fake rates.

There are multiple challenges involved in measuring the muon FRs. These include tuning the isolation requirement on the anti-selected muons to appropriately select jets faking muons but not muons produced inside jets and understanding the interaction of trigger-based and offline isolation requirements.

7.3.5 MC CORRECTIONS

The FF method derivation described in Section 7.2.1 relies on the assumption that all true objects pass their corresponding selection requirements. It is additionally assumed that all objects passing selection requirements in the Fake Region are fakes. Neither assumption is strictly true but they can both be easily accounted for in the final measurement.

To account for these assumptions, the fake factor (previously $f(p_T, \eta) = \frac{N_S}{N_A}$) is redefined as:

$$f(p_T, \eta) = \frac{N_S^{data} - N_S^{MC}}{N_A^{data} - N_A^{MC}}.$$

The MC terms are measured by combining truth-matched prompt objects from all physics processes that could pass the signal region selections (Chapter 6). In practice these are primarily di-boson processes. The MC-based subtraction is included in the numerator to remove contamination of true

objects in the Fake Region and in the denominator to account for true objects that do not pass the object selection criteria.

These corrections are not included in the initial FR measurements presented above but will be incorporated in the final background model.

7.4 FAKE FACTOR SOFTWARE

In order to measure the fake background contributions to the final analysis it is necessary to apply the measured FFs to the data in each analysis channel according to the equations derived in Section 7.2.1. In practice, this requires a complex software implementation. The FF software must first incorporate modified analysis channel selections that do not impose object identification requirements (or isolation requirements in the muon case). It must then construct all possible allowed combinations of final state objects for each data event and save each combination as a new event that is weighted appropriately according to the FF equations. These weighted events then populate the background model in each analysis category.

For the $VH, H \rightarrow \tau\tau$ analysis in particular, in order to accommodate all the studies necessary to validate the FF background model, the FF software must be highly flexible. It must be possible to consider multiple combinations of fake object types (as few as one or as many as three), different numbers of final state objects that are faked, and various numbers of fake candidates considered per event. Furthermore, the FF software must interface with the pre-selection and selection criteria for all four analysis channels as well as multiple orthogonal control regions that are used to validate the background model.

Other ATLAS analyses have implemented a data-based FF background model (see e.g. [175] and [176]) though they typically do not consider the multiple fake object types, numerous fake candidates, or relatively high number of final state objects necessary for the $VH, H \rightarrow \tau\tau$ analysis. An

extensive study of existing ATLAS FF software was conducted which ultimately indicated it was necessary to develop a dedicated $VH, H \rightarrow \tau\tau$ FF software package. This package currently allows up to three fake objects in the final state, any number of fake object candidates per event, and to include fake electrons, taus, or both in the background model; additional functionality is being developed.

7.4.1 CLOSURE TESTS

The FF background model and software are validated independently for each analysis channel using closure tests. Closure tests compare the data to the background model in the pre-selection regions and orthogonal control regions. Closure tests can additionally provide information for tuning the final fake background model including which types of fake objects need to be considered and if the binning used to measure the object FFs accurately capture the fake distributions.

An example of a Run-1 closure test for the $Z \rightarrow \tau\tau$ control region in the WH lep-had channel is shown in Figure 7.9. Fakes are the primary source of background in all analysis channels and thus it is expected that in the closure test distributions the fake background approximates the data distribution.

7.4.1.1 WH HADHAD CHANNEL

Initial closure tests for the WH had-had channel using a subset of 2018 data corresponding to a luminosity of roughly 30 fb^{-1} are presented in the following two sections. For these tests, only fake taus are considered and all tau candidates in an event are considered.

7.4.1.1.1 PRE-SELECTION

The WH had-had pre-selection region is defined by including all the analysis channel requirements described in Chapter 6 with the exception of the b-jet veto and the lepton-MET $p_T > 20$

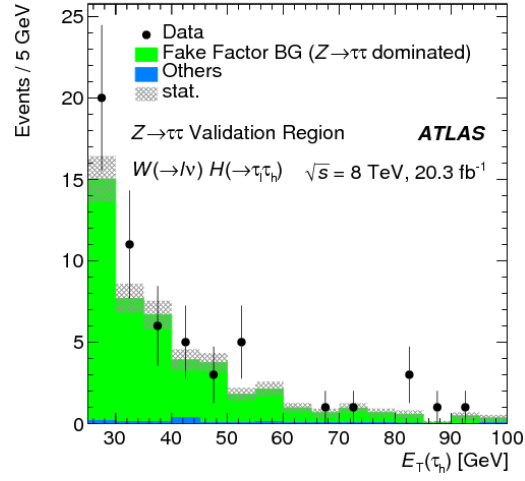


Figure 7.9: Run-1 distribution of the hadronic tau E_T as closure test for the fake factor method in the WH lep-had channel [171].

GeV requirement. Results of the closure test in this region for a variety of kinematics are shown in Figures 7.10 - 7.12. Here, tau-1 refers to the highest p_T tau attributed to the Higgs.

These closure tests unfortunately have low statistics (and correspondingly large uncertainties) due to considering only a small subset of the full Run-2 data-set. Nevertheless, the fake background distributions generally follow the pattern of the data distributions and thus these can be considered a ‘proof of principle’ for the FF software.

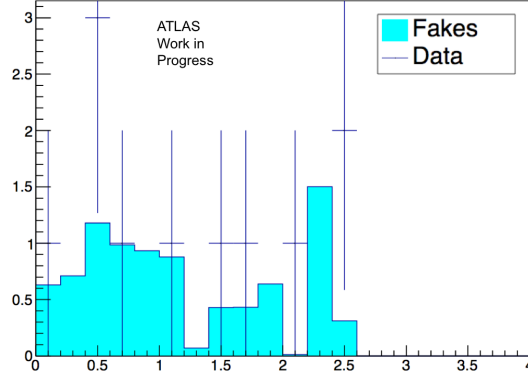


Figure 7.10: WH had-had pre-selection closure test $\tau_1 \eta$ distribution.

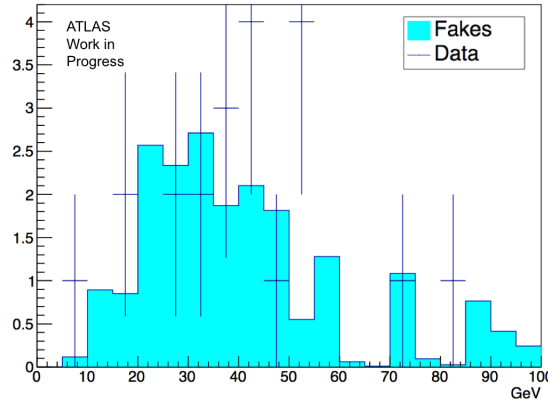


Figure 7.11: WH had-had pre-selection closure test lepton p_T distribution.

7.4.1.1.2 $Z \rightarrow \tau\tau$ CONTROL REGION

The $Z \rightarrow \tau\tau$ control region for the WH had-had channel is constructed by applying the pre-selection criteria described above with the exception of the oppositely charged tau requirement and with the additional requirements of $\text{lep-MET } p_T < 40 \text{ GeV}$ and $M_{2T} < 60 \text{ GeV}$. Preliminary results of the closure tests in this region are shown in Figures 7.13 and 7.14. As expected, the distributions exhibit the same low statistics effects seen in the pre-selection region. Additionally, as allowed by the FF equations, some events in the background distribution receive negative weights.

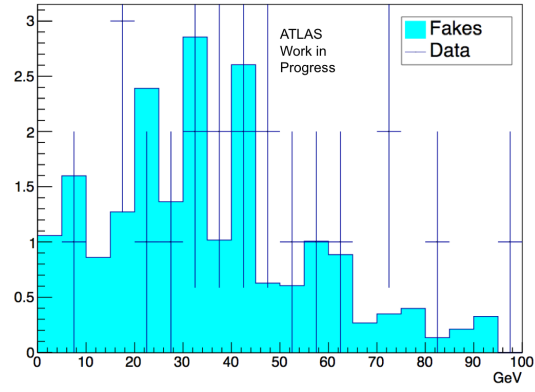


Figure 7.12: WH had-had pre-selection closure test MET distribution distribution.

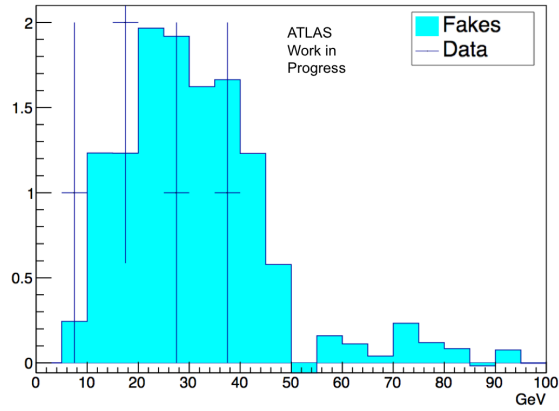


Figure 7.13: WH had-had $Z \rightarrow \tau\tau$ closure test lepton p_T distribution.

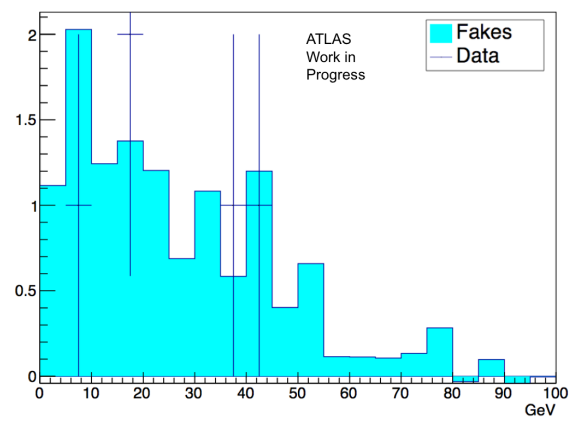


Figure 7.14: WH had-had $Z \rightarrow \tau\tau$ closure test MET distribution.

*I fly a star ship across the universe divide and when I
reach the other side; I'll find a place to rest my spirit if
I can. Perhaps, I may become a highwayman again or
I may simply be a single drop of rain but I will remain
and I'll be back again and again and again and again
and again and again.*

Willie Nelson

8

Conclusions

This thesis presented the role that computing, machine learning, and electronics play in ATLAS and the $VH, H \rightarrow \tau\tau$ analysis specifically.

Chapter 1 introduced the Standard Model and the particles and forces it describes. It also laid out the basic mathematical formalism, in terms of Quantum Field Theory, of the Standard Model and how that formalism can be used to make predictions that are then measurable at the Large Hadron Collider.

Chapter 2 described the design and operation of the LHC and detailed the various physical components of the ATLAS detector and how they are optimized for their individual physics tasks.

Chapter 3 explained the flow of data within ATLAS data flow from the trigger level through storage on the Worldwide LHC Computing Grid. It also described the various software packages

used to simulate MC ATLAS events and the algorithms used to reconstruct, identify, and calibrate individual physics objects.

Chapter 4 introduced the core Machine Learning concepts appropriate for physics data analyses and provided an overview of the current status of ML in LHC experiments.

Chapter 5 extensively detailed the design, implementation, and functionality of the ATLAS electron identification algorithm. It also presented novel work to provide a fully data-driven ID, extend the p_T range of the algorithm, and use image-based ML to design an improved ID.

Chapter 6 described the basic components of the ATLAS Run-2 $VH, H \rightarrow \tau\tau$ analysis including the object criteria, analysis category definitions, and mass estimation techniques. It also presented the results of the corresponding Run-1 analysis and introduced planned improvements in the Run-2 workflow.

Finally, Chapter 7 focused on the difficult process of modeling the fake backgrounds in the $VH, H \rightarrow \tau\tau$ analysis. It presented a mathematical derivation of the Fake Factor Method, described the process of measuring object fake rates, and showed preliminary electron, muon, and tau fake rates. It also motivated the need for an analysis specific FF software, described its implementation, and presented preliminary validation studies.

8.1 $VH, H \rightarrow \tau\tau$ ANALYSIS

8.1.1 ANALYSIS STATUS

The ATLAS $VH, H \rightarrow \tau\tau$ analysis is on-going with a small but dedicated team who interface with and receive support from the larger ATLAS $H \rightarrow \tau\tau$ analysis group. Although the studies presented here were done with only one year of data (either 2017 or 2018) the final analysis will incorporate the full four year, Run-2 139fb^{-1} data-set.

At the time of writing, the Run-2 analysis category selections (and corresponding Control Re-

gion selections) have been implemented. Their efficiency is being evaluated and small improvements (like the electron p_T floor reduction described in Chapter 6) are being explored. The mass estimation methods (MMC and M_{2T} , Chapter 6) have also been implemented. Preliminary fake rates and fake factors have been measured for electrons, muons, and taus (Chapter 7). Finally, as described in Chapter 7, the core functionality of the analysis specific Fake Factor application software has been built and initial closure test studies have begun.

8.1.2 EXPECTED EVENTS

The number of expected $VH, H \rightarrow \tau\tau$ in the full Run-2 data set can be calculated for ZH and WH channels. The basic equation used for this calculation is $n = \mathcal{L} \times \sigma_{VH} \times \text{BR}(H \rightarrow \tau\tau) \times \text{BR}(\text{allowed } \tau \text{ decays}) \times \text{BR}(V \rightarrow \text{light } l)$ where the allowed τ decays are all combinations of two τ decay modes except lep-lep. The predictions from this equation using cross-section from the 2018 CERN Higgs Yellow Report [177] and branching ratios from the Particle Data Group [12] are shown below. Notably, the cross-section differs from the Run-1 cross-section due to the higher center-of-mass energy.

Analysis Channel	Expected Events in Run-2
ZH	161.6
WH	123.8

Table 8.1: Expected event yield in the full Run-2 data-set for the ZH and WH channels of the $VH, H \rightarrow \tau\tau$ analysis.

Despite the large amount of Run-2 data recorded by ATLAS, the expect event numbers are quite low. This underscores the importance of building highly efficient analysis channel selections.

8.1.3 MACHINE LEARNING IN THE ANALYSIS

There are several ways ML could improve the Run-2 $VH, H \rightarrow \tau\tau$ analysis. One method is developing an ML-based classifier to remove the irreducible backgrounds. This strategy is well motivated: a Run-1 study showed that a BDT trained to separate signal from background gave a potential upper limit improvement of 30% (when fitting all analysis categories separately). Furthermore, the CMS Run-2 $H \rightarrow \tau\tau$ analysis utilized a set of NNs (one for each of the three final states for each year of data) to distinguish between two signal categories and several background categories. The networks achieved substantial separation between all classes and considering each background process separately improved signal purity [178].

This technique is currently being studied for use in the $VH, H \rightarrow \tau\tau$. A set of NNs (one for each analysis category) has been implemented to separate signal from diboson backgrounds. They require additional MC statistics to train properly and these samples are currently in production. An extension of this study would aim to additionally separate signal from fake backgrounds by training the NNs on a combination of diboson MC, signal MC, and data processed through the FF software. If this method is used in the final analysis, the NN output distributions would be fit as described in [178], eliminating the dependence on MMC and M_{2T} to estimate the di-tau mass spectrum.

In the case that the above method is unsuccessful or not incorporated into the final analysis, ML could also be used to improve the mass estimation techniques. As described in Chapters 4 and 7, BDTs, NNs, and imaged-based CNNs have been used successfully to calibrate particle energy and mass in other instances. Rather than relying on geometric reconstructions and probability distributions to approximate the missing mass, an ML algorithm could be used to predict a mass correction factor. It is conceivable that this technique could make use of additional event information that is not used in the MMC and M_{2T} methods.

8.2 FINAL REMARKS

This work has demonstrated the central role that computing and software play in the design and functioning of the LHC and ATLAS and HEP physics analyses in general. The importance of efficient algorithm design will only increase as the field begins to address storage and processing challenges of the HL-LHC.

The LHC physics community can, and should, benefit from strengthened collaborations with other computing research areas, including ML. The domain specific constraints of particle physics experiments yield interesting computing problems and the opportunity for both fields to develop exciting innovations. It has been inspiring to see how these collaborations have already begin to grow during my time in graduate school. There are many people on both sides doing excellent work to train researchers, homogenize tools across experiments, and construct more flexible software tools and data formats.

This work requires continued investment and shifts in perception, but I am excited to see it continue to develop.



Author's Individual Contributions

As is clear from this thesis, ATLAS work is highly collaborative. I am lucky to have worked with many groups with ATLAS and CERN more broadly, including both Professor Keith Baker's and Professor Sarah Demer's research groups at Yale, the ATLAS E/Gamma Working Group, the ATLAS Tau Working Group, the ATLAS HLep Physics Analysis Group, and the CERN Inter-experimental Machine Learning Working Group (IML). I list below some of my individual contributions to these efforts.

A.1 ELECTRON ID

- I supported (with another colleague) the electron ID for two and a half years. This involved answering questions from users, providing retunings of the ID when ATLAS software was

updated, and providing documentation and training materials to ATLAS physicists.

- I was fundamental in introducing ML techniques to the E/Gamma software development group. In particular, I developed the idea of using image-based CNNs for electron ID.
- I helped develop the software to produce images from EM calorimeter cells.
- I built an initial CNN to discriminate electron images from background images, and have been studying the feasibility of using a 3D convolution in this architecture.
- I developed the software to produce a data-driven $J/\psi \rightarrow ee$ -based electron ID for low p_T electrons. Additionally, I implemented the functionality necessary to provide electron ID support down to $p_T = 4.5$ GeV, compared to 7 GeV in Run-1.
- I performed studies to optimize the low p_T electron ID Likelihood variables including implementing new variables.

A.1.1 RELATED PRESENTATIONS

- ‘Electron ID Optimization at Low p_T ’: invited talk at ATLAS E/Gamma Workshop (November 2016)
- ‘Electron ID Optimization at Low p_T ’: invited talk at ATLAS TRT Days (February 2017)
- ‘Electron ID in ATLAS Run 2’: poster presented at LHCP (May 2017)
- ‘Electron ID in ATLAS Run 2’: contributed talk at US ATLAS Meeting (July 2017)
- ‘The Future of Electron ID’: invited talk at ATLAS E/Gamma Workshop (November 2017)
- ‘Convolutional Neural Networks for Electron ID in the ATLAS Detector’: poster presented at Women in Machine Learning (December 2017)
- ‘Electron ID in ATLAS Run 2’: poster presented at HEP2018 (January 2018)
- ‘Visualizing Electrons in ATLAS’: invited student talk at USLUA Meeting (October 2018), lightning round winner

- ‘Imaging Electrons in ATLAS’: contributed talk at APS April Meeting (April 2019)
- Publication: “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s}=13\text{TeV}$ ”, 2018, submitted to European Physics Journal
- Publication: “Electron efficiency measurements with the ATLAS detector using 2015 LHC proton-proton collisions”, 2016, ATLAS-CONF-2016-024

A.2 $VH, H \rightarrow \tau\tau$ ANALYSIS

- I helped produce simulation and data samples for the analysis.
- I demonstrated potential increased signal yield in lep-had channels by lowering the minimum electron p_T .
- I measured the electron fake rates and fake factors for the ZH and WH analysis categories
- I conducted electron origin studies to ensure that the ZH and WH fake region selections accurately approximate the fake composition in the signal region.
- I derived the FF application formulas for two, three, and four final state objects.
- I developed the FF application software including modified signal and control region selections.

A.2.1 RELATED PRESENTATIONS

- ‘Lepton Fake Rates in $VH, H \rightarrow \tau\tau$ ’: contributed talk at ATLAS Tau Workshop (October 2017)
- ‘Solutions and Improvements for $VH, H \rightarrow \tau\tau$ Run 2 Analysis’: invited student talk at USLUA Meeting (November 2017)

- ‘VH Analysis Status’: invited talk at HLeptons Workshop (November 2018)

A.3 MACHINE LEARNING

- I demonstrated improved signal selection in the $H \rightarrow ZZ_d \rightarrow lll$ analysis using BDTs and NNs.
- I co-organized a workshop on Deep Learning for Physical Sciences at NeurIPS 2017.
 - Submitted a proposal for Deep Learning and Physical Science for NeurIPS 2019.
- I’ve been actively involved in the ATLAS ML Working Group and the CERN IML working group including giving talks, developing tutorials, and providing documentation.
- This summer I will be teaching a hands-on ML course at the Princeton CoDAS summer school.

A.3.1 RELATED PRESENTATIONS

- ‘Machine Learning for Jet Physics’: group project at the SLAC Summer Institute, winner of the school-wide competition
- ‘Search for a Dark Z Boson with Machine Learning at ATLAS’: contributed talk at Dark Interactions Workshop (October 2016)
- ‘Machine Learning in ATLAS’: invited talk at SYNPA (December 2017)
- ‘Machine Learning in ATLAS’: Colloquium at Hanyang University (December 2017) item - ‘NeurIPS 2017 Summary’: invited talk in the IML forum
- Publication: “Machine Learning in High Energy Physics Community White Paper”, Proceedings, 18th International Workshop on Advance Computing and Analysis Techniques in Physics Research, 2018

- ‘Modeling Opioid Abuse Indicators and Interventions in Appalachia’: poster presented at ICLR AI for Social Good Workshop (May 2019)

A.4 OUTREACH

- I’ve helped in developing content for and managing the ATLAS social media accounts for the past 4 years, including being the primary manager of the Instagram account.
- I developed the ‘Physicist Fridays’ series to highlight young members of the collaboration.
- I helped promote ATLAS outreach efforts including coloring books, web material, publications, and teacher resources.
- I helped develop science communication trainings for HEP physicists.
- I participated in annual trips to DC to advocate for continued federal investments in basic science research.

A.4.1 RELATED PRESENTATIONS

- ‘Data Collection and Analysis at the ATLAS Detector’: poster at Yale Day of Data (November 2016)
- ‘Engaging Younger Audiences Using Instagram’: invited talk at ATLAS Week (February 2018)
- ‘ATLAS Social Media’: invited talk at US ATLAS Meeting (August 2018)
- Publication: “Social Media Strategy for the ATLAS Experiment”, 2016, Proceedings of Science 38th International Conference on High Energy Physics

References

- [1] David J Griffiths. *Introduction to elementary particles; 2nd rev. version*. Physics textbook. Wiley, New York, NY, 2008.
- [2] Mark Thomson. *Modern particle physics*. Cambridge University Press, New York, 2013.
- [3] Matthew D. Schwartz. *Quantum Field Theory and the Standard Model*. Cambridge University Press, 2014.
- [4] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to quantum field theory*. Addison-Wesley, Reading, USA, 1995.
- [5] Mark Srednicki. *Quantum Field Theory*. Cambridge Univ. Press, Cambridge, 2007.
- [6] The ATLAS Collaboration. Summary plots from the atlas standard model group. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/>.
- [7] G. G. Raffelt. Dark matter: Motivation, candidates and searches. In *1997 European School of High-energy physics, Menstrup, Denmark, 25 May-7 Jun, 1997: Proceedings*, pages 235–278, 1997.
- [8] Edmund J. Copeland et al. *International Journal of Modern Physics D*, 15(11):1753–1935, 2006.
- [9] Frederik Rühr. Prospects for BSM searches at the high-luminosity LHC with the ATLAS detector. *Nucl. Part. Phys. Proc.*, 273-275:625–630, 2016.
- [10] Julian S. Schwinger. A Theory of the Fundamental Interactions. *Annals Phys.*, 2:407–434, 1957.
- [11] Hilary Greaves and Teruji Thomas. On the CPT theorem. *Stud. Hist. Phil. Sci.*, B45:46–65, 2014.
- [12] M. et al Tanabashi. Review of particle physics. *Phys. Rev. D*, 98:030001, Aug 2018.
- [13] D. Decamp et al. Determination of the Number of Light Neutrino Species. *Phys. Lett.*, B231:519–529, 1989.
- [14] Abdus Salam. The electroweak force, grand unification and superunification. *Physica Scripta*, 20(2):216–226, aug 1979.
- [15] Peter W. Higgs. Broken symmetries and the masses of gauge bosons. *Phys. Rev. Lett.*, 13:508–509, Oct 1964.

- [16] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Phys. Rev. Lett.*, 13:321–323, Aug 1964.
- [17] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. Global conservation laws and massless particles. *Phys. Rev. Lett.*, 13:585–587, Nov 1964.
- [18] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [19] Serguei Chatrchyan et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett.*, B716:30–61, 2012.
- [20] John Ellis. Higgs Physics. (arXiv:1312.5672. KCL-PH-TH-2013-49. LCTS-2013-36. CERN-PH-TH-2013-315):117–168. 52 p, Dec 2013. 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report.
- [21] Bruce Mellado. Prospects for higgs boson searches at the large hadron collider. *Pramana*, 72:15–22, 01 2009.
- [22] The ATLAS Collaboration. Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector. *Phys. Lett.*, B784:173–191, 2018.
- [23] The ATLAS Collaboration. Observation of $H \rightarrow b\bar{b}$ decays and VH production with the ATLAS detector. *Phys. Lett.*, B786:59–86, 2018.
- [24] M. Aaboud et al. Search for the Decay of the Higgs Boson to Charm Quarks with the ATLAS Experiment. *Phys. Rev. Lett.*, 120(21):211802, 2018.
- [25] Cross-section measurements of the Higgs boson decaying to a pair of tau leptons in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. Technical Report ATLAS-CONF-2018-021, CERN, Geneva, Jun 2018.
- [26] M. Cepeda et al. Higgs Physics at the HL-LHC and HE-LHC. 2019.
- [27] The ATLAS Collaboration. Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector. *Phys. Rev.*, D92(1):012006, 2015.
- [28] Mila Pandurović. Measurement of Higgs decay to WW^* in Higgsstrahlung at $\sqrt{s}=500$ GeV ILC and in WW -fusion at $\sqrt{s}=3$ TeV CLIC. In *Proceedings, International Workshop on Future Linear Colliders 2016 (LCWS2016): Morioka, Iwate, Japan, December 05-09, 2016*, 2017.
- [29] The ATLAS Collaboration. Searches for the $Z\gamma$ decay mode of the Higgs boson and for new high-mass resonances in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:112, 2017.

- [30] Frank F. Deppisch, Wei Liu, and Manimala Mitra. Long-lived Heavy Neutrinos from Higgs Decays. *JHEP*, 08:181, 2018.
- [31] Arindam Das, Yu Gao, and Teruki Kamon. Heavy Neutrino Search via the Higgs boson at the LHC. 2017.
- [32] Rolf Landau. Particle physics - a historical overview. <https://indico.cern.ch/event/3324/>, July 2006.
- [33] J. L. Hewett et al. Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 2: Intensity Frontier. In *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2014.
- [34] R. Brock et al. Planning the Future of U.S. Particle Physics (Snowmass 2013): Chapter 3: Energy Frontier. In *Proceedings, 2013 Community Summer Study on the Future of U.S. Particle Physics: Snowmass on the Mississippi (CSS2013): Minneapolis, MN, USA, July 29-August 6, 2013*, 2014.
- [35] Michael Benedikt and Frank Zimmermann. Proton Colliders at the Energy Frontier. *Nucl. Instrum. Meth.*, A907:200–208, 2018.
- [36] W. James Stirling. Progress in Parton Distribution Functions and implications for LHC. In *Proceedings, 38th International Symposium on Multiparticle Dynamics (ISMD 2008): Hamburg, Germany, September 15-20, 2008*, pages 46–55, 2009. [,46(2008)].
- [37] C. Lefevre. LHC Guide. Mar 2017.
- [38] "CERN". "our member states". <https://home.cern/about/who-we-are/our-governance/member-states>, 2019.
- [39] Oliver Sim Brüning, Paul Collier, P Lebrun, Stephen Myers, Ranko Ostojic, John Poole, and Paul Proudlock. *LHC Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2004.
- [40] M. Lamont. LHC: Status and commissioning plans. In *QCD and high energy interactions. Proceedings, 44th Rencontres de Moriond, La Thuile, Italy, March 14-21, 2009*, pages 21–26, 2009.
- [41] "The ATLAS Collaboration". Public Results: Luminosity Run 2. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- [42] "The ATLAS Collaboration". "lhc commissioning with beam". <http://lhc-commissioning.web.cern.ch>.
- [43] Lyndon Evans and Philip Bryant. LHC Machine. *JINST*, 3:So8001, 2008.

- [44] J. W. Harris. Introduction to Hard Scattering Processes and Recent Results from Hard Probes at RHIC and LHC. In *Journal of Physics Conference Series*, volume 630 of *Journal of Physics Conference Series*, page 012052, July 2015.
- [45] Zachary Marshall. Simulation of Pile-up in the ATLAS Experiment. *J. Phys. Conf. Ser.*, 513:022024, 2014.
- [46] Claus Grupen. Physics of particle detection. *AIP Conference Proceedings*, 536(1):3–34, 2000.
- [47] The ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1999.
- [48] The ATLAS Collaboration. *ATLAS detector and physics performance: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1999.
- [49] The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:So8003, 2008.
- [50] The ATLAS Collaboration. Study of the material of the ATLAS inner detector for Run 2 of the LHC. . *JINST*, 12(CERN-EP-2017-081):P12009. 71 p, Jul 2017.
- [51] F. Hugging. The ATLAS Pixel Insertable B-Layer (IBL). *Nucl. Instrum. Meth.*, A650:45–49, 2011.
- [52] A Vogel. ATLAS Transition Radiation Tracker (TRT): Straw Tube Gaseous Detectors at High Rates. Technical Report ATL-INDET-PROC-2013-005, CERN, Geneva, Apr 2013.
- [53] Stefanie Morgenstern. ATLAS LAr Calorimeter Performance in LHC Run-2. Technical Report ATL-LARG-PROC-2018-004, CERN, Geneva, Jun 2018.
- [54] Ana Maria Henriques Correia. The ATLAS Tile Calorimeter. Technical Report ATL-TILECAL-PROC-2015-002, CERN, Geneva, Mar 2015.
- [55] *ATLAS muon spectrometer: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1997.
- [56] The ATLAS Collaboration. Performance of the ATLAS muon trigger in pp collisions at $\sqrt{s} = 8$ TeV. *Eur. Phys. J. C*, C75(CERN-PH-EP-2014-154. CERN-PH-EP-2014-154):120. 19 p, Aug 2014.
- [57] ATLAS magnet system: Technical design report. 1997.
- [58] Ryan Reece and Kyle Cranmer. Big-picture flow of atlas data. <https://reece.scipp.ucsc.edu/>.
- [59] The ATLAS Collaboration. "trigger and data acquisition webpage". <https://atlas.cern/discover/detector/trigger-daq>, 2019.

- [60] The ATLAS Collaboration. Performance of the ATLAS Trigger System in 2015. *Eur. Phys. J.*, C77(5):317, 2017.
- [61] Trigger Menu in 2017. Technical Report ATL-DAQ-PUB-2018-002, CERN, Geneva, Jun 2018.
- [62] The ATLAS Collaboration. "trigger operation public results". <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TriggerOperationPublicResults>, 2019.
- [63] Worldwide LHC Computing Grid. Worldwide lhc computing grid homepage. <http://wlcg.web.cern.ch/>.
- [64] John S. Jr. De Stefano. The us atlas computing facility. <https://www.racf.bnl.gov/experiments/usatlas>.
- [65] Christoph Eck et al. *LHC computing Grid: Technical Design Report. Version 1.06 (20 Jun 2005)*. Technical Design Report LCG. CERN, Geneva, 2005.
- [66] Rucio Development Team. Rucio: Scientific data management. <https://rucio.cern.ch/>.
- [67] Martin-Stefan Barisits et al. The ATLAS Data Management System Rucio: Supporting LHC Run-2 and beyond. Technical Report ATL-SOFT-PROC-2017-064. 3, CERN, Geneva, Nov 2017.
- [68] The ATLAS Collaboration. The ATLAS Simulation Infrastructure. *Eur. Phys. J. C*, 70(arXiv:1005.4568. CERN-PH-EP-2010-044):823–874. 53 p, May 2010. Submitted to *Eur. Phys. J. C*.
- [69] T. Gleisberg et al. Event generation with SHERPA 1.1. *JHEP*, 02:007, 2009.
- [70] Torbjörn Sjöstrand et al. An Introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015.
- [71] M. Bahr et al. Herwig++ Physics and Manual. *Eur. Phys. J.*, C58:639–707, 2008.
- [72] Johan Alwall et al. MadGraph 5 : Going Beyond. *JHEP*, 06:128, 2011.
- [73] HepMC Collaboration. Hepmc3 event record library. <http://hepmc.web.cern.ch/hepmc/>.
- [74] Stanislaw Jadach, Johann H. Kuhn, and Zbigniew Was. TAUOLA: A Library of Monte Carlo programs to simulate decays of polarized tau leptons. *Comput. Phys. Commun.*, 64:275–299, 1990.
- [75] D. J. Lange. The EvtGen particle decay simulation package. *Nucl. Instrum. Meth.*, A462:152–155, 2001.

- [76] Michelangelo L. Mangano, Mauro Moretti, Fulvio Piccinini, Roberto Pittau, and Antonio D. Polosa. ALPGEN, a generator for hard multiparton processes in hadronic collisions. *JHEP*, 07:001, 2003.
- [77] Stefano Frixione, Paolo Nason, and Bryan R. Webber. Matching NLO QCD and parton showers in heavy flavor production. *JHEP*, 08:007, 2003.
- [78] Borut Paul Kersevan and Elzbieta Richter-Was. The Monte Carlo event generator AcerMC versions 2.0 to 3.8 with interfaces to PYTHIA 6.4, HERWIG 6.5 and ARIADNE 4.1. *Comput. Phys. Commun.*, 184:919–985, 2013.
- [79] S. Agostinelli et al. GEANT4: A Simulation toolkit. *Nucl. Instrum. Meth.*, A506:250–303, 2003.
- [80] D. Costanzo, A. Dell’Acqua, A Di Simone, M. Gallas, A. Nairz, A. Rimoldi, J. Boudreau, and V. TSulaia. ATLAS detector simulation. 2005.
- [81] E. Barberio et al. The Geant4-Based ATLAS Fast Electromagnetic Shower Simulation. In *Astroparticle, particle and space physics, detectors and medical physics applications. Proceedings, 10th Conference, ICATPP 2007, Como, Italy, October 8-12, 2007*, pages 802–806, 2008.
- [82] K. Edmonds et al. The fast ATLAS track simulation (FATRAS). 2008.
- [83] The ATLAS Collaboration et al. The simulation principle and performance of the ATLAS fast calorimeter simulation FastCaloSim. Technical Report ATL-PHYS-PUB-2010-013, CERN, Geneva, Oct 2010.
- [84] V Lacuesta. Track and vertex reconstruction in the ATLAS experiment. *Journal of Instrumentation*, 8(02):C02035–C02035, feb 2013.
- [85] The ATLAS Collaboration. Performance of the atlas track reconstruction algorithms in dense environments in lh run 2. *Eur. Phys. J.*, C77(10):673, 2017.
- [86] R. Fruhwirth. Application of Kalman filtering to track and vertex fitting. *Nucl. Instrum. Meth.*, A262:444–450, 1987.
- [87] Jovan Mitrevski. Preparing ATLAS reconstruction software for LHC’s run 2. *Journal of Physics: Conference Series*, 664(7):072034, dec 2015.
- [88] Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment. Technical Report ATLAS-CONF-2012-042, CERN, Geneva, Mar 2012.
- [89] The ATLAS Collaboration. Vertex performance in 2018 data. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/IDTR-2018-006/>.

- [90] W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. Technical Report ATL-LARG-PUB-2008-002. ATL-COM-LARG-2008-003, CERN, Geneva, Apr 2008.
- [91] The ATLAS Collaboration. Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1. *Eur. Phys. J.*, C77:490, 2017.
- [92] The ATLAS Collaboration. Electron and photon energy calibration with the ATLAS detector using 2015–2016 LHC proton-proton collision data. *JINST*, 14(03):P03017, 2019.
- [93] The ATLAS Collaboration. Calibration of the tile hadronic calorimeter of ATLAS at LHC. *Journal of Physics: Conference Series*, 587:012009, feb 2015.
- [94] The ATLAS Collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett.*, B716:1–29, 2012.
- [95] The ATLAS Collaboration. Measurement of the ZZ Production Cross Section in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 116(10):101801, 2016.
- [96] The ATLAS Collaboration. Search for new particles in events with one lepton and missing transverse momentum in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *JHEP*, 09:037, 2014.
- [97] The ATLAS Collaboration. Muon reconstruction performance of the ATLAS detector in proton–proton collision data at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C76(5):292, 2016.
- [98] Ryan Atkin. Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645:012008, oct 2015.
- [99] The ATLAS Collaboration. ATLAS jet and missing-ET reconstruction, calibration, and performance. *Nucl. Part. Phys. Proc.*, 273-275:1121–1126, 2016.
- [100] The ATLAS Collaboration. Performance of b -Jet Identification in the ATLAS Experiment. *JINST*, 11(04):P04008, 2016.
- [101] The ATLAS Collaboration. Performance of top-quark and W -boson tagging with ATLAS in Run 2 of the LHC. 2018.
- [102] Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at $\sqrt{s} = 13$ TeV. Technical Report ATLAS-CONF-2017-029, CERN, Geneva, May 2017.
- [103] The Atlas Collaboration. Reconstruction of hadronic decay products of tau leptons with the atlas experiment. *The European Physical Journal C*, 76(5):295, May 2016.

- [104] Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment for Run-2 of the LHC. Technical Report ATL-PHYS-PUB-2015-045, CERN, Geneva, Nov 2015.
- [105] The ATLAS Collaboration. Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at $\sqrt{s} = 13$ TeV. *Eur. Phys. J.*, C78(11):903, 2018.
- [106] F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors. *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, USA, 2012. Curran Associates Inc.
- [107] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. *Deep Learning for Medical Image Processing: Overview, Challenges and the Future*, pages 323–350. Springer International Publishing, Cham, 2018.
- [108] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. *CoRR*, abs/1411.4555, 2014.
- [109] Karl Moritz Hermann, Tomáš Kociský, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. *CoRR*, abs/1506.03340, 2015.
- [110] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space, 2013.
- [111] Hongzi Mao, Mohammad Alizadeh, Ishai Menache, and Srikanth Kandula. Resource management with deep reinforcement learning. In *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, HotNets'16, pages 50–56, New York, NY, USA, 2016. ACM.
- [112] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang. Machine learning and deep learning methods for cybersecurity. *IEEE Access*, 6:35365–35381, 2018.
- [113] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, New York, NY, USA, 2016.
- [114] Xinyang Deng, Qi Liu, Yong Deng, and Sankaran Mahadevan. An improved method to construct basic probability assignment based on the confusion matrix for classification problem. *Information Sciences*, 340-341:250 – 261, 2016.
- [115] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. 14, 03 2001.
- [116] Jan Kukacka, Vladimir Golkov, and Daniel Cremers. Regularization for deep learning: A taxonomy. *CoRR*, abs/1710.10686, 2017.

- [117] Harris Drucker and Corinna Cortes. Boosting decision trees. volume 8, pages 479–485, 01 1995.
- [118] Jerome H. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4):367–378, February 2002.
- [119] Mahsa Shoaran, Benyamin Allahgholizadeh Haghi, Milad Taghavi, Masoud Farivar, and Azita Emami. Energy-efficient classification for resource-constrained biomedical applications. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, PP:1–1, 06 2018.
- [120] I.A Basheer and M Hajmeer. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1):3 – 31, 2000. Neural Computing in Micrbiology.
- [121] P.J. Werbos. *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*. Harvard University, 1975.
- [122] Sepp Hochreiter. The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 06(02):107–116, 1998.
- [123] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [124] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.
- [125] The Super Data Science Team. The ultimate guide to convolutional neural networks. <https://www.superdatascience.com/blogs/the-ultimate-guide-to-convolutional-neural-networks-cnn>.
- [126] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Neurocomputing: Foundations of research. chapter Learning Representations by Back-propagating Errors, pages 696–699. MIT Press, Cambridge, MA, USA, 1988.
- [127] Gilles Louppe, Michael Kagan, and Kyle Cranmer. Learning to Pivot with Adversarial Networks. 2016.
- [128] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep Learning and its Application to LHC Physics. *Ann. Rev. Nucl. Part. Sci.*, 68:161–181, 2018.
- [129] M Feikert et al. Hep ml resources. <https://github.com/iml-wg/HEP-ML-Resources>.

- [130] Kim Albertsson et al. Machine Learning in High Energy Physics Community White Paper. *J. Phys. Conf. Ser.*, 1085(2):022008, 2018.
- [131] Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s} = 13$ TeV. Technical Report ATL-PHYS-PUB-2017-004, CERN, Geneva, Apr 2017.
- [132] Andrew J. Larkoski, Ian Moulton, and Benjamin Nachman. Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning. 2017.
- [133] M. Aaboud et al. Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2. *Eur. Phys. J.*, C77(10):673, 2017.
- [134] David Rousseau. TrackML : The High Energy Physics Tracking Challenge. TrackML : The High Energy Physics Tracking Challenge. Mar 2018.
- [135] A. Rogozhnikov. Reweighting with Boosted Decision Trees. *J. Phys. Conf. Ser.*, 762(1):012036, 2016.
- [136] Stefanie Morgenstern. Electron and photon energy measurement calibration with the ATLAS detector. Technical Report ATL-PHYS-PROC-2018-138, CERN, Geneva, Oct 2018.
- [137] Josh Cogan, Michael Kagan, Emanuel Strauss, and Ariel Schwartzman. Jet-images: computer vision inspired techniques for jet tagging. *Journal of High Energy Physics*, 2015(2):118, Feb 2015.
- [138] Luke de Oliveira, Michael Kagan, Lester Mackey, Benjamin Nachman, and Ariel Schwartzman. Jet-images — deep learning edition. *JHEP*, 07:069, 2016.
- [139] Shannon Egan, Wojciech Fedorko, Alison Lister, Jannicke Pearkes, and Colin Gay. Long Short-Term Memory (LSTM) networks with jet constituents for boosted top tagging at the LHC. 2017.
- [140] Identification of Jets Containing b -Hadrons with Recurrent Neural Networks at the ATLAS Experiment. Technical Report ATL-PHYS-PUB-2017-003, CERN, Geneva, Mar 2017.
- [141] The ATLAS Collaboration. Tau trigger plots for lhcp 2019. https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TauTriggerPublicResults#Plots_for_LHCP_2019.
- [142] Gilles Louppe, Kyunghyun Cho, Cyril Becot, and Kyle Cranmer. QCD-Aware Recursive Neural Networks for Jet Physics. *JHEP*, 01:057, 2019.
- [143] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014.

- [144] The ATLAS Collaboration. Search for new light gauge bosons in higgs boson decays to four-lepton final states in pp collisions at $\sqrt{s} = 8\text{TeV}$ with the atlas detector at the lhc. *Phys. Rev. D*, 92:092001, Nov 2015.
- [145] Sanjoy Biswas, Emidio Gabrielli, Matti Heikinheimo, and Barbara Mele. Dark-Photon searches via Higgs-boson production at the LHC. *Phys. Rev.*, D93(9):093011, 2016.
- [146] Luke de Oliveira, Michela Paganini, and Benjamin Nachman. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis. *Comput. Softw. Big Sci.*, 1(1):4, 2017.
- [147] Michela Paganini, Luke de Oliveira, and Benjamin Nachman. Accelerating Science with Generative Adversarial Networks: An Application to 3D Particle Showers in Multilayer Calorimeters. *Phys. Rev. Lett.*, 120(4):042003, 2018.
- [148] Chase Shimmmin, Peter Sadowski, Pierre Baldi, Edison Weik, Daniel Whiteson, Edward Goul, and Andreas Søgaard. Decorrelated Jet Substructure Tagging using Adversarial Neural Networks. *Phys. Rev.*, D96(7):074034, 2017.
- [149] Christos et al Anastopoulos. Support Note for Electron ID: electron reconstruction and identification. Technical Report ATL-COM-PHYS-2017-1012, CERN, Geneva, Jul 2017.
- [150] ATLAS Collaboration. Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at $\sqrt{s} = 13\text{ TeV}$. *Submitted to: Eur. Phys. J.*, 2019.
- [151] ATLAS Collaboration. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data. *Eur. Phys. J.*, C77(3):195, 2017.
- [152] J. Alison, K. Brendlinger, S. Heim, J. Kroll, and C. M. Lester. Description and performance of the electron likelihood tool at atlas using 2012 lhc data.
- [153] The ATLAS collaboration. Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton-proton collision data. 2016.
- [154] The ATLAS Collaboration. Electron efficiency measurements in 2017 data and electron identification discriminating variables from 2016 data. <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2018-002/index.html>.
- [155] K. Albertson et al. Tmva 4: Toolkit for multivariate analysis. root.cern/tmva.
- [156] Julien Maurer. Search for compressed SUSY scenarios with the ATLAS detector. Technical Report ATL-PHYS-PROC-2017-174, CERN, Geneva, Sep 2017.

- [157] ATLAS Collaboration. Search for dark matter at $\sqrt{s} = 13$ TeV in final states containing an energetic photon and large missing transverse momentum with the ATLAS detector. *Eur. Phys. J.*, C77(6):393, 2017.
- [158] ATLAS Collaboration. Search for dark matter and other new phenomena in events with an energetic jet and large missing transverse momentum using the ATLAS detector. *JHEP*, 01:126, 2018.
- [159] ATLAS Collaboration. Measurement of inclusive and differential cross sections in the $H \rightarrow ZZ^* \rightarrow 4\ell$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 10:132, 2017.
- [160] Keras Team. Keras. <https://github.com/charlespwd/project-title>, 2013.
- [161] Martín Abadi et al. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [162] Combined measurements of Higgs boson production and decay using up to 80 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13$ TeV collected with the ATLAS experiment. Technical Report ATLAS-CONF-2019-005, CERN, Geneva, Mar 2019.
- [163] The ATLAS Collaboration. Measurement of the production cross section for a Higgs boson in association with a vector boson in the $H \rightarrow WW^* \rightarrow \ell\nu\ell\nu$ channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. 2019.
- [164] The ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D99:072001, 2019.
- [165] The ATLAS Collaboration. Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a tau pair with the ATLAS detector. Technical Report ATL-COM-PHYS-2014-487, CERN, Geneva, May 2014.
- [166] The ATLAS Collaboration. Hadronic tau performance public results. <https://twiki.cern.ch/twiki/bin/viewauth/AtlasPublic/TauPublicResults>.
- [167] The ATLAS Collaboration. Muon performance public results. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/MuonPublicResults>.
- [168] The ATLAS Collaboration. Electron photon performance public results. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/ElectronGammaPublicCollisionResults>.
- [169] The ATLAS Collaboration. Jet and missing energy performance public results. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/JetEtmissPublicResults>.

- [170] A. Elagin, P. Murat, A. Pranko, and A. Safonov. A New Mass Reconstruction Technique for Resonances Decaying to di-tau. *Nucl. Instrum. Meth.*, A654:481–489, 2011.
- [171] The ATLAS Collaboration. Search for the Standard Model Higgs boson produced in association with a vector boson and decaying into a tau pair in pp collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector. *Phys. Rev.*, D93(9):092005, 2016.
- [172] A. J. Barr, T. J. Khoo, P. Konar, K. Kong, C. G. Lester, K. T. Matchev, and M. Park. Guide to transverse projections and mass-constraining variables. *Phys. Rev.*, D84:095031, 2011.
- [173] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J.*, C71:1554, 2011. [Erratum: *Eur. Phys. J.* C73,2501(2013)].
- [174] Y Rozen. Data-driven background estimation in ATLAS. Technical Report ATL-PHYS-PROC-2013-023, CERN, Geneva, Jan 2013.
- [175] Marzieh Bahmani. Data-driven estimation of fake τ background in Higgs searches in ATLAS. Technical Report ATL-PHYS-PROC-2019-006, CERN, Geneva, Jan 2019.
- [176] The ATLAS Collaboration. Cross-section measurements of the Higgs boson decaying into a pair of τ -leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Rev.*, D99:072001, 2019.
- [177] D. de Florian et al. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. Oct 2016. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG>.
- [178] CMS Collaboration. Measurement of Higgs boson production and decay to the $\tau\tau$ final state. 2019.